Methods for monitoring influenza surveillance data

Benjamin J. Cowling, PhD*,

Irene O. L. Wong, MPhil*,

Lai-Ming Ho, PhD*,

Steven Riley, DPhil*,

Gabriel M. Leung, MD*

*Department of Community Medicine and School of Public Health, University of Hong Kong.

Word count:

Abstract: 231 words

Main text: 3559 words

Corresponding author:

Dr Benjamin J Cowling

Department of Community Medicine, the University of Hong Kong

21 Sassoon Road, Pokfulam, Hong Kong

tel: +852 2819 9141; fax: +852 2855 9528;

e-mail: bcowling@hku.hk

Summary

Background: A variety of Serfling-type statistical algorithms requiring long series of historical data, exclusively from temperate climate zones, have been proposed for automated monitoring of influenza sentinel surveillance data. We evaluated three alternative statistical approaches where alert thresholds are based on recent data in both temperate and subtropical regions.

Methods: We compared time series, regression, and cumulative sum (CUSUM) models on empirical data from Hong Kong and the US using a composite index (range = 0-1) which consisted of the key outcomes of sensitivity, specificity, and time to detection (lag). The index was calculated based on alarms generated within the first 2 or 4 weeks of the peak season respectively.

Results: We found that the time series model was optimal in the Hong Kong setting, while both the time series and CUSUM models worked equally well on US data. For alarms generated within the first 2 weeks (4 weeks) of the peak season in Hong Kong, the maximum values of the index were: time series 0.77 (0.86); regression 0.75 (0.82); CUSUM 0.56 (0.75). In the US data the maximum values of the index were: time series 0.81 (0.95); regression 0.81 (0.91); CUSUM 0.90 (0.94).

Conclusions: Automated influenza surveillance methods based on short-term data, including time series and CUSUM models, can generate sensitive, specific and timely alerts, and can offer a useful alternative to Serfling-like methods that rely on long-term, historically-based thresholds.

Keywords: Influenza, public health, detection, population surveillance.

Sentinel practices have been deployed in influenza surveillance in some western countries for decades and more recently in many others mostly outside of the temperate climate zone. Hong Kong, an advanced economy geographically situated in the epicenter of the influenza basin in southern China, established a sentinel surveillance network for influenza-like-illness (ILI) in the late 1990s which began reporting in 1998. Like elsewhere, the peak influenza season is associated with higher health care utilization in Hong Kong.[1, 2] Thus it would be useful to have a valid and reliable way to alert the onset of the peak season to enhance case detection and diagnosis, and to allow timely initiation of precautionary measures in vulnerable populations such as the elderly.[3]

Recent developments in computer-assisted outbreak detection offer a range of approaches to infectious disease monitoring.[4-15] A widely-used approach is based on a seasonal regression model originally proposed by Serfling.[16] Under this model data from three or more previous years are used to calculate a time-varying threshold and an alert is generated if current data surpass the threshold. Similar approaches incorporating historical data have been used in the US,[4, 5] the UK,[6] France,[7] Australia[8] and the Netherlands.[9] Simpler related methods using the same fixed threshold throughout the year are sometimes used due to their ease in application.[13, 14]

An alternative set of approaches to surveillance have instead set thresholds based on short-term data from recent weeks. Briefly, surveillance data as a type of time series data can be monitored with specialized methods such as Box Jenkins models[17] and dynamic linear models,[18] which are both part of a wider class of state space models.[19] Reis et al.[15] describe a hybrid monitoring method which uses 'cuscore' statistics based on forecast errors from Box Jenkins time series

models. The use of cumulative sum (CUSUM) statistics, originally developed for industrial quality control,[20] is growing in popularity for automated surveillance. CUSUMs may incorporate historical information in the threshold calculation,[5, 10] or be based only on recent data.[11, 12]

In this paper, the three methods in this latter group of statistical approaches requiring only data from recent weeks to generate alerts are compared on influenza sentinel surveillance data from Hong Kong and the US. This may be particularly interesting for and applicable to surveillance networks in subtropical settings or with fewer numbers of reporting sentinels, where reported inter-epidemic levels of influenza-like-illness are subject to more fluctuation and peak seasons can be less distinct.

METHODS

Hong Kong Surveillance Data

The sentinel surveillance data from Hong Kong are provided by a network of general practitioners and published online.[21] At the end of each week the sentinel practitioners report the number of consultations with patients complaining of ILI symptoms (defined as fever plus cough or sore throat), and the total number of consultations. Data are collated and analyzed by the following Wednesday or Thursday, thus the reporting delay is approximately one week. The sentinel surveillance system was initiated in 1998 with 18 practitioners, and a further 42 sentinels were added throughout 1998 and 1999 to reach the current level of 50 practitioners covering a population of 6.8 million (0.74 per 100,000) giving weekly reports.

The start of the influenza peak season may be determined from laboratory data on influenza isolates. Each month a median of 1300 specimens (inter-quartile range 970-1850) were sent to the Government Virus Unit of the Department of Health primarily from hospitals, and the number of influenza A and B isolates were reported. We calculated the highest proportion of positive influenza isolations each season, and we define the onset of each peak season when the proportion of isolates positive for influenza surpassed 30% of the maximum seasonal level. To investigate the sensitivity of our results to this definition of the start of the peak season, we further adopted two alternative definitions that the onset of the peak season occurred when the proportion of influenza isolations passed 20% or 40% of the maximum seasonal level.

US Surveillance Data

Each week, approximately 1,000 sentinel health-care providers (0.38 per 100,000) from across the US report the total number of patients seen and the number of those patients with ILI (similarly defined as in Hong Kong). The sentinel data from 1997 onwards are reported online.[22] However other than in 2003 sentinel data were not available during the periods of low influenza activity from June to September each year. Careful analysis of the period June to September 2003 showed a low degree of homogeneous variation in ILI reports around a constant level, with no autocorrelation between successive weeks. Where data were missing in other years between June and September, simulated values were randomly generated from a Normal distribution with mean equal to the level of the observed data at the start of October that year and variance equal to that observed in the data from June to September 2003.

Weekly laboratory surveillance data are also published for the same period, and may be used as the gold standard measure of the onset of the peak influenza season each year as described above for the Hong Kong laboratory data.

Methods for Generating Early Alerts

We considered three alternative approaches to the early detection of the onset of the peak season, where no method required more than a maximum of 9 weeks baseline data. In the first two approaches an alert is generated only if the current observation falls outside a forecast interval calculated from previous weeks' data. The third approach uses CUSUMs.

The first approach is a time series technique, the dynamic linear model.[18] This model will have similar performance to Box Jenkins methods[19] and is typically simpler to implement since it does not require specialized statistical software and can be directly applied to raw data. The proposed model for the series of observations y_t is described by the equations

$$y_t = \theta_t + v_t \quad \text{where } v_t \sim N(0,V),$$

and $$\theta_t = \theta_{t-1} + w_t \text{ where } w_t \sim N(0,W),$$

where the series of unobserved system parameters θ_t describe the correlation between successive weeks. The errors v_t and w_t are internally and mutually independent, and the variance V can be estimated from the data. The parameter W must be prespecified, and represents the assumed smoothness of the changes in influenza prevalence in the community from week to week. In this model information from all preceding weeks is used to construct a 100(1-α)% forecast interval and if the data for the current week falls outside this forecast interval then an alert is raised.

Further details on the time series method, and an implementation in MS Excel, are available in the online supplementary materials.

The second approach is a simple regression model.[4] An alert is generated if data from the current week fall outside a 100(1-α)% forecast interval from a Normal distribution with 'running' mean $\tilde{y}_{(m)}$ and running sample variance $\tilde{s}^2_{(m)}$ calculated from the preceding *m* weeks. The forecast interval is calculated as $\tilde{y}_{(m)} \pm t_{m-1,1-\alpha/2}\tilde{s}_{(m)}\sqrt{1+1/m}$, where $t_{m-1,1-\alpha/2}$ is the 100(1-α)th percentile of Student's t-distribution with *m*-1 degrees of freedom.

The final approach is the CUSUM method. For the series of observations y_t, $t = 1,2,\ldots$, we define the *d*-week upper CUSUM at time *t*, C_t^+, as

$$C_t^+ = \max\left\{0, \frac{y_t - \tilde{y}_{(7)}}{\tilde{s}_{(7)}} - k + C_{t-1}^+\right\},$$

with $C_{t-d}^+ = 0$.[23] The running mean $\tilde{y}_{(7)}$ and running variance $\tilde{s}^2_{(7)}$ are calculated from the series of seven weeks $y_{i-d-7},\ldots,y_{i-d-1}$ preceding the most recent *d* weeks. The alarm is raised if the upper CUSUM C_t^+ exceeds a pre-specified threshold of $\Phi_{1-\alpha/2}$ for some α, where Φ is a standard Normal deviate (z value). The parameter *k* represents the minimum standardized difference from the running mean which is not ignored by the CUSUM calculation.

Metrics for Comparing the Methods

The performance of an early warning system can be measured by three relevant indices, namely the sensitivity, specificity, and timeliness of the alarms that are generated.[11, 24] Sensitivity is

defined as whether there was at least one alarm during the peak season.[11] Specificity is defined as 1-*r*/*n*, where *r* is the number of alarms outside the peak season periods (i.e. false alarms) and *n* is the total number of weeks outside the peak season periods.[11] Timeliness, or lag, is defined as the number of weeks between the beginning of the peak season and the first week that an alarm was raised.[11] The most desirable method will have maximum sensitivity and specificity, and minimum lag.

These three measures of sensitivity, specificity and timeliness may be combined in a single metric analogous to the area under the receiver operating characteristic (ROC) curve for sensitivity and specificity. By adding information on timeliness as a third dimension to the traditional ROC curve, the resulting volume under the ROC surface (VUTROCS) provides an overall measure of performance.[25] The VUTROCS for a particular method can be calculated as follows. Incorporating only alerts from the first week of the peak season, compute the sensitivity and specificity given a range of thresholds, and calculate the corresponding area under the ROC curve. Repeat this procedure up to the longest time for which an alarm is considered useful, and average the areas to estimate the VUTROCS. A higher VUTROCS would indicate superior performance, and the maximum VUTROCS of 1 would indicate that alarms are generated with perfect sensitivity and specificity, and at the soonest possible moment (i.e. the same week as the start of the peak season). Given the range of VUTROCS values is from 0 to 1, a difference of greater than 0.1 could be considered meaningful.

For sentinel surveillance, the usefulness of an early warning of the onset of the peak season is highly dependent on how early that warning is. In Hong Kong, the majority of excess hospital

admissions associated with the influenza peak season occur during the first 8-10 weeks of the peak season.[1] For an early warning to have any impact, it needs to be issued ideally in the first three weeks and at most within the first five weeks. Acknowledging a potential 1-week reporting delay due to data collection, collation and analysis, we consider two versions of the VUTROCS, the first where evaluation is limited to data on the first 2 weeks of the peak season, and the second where it is limited to the first 4 weeks.

The three early warning detection approaches were implemented on both sets of empirical data each with four choices of parameter combinations as given in Table 1. For each parameter combination, a set of 10 different thresholds were used in order to give broad coverage in terms of sensitivity, specificity and timeliness. These 10 resulting triplets could then be combined into a single VUTROCS for each of the parameter combinations. The sensitivity and timeliness of each method for a fixed specificity of 0.95, and the threshold required to obtain this specificity, were calculated by linear interpolation.

(Table 1 here)

All analyses were conducted in R version 2.1.0 [26].

RESULTS

The seven annual cycles of Hong Kong surveillance data are shown in Figure 1. The sentinel data typically varied around a level of 40 or 50 ILI diagnoses per 1000 consultations before the start of the peak season. After the peak season onset the sentinel data rapidly rose from baseline

levels to a maximum within 3-5 weeks. This was followed by a decline back to baseline levels after about three months. The eight annual cycles of sentinel data from the US in the years 1997-2005 are presented in Figure 2. Compared to epidemiologic patterns in the temperate climate zones, Hong Kong's seasonal swings were much less distinct where there appeared to be a much smaller but definite summer peak during some years and/or a long plateau between the winter and summer surges. In the years where the milder H1N1 strain dominated, peaks in the sentinel data were slightly smaller.

The results for the two alternative versions of the VUTROCS (2-week and 4-week evaluation) are summarized in Table 2. For the Hong Kong data, the time series method seems most optimal under both scenarios, and the CUSUM method performs particularly poorly. Conversely for the US data, there appears to be much less difference between the three methods. In particular the CUSUM method is optimal under the 2-week VUTROCS, while the performance of the time series and CUSUM methods is similar when measured by the 4-week VUTROCS.

(Table 2 here)

While the VUTROCS summarizes overall performance, table 3 shows the sensitivity and timeliness of each method and parameter combination for a fixed specificity of 0.95. The thresholds required to achieve this specificity are also presented in terms of the parameter α. For the Hong Kong data, the time series method with 88% forecast intervals (i.e. α=0.12) could achieve specificity of 0.95 with sensitivity 1 and timeliness around 1.4 to 1.5 weeks. The regression method could also achieve high sensitivity, but had worse timeliness. The best

parameter combination for the CUSUM method had sensitivity 0.86 and timeliness 1.91 weeks. For the US data, the time series method again had superior performance.

(Table 3 here)

In the sensitivity analyses (Figure 3), we found that our conclusions were unchanged by alternative definitions of the start of the peak season. In the Hong Kong data (Figures 3a and 3b), the time series method outperformed the other methods, although the difference was less under the stricter definition of the onset of the peak season requiring laboratory isolations above 40% of peak levels. In the US data (Figures 3c and 3d) there was little discernable difference between the three methods under comparison.

DISCUSSION

Our findings suggest that the time series approach is superior to the CUSUM method for the Hong Kong data with the regression model having intermediate performance, but that there is little difference between the time series and CUSUM methods on the US data while both outperformed the regression-based technique. This apparent divergence of results may be explained by further consideration of the underlying epidemiologic patterns of influenza and the characteristics of the sentinel systems. The Hong Kong data are provided by far fewer sentinels, in absolute terms, compared to the network in the US, and the Hong Kong data are noticeably more variable (Figure 1) than the US data (Figure 2). Secondly, the Hong Kong influenza peak typically arrives abruptly with the sentinel data rising from baseline levels to peak within a period of 3-5 weeks, whereas the American ILI activity typically rises more slowly and more

exponentially to the peak over 8-10 weeks. Thus for the latter, the CUSUM method is expectedly better at aggregating a number of initially smaller increases in the early weeks of the peak season to detect a significant change, whereas none of those smaller rises fall outside the forecast intervals of the time series and regression models. Furthermore the annual peaks in Hong Kong were typically sharper and more sudden than in the American data where peaks slowly emerged, perhaps because the former represents one city whereas the latter is a wide geographical area where the peak season may emerge earlier in some geographical areas than others while data are summarised across the whole country. The peaks in sentinel data in years dominated by the milder influenza A (H1N1) strain seemed slightly smaller; however there was too little data to investigate this thoroughly.

Of the two methods based on forecast intervals evaluated here, it is perhaps unsurprising that time series typically outperformed simple regression. This was most likely because the time series model was better at dynamically adapting to changes in the underlying level of reports, and unlike the regression method could exploit the correlation structure in previous reports. Nevertheless, the results suggest that the simpler regression approach can under most circumstances produce reasonably useful and timely alarms, albeit inferior to more sophisticated approaches. We note that the regression model performed better in the American setting than in Hong Kong, probably because the seasonal patterns in the American data are more clear, and the peak more distinct rather than the subtropical seasonality observed in Hong Kong.[27, 28]

It is important to recognize that the calibration of alarm thresholds depends on the inherent tradeoff between the cost of false alarms (specificity) and the expected benefit of earlier

detection (sensitivity and timeliness).[24] It is thus important to formally and explicitly calibrate the parameters which affect the decision limit. A typical choice for regression models is to use two standard deviations from the mean, corresponding to a 95% forecast interval,[8, 12] or three standard deviations, corresponding to a 99.9% forecast interval.[11] Our results suggest that for the time series method high specificity of 0.95 could be obtained by using 90% forecast intervals, and this would allow alarms to be generated after an average of 1.4 to 1.5 weeks in Hong Kong (table 3), corresponding to warnings within 2-3 weeks allowing for short reporting delays. In any novel application of these methods it would be important to appropriately calibrate the threshold since the sensitivity or specificity of a given threshold will vary in different settings.

In this study we have focused on methods which specify thresholds based on short-term data (i.e. from the current year only), rather than historical threshold methods such as Serfling. One reason that short-term methods may be more useful in Hong Kong is the larger degree of variation between weekly reports, and also between seasons (Figure 1). For example, the 'baseline' level of reports of 5-6% in May-December 1999 was barely exceeded by the levels of reports during the peak seasons of 2001, 2003 and 2004. A recent study by the CDC[11] suggested that regression and CUSUM approaches based on short-term data may be equivalent or superior to regression approaches based on historical data across a wide range of syndromes with daily reporting. However, this study did not include other statistical approaches (e.g. time series), and furthermore daily reported data can have different statistical properties to weekly reported data: in particular there is likely to be higher variance and higher autocorrelation in daily series.

We further note the difference in objective between prospective surveillance, studied here, and the quantification of excess mortality due to influenza for which purpose the Serfling method was originally designed. A prospective surveillance algorithm should be highly sensitive and specific, that is it should track the data before the peak season while generating very few false alarms, and be able to quickly generate an alarm when the peak season starts, therefore interest is primarily in the performance of forecast intervals. Whereas quantifying the excess mortality due to influenza requires methods which can retrospectively fit seasonality in observed data, and interest is primarily in the residuals, or the difference between observed data and the mean levels predicted by the model. Recent estimates of the number of excess influenza-associated hospitalisations[28] and deaths[27] in Hong Kong used Poisson regression methods allowing for historical trends, where Poisson regression was used in preference to multiple linear (Normal) regression due to small event counts.

A potential caveat of our analysis is the small number of annual cycles of sentinel data available for study. However with each cycle taking a year to generate it will be many years before a larger dataset is available. In the meantime sentinel systems are being introduced in more countries, while there is a lack of evidence-based information on how best to generate timely and reliable intelligence on which to base public health policy. A further limitation is that there remains no agreed gold standard measure of when the peak influenza season starts every year. In this paper we have used laboratory data which should provide a reliable measure of when influenza begins to circulate in the community, and defined the onset of the peak season when laboratory levels surpassed 30% (20%-40%) of peak levels. An alternative choice of gold standard would have been to use mortality data or hospital admission data to define the onset of

the peak season. But coding difficulties and misclassification bias often make these data even more unreliable, and moreover, there is an unspecified lead time lag between the onset of the influenza season and when one would expect to observe corresponding increases in morbidity and mortality. Lastly, there is some evidence to suggest that influenza-associated hospitalization and deaths are seriously under-coded in Hong Kong.[27] However it is possible that a useful sentinel surveillance system could be implemented based on rapid reporting of chief complaints in accident and emergency departments, or at the time of admission. One final caution is that we have evaluated specific methods with only a few chosen parameter combinations. However we have tried to find a balance between investigating a range of practical parameter combinations without overfitting the models to the data. Given the broadly similar results for the different parameter combinations within each method, we believe that our conclusions are robust to a range of parameter sets.

In conclusion, this study has described three different methods for automated monitoring of surveillance data, including a time series approach which has not previously appeared in the surveillance literature. These results may be useful to other subtropical countries with varying levels of influenza activity outside the peak seasons, or for developing surveillance systems with fewer sentinels. We should compare results on data from these places to confirm the generalizability of our findings. As few as 10 sentinels could potentially provide useful data on trends in influenza incidence. If data were collected within say a week, and our results were applicable, the methods outlined here could be utilised to detect the onset of an annual peak season within 2-3 weeks. Dissemination of alerts could facilitate enhanced case detection and diagnosis, and could allow timely initiation of precautionary measures in vulnerable populations.

Acknowledgements

We gratefully acknowledge the sentinel practitioners who through their own goodwill have been providing weekly data to the Hong Kong Centre for Health Protection, for the purpose of infectious disease surveillance. We acknowledge the sentinel practitioners in the US who provide data to the Center for Disease Control and Prevention. This research was in part funded by the Research Fund for the Control of Infectious Diseases of the Health, Welfare and Food Bureau of the Hong Kong SAR Government (grant no. 04050102). This research was also supported by the University of Hong Kong Vice Chancellor's Development Fund in influenza research. We thank two anonymous referees for their helpful comments.

Key messages

In settings where sentinel networks are established to detect the start of the annual influenza peak season, it is important to use appropriate methodology to detect significant increases in disease incidence. Acknowledging that such networks may not have long series of historical data, we investigate the performance of methods where specification of alert thresholds only requires recent data. We find that for weekly surveillance data, such methods can generate sensitive, specific and timely alerts.

References

[1] Yap FHY, Ho PL, Lam KF, Chan PKS, Cheng YH, Peiris JSM. Excess hospital admissions for pneumonia, chronic obstructive pulmonary disease, and heart failure during influenza seasons in Hong Kong. *J Med Virol*. 2004;73:617-23.

[2] Chiu SS, L. LY, H. CK, Wong WHS, Peiris JSM. Influenza-related hospitalisations among children in Hong Kong. *N Engl J Med*. 2002;347(26):2097-103.

[3] Quigley C, Sopwith W, Ashton M. How to deal with influenza: Worthwhile surveillance system is in action. *BMJ*. 2004 November 20, 2004;329(7476):1239.

[4] Stroup DF, Williamson GD, Herndon JL, Karon JM. Detection of aberrations in the occurrence of notifiable diseases surveillance data. *Stat Med*. 1989;8(3):323-9; discussion 31-2.

[5] Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting Salmonella outbreaks. *Emerg Infect Dis*. 1997;3(3):395-400.

[6] Farrington CP, Andrews NJ, Beale AD, Catchpole MA. A statistical algorithm for the early detection of outbreaks of infectious disease. *J Roy Statist Soc, Series A*. 1996;159(3):547-63.

[7] Costagliola D, Flahault A, Galinec D, Garnerin P, Menares J, Valleron AJ. A routine tool for detection and assessment of epidemics of influenza-like syndromes in France. *Am J Public Health*. 1991;81(1):97-9.

[8] Stern L, Lightfoot D. Automated outbreak detection: a quantitative retrospective analysis. *Epidemiol Infect*. 1999;122(1):103-10.

[9] Widdowson MA, Bosman A, van Straten E *et al.* Automated, laboratory-based system using the Internet for disease outbreak detection, the Netherlands. *Emerg Infect Dis*. 2003;9(9):1046-52.

[10] O'Brien SJ, Christie P. Do CuSums have a role in routine communicable disease surveillance? *Public Health*. 1997;111(4):255-8.

[11] Hutwagner LC, Browne T, Seeman GM, Fleischauer AT. Comparison of aberration detection methods using simulated data. *Emerg Infect Dis*. 2005;11(2):314-6.

[12] Hutwagner LC, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). *J Urban Health*. 2003;80(2 Suppl 1):i89-96.

[13] Hashimoto S, Murakami Y, Taniguchi K, Nagai M. Detection of epidemics in their early stage through infectious disease surveillance. *Int J Epidemiol*. 2000;29(5):905-10.

[14] Watts CG, Andrews RM, Druce JD, Kelly HA. Establishing thresholds for influenza surveillance in Victoria. *Aust N Z J Public Health*. 2003;27(4):409-12.

[15] Reis BY, Pagano M, Mandl KD. Using temporal context to improve biosurveillance. *Proc Natl Acad Sci U S A*. 2003;100(4):1961-5.

[16] Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Rep*. 1963;78:494-506.

[17] Chatfield C. *Time-series forecasting*. Boca Raton: Chapman & Hall; 2001.

[18] West M, Harrison J. *Bayesian forecasting and dynamic models*. 2nd ed. New York: Springer; 1997.

[19] Durbin J, Koopman SJ. *Time series analysis by state space methods*. New York: Oxford University Press; 2001.

[20] Box GEP, Luceno A. *Statistical control by monitoring and feedback adjustment*. New York: Wiley; 1997.

[21] Centre for Health Protection. Sentinel Surveillance. [cited 2005 1 October]; Available from: http://www.chp.gov.hk/dns_submenu.asp?id=44&pid=26&lang=en

[22] Centers for Disease Control and Prevention. United States Surveillance Data: 1997-1998 through 2002-2003 Seasons. [cited 2005 1 October]; Available from: http://www.cdc.gov/flu/weekly/ussurvdata.htm

[23] Montgomery DC. *Introduction to statistical quality control*. Hoboken: John Wiley & Sons; 2005.

[24] Mandl KD, Overhage JM, Wagner MM *et al.* Implementing syndromic surveillance: a practical guide informed by the early experience. *J Am Med Inform Assoc*. 2004;11(2):141-50.

[25] Kleinman K, Abrams A. Comparing aberration detection systems: metrics for evaluation through simulation. Program and abstracts of the fifth International, Interdisciplinary Conference on Geomedical Systems (GEOMED); 2005; Cambridge, UK; 2005.

[26] R Development Core Team. *R: A language and environment for statistical computing*: R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org; 2005.

[27] Wong CM, Chan KP, Hedley AJ, Peiris JS. Influenza-associated mortality in Hong Kong. *Clin Infect Dis*. 2004 Dec 1;39(11):1611-7.

[28] Wong CM, Yang L, Chan KP *et al.* Influenza-associated hospitalization in a subtropical city. *PLoS Med*. 2006;3(4):e121.

Table 1: Choice of methods and parameter combinations

Method	Parameter	Description	Range
Time series	*W*	Represents the assumed smoothness of the underlying system	0.025, 0.050, 0.075 or 0.100
Regression	*m*	Represents the number of prior weeks used to calculate the 'running' mean and variance	3, 5, 7, 9
CUSUM	*d*	Represents the number of days to sum over	2, 3
	k	Represents the minimum standardized difference which must be exceeded for a data point to be included in the CUSUM calculation	1, 2

Footnote: Each method also has a parameter α which defines the height of the threshold (a higher value of α would indicate a stricter threshold).

Table 2: Performance of the time series, regression and CUSUM methods.

Method	Parameter Combinations	Hong Kong Data		US Data	
		2-week VUTROCS[a]	4-week VUTROCS[b]	2-week VUTROCS[a]	4-week VUTROCS[b]
Time series	W=0.100 (least smooth)	0.77	0.82	0.81	0.95
Time series	W=0.075	0.77	0.86	0.81	0.93
Time series	W=0.050	0.77	0.86	0.80	0.92
Time series	W=0.025 (most smooth)	0.71	0.83	0.71	0.88
Regression	3-day running mean	0.69	0.78	0.75	0.84
Regression	5-day running mean	0.75	0.82	0.76	0.86
Regression	7-day running mean	0.70	0.80	0.78	0.89
Regression	9-day running mean	0.68	0.79	0.81	0.91
CUSUM	k=1, 2-week	0.56	0.75	0.80	0.90

	CUSUM				
CUSUM	k=2, 2-week	0.52	0.75	0.85	0.91
	CUSUM				
CUSUM	k=1, 3-week	0.55	0.73	0.83	0.93
	CUSUM				
CUSUM	k=2, 3-week	0.52	0.74	0.90	0.94
	CUSUM				

[a] The 2-week VUTROCS evaluates the volume under the ROC surface (higher being better) of a method for producing an alarm more quickly after the start of a peak season and at most within 2 weeks, weighed against the false positive rate of the method.

[b] The 4-week VUTROCS evaluates the volume under the ROC surface (higher being better) of a method for producing an alarm more quickly after the start of a peak season and at most within 4 weeks, weighed against the false positive rate of the method.

Table 3: Sensitivity and timeliness of time series, regression and CUSUM methods for fixed specificity of 0.95.

Method	Parameter Combinations	Hong Kong Data			US Data		
		α^a	Sensitivity	Timeliness (weeks)	α^a	Sensitivity	Timeliness (weeks)
Time series	W=0.100 (least smooth)	0.11	1.00	1.56	0.11	1.00	0.75
Time series	W=0.075	0.12	1.00	1.40	0.13	1.00	0.88
Time series	W=0.050	0.12	1.00	1.40	0.13	1.00	0.83
Time series	W=0.025 (most smooth)	0.12	1.00	1.52	0.14	1.00	0.96
Regression	3-day running mean	0.07	0.57	2.60	0.05	0.52	2.06
Regression	5-day running mean	0.10	1.00	1.72	0.03	0.57	2.11
Regression	7-day	0.09	0.95	1.82	0.02	0.65	1.83

	running mean						
Regression	9-day	0.09	0.97	1.65	0.02	0.90	1.45
	running mean						
CUSUM	k=1, 2-week	0.02	0.86	2.00	0.01	0.89	1.16
	CUSUM						
CUSUM	k=2, 2-week	0.28	0.86	1.91	0.02	0.88	1.25
	CUSUM						
CUSUM	k=1, 3-week	0.01	0.74	2.59	0.01	0.86	1.53
	CUSUM						
CUSUM	k=2, 3-week	0.19	0.85	2.00	0.01	0.82	1.51
	CUSUM						

[a] α represents the threshold required for each method to give specificity of 0.95.

Figure 1. Seven annual cycles (unbroken line) of sentinel surveillance data from Hong Kong, 1998-2005. The monthly proportions of laboratory samples testing positive for influenza isolates are overlaid as gray bars and the beginning of each peak season (inferred from the laboratory data) is marked with a vertical dotted line. The primary circulating subtype of influenza A is indicated above each peak.

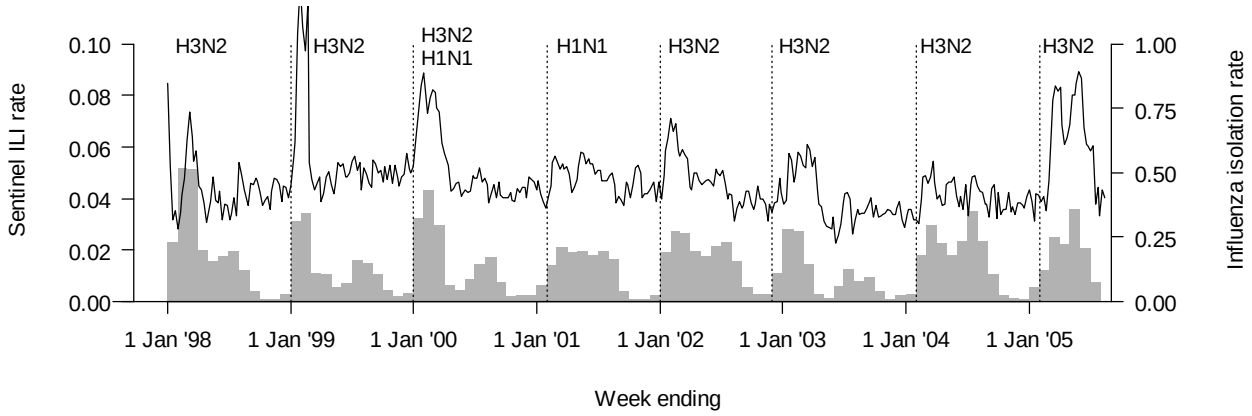

Figure 2. Eight annual cycles (unbroken line) of sentinel surveillance data from the US, 1997-2005, including simulated data (dashed lines) between June and September for all cycles except 2003-4, based on the empirical data for June to September 2003. The weekly proportions of laboratory samples testing positive for influenza isolates are overlaid as gray bars and the beginning of each peak season (inferred from the laboratory data) is marked with a vertical dotted line. The primary circulating subtype of influenza A is indicated above each peak.

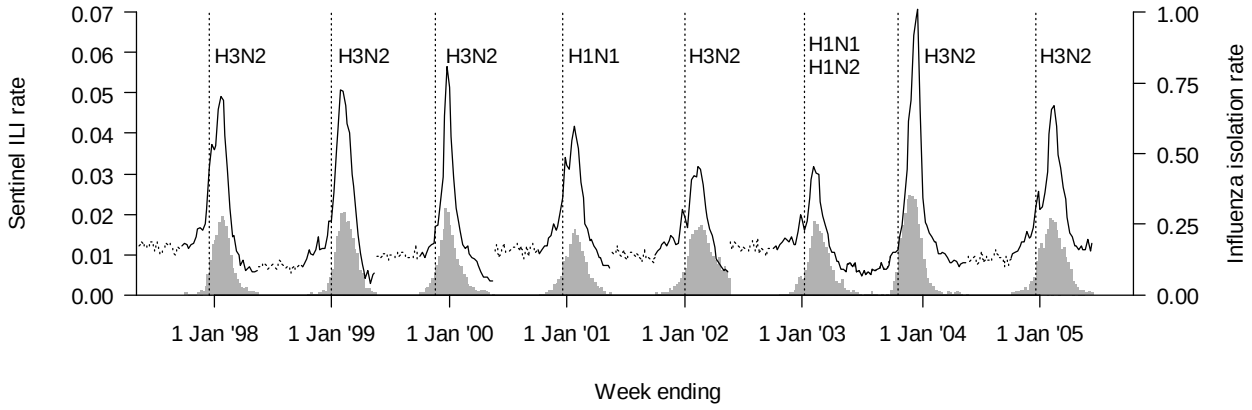

Figure 3. Sensitivity analysis. Each plot shows the estimated volume under the ROC surface for alternative definitions of the influenza peak season at 20% of seasonal peak levels or 40% of seasonal peak levels for the time series (black squares), regression (open circles) and CUSUM (black triangle) methods. (a) 2-week VUTROCS, Hong Kong data; (b) 4-week VUTROCS, Hong Kong data; (c) 2-week VUTROCS, US data; (d) 4-week VUTROCS, US data.