Author’s response to reviews

Title: Automated Real Time Constant-Specificity Surveillance for Disease Outbreaks

Authors:

Shannon C. Wieland (shann@mit.edu)
John S. Brownstein (john.brownstein@childrens.harvard.edu)
Bonnie Berger (bab@mit.edu)
Kenneth D. Mandl (kenneth.mandl@childrens.harvard.edu)

Version: 2 Date: 4 May 2007

Author’s response to reviews:

Dear Editor,

We thank both reviewers for their detailed and thoughtful analysis of our manuscript, and we feel that their comments have led to significant improvements. Please find their comments and our responses below.

First Reviewer’s Major Compulsory Revisions:

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.

On page 2, we have included the mathematical definition of the specificity, and we have added an equation showing its relationship to the false alarm rate.

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.

We have added a paragraph about sensitivity on page 2, explaining that sensitivity depends on the both the outbreak and the specificity. We mention that there is a trade-off between sensitivity and specificity, which must be considered in the context of the outbreak of interest.

3. Methods Page 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. The Serfling model also assumes a normal distribution of model residuals, but given its long term acceptance in the public health community, at least in the US, its inclusion is well justified...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.

To address this point, we have added a generalized linear model that uses a Poisson distribution function to all the analyses. We devote part of the discussion to the assumption implicit in the autoregressive, Serfling, trimmed seasonal and wavelet models that the errors are normally distributed.

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

We have clarified that the choice of 32 days was not ours but derived from the method described in Zhang et al. We have also added a brief explanation of the mathematical expedience of that choice.

5. 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.

As the reviewer points out, the raw data were hard to fit using common models including a Poisson model. Instead, we modeled both the expectation and the variance as functions of trend, day of year, and day of week. The full model for the number of visits is a Gaussian that incorporates these models, an approach that has been used before with generalized additive models. We have explained the model more fully in the methods section by adding additional text and equations, and we have also added citations to previous work.

6. 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?

We wished to show that existing methods, as they are reported in the literature, do not have constant specificity. However, it may indeed be possible to apply a model of the variance to these methods to obtain constant specificity. We did not test this approach to see if it leads to constant specificity, but we have added a paragraph in the discussion section suggesting this alternative.

7. Was the GAM model robust -- i.e. did it converge and produce meaningful results over the whole period of analysis?

We now mention in the description of the expectation-variance model that convergence of the backfitting procedure is required, and we include a reference to Hastie and Tibshirani, in which the procedure is described. We also mention in the results section that the model always converged, that the predictions for the expectation and variance were positive throughout the study period, and that the predicted visits were close to the actual visits.

8. 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 U.S. 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.

We now mention in the discussion that outbreaks of diseases of interest are not present in the training set. In this particular case, we are not interested in endemic or seasonal infectious diseases like flu and norovirus; instead we are concerned with unusual events like bioterrorism. We realize that Serfling is commonly used for detecting influenza epidemics. In this case, we have applied it instead as a general aberration detection method for unexpected outbreaks. We have detailed this distinction and the important limitations of our approach in the discussion.

9. Simulated outbreaks Page 6/7. 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.

Unfortunately, it would be very difficult to find a limited set of outbreaks that capture all the essential features of any outbreak that a public health practitioner might want to detect. Instead, we have used a very limited set of simple artificial outbreaks, as previously proposed by Mandl, Reis and Cassa, “Measuring Outbreak-Detection Performance by Using Controlled Feature Set Simulations” MMWR Morb Mortal Wkly Rep 2004, 53(suppl):130-136. We have added this to the limitations of the study at the end of the discussion.

10. Results Page 7. 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.

Thank you for the suggestion. We have added this as Figure 1.

11. Page 10 discussion. 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.

In the discussion, we have added a paragraph about the dependence of the cost of an alarm strategy on the specificity.

First Reviewer's Discretionary Revisions

1. 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.

We mention in the data section of the methods that previous studies have found that visits exhibit a dependence on the day of week and the season. In the description of the Serfling method, we mention that the sine and cosine terms capture seasonal effects in the data. In the description of the trimmed seasonal model, we explain that the day of year is considered to remove seasonal effects.

2. Page 3. I suggest describing the autoregressive model in a new paragraph rather than at the end of the first paragraph of the methods.

We agree and have made this change.

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

We have added a short explanation of our choice of Gaussian kernel smoother, and of our use of circular boundaries for the day-of-year smoother.

Second Reviewer's Major Compulsory Revisions

1. To put the study in context the results section would benefit from a few lines of description of the 12 years training/test data, e.g. what was the mean, maximum, minimum and variance of daily respiratory complaints over the 12 year period; & approximate population served by the hospital.

Thank you for the suggestion. We have added a short paragraph near the beginning of the data section describing the raw data, and referencing a figure of the raw time series.

2. It would also be useful to have some thought on what type of syndromes the expectation variance model is most suitable for monitoring, i.e. for common seasonal syndromes (e.g. respiratory complaints) or rare more specific syndromes (e.g. encephalitis)?

We have added to the limitations that it would only be useful for syndromes having visit patterns that depend on the three model covariates: trend, season, and day-of-week. It would not be useful for sporadic or rare syndromes.

3. "With constant specificity, public health practitioners can better evaluate cost effectiveness of surveillance systems." This statement is made as the last line of the abstract but is not mentioned or discussed explicitly in the main body of text. I think this conclusion is valid but needs to be backed up in the discussion.

We have added a section to the discussion about the cost of an outbreak, and how estimating it accurately depends in part on an accurate estimate of the specificity. We include an equation showing how to compare the cost of two outbreak strategies.

Second Reviewer's Discretionary Revisions

1. What are the resource implications for Public Health Departments to implement this methodology? Will this method require additional time, expertise, hardware/software to implement?

We mention in the discussion that the resources required are minimal for public health departments with surveillance data.
2. The summary of the type of variation in specificity (i.e. by day, month, year...) for each of the 4 existing methods is a useful one (Results: Paragraph 1). could the authors summarise this in a table (e.g. for each method when is highest, mean, lowest specificity observed. Or alternatively merely add the results for the 'trimmed seasonal' and 'wavelet' approach to Figure 2 to complete the picture.

Thank you for the suggestion. We have added the other comparison models to the figure (now figure 3). We thank you for your consideration.

Sincerely,
Shannon Wieland