Reviewer’s report

Title: Using Web Data to Improve Surveillance for Heat Sensitive Health Outcomes

Version: 0 Date: 10 Apr 2019

Reviewer: Loukas Samaras

Reviewer's report:

Overall:
It is a very good and interesting research. It is well described and the results are sufficiently justified. There is also an adequate and very good analysis of the data and the used methods both for Regressions and the ARMA models.

Therefore, I recommend it for publication with a minor revision on some specific parts.

Specific Comments for minor revision:

Abstract:
i) line 75: "This paper also proves.. ": You’d better write "shows" or something similar. Although the results are clear, it’s better to avoid "prove", since the data sample include one summer period and a geographic region (Florida).

1. Introduction:
ii) line 107: "..syndromic surveillance systems include a lack of resources...":
It is true, but not only that. There are other reasons too, e.g. delay of official government data and results, complexed or expensive procedures to gather health data from various sources., etc. If possible, you can mention the above.

2. Methodology:
iii) line 167: "accessible over the preceding seven days":
Using scripts, we can obtain weekly or daily data (mainly past data) through API's (C#, Python, etc.) from Google as well. Twitter may certainly has more advantages over Google in capturing data (time fraction, computer automation etc.)
If possible, you can modify these lines appropriately to clearly show that
iv) lines 237, 393: ARMA-GARCH models:
These models are referenced as ARIMAX models (with an exogenous explanatory variable). ARMA is a special case of a two-parameter ARIMA model without Integration. If this the case, you can mention the above.
3. Results:
v) line 293: "Nearly all Twitter keywords were significantly related (p < 0.05)". What is the overall Pearson Coefficient (for each case you examine)? You mention in line 305 that "the consistency of the association, < 0.60 Pearson correlation coefficient". Does this refer to Twitter results or Google search results? You can clarify this.

4. Discussion:
vi) Which web tool gives the best results in terms of association and correlation of data, Twitter or Google? You can make it more clear.

You mention in lines 423-424 that "We hypothesized that more frequent Twitter tweets or Google search that results mentioning heat-related keywords are positively associated with health outcomes":

How much is this association? Did you find better results in Twitter or not? If possible, you can refer to that.

**Level of interest**
Please indicate how interesting you found the manuscript:

An article of importance in its field

**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 interest.

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