Author's response to reviews

Title: Rapid detection of pandemic influenza in the presence of seasonal influenza

Authors:

Brajendra K. Singh (brajendra.singh@ed.ac.uk)
Nicholas J. Savill (nick.savill@ed.ac.uk)
Neil M. Ferguson (n.ferguson@imperial.ac.uk)
Chris Robertson (chris@stams.strath.ac.uk)
Mark E. J. Woolhouse (mark.woolhouse@ed.ac.uk)

Version: 2 Date: 26 September 2010

Author's response to reviews: see over
Cover Letter

To
The Editor
BMC Public Health

Dear Sir/Madam,

Thanks for your kind message inviting us to respond to the comments made by the two reviewers. I am very grateful for the positive reports on this work by both reviewers. I think both reviewers have made a number of valid criticisms and comments on our work. Many of the comments are highly relevant for this work. In the following we have addressed all the comments point-by-point. Also, the necessary changes appear in the manuscript.

With kind regards,

Yours sincerely,

Brajendra K. Singh
On the comments of Reviewer 1

We thank the reviewer for the recognising of the presence of a plethora of heterogeneities in biological data, and their influence on the limitation imposed on the success of a statistical method employed to analyse the data. Needless to say, the Scottish seasonal ILI (influenza-like illness) data are full of noise, induced by the non-adherence to a uniform reporting protocol of ILI cases by sentinel GPs. Also, we think the point that “a combination of methods will (should) be used to assess an unusual public health situation” is very well made by the reviewer. The general points made by this reviewer now have been incorporated in the discussion section.

Minor Essential Revisions

Background: Last sentence, 1st paragraph.

Please rephrase or remove the reference to the ‘containment of an outbreak at its source’. I know this has been advocated by one of the co-authors (Neil Ferguson), but this has been shown not to work in most situations. The rather desperate attempts to contain local outbreaks of the pandemic virus using mass prophylaxis has led to unnecessarily large numbers of adverse reactions to the antiviral drug oseltamivir that was used for this purpose early on in the 2009 pandemic, particularly in children (see refs below). Now, in retrospect, this type of mass prophylaxis needs to be carefully balanced against the severity of illness induced by the virus in question. Not only may this be unnecessary in terms of adverse effects on individuals, it is also extremely expensive for governments and public health agencies to fund.


Authors’ response: We have rephrased the reference to the ‘containment of an outbreak at its source.’

Discussion: 2nd paragraph

This paragraph seems a bit odd. Earlier in their Results, the authors propose that with a 1% case reporting rate, their WCR approach with a mean detection time (MDT) of 4 weeks is superior to the Cusum or ILI threshold methods with MDTs of 5 weeks each. The in this 2nd paragraph of the Discussion, they then question whether, in fact, a delay of up to 12 weeks is, in fact, poor? If it is not, then presumably whether the MDT is 4 or 5 does not really matter, so all the methods (WCR, Cusum and ILI threshold) are equivalent? The reasons they give thereafter to explain the poor
performance of their WCR approach can be applied to all 3 approaches to various extents.

There is no need for the authors to feel 'embarrassed' about the poor performance of their model using the real H1N1/2009 data, which is how this paragraph seems to read to this reviewer. Rather, I would ask the authors to incorporate some of my more general points above and re-write this 2nd paragraph to be more objective and suggest that, in any case, on a practical, daily surveillance level, all 3 statistical methods may well be used in combination if there is any suspicion of unusual influenza activity. This paper just offers an additional tool in the influenza surveillance armamentarium - it does not have to be significantly superior to all other methods (which it is not) to be of use, but can act as a check on the results from other, older statistical methods used to analyse such data.

Authors’ response: We have rewritten this 2nd paragraph and have included some of the reviewer’s general point in a newly-added second paragraph in the Conclusions section.

On the comments of Reviewer 2

Apart from the minor essential comments, this reviewer has several comments on the major part of our work: the joint probability distribution of (WCR, N_{HB}) and detection algorithm based on it. We think this set of comments has given us opportunity to clarify these issues which the journal readership might have found difficult to understand. For this reason we thank the reviewer for the comments.

Major comments/issues

1 Proposed detection algorithm

(1a) The use of the joint probability distribution of WCR and N_{HB} begs some questions. The joint distribution summary given in Figure 2 is odd. It implies that having both WCR and N_{HB} close to zero is fairly common, WCR is usually less than 1 and is rarely bigger than 4. These results are surprising to me. WCR is defined as a ratio of cases in week t and week t-1. I would have thought values near 1 would be common. Also, N_{HB} is the number of health boards reporting an increase. As there are 13 health boards in total I would have thought values near 6 would not be very uncommon.

Authors’ response: Although there are 13 health boards (HBs) participating the SERVIS scheme of seasonal ILI reporting through sentinel GPs, there are 10 (median number, with a range of 9 to 12) HBs that have, at least, a sentinel GP in any given past influenza seasons. This is the reason that the average value of N_{HB} lies around 4 and not 6.

Furthermore, for ILI cases in GP practices, and bearing in mind the small sampling fraction of GP practices (that is, 20 to 44 out of 1000 practices) in the surveillance system there is a discrete nature to WCR. When WCR is 1 this means that the
number of cases is the same one week to the next and this usually occurs when there are 0, 1, 2 cases reported. In such a situation we have one or two boards with an increase and one or two with a decrease and no change in the rest. This leads to the distribution presented in the manuscript as the $N_{HB}$ is the number of boards with an INCREASE.

We agree that if the number of cases per week were larger, say around 10 to 20 at least, and WCR was exactly 1 then you would expect a different shape to the distribution. For this reason the joint distribution of WCR and $N_{HB}$ requires to be evaluated for each system and for each response.

(1b) How much benefit is provided by using both WCR and $N_{HB}$? What about just WCR? $N_{HB}$ is an unusual choice as it gives equal weight to each health board even the ones representing small populations. With the small health boards whether there is an increase or decrease can be quite dependent on the inherent noise.

Authors’ response: We had checked whether the use of WCR alone would be sufficient. The efficiency of the WCR method using the information on WCR alone proves worse than the other two methods. This is not surprising though. The synchronous increases of ILI cases seem to be a key ingredient in simulated pandemics (though this behaviour was not apparent in the 2009 pandemic influenza A(H1N1v) data from SERVIS sentinel GPs - see Discussion section of the MS).

There is an issue about the size of the health board if the surveillance system was based upon using GP practices in proportion to the size of the board as there would be much more variation in the small boards relative to the large ones. In Scotland the largest board has in excess of 1.3 m people while the smallest ones have around 20,000. However the SERVIS system does not recruit GPs in direct proportion to the size of the board. The smallest health boards have one or two practices and the larger ones have 6 to 9, making the surveillance system highly prone to random effects. Thus the effect of small boards relative to large ones is not as great as would be the case if there was sampling in direct proportion to the board size.

(1c) You determine pandemic signals based on $P(WCR=xx \text{ and } N_{HB}=y)$. However, this probability could be (is?) small in a number of situations that don’t correspond to pandemics, e.g. for

- small values of $xx$ or $y$
- $y$ small and $xx<1$

Please explain more fully why this choice seems to work.

Authors’ response: We are aware of this situation. We agree that the WCR algorithm is not full-proof of generating false alarms. In order to keep the number of false alarms at a minimum level we set the specificity of detection at 95% and above. (We are tolerating a maximum of 1 to 2 false alarms per influenza season of a typical length of 33 weeks.) As shown in the figure, the situation in bullet points is (almost) never encountered for specificity (Sp) of 99%. Also, except on a few occasions, a specificity of 95% is sufficient to enable the detection method to avoid the situation.
Figure: Plots of $x \sim y$ in the ($WCR - N_{HB}$) joint-probability space at the time of pandemic detection (that is, when $P(WCR=x\text{ and }N_{HB}=y)$ is less than the detection thresholds). The plots are shown for a set of 1800 ($6 \times 300$) runs of pandemics starting on week 1. Detection specificity (Sp) and pandemic case reporting (CRP) rate, in percentage, are specified in each plot.

It would certainly be possible to modify our use of the joint distribution of $WCR$ and $N_{HB}$ to exclude the situations highlighted for a signal for pandemic flu. This would serve to increase the sensitivity of our system relative to the others. We did not do so as these events occur with extremely small probability and in a pandemic are even more unlikely. Also, $WCR$ equal to 1 with $N_{HB}$ being zero (the bottom-left corner in the main plot shown in Figure 2) is excluded from our pandemic detection criteria.

(1d) The joint distribution of $WCR$ and $N_{HB}$ will depend on time. This is not taken into account and should be explicitly mentioned.

Authors’ response: The point about time-dependency in the WCR method is well made and we had thought about it as we have already noted in the Discussion section. However, we would like to point out that the historical seasonal ILI data are so patchy that the weekly time-varying joint-probability of ($WCR$ and $N_{HB}$) would not be of any use.
I assume an influenza pandemic could start at any time of year, i.e. even when there is no seasonal influenza. What happens in that case? In your proposed method you seem to only get ILI reports for 33 (or so) weeks per year.

Authors’ response: That an influenza pandemic could start at any time of the year was the reason that we did consider different pandemic starting weeks, sliding through weeks – from the first to the last – of a typical influenza period. Since we are adding seasonal influenza cases to the sampled pandemic ones in our analysis in order to see how much delay/masking can occur due to seasonal cases, we limited our analysis for a typical influenza period of 33 weeks only. We think it is easy to see that in the absence of seasonal cases, the WCR method would generate pandemic alarm relatively early. (This effect is conspicuously present in the distribution of detection times of Figure 5 towards the end of the influenza seasonal period.)

2 Simulation study

(2a) The use of the pandemic model is a strength of this work. However, it is not clear to me from the titles of the listed references (I did not look at them closely) that the model has been shown to be a reasonable approximation of a real pandemic. Can you provide some assurance?

Authors’ response: The pandemic model used was published in Nature, and before the 2009 influenza A(H1N1v) pandemic, was the best available model of the likely spread of a pandemic throughout the UK. A characteristic of this model, and others (HPA) was the relatively synchronous increase in a number of areas and this is one feature used in our proposed detection system. This synchrony is caused by the assumptions about seeding of infection into the UK made by the model, namely that cases would be imported from outside the UK into all regions at a constant per-capita rate per region. In the event, early seeding of the 2009 H1N1 pandemic was more heterogeneous, due to the clustering of seeding associated with the return of tourist groups from Mexico and then the US. In addition, the asynchrony between regions in the 2009 pandemic was amplified by the relatively low reproduction number (R=1.4) compared with past pandemic and the interruption of transmission caused by school summer holidays, (which occur earlier in Scotland than in the rest of the UK). By contrast, the pandemic simulation assumed a single uninterrupted wave with a higher reproduction number (1.7), which leads to greater synchrony (the simulation matches well spatiotemporal incidence trends seen in the 1957 and 1968 pandemics). Future work could examine the performance of the proposed algorithm with a wider range of simulated pandemic scenarios, including low-transmissibility ‘mild’ pandemics like that seen in 2009, and will consider the impact of geographically variable interruptions of transmission caused by seasonal factors, school holidays or interventions.

(2b) Why do you use 30 samples from each of the 10 simulated pandemics rather than one sample from each of 300 simulated pandemics? Is this a computational issue? It seems to me using 300 simulated pandemics would be preferred as the results for different samples from the same pandemic will be correlated. Also, where do the three different values of come in? Do you repeat the whole simulation process for each?
Authors’ response: This is also a good point and we should have checked our analysis with 300 pandemic runs. Note, however, that the pandemic model was simulated for the whole of Great Britain, with infection being seeded into the UK from outside at a geographically constant but temporally increasing per-capita rate. Simulated pandemics vary a lot in their take-off phase. But once the pandemic gets going, the time profiles of pandemics in different runs do not differ much from one another. Since our use of simulated pandemic ILI infections are related with the time-period in which the first Scottish cases were reported (by which time the pandemic cases get in to exponentially rising phase), the use of 300 samples of pandemic ILI cases in the form of either 30 samples from 10 simulated pandemics or one each from 300 simulated pandemics we think will give the same results.

3 Chart Setup

(3a) On page 5 you discuss estimating the background pattern of the seasonal ILI cases. However, it seems here you are referring to only the distribution of weekly number of cases without regard to the time series pattern. While this is fine since I imagine you can argue that from year to year the exact date when the seasonal increase in influenza starts is hard to predict (certainly if you only have data from 6 different years). What you mean by the background pattern should be more clearly defined.

Authors’ response: By the background pattern of the seasonal ILI cases we mean that what is the information that one can sift out of the seasonal ILI time series from different influenza seasons. The time series alone are of no use. There are no discernible patterns. But when looked at the joint-probability density of \((WCR, N_{HB})\), we think that the seasonal ILI activities could be very well summarised by this distribution. Thus, we are referring to the joint-probability distribution.

(3b) I don’t really understand your rationale for excluding the 2007-2008 data from the analysis. While the baseline activity was low for this year it is still representative of a possible seasonal influenza pattern.

Authors’ response: As we have tried to make this point clear above in our response here (see 1e) as well as in the main text that the inclusion of seasonal ILI data from a low-activity influenza season would only enhance the performance of the WCR method. (Note the ILI rate threshold method, which closely competes with the WCR method, will perform poorly in such low-activity influenza season.) In order to keep the performance estimate for the WCR method at a conservative level, we have excluded the 2007-08 data.

Minor Essential Revisions

Page 2, results, it should be mentioned that the results reported were observed in your simulation study. Presumably your method is not really 100% sensitive.

Authors’ response: We add a sentence to accommodate this fact that most of the results reported were observed in our simulation work. Also, as we mentioned there,
the WCR method is 100% sensitive for case reporting rate of 1% or above. Of course, its sensitivity is down to 98%, if the case reporting rate is 0.5%.

**Page 3, last paragraph.** please define the “basic reproduction ratio” and the “weekly case ratio.” WCR is defined later, but it should be defined earlier.

**Authors’ response:** We have now defined the “basic reproduction ratio” as well as “weekly case ratio”.

**Page 5, line 5.** referring to the “6 seasons” is somewhat confusing. Why not call them years instead?

**Authors’ response:** We changed it to “6 influenza seasons”. An influenza season in a temperate country like the UK is distinct from normal seasons of a year.

**Page 7, top.** is there not a range of values that give the same probability?

**Authors’ response:** Yes. There is a range of values that gives the same probability. We have chosen the lowest end of this range.

**Page 7, and throughout.** The method described on page 7 is only one way to implement a CUSUM chart. As such it is more appropriate to refer to it as the Cowling CUSUM (or something like that) rather than suggest it covers all CUSUM based methods.

**Authors’ response:** We are now calling it a moving-average Cusum method.

**Table 1** type “case reproting” typo

**Authors’ response:** The typo has been corrected.

**Figure (1a)** in 2002-2003 there are more than 33 weeks of data. This is not commented on in the paper. Same thing in Figure 3.

**Authors’ response:** We had given the reason as to why the 2002-03 influenza season was longer than a typical one in the figure caption. However, we didn’t repeat the same in the figure caption of Figure 3.

**Figure (1b)** You provide data plots for 10 health boards. Are there not 13?

**Authors’ response:** We did mention why there were only 10 health boards.

**Figure (3) and (4),** the label on the vertical axis should tell me the probability of what?

**Authors’ response:** We have now changed the axis label from “Probability” to “Probability of (WCR, N_{HB})”.