Reviewer’s report

Title: Respiratory Syncytial Virus tracking using Internet Search Engine Data

Version: 0 Date: 13 Dec 2017

Reviewer: Ryan Malosh

Reviewer's report:

This manuscript extends an established method for predicting influenza circulation to respiratory syncytial virus. This is a thoughtful analysis with rigorous methodology and the conclusions are appropriate to the findings. This reviewer is unable to assess the machine learning aspect of the research, thus I suggest further statistical review. In addition, I believe the following comments should be addressed:

Specific Comments

Page 3, Line 19-23: I respectfully disagree that there is low urgency or recognition of the burden of RSV disease. This statement may be accurate for certain aspects of the population or specific geographic regions, but certainly not all. In particular the burden of disease in children aged < 1 year and adults > 65 years is recognized. Evidence for this recognition and urgency on the part of public health researchers is evidenced by the over 40 vaccine products in various stages of clinical testing (see the PATH RSV vaccine snapshot).

Page 3, Line 51-53: My understanding is that most influenza forecasting models have limited utility in long term prediction (the best results tend to be a couple of weeks in advance and prediction seems to be worse near the peak). One of the benefits of improved forecasting of RSV circulation that the authors fail to mention is the implementation of prevention strategies (timing of vaccine administration, once licensed or administration of palivizumab). It would be of great interest to hear if/how the authors believe these findings could impact those decisions.

Page 5 Line 20: Did the authors consider collecting data for RT-PCR confirmation of RSV infection? The use of clinical multiplex PCR to test for a panel of respiratory viruses (and
bacteria) is expanding rapidly. As PCR is highly sensitive this could be relevant as a sensitivity analysis, especially in more recent years.

Page 6, Line 16: Why were lungs not included in the affected body parts? This seems particularly relevant for parents of young children who observe cough and wheezing and for older adults with underlying chronic respiratory conditions.

Page 7, Line 22: A reference is needed for this bootstrap method.

Page 16, Lines 22-27: I think the limitations related to RSV and Influenza co-circulation warrants further discussion. While I agree that at the weekly level this may be less of a concern the amount of co-circulation will be highly variable by year. Further, influenza and RSV are difficult to distinguish clinically and influenza is included in the search terms. Perhaps a sensitivity analysis looking at years with substantial overlap in lab confirmed RSV and influenza and examining the model with and without influenza in the search terms could shed more light on this issue.

Figure 1: Do the authors have hypothesized explanations for the model not dropping to zero in years 2007-2011 in OH and PA? Or for the large underestimates in CA and MI in 2010? The titles for Figure 2 and 3 seem to be flipped. Additionally the panels in figure 3 are not labeled.

Are the methods appropriate and well described? 
If not, please specify what is required in your comments to the authors.
Yes

Does the work include the necessary controls? 
If not, please specify which controls are required in your comments to the authors.
Yes

Are the conclusions drawn adequately supported by the data shown? 
If not, please explain in your comments to the authors.
Yes
Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?

If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I recommend additional statistical review

Quality of written English

Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests

Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons
CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal