Reviewer's report

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?

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).

**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.