Statistics, Politics, and Policy

Volume 2, *Issue* 1 2011 *Article* 1

Assessing the Early Aberration Reporting System's Ability to Locally Detect the 2009 Influenza Pandemic

Katie S. Hagen, *U.S. Navy*
Ronald D. Fricker Jr., *Naval Postgraduate School*
Krista D. Hanni, *Monterey County Health Department*
Susan Barnes, *Monterey County Health Department*
Kristy Michie, *Monterey County Health Department*

Recommended Citation:

Hagen, Katie S.; Fricker, Ronald D. Jr.; Hanni, Krista D.; Barnes, Susan; and Michie, Kristy (2011) "Assessing the Early Aberration Reporting System's Ability to Locally Detect the 2009 Influenza Pandemic," *Statistics, Politics, and Policy*: Vol. 2: Iss. 1, Article 1.
DOI: 10.2202/2151-7509.1018
Available at: http://www.bepress.com/spp/vol2/iss1/1

Assessing the Early Aberration Reporting System's Ability to Locally Detect the 2009 Influenza Pandemic

Katie S. Hagen, Ronald D. Fricker Jr., Krista D. Hanni, Susan Barnes, and Kristy Michie

Abstract

The Early Aberration Reporting System (EARS) is used by some local health departments (LHDs) to monitor emergency room and clinic data for disease outbreaks. Using actual chief complaint data from local public health clinics, we evaluate how EARS—both the baseline system distributed by the CDC and two variants implemented by one LHD—perform at locally detecting the 2009 influenza A H1N1 pandemic. We also compare the EARS methods to a CUSUM-based method. We find that the baseline EARS system performed poorly in comparison to one of the LHD variants and the CUSUM-based method. These results suggest that changes in how syndromes are defined can substantially improve EARS performance. The results also show that incorporating algorithms that use more historical data will improve EARS performance for routine surveillance by local health departments.

KEYWORDS: biosurveillance, syndromic surveillance, H1N1, influenza

Author Notes: We which to express our appreciation to the Technology Innovation Office of the Defense Threat Reduction Agency for funding that supported K. Hagen's research and travel. We also wish to express our appreciation to the editors and three anonymous reviewers for the time and effort they put into their reviews; their comments helped us to substantially improve the paper. Funding for the implementation and operation of EARS at the Monterey County Health Department is provided by the Centers for Disease Control and Prevention Public Health Emergency Preparedness Program, the California State General Fund Pandemic Influenza Planning Program, and the U.S. Department of Health and Human Services Assistant Secretary for Prevention and Response Hospital Preparedness Program.

1 Introduction: Biosurveillance

Homeland Security Presidential Directive 21 (HSPD-21) defines *biosurveillance* as "the process of active data-gathering with appropriate analysis and interpretation of biosphere data that might relate to disease activity and threats to human or animal health – whether infectious, toxic, metabolic, or otherwise, and regardless of intentional or natural origin – in order to achieve early warning of health threats, early detection of health events, and overall situational awareness of disease activity" (U.S. Government, 2007). One type of biosurveillance is *epidemiologic surveillance* which HSPD-21 defines as "the process of actively gathering and analyzing data related to human health and disease in a population in order to obtain early warning of human health events, rapid characterization of human disease events, and overall situational awareness of disease activity in the human population." Thus, epidemiologic surveillance addresses that subset of biosurveillance as it applies to human populations. *Syndromic surveillance* is a specific type of epidemiologic surveillance that has been defined as "the ongoing, systematic collection, analysis, interpretation, and application of real-time (or near-real-time) indicators of diseases and outbreaks that allow for their detection before public health authorities would otherwise note them" (Sosin, 2003). Syndromic surveillance is epidemiologic surveillance restricted to general illness categories. For additional background and discussion of biosurveillance in general, and syndromic surveillance in particular, see Shmueli & Burkom (2010), Fricker (2010b, 2008, 2007), and Fricker & Rolka (2006).

The medical and public health communities are developing *biosurveillance systems* designed to proactively monitor populations for possible disease outbreaks[1]. One goal of these systems is to improve the likelihood that a disease outbreak, whether man-made or natural, is detected as early as possible so that the medical and public health communities can respond as quickly as possible. This goal is often referred to as *early event detection* (EED). As shown in Figure 1, a biosurveillance system has four main functions: data collection, data management, analysis, and reporting. The ideal biosurveillance system analyzes population health-related data in near-real time to identify subtle trends not visible to individual physicians and clinicians. Many of these systems use one or more statistical algorithms to assess data for anomalies. Systems are monitored for anomalies which may then trigger detection, alerting, investigation, quantification, and localization of potential public health problems.

[1]In a strict epidemiological context, the term "outbreak" implies an identified chain of transmission. Here we use the term "outbreak" to denote a sudden increase in the incidence

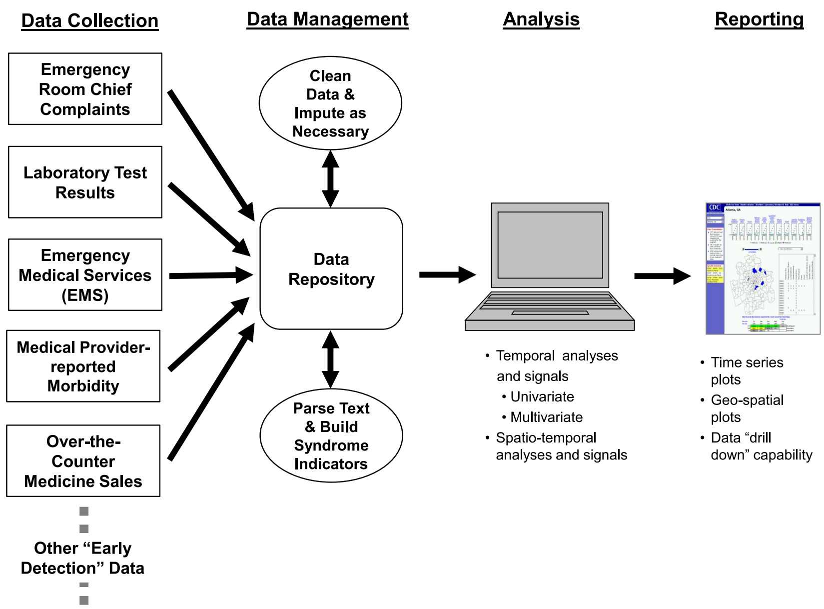

Figure 1: A biosurveillance system has four main functions: data collection, data management, analysis, and reporting. Raw data enters the system at the left and flows through the system to become actionable information at the right.

Effective early event detection depends on sensitive statistical algorithms, but it also depends on accurate data (which are often daily counts of individuals classified into one or more syndrome categories). “Accurate” in this context means the syndrome counts mirror in behavior the underlying disease that the biosurveillance system is intended to monitor. In particular, at a minimum the syndrome counts should show increases when the prevalence of the underlying disease increases and ideally they should show increases that precede increases in diagnosed cases. Manipulating both the syndrome definitions and the statistical algorithms can affect the performance of a biosurveillance system, where a perfect system would quickly and correctly identify an increase in disease when one is occurring, yet not falsely signal an outbreak or epidemic in the absence of one.

This paper looks at how changes in both statistical algorithms and syndrome definitions affect how one biosurveillance system would have de-

of a disease within a defined population.

tected the local arrival of the 2009 influenza A H1N1 virus pandemic (often colloquially referred to as swine flu) in Monterey County, California. That is, we retrospectively mimic a prospective surveillance system, assessing how variations in the syndrome definitions and modifications to the statistical algorithms affect the system's ability to detect a known outbreak.

Comparisons between biosurveillance systems or the algorithms they use have been not been published in the literature as frequently as some might desire (Fricker, 2010b). Those related to the system evaluated in this paper include Hutwagner *et al.* (2003), Hutwagner *et al.* (2005), Tokars *et al.* (2009), and Fricker *et al.* (2008a,b). The latter two are the most closely related to this work, where the statistical algorithms were compared using simulated data. This work expands on and extends Fricker *et al.* (2008a,b) in two ways: (1) it uses actual Monterey County data rather than simulated data, and (2) it explores how changes in syndrome definitions affect biosurveillance system performance.

This paper is organized as follows. In Section 2 we describe the biosurveillance system used in Monterey County, definitional variants for the influenza-like illness (ILI) syndrome, how the ILI syndrome counts are calculated, and the various EED algorithms we evaluated. In Section 3 we present our results, comparing both the performance of the statistical algorithms as well as the impacts of changing the syndrome definitions. Finally, in Section 4 we discuss our findings and conclusions with a focus on particular improvements that have the potential to dramatically improve biosurveillance system performance.

2 Biosurveillance in Monterey County, California

The Early Aberration Reporting System (EARS) is a biosurveillance system that was and continues to be developed by the Centers for Disease Control and Prevention (CDC, 2007b). Written in SAS, it was originally designed as a drop-in surveillance system for large-scale events where little or no baseline data are available (CDC, 2007a). For example, the EARS system was used in the aftermath of Hurricane Katrina to monitor communicable diseases in Louisiana (Toprani *et al.*, 2006), for syndromic surveillance at the 2001 Super Bowl and World Series, as well as at the Democratic National Convention in 2000 (Hutwagner *et al.*, 2003). Though developed as a drop-in surveillance system, EARS is now used for routine biosurveillance by many state and local health departments (LHDs), including Monterey County.

EARS conducts EED by monitoring for increases in syndromes derived from chief complaints. A *syndrome* is “a set of symptoms or conditions that occur together and suggest the presence of a certain disease or an increased chance of developing the disease” (International Foundation for Functional Gastrointestinal Disorders, 2010). In the context of syndromic surveillance, a syndrome is a set of non-specific pre-diagnosis medical and other information that may indicate the release of a bioterrorism agent or natural increase in disease. Syndromes monitored by Monterey County include ILI, gastrointestinal, upper respiratory, lower respiratory, and neurological. A *chief complaint* is a brief summary of the reason or reasons that an individual presents at a medical facility. Written by medical personnel, chief complaints are full of jargon, acronyms, and abbreviations for use by other medical professionals. To distill the chief complaints down into syndrome indicators, the text is searched and parsed for key words, often of necessity including all the ways a particular key word can be misspelled, abridged, and otherwise abbreviated (Fricker, 2010b).

The Monterey County Health Department (MCHD) is an LHD that uses EARS V4.5 to monitor chief complaint data from four hospital emergency rooms (ERs) and six public health clinics, particularly as an alert system for various types of disease outbreaks which may include those naturally occurring (e.g., influenza), accidental (e.g., fire-related illnesses), or intentional (e.g., bioterrorism). Table 1 gives examples of actual chief complaints taken from Monterey County clinic data. The results of this study are based on the clinic data only.

While there are a number of biosurveillance systems available for use, MCHD uses EARS because it allows them to maintain local control of the data and because of the system’s flexibility. In particular, MCHD values the ability in EARS to develop syndromes for unique, local circumstances such as agriculture pesticide spraying and fire-related illness tracking (Hanni, 2011; Fricker & Hanni, 2010). While this flexibility is considered a significant benefit of the system, the effects of changes to EARS syndrome definitions have not been published in the literature and, thus, the effects of such changes on the EED ability of the system are unknown.

2.1 Calculating Syndrome Counts

EARS monitors various syndromes for outbreaks based on the presence of key words in chief complaint records. This is accomplished in two steps. First, the presence of particular words, word variants, common typos, and associated medical abbreviations and jargon are searched for in the chief complaint text and linked to specific symptoms. Second, the symptoms are then analyzed to

FU ANEMIA	chdp
F/U ASTHMA, FEVER AND COUGH	FEVER,PHLEGM
fever x3 days cough	4WK FU OB
FEVER,WHEEZING	FEVER
VOMITING,FEVER,POSS EAR INFECT	PAP
CHDP	COUGH,DXd W/ ASTHMA
WI C/O HA//MM	CHDP
PAP	DEPO
walk-in hospital fu	4WK FU OB ..OVBK
COLD	ABD PAIN CONJESTION
WALK IN BURN TO R-HAND	new born with mom
FU RESULTS	RASH
FU OB	FU WT CHECK
SHLDR PAIN FOR 1 WK	PAP

Table 1: Examples of actual chief complaints taken from Monterey County's clinic data.

determine the presence of a syndrome. Ultimately, for each syndrome, every individual in the data is categorized as either having that syndrome or not (so that individuals can be categorized in multiple syndrome categories).

For example, the flu symptom is the simplest, where the existence of the word "flu" in an individual's chief complaint text results in the flu symptom being set for that individual. More complicated is the fever symptom, where the presence of the word "fever", "fver", "780.6", or any of another 105 terms result in the fever symptom being set for the individual. A similar approach is taken for other symptoms such as sore throat, cold, cough, etc., where for 76 symptoms in the EARS system, there are 9,421 total terms that are searched for in the chief complaint text, or an average of 124 terms per symptom. At one extreme is the flu symptom with only one term ("flu") that is searched for, while at the other extreme is the abscess symptom with 488 terms.

One of the ways in which EARS can be manipulated at the local level is by changing how a syndrome is defined. As illustrated in Table 2, MCHD defined the ILI syndrome in three different ways. According to the EARS baseline ILI syndrome definition, a record is flagged for ILI when the chief complaint field contains any one or more of the following symptoms: "sore throat" or "cold" or "cough" (where the quotation marks are intended to emphasize that each symptom involves searching through chief complaint text for a variety of terms).

ILI Definitions	**Symptom Combination Logic**
EARS Baseline:	"cold" *or* "cough" *or* "sore throat"
MCHD Expanded:	"cold" *or* "cough" *or* "fever" *or* "chills" *or* "muscle pain" *or* "headache" *or* ("flu" *and not* "shot")
MCHD Restricted:	("fever" *and* "cough") *or* ("fever" *and* "sore throat") *or* ("flu" *and not* "shot")

Table 2: ILI definitions. The EARS baseline definition is what is used by EARS V4.5. The MCHD expanded and restricted definitions are variants created by the Monterey County Health Department. The quotation marks around the symptoms are intended to emphasize that each symptom involves searching through chief complaint text for a series of terms, from just one for the flu symptom to 236 for the sore throat symptom.

Prior to emergence of the 2009 H1N1 virus, Monterey County implemented the expanded ILI syndrome definition to increase the probability their system would signal during an outbreak. As shown in Table 2, this expanded ILI syndrome definition added in fever, chills, muscle pain, headache, and flu symptoms while deleting the sore throat symptom in the EARS baseline definition. In addition, it did not count records that included the word "shot" so as to not count individuals who had received a flu shot (who otherwise would have been incorrectly included in the ILI syndrome count by virtue of the presence of the word "flu" in their chief complaint text). This expanded ILI syndrome definition generated a substantial increase in the daily ILI syndrome counts and resulted in an estimated rate of ILI that significantly exceeded what was being reported at the state level via the California Sentinel Provider system.

In October 2009, MCHD subsequently revised the ILI syndrome definition to better match the California Sentinel Provider system results. Instead of simply looking for the existence of one or more symptoms, the restricted ILI syndrome definition now requires more evidence where, as shown in Table 2, two or more symptoms need to be present (fever and cough, for example). The goal was to better adjust the observed counts to match the California Sentinel Provider data, where the restricted definition limits incorrectly classifying individuals with the ILI syndrome who do not actually have the flu, but this strategy comes at the cost of a greater chance of failing to count those with the flu in the ILI syndrome.

The impact of changing the syndrome definitions is quite dramatic, at least in terms of the total number of individuals classified with the ILI syndrome. For example, for one year of Monterey County data, the EARS baseline definition resulted in just under six percent of the record being classified with the ILI syndrome (9,093 out of 153,696 records). In contrast, the MCHD expanded definition resulted in a 53% increase in the number of records classified as the ILI syndrome (13,956 records or nine percent of the total) while the MCHD restricted ILI resulted in a 92% reduction of the number of records (734 or 0.5 percent of the total). Clearly, the choice of syndrome definition can have a significant effect on the daily syndrome counts, on which the EARS' detection algorithms rely.

2.2 Early Event Detection Methods

In this section, we define the EARS methods used in EARS V4.5, followed by the CUSUM methodology, and then we describe how we applied each to the MCHD data. Subsequent to this research the CDC released a new version of EARS in which a "4th algorithm choice was added" that is intended to improve the chance of correctly detecting outbreaks while also decreasing the likelihood of false positive signals (CDC, 2010). We do not assess the performance of this new algorithm in this work.

2.2.1 EARS' C1, C2, and C3

The EARS C1, C2, and C3 methods were intended to be CUSUM-like methods (Fricker, 2010a; Hutwagner *et al.*, 2003) and, in fact, the EARS documentation, SAS code, and at least one paper (Zhu *et al.*, 2005) explicitly refer to them as CUSUMs. However, the C1 and C2 are actually Shewhart variants that use a moving sample average and sample standard deviation to standardize each observation. (See Shewhart, 1931, or Montgomery, 2009, for more detail on the Shewhart method and the next section for the definition of the CUSUM.) The C1 uses the seven days prior to the current observation to calculate the sample average and sample standard deviation. The C2 is similar to the C1 but uses the seven days prior to a two-day lag. The C3 combines information from C2 statistics as described below.

Let Y_t be the observed count for period t representing, for example, the number of individuals classified with a particular syndrome at all the public clinics in Monterey County on day t. The C1 calculates the statistic $\mathrm{C1}_t$ as

$$\mathrm{C1}_t = \frac{Y_t - \bar{Y}_{1,t}}{S_{1,t}} \tag{1}$$

where $\bar{Y}_{1,t}$ and $S_{1,t}$ are the moving sample mean and standard deviation, respectively:

$$\bar{Y}_{1,t} = \frac{1}{7}\sum_{i=t-1}^{t-7} Y_i \text{ and } S^2_{1,t} = \frac{1}{6}\sum_{i=t-1}^{t-7}\left[Y_i - \bar{Y}_{1,t}\right]^2.$$

As implemented in the EARS system, the C1 signals at time t when its statistic exceeds a threshold h, which is fixed at three sample standard deviations above the sample mean: $\mathrm{C1}_t > 3$.

The C2 is similar to the C1, but incorporates a two-day lag in the mean and standard deviation calculations. Specifically, it calculates

$$\mathrm{C2}_t = \frac{Y_t - \bar{Y}_{3,t}}{S_{3,t}} \tag{2}$$

where

$$\bar{Y}_{3,t} = \frac{1}{7}\sum_{i=t-3}^{t-9} Y_i \text{ and } S^2_{3,t} = \frac{1}{6}\sum_{i=t-3}^{t-9}\left[Y_i - \bar{Y}_{3,t}\right]^2,$$

and in EARS it signals when $\mathrm{C2}_t > 3$.

The C3 uses the C2 statistics from day t and the previous two days, calculating the statistic $C_3(t)$ as

$$\mathrm{C3}_t = \sum_{i=t}^{t-2} \max\left[0, \mathrm{C2}_i - 1\right]. \tag{3}$$

In EARS it signals when $\mathrm{C3}_t > 2$. For additional information on EARS and EARS methods see Fricker (2011, 2008), Fricker *et al.* (2008a,b), CDC (2006), and Hutwagner *et al.* (2005, 2003).

2.2.2 CUSUM

The CUSUM method of Page (1954) and Lorden (1971) is a well known statistical process control methodology. In that literature it is often referred to as the CUSUM control chart. Montgomery (2009) is an excellent introduction to the CUSUM in an industrial statistical process control setting and Hawkins & Olwell (1998) provides a comprehensive treatment of the CUSUM.

Formally, the CUSUM is a sequential hypothesis test for a change from a known in-control density f_0 to a known alternative density f_1. The method monitors the statistic C_t, which satisfies the recursion

$$\mathrm{C}_t = \max[0, \mathrm{C}_{t-1} + L_t], \tag{4}$$

DOI: 10.2202/2151-7509.1018

where the increment L_t is the log likelihood ratio

$$L_t = \log \frac{f_1[Y_t]}{f_0[Y_t]}.$$

The method is usually started at $\mathrm{C}_0 = 0$; it stops and concludes that $Y_t \sim f_1$ at the first time when $\mathrm{C}_t > h$, for some threshold h that achieves a desired *average time between false signals* (ATFS) when $Y_t \sim f_0$ (i.e., when no outbreak is present).

The ATFS is the mean number of time periods it takes for an EED method to first signal, starting from some initial state, given there are no outbreaks. That is, roughly speaking, the ATFS is the expected time to the first false signal. If the CUSUM is reset to its initial state after each signal, it is equivalent to the in-control average run length (ARL_0) metric used in statistical process control. For further discussion of the ATFS metric, see Fricker (2010a,b) and Fricker *et al.* (2008a).

If f_0 and f_1 are normal densities with means μ and $\mu + \delta$, with $\delta > 0$ and unit variances, then Equation 4 reduces to

$$\mathrm{C}_t = \max[0, \mathrm{C}_{t-1} + Y_t - \mu - k], \tag{5}$$

with $k = \delta/2$, where k is commonly referred to as the *reference value*. If the Ys are independent and identically distributed according to f_0 before some unknown change point and according to f_1 after the change point, then the CUSUM has certain optimality properties. See Moustakides (1986) and Ritov (1990).

Equation 5 is the CUSUM form routinely used, even when the underlying assumptions are only approximately met. Also, Equation 5 is a one-sided CUSUM, meaning that it will only detect increases in the mean. If it is important to detect both increases and decreases in the mean, a second CUSUM must be used to detect decreases. In syndromic surveillance, since it is only important to quickly detect increases in disease incidence, the second CUSUM is generally unnecessary.

In industrial settings, the CUSUM is often applied directly to the observations because some control is exhibited over the process such that it is reasonable to assume f_0 is stationary. In syndromic surveillance this is generally not the case as the data often have uncontrollable systematic trends, such as seasonal cycles and day-of-the-week effects. One solution is to model the systematic component of the data, use the model to forecast the next day's observation, and then apply the CUSUM to the forecast errors (Montgomery, 2009, pp. 450-457).

2.2.3 Applying the CUSUM to Adaptive Regression Forecast Errors

We used the "adaptive regression model with sliding baseline" of Burkom *et al.* (2006) to model the systematic component of syndromic surveillance data. The basic idea is as follows. Let Y_t be an observation, say the syndrome count, on day t. Regress the observations for the past n days on time relative to the current period. Then use the model to predict today's observation and apply the CUSUM to the difference between the predicted value and today's observed value. Repeat this process each day, always using the most recent n observations as the sliding baseline in the regression to calculate the forecast error. For $t > n$, and assuming a linear formulation with day-of-the-week effects, the model is

$$Y_i = \beta_0 + \beta_1 \times (i - t + n + 1) + \beta_2 I_{\text{Mon}} + \beta_3 I_{\text{Tues}} + \beta_4 I_{\text{Wed}} + \beta_5 I_{\text{Thurs}} + \epsilon \quad (6)$$

for $i = t-1, \ldots, t-n$ and where MCHD clinics are only open on weekdays, so covariates for the weekend days are not required. The Is are indicators, where $I = 1$ on the relevant weekday and $I = 0$ otherwise, and ϵ is the error term which is assumed normally distributed. Of course, as appropriate, the model can also be adapted to allow for nonlinearities by adding a quadratic term into Equation 6.

Burkom *et al.* (2006) used an 8-week sliding baseline ($n = 56$ based on a 7-day week). We compared the performance for a variety of n values and between a linear and quadratic form of the model, similar to Fricker *et al.* (2008a), and found that $n = 35$ (a 7-week sliding baseline based on a 5-day week) without a quadratic term worked best. Our judgement of "best" was based on: (1) the number of weeks in the sliding baseline was similar to that recommended in Burkom *et al.* (2006) and used in Fricker *et al.* (2008a,b), and, (2) the forecast errors were normally distributed.

The model is fit using ordinary least squares, regressing (at each time t) $Y_{t-1}, \ldots, Y_{t-n}$ on $n, \ldots, 1$, where n must be greater than p, the number of covariates to be estimated in the model. Having fit the model, the forecast error when day t is a Friday is

$$R_t = Y_t - \left[\hat{\beta}_0 + \hat{\beta}_1 \times (n+1)\right],$$

where $\hat{\beta}_0$ is the estimated slope and $\hat{\beta}_1$ is the estimated intercept. For any other day of the week the forecast error is

$$R_t = Y_t - \left[\hat{\beta}_0 + \hat{\beta}_1 \times (n+1) + \hat{\beta}_j\right],$$

DOI: 10.2202/2151-7509.1018

$j = 2, 3, 4, 5$, where $\hat{\beta}_2$ is the estimated day-of-the-week effect for Monday, $\hat{\beta}_3$ is for Tuesday, $\hat{\beta}_4$ is for Wednesday, and $\hat{\beta}_5$ is for Thursday. Standardizing, we have $Z_t = R_t/\hat{\sigma}_t$ where, following Fricker *et al.* (2008a) and assuming the forecast errors are independent with expected value zero (reasonable assumptions based on an analysis of the actual forecast errors; see Hagen, 2010), we estimated the variance of the forecast error at time t as

$$\hat{\sigma}_t^2 = \frac{1}{33} \sum_{i=t-35}^{t-1} R_i^2 \left(1 + \mathbf{x_0'}(\mathbf{X'X})^{-\mathbf{1}}\mathbf{x_0}\right). \tag{7}$$

In Equation 7, $\mathbf{x_0}$ is the covariate vector for day t, where for example for Monday $\mathbf{x_0} = \{1, 36, 1, 0, 0, 0\}$, and $\mathbf{X}$ is the 35×6 associated matrix of covariates for the previous 35 days, for this example:

$$\mathbf{X} = \begin{pmatrix} 1 & 35 & 0 & 0 & 0 & 0 \\ 1 & 34 & 0 & 0 & 0 & 1 \\ 1 & 33 & 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 3 & 0 & 0 & 1 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

The CUSUM applied to this problem is thus

$$\mathrm{C}_t = \max[0, \mathrm{C}_{t-1} + Z_t - k], \tag{8}$$

where each day the adaptive regression is re-fit to the past 35 days of data, the current day's forecast error is calculated and standardized, and the resulting Z_t is used in the CUSUM. Intuitively the idea is that an outbreak will result in larger than expected positive forecast errors and their values will accumulate in the CUSUM and eventually result in a signal.

In applying the CUSUM, we used three variants – based on the values of k and h – to illustrate a range of performance. As shown in Table 3, we called these "aggressive," "moderate," and "routine." CUSUM1 is called aggressive because, based on $k = 0.5$, it will signal quickly for a one standard deviation increase in the forecast errors, and with $k = 0.5$ and $h = 0.365$ it has an ATFS of 5 days. Intuitively, one can think of small ATFS values giving the CUSUM a higher probability of detecting outbreaks, but at the expense of a higher false positive signal rate, where an ATFS of 5 days means that there will be a false positive signal once a week on average (assuming the CUSUM is reset after each signal).

Type	**Label**	k	h	**ATFS**
Aggressive:	CUSUM1	0.5	0.365	5
Moderate:	CUSUM2	1.0	0.695	20
Routine:	CUSUM3	1.0	1.200	60

Table 3: Parameters for the three CUSUM variants we used. The choices were based on the average time to first signal (ATFS) metric, where CUSUM1 is designed to have a high probability of signaling an actual outbreak and, concomitently, a higher false positive signal rate. At the other extreme, CUSUM3 is designed to have a low false positive rate as well as a lower probability of signaling an actual outbreak.

In comparison, CUSUM2 is called moderate because, based on $k = 1.0$, it will signal quickly for a two standard deviation increase in the forecast errors, and with $k = 1.0$ and $h = 0.695$ it has an ATFS of 20 days. Thus CUSUM2 will be less sensitive than CUSUM1 but it will also have fewer false positive signals, with only one per month on average (four times less than CUSUM1). Finally, CUSUM3 is called routine because, while it will signal quickly for a two standard deviation increase in the forecast errors like CUSUM2, with $k = 1.0$ and $h = 1.2$ it has an ATFS of 60 days. Thus CUSUM3 will be less sensitive than either CUSUM1 or CUSUM2, but it will also only have a false positive signal once per quarter on average.

3 Results

In this section, we first discuss how we determined when the seasonal ILI and 2009 H1N1 pandemic outbreaks occurred in Monterey County. We then apply the EARS' and CUSUM-based detection algorithms to the three sets of ILI syndrome data (CDC baseline, MCHD expanded, MCHD restricted).

3.1 Determining the Outbreak Periods

In order to judge how well the various detection methods perform, we first sought to establish when the seasonal ILI and 2009 H1N1 pandemic outbreaks actually occurred within Monterey County. As anyone who has attempted to do this will recognize, establishing some sort of universal "ground truth" about precisely when an outbreak started is often elusive. Not only can the timing of the outbreak vary by geography and subpopulation, but the data can be quite imprecise. Furthermore the determination can at times be circular

DOI: 10.2202/2151-7509.1018

in the sense that knowledge of the start of an outbreak is required to judge algorithm performance but the algorithms are sometimes the most effective way to determine the outbreak start.

In this case, as shown in Figure 2, we looked at four sources of information. The figure shows the reported weekly percentage of patients classified with influenza-like illness (ILI) from September 28, 2008 to January 2, 2010, along with locally-weighted smoothing lines to better show the underlying trends, using data from: (1) the California Sentinel Provider Influenza Surveillance Program, (2) Monterey County hospital ERs, and (3) Monterey County public health clinics, as well as (4) laboratory-confirmed, hospitalized cases of 2009 H1N1 in Monterey County. These four sets of data reflect four different populations.

The first population consists of patients seen by medical providers throughout the state of California who voluntarily conduct surveillance for ILI and report weekly to the CDC. As described on the California Department of Public Health's web site (California Department of Public Health, 2010), the case definition for ILI is any illness with fever greater than 100°F and cough and/or sore throat (in the absence of a known cause). As such, the circles in Figure 2 are a measure of ILI activity throughout the entire state. The solid black line is locally-weighted smoothing line to better show the underlying trends in the state-level data.

The second population, represented by the triangles and associated dashed line in Figure 2, are individuals who went to an emergency room of one of the four Monterey County hospitals and who were subsequently classified with ILI by EARS using the MCHD restricted definition. The third population, represented by the crosses and the associated dashed line, are individuals who went to one of six Monterey County public clinics and who were subsequently classified with ILI by EARS using the restricted definition.

Finally, the fourth population is the entire population of Monterey County and the data are the laboratory-confirmed, hospitalized cases of 2009 H1N1 in Monterey County. These are shown as the black diamonds at the bottom of the plot, where each diamond represents one person and is plotted for the week the individual first became symptomatic due to 2009 H1N1 virus infection. A note about these data is in order: At the epidemic's onset, medical providers were required to report all laboratory-confirmed cases of 2009 H1N1 to their local health jurisdictions under Title 17 of the California Code of Regulations. In May 2009, the provider reporting requirements were restricted to fatal and/or hospitalized, laboratory-confirmed 2009 H1N1 cases. This allowed health officials to focus on the determinants of severe illness. Providers were encouraged to use clinical presentation rather than laboratory testing

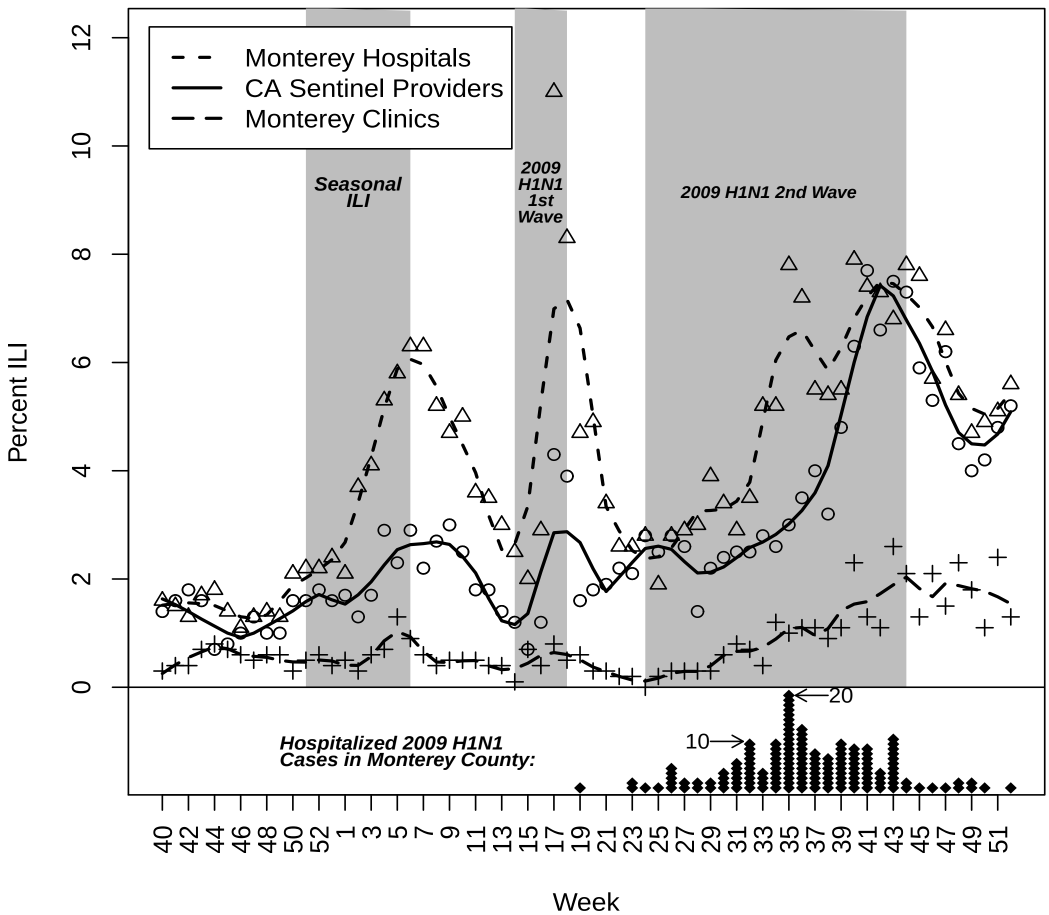

Figure 2: Percentage of patients classified with ILI from September 28, 2008 (week 40) to January 2, 2010 (week 52) for: (1) the California Sentinel Provider system, (2) Monterey hospital ERs, and (3) Monterey public health clinics. The diamonds are laboratory-confirmed, hospitalized cases of 2009 H1N1 in Monterey County, where each diamond represents one person and is plotted for the week the individual first became symptomatic due to 2009 H1N1 infection. The arrows show that week 32 had 10 cases and week 35 had 20 cases.

to guide management of patients. Widespread laboratory testing was not recommended. The reporting requirements were further restricted in May 2010 to laboratory-confirmed cases resulting in Intensive Care Unit admission and/or death. Therefore, no centralized database of outpatient 2009 H1N1 cases, which represented the majority of 2009 H1N1 infections, exists with which we could compare ILI reports.

In looking at the smoothed curves in Figure 2, we expected the MCHD hospital ER and clinic ILI syndrome trends would closely match the California Sentinel Provider data, and the three time series do show similar patterns. However, there are also some differences. For example, as highlighted by the left-most shaded region in Figure 2, the seasonal ILI pattern is visible with similar trends in both the sentinel provider and hospital data, starting late in week 50 and peaking in week 6, but in the clinic data seasonal ILI is much less evident and seems to start later, showing up as a slight increase starting around week 2 and peaking in week 5. Further, note that the statewide seasonal ILI pattern peaks from weeks 5 to 9 while the Monterey County hospital ERs and clinics have a much sharper peak around week 5 or 6 after which the ILI incidence decreases substantially. This pattern is consistent with the fact that the California Sentinel Provider data are for the whole state, where the longer peak likely reflects the outbreak occurring in different times and parts of the state. Monterey County is a small geographic location, and it appears the ILI had peaked early on in this location compared to the entire state.

As for the second outbreak period in Figure 2 (the “2009 H1N1 1st Wave”), there is consistency across all three time series, with the 2009 H1N1 pandemic starting in week 14 and peaking in week 18 of 2009. Subsequent to the first wave, a second wave of the H1N1 pandemic (“2009 H1N1 2nd Wave” in the figure) may have started as early as week 24 and peaked somewhere between weeks 35 and 44. This is where there is a bit of divergence between the three time series. The hospital ERs show an initial spike at weeks 34 and 35 followed by a larger peak around week 42 and then a subsequent decline. The California Sentinel Provider data and the clinic data are consistent with this later peak around weeks 42 to 44, but they both show a more gradual increase to the peak and no spike at weeks 34 and 35. Perhaps the difference is that the 2009 H1N1 virus spread slightly differently in the population served by the Monterey County hospitals or perhaps during the weeks 34-35 peak people were more likely to go to the ER than the public clinics. These differences may also be due to how people use hospitals versus clinics and the severity of or worry about their symptoms. The laboratory-confirmed, hospitalized cases at the bottom show a combination of the two trends, where we see that weeks 26, 32 and 34 through 36 had spikes in cases, but the entire “2009 H1N1 2nd Wave” period shows substantial 2009 H1N1 activity.

What Figure 2 illustrates is that the outbreak indications are quite similar across the three populations. However, it is important to note that the clinic and hospital percentage of ILI are based on MCHD's restricted ILI syndrome definition. As such, it is not obvious that these observed trends will manifest in the same way in the raw daily count data and under the baseline and expanded ILI definitions. However, as shown in Figure 3, the dates do in fact match up fairly well. Thus, regardless of the definition used, Monterey County clinic ILI syndrome data followed the sentinel provider trends fairly closely. Thus, for the remainder of this paper, "ground truth" will be taken to be the three periods of rising counts shown via the shaded areas in Figures 2 and 3. These correspond to:

- Seasonal ILI outbreak period: 12/12/2008 (week 50) – 2/13/2009 (week 6),
- First 2009 H1N1 pandemic outbreak period: 4/6/2009 (week 14) – 5/8/2009 (week 18),
- Second 2009 H1N1 pandemic outbreak period: 6/15/2009 (week 24) – 11/6/2009 (week 44),

where by "outbreak period" we mean that period of time in which the syndrome counts were increasing from their nominal state up to some peak. We focus on this period as the point of EED is to identify the start of the outbreak as soon as possible.

3.2 Assessing Performance

To assess the performance of the algorithms for the various syndrome definitions, we ran the EARS methods and the three CUSUMs (as described in Section 2) on the daily counts derived for the baseline, expanded, and restricted ILI syndrome definitions and compared the resulting signals to the outbreak periods. In so doing, we followed current MCHD practice of not resetting the algorithms after each signal. See Fricker (2010a,b) for additional discussion about the pros and cons of such practice.

3.2.1 Using EARS Baseline ILI Definition

Figure 4 compares the results of the six EED algorithms under the CDC baseline ILI syndrome definition. As in Figure 3, the small circles on the graph are the aggregate daily ILI counts for Monterey County clinics and the

DOI: 10.2202/2151-7509.1018

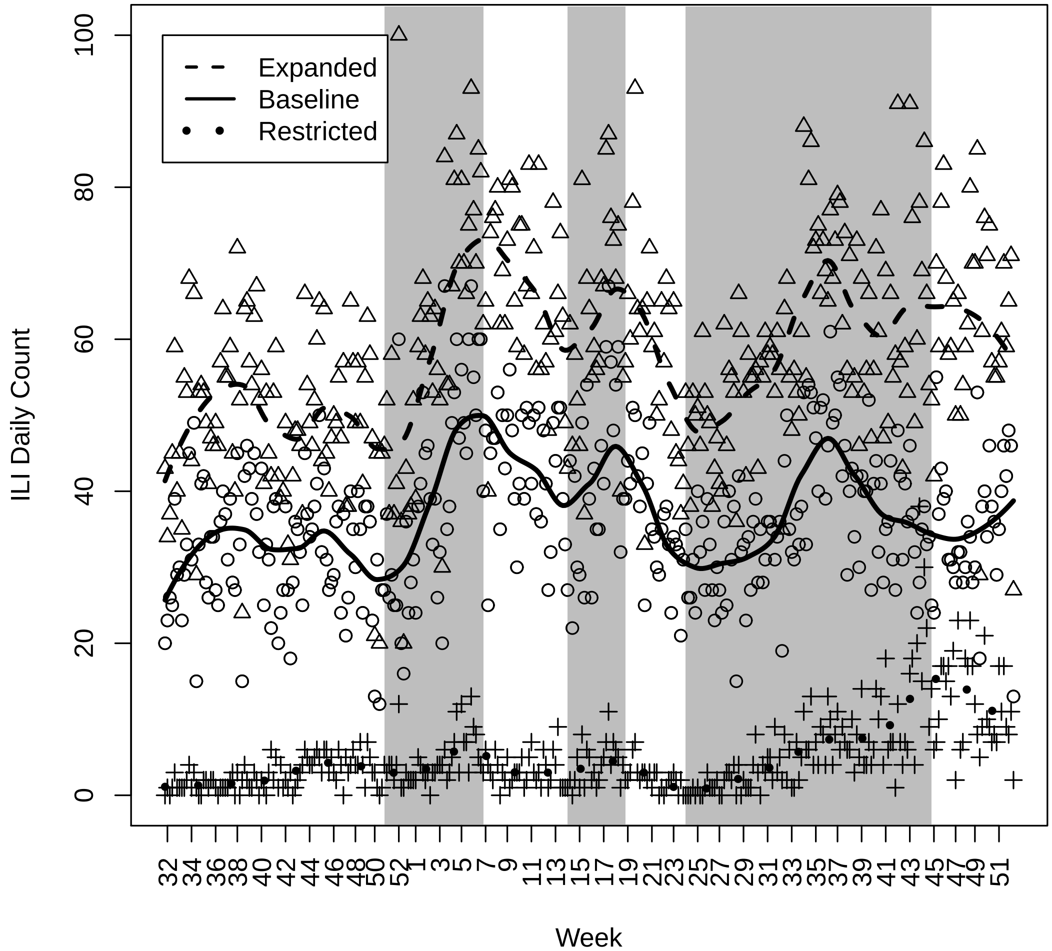

Figure 3: Comparison of the estimated ILI counts using the baseline, expanded, and restricted definitions. The shaded areas, which match those of Figure 2, show that the three outbreak periods are largely consistent across the different populations and ILI syndrome definitions.

black line is a locally-weighted smoothing line to show the underlying trends in the data. The shaded areas denote the three outbreak periods that were just defined: the seasonal ILI followed by the two 2009 H1N1 waves. Finally, at the top of the plot are the daily signals for the six detection algorithms. A signal on a particular day is denoted by a vertical line “|”, where heavier black bars simply indicate a sequence of daily signals.

Figure 4 shows that, right at the start of the seasonal ILI outbreak (i.e., at week 51 or December 15, 2008), all EED methods with the exception of CUSUM3 signaled (and, in fact, CUSUM1 signaled for three consecutive days). Subsequent to the initial signal, the EARS C1 and C3 methods each only signaled one additional time during the outbreak period. In comparison, the CUSUM methods continued to signal periodically throughout the outbreak period and in a manner consistent with their design. That is, CUSUM1 was designed to be the most sensitive, CUSUM2 less so, and CUSUM3 the least sensitive. Of course, this comes with the trade-off that the more sensitive the CUSUM the more it also signals in the non-outbreak periods as well.

For the first 2009 H1N1 outbreak period (e.g., weeks 14-19), none of the methods signalled at the outset of the outbreak period – though the fact that CUSUM1 signals two days prior and CUSUM2 signals three days prior might be an indication that the outbreak period started a few days earlier than the shading shows. What is clear is that the EARS C1 and C2 methods completely miss the outbreak while the C3 only signals once at the peak of the outbreak. In contrast, the CUSUM methods all signal more consistently and regularly and, with the exception of CUSUM3, earlier than C3. Finally, for the second 2009 H1N1 outbreak period (weeks 24-44), CUSUM1 signals right at the outset of the outbreak period with CUSUM2 and C3 following five and seven days later, respectively. However, C2 fails to signal at all while C1 takes 16 days to signal and CUSUM3 takes 22 days.

3.2.2 Using MCHD Expanded ILI Definition

Figure 5 compares the performance of the six EED algorithms under the MCHD expanded ILI syndrome definition. What is most striking in this plot is the complete lack of signals over all three outbreak periods for the C1 and C2 methods. In particular, note the large observation of $y = 100$ in week 52 (that occurred on December 22nd) where, for this particular day, the estimated standard deviations (S_1 in Equation 1 and S_3 in Equation 2) are so large that the resulting statistics are just below the signaling threshold. And, while the C3 method does signal for the 2009 H1N1 outbreak periods, the initial signals are 17 and 18 days after the start of the outbreak, respectively, which is more than a three week delay. The CUSUM methods seem to do better, though CUSUM1 and CUSUM2 each have delays of six days for the first 2009 H1N1 outbreak and 7 days for the second 2009 H1N1 outbreak, and the CUSUM3 does not perform any better than the C3 in terms of delay.

DOI: 10.2202/2151-7509.1018

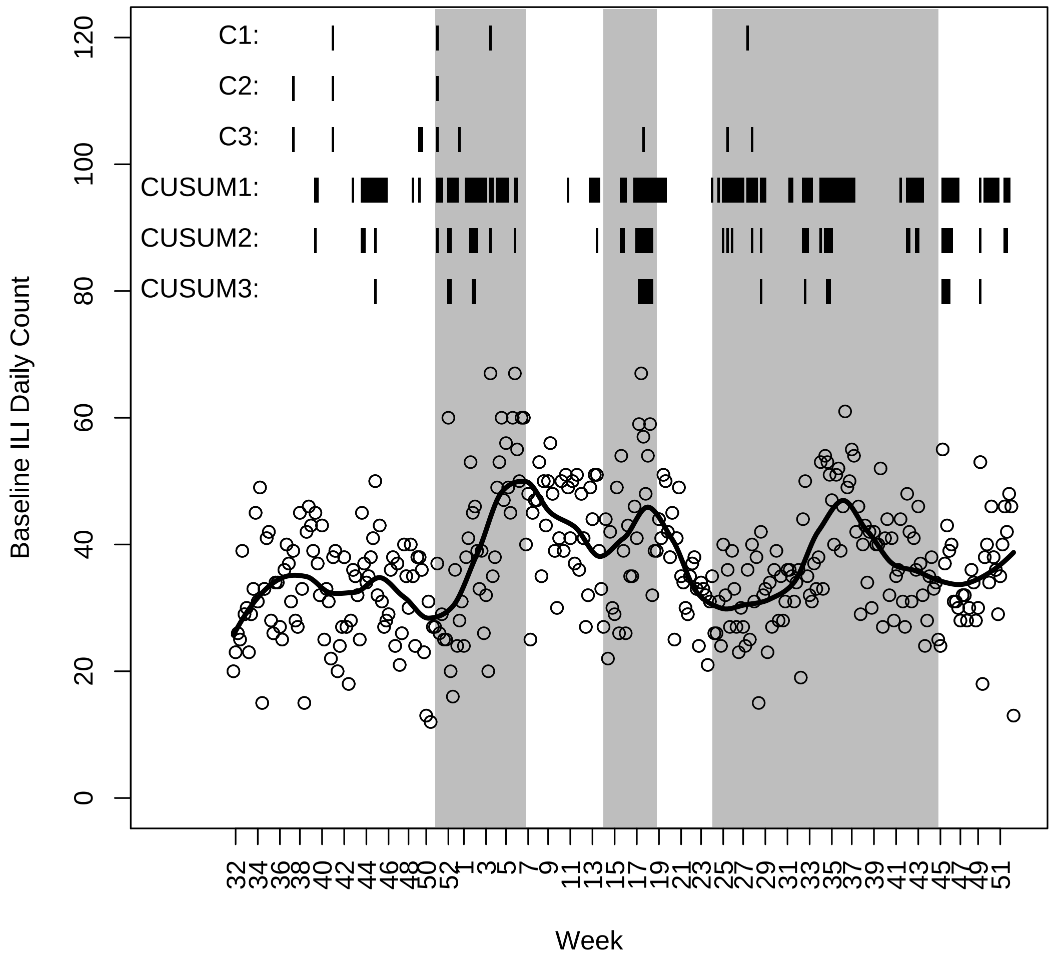

Figure 4: Algorithm signal times using the CDC baseline ILI syndrome definition. A signal on a particular day is denoted by a vertical line "|" and the heavier black bars indicate a sequence of daily signals. The circles are the aggregate daily ILI counts for Monterey County clinics (based on the CDC baseline ILI syndrome definition) and the black line is a locally-weighted smoothing line to show the underlying trends.

Thus, the most important result is that all of the methods perform substantially worse using the MCHD expanded ILI syndrome definition compared to the CDC baseline definition. This is surprising because, in implementing this definition, MCHD intended to make the EARS system more sensitive to detecting outbreaks yet, at least for these three outbreak periods, the expanded definition does just the opposite. The explanation for this outcome, which is clear in hindsight, is that the expanded definition introduced excessive noise into the data. That is, it classifies individuals with ILI who should not have been and thus masks the outbreak signals with noise. This introduction of noise is evident in Figure 3 where the MCHD expanded ILI syndrome curve essentially mirrors the CDC baseline curve, except it is shifted upwards.

3.2.3 Using MCHD Restricted ILI Definition

Figure 6 compares the performance of six EED algorithms under the MCHD restricted ILI syndrome definition. Here we see that the CUSUM methods perform better than the EARS methods using the other ILI definitions in the sense that they more regularly signal during the outbreak periods. Furthermore, many of the EARS and CUSUM methods' signals tend to align temporally suggesting that all the methods are detecting similar aberrations in the restricted data.

Comparing back to Figure 4, with the exception of CUSUM3, it appears that all of the methods are slower at detecting the seasonal ILI outbreak. However, this conclusion is confounded by the fact that the shaded area better corresponds to when the baseline and expanded data show an up-tick. The restricted data do not show an increase in ILI counts until week 52 or so, which is when the CUSUMs signal. Whether the outbreak actually began in week 50, 51, or 52 for the population served by the clinics is simply unknowable. It is clear, however, that the CUSUM methods signal the seasonal ILI earlier than the EARS methods.

3.2.4 Summarizing the Results

Visually, the restricted ILI syndrome definition seems to result in better algorithm performance, particularly for the EARS methods. To more formally and quantitatively compare between ILI syndrome definitions and detection algorithms, we define the following four metrics. First, we define *sensitivity* as the number of outbreak period days with a signal divided by the number of outbreak period days. That is,

$$\begin{aligned}\text{sensitivity} &= \frac{\text{\# outbreak period days with signal}}{\text{\# outbreak period days}} \\ &= \frac{\text{\# outbreak period days with signal}}{170}.\end{aligned}$$

DOI: 10.2202/2151-7509.1018

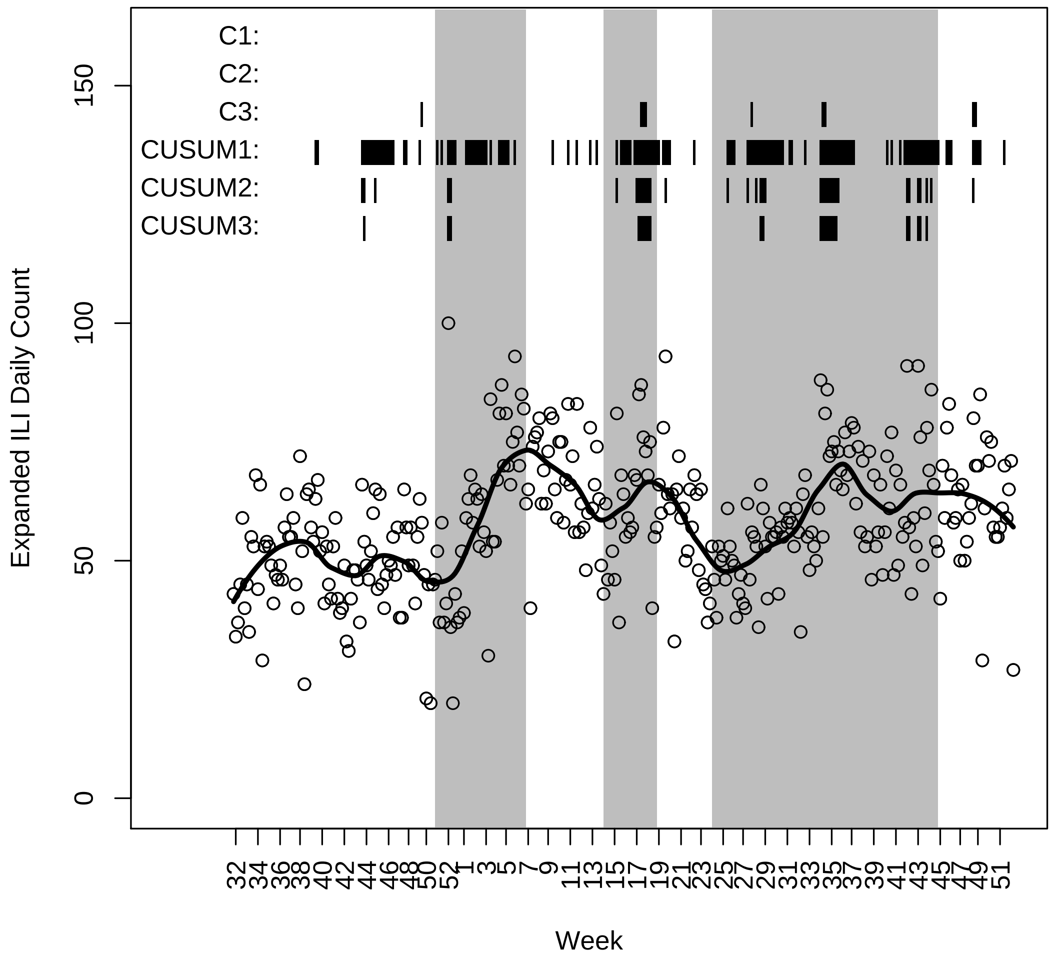

Figure 5: Algorithm signal times using the MCHD expanded ILI syndrome definition. A signal on a particular day is denoted by a vertical line "|" and the heavier black bars indicate a sequence of daily signals. The circles are the aggregate daily ILI counts for Monterey County clinics (based on the MCHD expanded ILI syndrome definition) and the black line is a locally-weighted smoothing line to show the underlying trends.

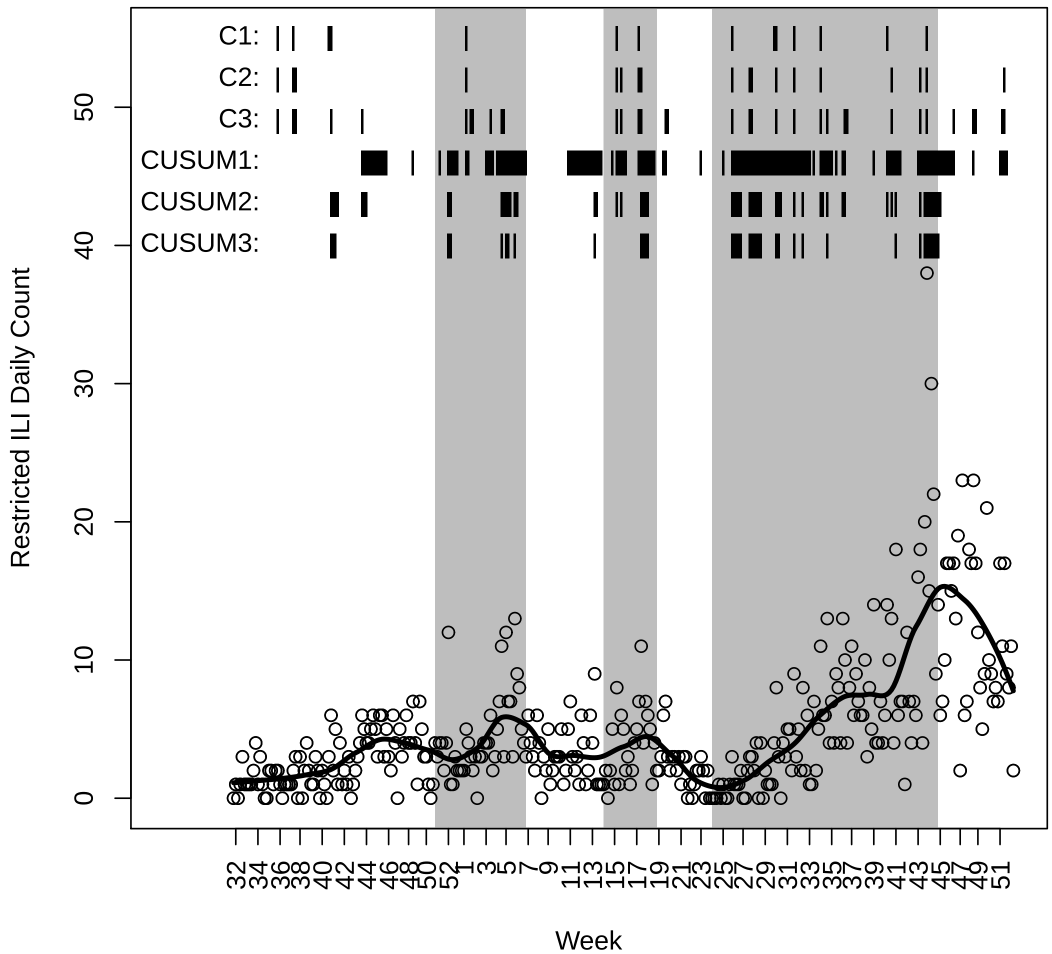

Figure 6: Algorithm signal times using the MCHD restricted ILI syndrome definition. A signal on a particular day is denoted by a vertical line "|" and the heavier black bars indicate a sequence of daily signals. The circles are the aggregate daily ILI counts for Monterey County clinics (based on the MCHD restricted ILI syndrome definition) and the black line is a locally-weighted smoothing line to show the underlying trends.

Second, we define *specificity* as the number of non-outbreak period days without a signal divided by the number of non-outbreak period days. That is,

$$\begin{aligned}\text{specificity} &= \frac{\#\text{ non-outbreak period days without signal}}{\#\text{ non-outbreak period days}} \\ &= \frac{\#\text{ non-outbreak period days without signal}}{183}.\end{aligned}$$

Note that we have specifically defined sensitivity and specificity in terms of the outbreak periods, which are those periods in which syndrome counts are increasing from their nominal levels. As such, these metrics are intended to measure how well the EED methods signal during those periods of time when an outbreak has started and is increasing.

Third, we define the *average delay*, denoted $\bar{d}_1$, as the average time it takes an algorithm to signal from the start of the outbreak period, where for a perfect algorithm that signalled on the first day of all three outbreak periods $\bar{d}_1 = 0$. Finally, we define the *average delay from first signal*, denoted $\bar{d}_2$, as the average time it takes an algorithm to signal from the time of the earliest signal of all six algorithms within a given outbreak period, where if an algorithm consistently signalled first for all three outbreak periods then $\bar{d}_2 = 0$.

Given these metrics, Table 4 quantifies the performance of the six EED algorithms under three ILI definitions. Note that the "+" sign after some of the average delay measures in the table indicates that the algorithm failed to signal during one or more outbreak periods. When this happened, the length of the outbreak period was used in place of the (nonexistent) delay and hence the average delay shown is an underestimate of the actual average delay.

Table 4 clearly demonstrates the benefit of the MCHD restricted ILI definition. In particular, for the EARS methods it both improves sensitivity (at a very modest cost to specificity) and the delay times. Simply put, under the restricted ILI definition, the EARS methods had a higher probability of signalling during an outbreak period and they signalled faster. The restricted ILI definition also improved the performance of the CUSUM methods, again increasing sensitivity and generally decreasing the delay.

	CDC Baseline				MCHD Expanded				MCHD Restricted			
Algorithm	Sens.	Spec.	$\bar{d}_1$	$\bar{d}_2$	Sens.	Spec.	$\bar{d}_1$	$\bar{d}_2$	Sens.	Spec.	$\bar{d}_1$	$\bar{d}_2$
C1	0.02	0.99	14+	11+	0.00	1.00	57+	52+	0.06	0.98	9.7	6.0
C2	0.01	0.99	43+	40+	0.00	1.00	57+	52+	0.08	0.98	9.7	6.0
C3	0.03	0.98	8.7	5.7	0.04	0.98	26+	21+	0.13	0.93	9.7	6.0
CUSUM1	0.55	0.75	3.0	0.0	0.58	0.77	4.7	0.0	0.62	0.76	3.7	0.0
CUSUM2	0.21	0.93	4.7	1.7	0.18	0.97	6.3	1.7	0.28	0.95	7.0	3.3
CUSUM3	0.09	0.97	14.7	11.7	0.14	0.99	14.7	10.0	0.21	0.98	10.7	7.0

Table 4: Performance of the six EED algorithms under the three ILI syndrome definitions. The "+" sign after some of the average delay measures means the algorithm failed to signal during one or more outbreak periods. When this happened, the length of the outbreak period was used in place of the delay and hence the average delay shown is an underestimate of the actual average delay.

When comparing between the EARS and CUSUM methods, the CUSUM is clearly superior in this application. However, this should not be surprising for a number of reasons. First, the CUSUM as implemented here has the advantage of using much more historical data than the EARS methods: 35 days versus 7 days. This gives the CUSUM more power to detect changes. Second, the CUSUM is inherently able to detect smaller (mean) changes than the EARS methods because the CUSUM is designed to accumulate evidence over time. In the statistical process control literature, this is a well-known property of the CUSUM when compared to Shewhart methods such as the EARS' C1 and C2. Third, EARS v4.5 is designed for a 7-day week and thus the lack of clinic weekend data actually inhibits EARS performance in this particular application. Finally, fourth, we allowed the CUSUM to have adjustable parameters (h and k), so we were able to in a sense fine tune the CUSUM to the conditions. In comparison, in EARS v4.5 the C1, C2, and C3 thresholds are fixed at values that are unlikely to be preferred under all conditions.

When comparing among the CUSUM methods, CUSUM1 clearly had the best performance in terms of average delay, followed by CUSUM2, and then CUSUM3. This is not surprising since that is how the CUSUMs were defined: aggressive, moderate, and routine. The speed-of-detection performance of the CUSUM1 does not come for free, however. The cost is in terms of the specificity, which characterizes the false signal rate. In particular, we see that CUSUM1 signals roughly one day out of every four when there is no outbreak.

This is likely to be unacceptably high. If so, then adjusting the CUSUM's parameters, such as with CUSUM2 and CUSUM3, can reduce the false signal rate, though this will come at the cost of additional delay and reduced sensitivity. For example, under the restricted ILI definition, switching from CUSUM1 to CUSUM2 will decrease the rate of false signals from 1 per 4 days to 1 per 20, but it will also add an additional three days of delay or so. Ultimately, these sorts of trade-offs should be made by the public health practitioner in the context of the public health threats being faced and the resources available to investigate biosurveillance signals.

4 Conclusions & Discussion

Biosurveillance systems have great promise as a public health tool for improving population health and well-being. They also have the potential to improve public health response to natural disease outbreaks and bioterrorism. However, continuing research is necessary to better understand how to most effectively design and employ them.

For example, as these results have shown, biosurveillance system early event detection performance can be improved with changes in syndrome definitions. This idea is simple: to the extent that noise can be eliminated from the data, it will be easier for detection algorithms to identify anomalies in the data. To date more research has been focused on developing complicated and sophisticated detection algorithms rather than improving the data upon which the algorithms are run. However, arguably, better data is the "low hanging fruit" with the potential to significantly improve biosurveillance performance. Greater emphasis should therefore be focused on improvements in the data: collection, management, text searching logic, syndrome definitions, etc. This is a non-trivial exercise, particularly for rarely occurring diseases and bioterrorism agents for which (thankfully) there are little to no data from which to assess detection performance.

In this research we had three clear outbreaks from which to assess performance, but that is not the usual case, and even in this research we could not be definitive about precise outbreak start times. Surprisingly, under the CDC baseline and MCHD expanded ILI definitions, EARS methods were of little to no value in signaling an outbreak. Furthermore, we note that while the MCHD restricted ILI definition performed well, at least in comparison to the baseline and expanded definitions, there could very well be other definitions that perform even better. This finding was enlightening for us and underscores the need for additional research into syndrome definitions.

This work has also clearly shown that there are alternatives to the EARS C1, C2, and C3 detection algorithms that have better performance characteristics. Ultimately, it was CUSUM1 that proved the most reliable at signaling alarms prior to and throughout the time when Monterey County was experiencing 2009 H1N1 cases. Whether the CDC incorporates a CUSUM-based method into EARS or not, it is clear that the design of the EARS algorithms for drop-in surveillance impedes their performance for routine surveillance when more historical data are available. For the EARS system, as well as other biosurveillance systems, the principle is simple: form should follow from function. Thus, for EARS routine surveillance implementations that have more historical data, future design modifications should allow local users to exploit all the information available in the data.

References

Burkom, H.S., Murphy, S.P., & Shmueli, G. 2006. Automated Time Series Forecasting for Biosurveillance. *Statistics in Medicine*, **26**, 4202–4218.

California Department of Public Health. 2010. Accessed at `www.cdph.ca.gov/programs/vrdl/Pages/CaliforniaSentinelProviderProgram.aspx`, on December 3, 2010.

CDC. 2006. *EARS v4.5 Users Guide*. Department of Health and Human Services, September 19, 2006.

CDC. 2007a. *BioSense*. Accessed at `www.cdc.gov/biosense` on April 30, 2007.

CDC. 2007b. *Early Aberration Reporting System*. Accessed at `www.bt.cdc.gov/surveillance/ears` on April 30, 2007.

CDC. 2010. **** New EARS Surveillance Tool Available ****. E-mail dated November 10, 2010 11:24 AM.

Fricker, R.D., Jr. 2007. Directionally Sensitive Multivariate Statistical Process Control Methods with Application to Syndromic Surveillance. *Advances in Disease Surveillance*, **3**(1), 1–17. Available on-line at `www.isdsjournal.org`.

Fricker, R.D., Jr. 2008. Syndromic Surveillance. *Pages 1743–1752 of:* Melnick, E., & Everitt, B. (eds), *Encyclopedia of Quantitative Risk Assessment*. John Wiley & Sons Ltd.

DOI: 10.2202/2151-7509.1018

Fricker, R.D., Jr. 2010a. *Introduction to Statistical Methods for Biosurveillance*. Cambridge University Press. Draft available online at http://faculty.nps.edu/rdfricke/OA4910.htm#book

Fricker, R.D., Jr. 2010b. Methodological Issues in Biosurveillance (with discussion and rejoinder). *Statistics in Medicine*, **30**, 403–441.

Fricker, R.D., Jr. 2011. Biosurveillance: Detecting, Tracking, and Mitigating the Effects of Natural Disease and Bioterrorism. *In:* Cochran, J.J. (ed), *Encyclopedia of Operations Research and Management Science*. John Wiley & Sons Ltd. Available on-line at http://onlinelibrary.wiley.com/book/10.1002/9780470400531.

Fricker, R.D., Jr., & Hanni, K. 2010. *Biosurveillance: Detecting, Tracking, and Mitigating the Effects of Natual Disease and Bioterrorism*. Online presentation, Military Operations Research Society, Monterey, CA, February 10, 2010. Available online at http://faculty.nps.edu/rdfricke/frickerpr.htm.

Fricker, R.D., Jr., & Rolka, H.R. 2006. Protecting Against Biological Terrorism: Statistical Issues in Electronic Biosurveillance. *Chance*, **19**, 4–13.

Fricker, R.D., Jr., Hegler, B.L., & Dunfee, D.A. 2008a. Comparing Biosurveillance Detection Methods: EARS' Versus a CUSUM-based Methodology. *Statistics in Medicine*, **27**, 3407–3429.

Fricker, R.D., Jr., Knitt, M.C., & Hu, C.X. 2008b. Directionally Sensitive MCUSUM and MEWMA Procedures with Application to Biosurveillance. *Quality Engineering*, **20**, 478–494.

Hagen, K.S. 2010. *Assessing the Effectiveness of the Early Aberration Reporting System (EARS) for Early Event Detection of the H1N1 (Swine Flu) Virus*. Naval Postgraduate School. Master's thesis.

Hanni, K.D. 2011. Comments on 'Some Methodological Issues in Biosurveillance'. *Statistics in Medicine*, **30**, 423–425.

Hawkins, D.M., & Olwell, D.H. 1998. *Cumulative Sum Charts and Charting for Quality Improvement*. Springer.

Hutwagner, L., Thompson, W., Seeman, G.M., & Treadwell, T. 2003. The Bioterrorism Preparedness and Response Early Aberration Reporting System (EARS). *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, **80**, 89i–96i.

Hutwagner, L.C., Browne, T., Seeman, G.M., & Fleischauer, A.T. 2005. Comparing Aberration Detection Methods with Simulated Data. *Emerging Infectious Diseases*, **11**, 314–316.

International Foundation for Functional Gastrointestinal Disorders. 2010. *Glossary.* Accessed at `www.iffgd.org/GIDisorders/glossary.html` on August 9, 2010.

Lorden, G. 1971. Procedures for Reacting to a Change in Distribution. *Annals of Mathematical Statistics*, **42**, 1897–1908.

Montgomery, D.C. 2009. *Introduction to Statistical Quality Control.* 6 edn. John Wiley & Sons.

Moustakides, G.V. 1986. Optimal Stopping Times for Detecting a Change in Distribution. *Annals of Statistics* **14**, 1379–1387.

Page, E.S. 1954. Continuous Inspection Schemes. *Biometrika*, **41**, 100–115.

Ritov, Y. 1990. Decision Theoretic Optimality of the CUSUM procedure. *The Annals of Statistics* **18**, 1464–1469.

Shewhart, W.A. 1931. *Economic Control of Quality of Manufactured Product.* D. van Nostrand Company, Inc.

Shmueli, G., & Burkom, H.S. 2010. Statistical Challenges Facing Early Outbreak Detection in Biosurveillance. *Technometrics*, **52**, 39–51.

Sosin, D.M. 2003. Syndromic Surveillance: The Case for Skillful Investment View. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, **1**, 247–253.

Tokars, J.I., Burkom, H., Xing, j., English, R., Bloom, S., Cox, K., & Pavlin, J.A. 2009. Enhancing Time-Series Detection Algorithms for Automated Biosurveillance. *Emerging Infeectious Diseases*, **15**, 533–539.

Toprani, A., Ratard, R., Straif-Bourgeois, S., Sokol, T., Averhoff, F., Brady, J., Staten, D., Sullivan, M., Brooks, J.T., Rowe, A.K., Johnson, K., Vranken, P., & Sergienko, E. 2006. Surveillance in Hurricane Evacuation Centers - Louisiana. *Morbidity and Mortality Weekly Report*, **55**, 32–35.

U.S. Government. 2007. *Homeland Security Presidential Directive 21: Public Health and Medical Preparedness.* Accessed on-line at `www.fas.org/irp/offdocs/ nspd/hspd-21.htm` on September 29, 2009.

Zhu, Y., Wang, W., Atrubin, D., & Wu, Y. 2005. Initial Evaluation of the Early Aberration Reporting System — Florida. *Morbidity and Mortality Weekly Report*, **54 (supplemental)**, 123–130.

DOI: 10.2202/2151-7509.1018