Reviewer's report

Title: Automated Real Time Constant-Specificity Surveillance for Outbreaks

Version: 1 Date: 14 February 2007

Reviewer: David Muscatello

Reviewer's report:

General

This is an interesting, well written paper that highlights an important dimension of disease surveillance using routinely collected health data sources: maintaining specificity to events of public health importance in an adaptive, prospective surveillance system.

The variable specificity reported for the various surveillance methods over time is very valuable information and probably not well understood by many surveillance personnel.

A very rigorous and reasoned approach has been taken to evaluate the methods.

-----------------------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. BACKGROUND Page 2
   It would be useful in the introduction to remind the reader who is unfamiliar with statistical process control of the relationship between the false alarm rate and specificity.

2. Sensitivity is another important dimension of outbreak detection which isn't mentioned. What assumptions about sensitivity are being made in this study? Maintaining specificity may be of no value if sensitivity is degraded. This is referred to later but may be useful mentioning up front.

Methods Page 3

3. This study uses count data, but all methods evaluated appear to assume that all models have a normal error distribution. There is no discussion of whether it is valid to make this assumption with the count data used. The Poisson assumption is a more natural choice for count data.

4. The Serfling model also assumes a normal distribution of model residuals, but given it's long term acceptance in the public health community, at least in the US, it's inclusion is well justified.

Page 4

5. My understanding is that ARMA models have the same difficulty with count data. It's unclear what the distribution is of the errors referred to in this model.

6. Please explain the choice of 32 days in the wavelet model for low frequency variation.

Page 5

7. There is a difficulty with the GAM approach described. If the 'expectation GAM' was an adequate model, then the model residuals would follow a normal distribution with constant variance, if a normal error distribution was assumed in the model. The need for the 'variance GAM' appears to arise because there is a poor model fit and there is unmodelled variation remaining in the residuals. The need for inclusion of linear time, day of year, and day-of-week variables in the 'variance model'highlights the ineffectiveness of including these variables in the original 'expectation GAM'. The only reason I can see for including the second GAM is because the raw data is difficult to fit using any common model, especially when the model is required to be prospectively adaptive - if so, this problem needs to be explicitly stated and the need for re-inclusion of the same variables in the second model needs to be better justified.

8. Why not apply a second expectation model to the other methods to ensure constant specificity? There may be computational advantages in using simpler models to achieve the same result? Why is this approach only suited to the GAM model? Was the GAM model robust - ie did it converge and produce
meaningful results over the whole period of analysis?

Page 6
9. An unstated assumption in determining the threshold lambda for providing a certain specificity from the training data is that there are no outbreaks in the training period. To calculate the false alarm rate using the method described, all alarms must be false. Is this a reasonable assumption? Seasonal influenza outbreaks occur to varying degrees from year to year and Serfling's method is commonly used in the US to identify these outbreaks from background mortality data. These outbreaks may also cause aberrations in the syndromic data used in this study. Are these outbreaks of interest to surveillance or are these to be excluded by design? Please be clear about what assumptions are being made and the limitations of those assumptions.

Simulated outbreaks Page 6/7
10. While simulation provides a flexible means of evaluating surveillance methods, these simulated outbreaks seem quite arbitrarily chosen - is there a public health justification for these choices? The relative impact of these choices will depend very much on the size and characteristics of the population under surveillance and the size of the background surveillance counts.

Results Page 7
11. It would seem natural to include a time series of the raw counts used in this study - to give a sense of the size and complexity of the data being modelled.

Page 10 Discussion
12. If I understand correctly, with 97% specificity public health users of the surveillance system would have to put up with 3 false alarms every 100 days, or approximately one false alarm each month. The importance of the choice of specificity on the costs associated with responding to false alarms should also be mentioned.

-------------------------------------------------------------------------------------------------------------------------
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)
-------------------------------------------------------------------------------------------------------------------------
Discretionary Revisions (which the author can choose to ignore)

13. Page 3: For readers less familiar with this topic, it would be helpful to explain why day of week and day of year are important features of the models, and that sine and cosine terms capture predictable seasonal patterns in the time series.


15. Page 5: It would be helpful to explain why the specific smoothers were chosen for each variable in the GAM model.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I declare that I have no competing interests