Author's response to reviews

Title: Detection of influenza-like illness aberrations by directly monitoring Pearson residuals of fitted negative binomial regression models

Authors:

Ta-Chien Chan (dachianpig@gmail.com)
Yung-Chu Teng (twseventh@gmail.com)
Jing-Shiang Hwang (hwang@sinica.edu.tw)

Version: 4 Date: 22 January 2015

Author's response to reviews: see over
Dear Editor,

Thank you for your help reviewing this paper. On behalf of all co-authors of the manuscript, I appreciate the two reviewers’ effort in providing constructive comments and suggestions. We have carefully made a major revision based on most of the reviewers’ concerns and suggestions. We have described our proposed approach more clearly as monitoring directly the standardized Pearson residuals of fitted negative binomial models to distinguish it from the traditional methods of control chart and those modifications. We have therefore slightly revised the title as “Detection of influenza-like illness aberrations by directly monitoring Pearson residuals of fitted negative binomial regression models”.

We have re-designed a simulation study to demonstrate that the proposed relatively simple method of aberration detection is promising in terms of early outbreak detection compared to the modified CUSUM methods. We agree that some other methods such as those proposed by Dr. Sparks may have great performance in some cases, but those methods may be too technical for general readers. In contrast, our approach of using standardized residuals directly is much easier to understand and use. We hope you find that the performance of this simple approach is acceptable.

These revisions are marked in the main text in red and summarized in the point-to-point responses to the reviewers’ comments on the following pages. The point-to-point responses are listed below. We hope that all unclear points are clarified in the current manuscript. We hope this manuscript can meet the standard of the journal and look forward to hearing from you soon.

Sincerely yours,

Jing-Shiang Hwang, Ph.D.
Distinguished Research Fellow
Institute of Statistical Science
Academia Sinica, Taiwan
Reviewer's report 1

Title: Detection of influenza-like illness aberrations using Pearson residuals of fitted negative binomial regression models

Version: 3  Date: 16 December 2014

Reviewer: Fernanda Dorea

Reviewer's report:

While I still believe this could be an interesting methodology to put forward, and appreciate the authors’ efforts into attending a number of minor and moderate corrections suggested, I don't think the authors have addressed the main methodological concerns of both reviewers in the first round of revision, as detailed above:

1) When I suggested that the model should be fitted to T-1 (Reviewer 1, comment 12), I did not mean that the model should be used for prediction. I meant that the authors should explain how (how else) they avoid over-fitting to current day, especially in case the current day has aberrations. Is there no guard-band at all?

Response: We thought there would not be a problem of over-fitting, because we have fitted a long series of size $T_0=3x365$. But we agree with your comment. In the revision, we have corrected it in the text such as shown below. We have also re-fitted the models and presented the new results which are almost the same as the old results.

“Let $\hat{\beta}$ and $\hat{\kappa}$ be the estimated vector of coefficients and dispersion parameter from the fitted model using the series of $y_i$ and $X_i$ for $i = t-1,...,t-T_0$. The expected number of visits for day $t$ is estimated by $\hat{y}_t = \exp(X_t^T \hat{\beta}).”$

2) In my previous comment 15, I point out that the authors keep comparing sensitivities in a table that lists very varied specificity. The authors’ response still doesn’t’ address this problem. If the CUSUM sensitivities are low in a scenario where specificities are higher than the method proposed, than it can’t be really compared. Specificities should be fixed, and then sensitivities compared. At least using simulated data, this should be possible. That is, the authors should find a CUSUM threshold that provides the same specificity as the method they propose, and only then compare sensitivity.

3) The authors repeatedly state, in their response to the second reviewer specially, that their intention was not to prove that their method was not superior to all popular methods, but to provide an alternative approach. I agree this is a valid point, but then the authors need to demonstrate WHEN this approach they are putting forward works, and in which cases (even if few or very particular) it would work better than something else. What types of data or in what surveillance scenarios can it be useful?
Compared to what is it better? If the authors’ goal is to prove only that it is better than this one method (CUSUM), then they must at least make a point of when it is better, paying attention to my comment 2 above. One important point to note, is that in the case of data with strong seasonal effects, the “popular” method would not be the CUSUM applied directly to the data, but to at least remove temporal effects first. In my view, the standardized residuals method should be compared to CUSUM applied to non-standardized residuals. That would be a more fair comparison, as it compares two methods that do try to handle seasonal and day-of-week effects, not one that deals with it, and one that ignores it. The need for dealing with those effects has already been extensively demonstrated in the literature. It’s just HOW to deal with those effects that we should still be investigating, and using standardized residuals directly, instead of monitoring raw residuals with a control chart, is indeed underused, but its value need to be demonstrate clearly (again, comment 2 above).

Response: We appreciate very much the comments 2) and 3). These comments motivated us to re-design the simulation study. In the revision, daily series of counts with seasonal and day-of-the-week effects were generated from negative binomial models. Exact counts generated with three different signal-to-noise ratios for a period of 40 days were added for detection performance evaluation. The proposed methods of directly monitoring Pearson residuals with two threshold values were compared to two modified CUSUM methods applied to original observed counts and Pearson residuals. The thresholds for the CUSUM methods were chosen by trial and error, so the false alarm rates were a little higher than those of our proposed method (as suggested in comment 2). The results showed that the CUSUM method applied to Pearson residuals was indeed better than the CUSUM method with original observations when the signals were not weak, but clearly no better than our proposed method in detection power. We concluded that the proposed monitoring of standardized residuals directly is relatively simple, and performed better in terms of both the probability of successfully detecting the epidemic and early detection than the CUSUM methods when the false alarm rates were controlled at comparable levels in the simulation study.

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.
Reviewer's report 2

Title: Detection of influenza-like illness aberrations using Pearson residuals of fitted negative binomial regression models

Version: 3 Date: 22 December 2014

Reviewer: Ross Sparks

Reviewer's report:

Monitoring methods:

My original comment: …In the discussion on the model selection and Table 1 – it is unclear what happened with days Tuesday to Friday (was that the same as Monday?). In addition, the seasonal terms is not specified, e.g., where harmonic used? As the paper currently stands the model appears to unspecified.

The authors’ response: We used the seasonal term to adjust the seasonal effect. Because the outpatient visits were mostly off on Sunday. The outpatient visits on Monday were significantly different from those on Tuesday to Friday.

My reply: The response is inadequate. Readers need to be able to replicate the models fitted in the paper, but this is not possible with the current version of the paper. The following issues are unclear:

• I assume that Tuesday to Friday are combined to represent one level for day of the week and in Table 1, and that this is used as the baseline level which Monday, Saturday and Sunday counts are compared with. This needs to be made clear.

Response: Yes, you are right. Tuesday to Friday were assumed to be one level of baseline. The effects of Monday, Saturday and Sunday listed in the original Table 1 (Table 2 in the revision) obtained in the regression models were compared to that baseline level. We have not listed the baseline levels for all factor variables selected in Table 2 to save space. But we have revised the text to make it clearer for each variable’s baseline level.

• “Seasonal term” is unusual because it will adjust for any local outbreak whether it is in season (i.e., usual) or out-of season (unusual) – I would preferred if it was labelled as an adjustment for known outbreaks rather than an adjustment for usual season outbreaks.

Response: The seasonality of influenza epidemic was often modeled with harmonic terms. We have found these terms of sine and cosine functions sometimes were not good enough to capture the seasonal pattern. In this study, we proposed an alternative term called moving month-of-the-year time-dependent variable, which is the difference between medians of visits in the past 30 days and visits in the past 365 days for seasonal adjustment in the model. This time-dependent variable is determined dynamically by the observed data. It can capture the seasonal pattern and therefore
can be used as an alternative term for adjusting usual season outbreaks.

• Visits a week ago is unclear – is this an indicator variable indicating that the patient visited ED within the past week or exactly one week ago?
Response: We revised it as “exactly one week earlier”.

My original comment: the unattractive feature remains the focus the paper has on standardised Pearson residuals is unnecessary because CUSUM plans can be derived directly (e.g., see Sparks, Keighley and Muscatello, 2010).
The authors’ response: Thanks. We tried to integrate the easy and familiar approach for the public health workers to implement.
My reply: Several authors have already criticised the EARS modified CUSUM statistic, (e.g., see Fricker et al. 2008 – they demonstrated that the CUSUM applied to model based prediction errors had advantages over C3). In addition, Sparks et al. (2010) used optimal CUSUM for Pearson residuals which the authors could easily apply in their application. I don’t think the argument of applying inefficient CUSUM approaches even if they are familiar is sound practice. I think the authors should at least advocate more efficient CUSUM methods (as possibly a third alternative) and the readers can judge for themselves which they will use. Note I am not suggesting that the authors apply the adaptive CUSUM in Sparks et al. (2010) that does not use the Pearson residuals. Note that Sparks et al. (2011) also proposed an EWMA plan using Pearson residuals when monitoring negative binomial counts.
Specific points
Response: Thanks for the comments and suggestions. We understand that the method of CUSUM applied to original observations is not an ideal choice for comparison, but a naïve one. We also understand that our proposed method may not perform as well as some available methods in the literature. But we also found that those efficient methods may be too technical for general readers. One of the main purposes of this study is to demonstrate that the simple approach of monitoring standardized residuals directly could also produce promising results. Motivated by the comments and suggestions of you and Dr. Dorea, we have re-designed a simulation study. In the revision, we included the CUSUM method applied to Pearson residuals from the same fitted models for comparison. The results first showed that indeed the performance of CUSUM applied to residuals is better than that of the CUSUM applied to observations when the signals were strong. And our proposed simple method was still better than the CUSUM method applied to the Pearson residuals. We concluded that the proposed monitoring of standardized residuals directly is relatively simple and performed better in terms of both probability of successfully detecting an epidemic
and early detection than the CUSUM methods in the simulation study.

• Page 12 middle of the page – please explain where the formula for sd(vi) comes from.
  
  Response:
  
  In the old simulation design, we have set dispersion parameter $\kappa_i = 0.2 / \mu_i$.
  
  $\text{Var}(v_i) = \text{Var}(u_i + e_i) = \mu_i + \kappa_i \mu_i^2 + \gamma^2 \mu_i^2 = 1.2 \mu_i + \gamma^2 \mu_i^2$.
  
  However, we have re-designed the simulation study. There is no such variable in the revision now. It is now $\text{Var}(w_i) = \mu_i + \kappa_i \mu_i^2 = 1.2 \mu_i$.

• Page 17 – the reference to others using Pearson residuals is not accurate for example both Sparks et al. (2010, 2011) used Pearson residuals for negative binomial counts.

  Response: We have added the following text in the discussion section.

  The use of negative binomial models and Pearson residuals for improving the performance of aberration detection has been also considered in the literature [Sparks et al. (2010, 2011)]. These methods are built on cumulative deviations in a traditional framework of control chart for monitoring aberration. Although these methods had great performance in some situations, they may be too technical and complex to be implemented into routine surveillance by general readers. In contrast, the proposed alternative method by monitoring the Pearson residuals directly is simple and relatively easy to use.