Author's response to reviews

Title: Accuracy of epidemiological inferences based on publicly available information: retrospective comparative analysis of line lists of human cases infected with influenza A(H7N9) in China

Authors:

Eric HY Lau (ehylau@hku.hk)
Jiandong Zheng (zhengjd@chinacdc.cn)
Tim K Tsang (matklab@hku.hk)
Qiaohong Liao (liaoqh@chinacdc.cn)
Bryan Lewis (blewis@vbi.vt.edu)
John S Brownstein (John.Brownstein@childrens.harvard.edu)
Sharon Sanders (flutrackers@earthlink.net)
Jessica Y Wong (ytwongj@hku.hk)
Sumiko R Mekaru (Sumiko.Mekaru@childrens.harvard.edu)
Caitlin Rivers (cmrivers@vbi.vt.edu)
Peng Wu (pengwu@hku.hk)
Hui Jiang (jianghui@chinacdc.cn)
Yu Li (liyu1@chinacdc.cn)
Jianxing Yu (yujx@chinacdc.cn)
Qian Zhang (zhangqian@chinacdc.cn)
Zhaorui Chang (changzr@chinacdc.cn)
Fengfeng Liu (liuff@chinacdc.cn)
Zhibin Peng (pengzb@chinacdc.cn)
Gabriel M Leung (gmleung@hku.hk)
Luzhao Feng (fenglz@chinacdc.cn)
Benjamin J Cowling (bcowling@hku.hk)
Hongjie Yu (yuhi@chinacdc.cn)

Version: 2
Date: 5 March 2014

Author's response to reviews: see over
Dear Dr Alam,

RE: Accuracy of epidemiological inferences based on publicly available information: retrospective comparative analysis of line lists of human cases infected with influenza A(H7N9) in China

Thank you for considering our manuscript for publication, and for forwarding the reviewer comments. Our specific point-by-point responses are given below.

We have addressed comments from reviewer 2 and improved clarity in describing the level of missingness in the data, definition of accuracy, and relative importance of different data sources. We also clarified how transmissibility was assessed though estimating quantitative measures such as the basic reproduction number is difficult with very limited understanding of the transmission dynamics. As suggested by the editor, we have now also uploaded all line lists from publicly available information as a supplementary data file. We have provided the emails of all authors in the title page. All authors have read and approved the final manuscript.

Yours sincerely,

Ben Cowling

On behalf of all authors

**Editor’s comment:**

We would encourage you to provide the line list with publicly available information that has been used for the analysis to be made available as supplementary material data file. This way this would enable reproducibility of the analysis.
Response: We have now provided the data for all line lists with publicly available information, at different dates as supplementary material data file (additional file 1), along with a description of the variables in a zipped file.

Reviewer 1 comments:

In this paper authors carried out various epidemiological inferences on the A/H7N9 avian influenza outbreak in humans in China, including changes in transmissibility and severity, using different line lists of A/H7N9 influenza cases in humans. These line lists were compiled from publicly available information. Authors then compared these epidemiological estimates from line-lists with those derived from official data from the Chinese CDC. Overall this is an important paper as their findings underscore the need to create a reliable online repository of epidemiological data that is updated in real-time for health emergencies like the ongoing A/H7N9 influenza in China and the Middle East Respiratory Syndrome in order to carry out quick and reliable epidemiological assessments. The paper is well-written and I have no major comments.

Response: We thank the reviewer for the comment.

However, I think the authors could point out that monitoring the evolving transmissibility of emerging influenza viruses is crucial and requires fairly accurate information about the onset of symptoms of cases in addition to reliable exposure history information. In particular, early assessments of the transmission potential of A/H7N9 influenza in humans were possible using official data from the Chinese CDC (see BMC Med 11:214, 2013). By contrast, a substantial number of cases (e.g., about 20% of cases) were missing symptoms onset information in the Flutrackers line list as of April 25, 2013, which hindered the estimation of reproduction numbers using this publicly available data.

Response: We thank the reviewer for the insightful comment. We have now included this point in the discussion, which reads,

"Monitoring the evolving transmissibility of emerging influenza viruses is crucial, but requires fairly accurate information about the onset of symptoms of cases in addition to reliable exposure history information, and the understanding of the transmission dynamics among poultry and from poultry to human. For the line lists using publicly available data these information were very limited, thus hindering quantification of transmissibility in terms of the basic reproduction number."

Reviewer 2 comments:
In their well written manuscript the authors present a systematic comparison of different line lists – all based on publicly available data – with the line list compiled by the Chinese Centre for Disease Control and Prevention (China CDC). The latter serves as “gold standard” containing detailed epidemiologic information on each laboratory confirmed case of influenza A(H7N9). This analysis points out the important role of line list to allow the identification and evaluation of critical epidemiological parameters, especially in the beginning of the spread of a novel pathogen.

Response: We thank the reviewer for the comments.

Major Compulsory Revisions

Abstract

1. P3, l16: “We characterized clinical severity and transmissibility...” The term “transmissibility” might be misleading, as the authors do not present data on R0 or risk of exposure, but rather did use the change in the incidence rates to demonstrate the impact of closure of poultry markets (as defined in the methods section, p9). Please rephrase.

Response: We agree with the reviewer that transmissibility is better quantified by $R_0$, which allows the assessment of the speed of transmission and comparison with other diseases. However currently we still have very limited information on the dynamics and relevant epidemiological parameters of transmission within poultry or from poultry-to-human. Under this situation, it is almost impossible to estimate $R_0$ even the poultry-to-human transmission dynamics is clearly understood.

Given this limitation, we assessed the impact of exposure to live poultry market, which is regarded as the major risk to infection based on most updated evidence on influenza A(H7N9) (Yu et al.). In this sense we respectfully disagree that the term “transmissibility” is inappropriate, at the same time we now addressed the limitation in the discussion, which reads,

“Monitoring the evolving transmissibility of emerging influenza viruses is crucial, but requires fairly accurate information about the onset of symptoms of cases in addition to reliable exposure history information, and the understanding of the transmission dynamics among poultry and from poultry to human. For the line lists using publicly available data these information were very limited, thus hindering quantification of transmissibility in terms of the basic reproduction number.”

Reference:

2. P3, l20: “..., but there were more missing data on dates...” please include overall quantitative information about the extent of this information missing for the five line lists analyzed.

Response: As the missingness of different variables varied for different line lists at different dates. To avoid providing too many specific details in the abstract, we now provided a general level of missingness, which reads:

“Demographic information were mostly complete (less than 10% missing for all variables) in different line lists, but there were more missing data on dates of hospitalization, discharge and health status (more than 10% missing for each variable).”

Background

3. The authors correctly mention that there is no official framework setup for early and rapid international exchange of epidemiological data. However, they should also address the role of WHO and the responsible national authorities.

Response: We agree with the reviewer that providing such background will give a more complete picture of various efforts from different parties. We have now added a description of the role of WHO in the background, which reads,

“The Chinese National Health and Family Planning Commission notified the World Health Organization in late March and joined forces for the prevention and control of the disease, along with other international animal health organizations [5].”

Methods

4. P7, l18: “Each line list was compiled based on...” The Table S1 in the supplementary material is very much appreciated. The authors should include information about the case definition used to include cases as confirmed by the other line lists. As the publication delay and information between different data sources differs, usually for constructing a line list a number of decisions have to be made about which information should be included from which source. If there is information about the ranking of different data sources by the five line lists, this should also be included in Table S1.

Response: We thank the reviewer for the suggestion. All line lists used the same case definition which is the reported laboratory-confirmed influenza A(H7N9) cases. With regard to different data sources, the official websites were generally regarded as more reliable than the press or online surveillance system, which were in turn regarded as more reliable than personal blogs. Except for HealthMap with well-trained algorithm to retrieve large amount of disease information from multiple data sources, most of the other line lists were constructed specifically for the H7N9 outbreak. For these line lists an exact
ranking is not available for inclusion in Table S1. However we now clarified the case definition and general ranking in the section “Sources of Data”, which reads:

“Each line list was compiled based on reports of laboratory-confirmed influenza A(H7N9) cases released by, in the order of importance, the national and provincial Ministry of Health websites or microblogs, World Health Organization, international online disease reporting systems and online Chinese news or blogs (Additional file 2: Table S1).”

5. P8, l3: Please specify which dates were selected for the estimation and explain the rational for the selection of these dates. E.g. for the analysis of onset to hospitalization May 1st 2013 was chosen, even though the analysis included cases until May 31st.

Response: As different line lists were updated independently at different dates, for comparison purpose the dates for analyses were chosen close to the time when most of the line lists were updated. Also, as the H7N9 epidemic stabilized after early May, 2013, the Virginia Tech, Bloomberg line lists stopped updating after the first week in May, while HKUSPH and FluTrackers have less regular updates. In Figure 1, we chose May 1 to include all line lists for comparison. This is now clarified in the Methods, which reads,

“As the line lists were updated independently at different dates, for comparison purpose the dates of analyses were chosen to match the time of updates for most line lists.”

6. P9, l2: Poisson regression models have been used. Had the data set been tested for overdispersion?

Response: We examined the mean and variance of the case count for each city. The variances were less than 1.4 times of the mean and didn’t indicate strong overdispersion. The ratios between residual deviance to degree of freedom of the poisson models were all smaller than 1. To further test for overdispersion, we now also fitted negative binomial regression models and found that the overdispersion parameters were all smaller than 0.001 for all cities.

(The cited Ref.#14 needs to be updated.)

Response: We have updated reference #14 (now ref #15).

Was the model exactly applied as described in the reference?

Response: We used a poisson regression under a frequentist approach rather than a Bayesian approach as in Ref #14 (now ref #15). The model was not exactly the same and we now added a reference (Shardell et al.) which adopted the same approach to assess the impact of interventions.
Reference:


What adaptations were made when using the line list data from the other sources (e.g. taking into account the reporting delay)?

Response: To mimic the real situation with data available at the time of analysis, we did not adjust for the reporting delay.

Did the model control for differences in the age and sex distribution between the different line lists?

As the demographic information were found to have very little differences across line lists (Figure S6), we did not control for age or sex differences. Also, the analysis focused on 3 cities with relatively small number of cases (range: 8 – 33) which gives very little power to control for other factors.

Results

7. P9, l19: “More than 90% of the cases could be matched...” What happened to the cases that could not be matched? Were these cases excluded from the analysis?

Response: We assumed that the analysts of the line lists based on publicly available information do not know if the cases can be matched to the official line list. Hence, all analyses were based on the full dataset including both matched and unmatched cases. This is now clarified in the second paragraph of the results, which read:

"We compared different epidemiological characteristics inferred from different line lists over time, for all cases irrespective of matching."

Discussion

8. Major limitations of publicly available information should be discussed. E.g. information published on individual cases from different sources cannot be matched without availability of a “gold standard”. Frequently there is the danger of duplicate reporting with differing dates, especially when family clusters are involved etc.
Response: We agree with the reviewer that duplicate reporting with inconsistent information can be a problem, especially if the size of the outbreak is large. In general, information released by the national or international health organizations was regarded as more reliable but there was not a well-defined rule for resolving inconsistencies. This is now described as a limitation in the discussion, which reads,

“Secondly, duplicate reporting from different data sources may have inconsistent epidemiological information. National or international health organizations were regarded as most reliable but there was no well-defined rules for resolving inconsistencies.”

Minor Essential Revisions

9. P3, l22: the authors refer to HFR as "hospital fatality risk" should it rather read "hospital fatality ratio"? (also in the rest of the text)

Response: We understand that fatality “risk”, “ratio” and “rate” have been used interchangeably in the literature. However we preferred to use the term “hospital fatality risk” to represent the probability (i.e. risk) of death among hospitalized cases which ranges from 0 to 1 (Kelly & Cowling, Epidemiology, 2013). In epidemiology, ratios are most commonly used to compare like quantities (e.g. odds ratios, risk ratios), where a ratio of 1 indicates no difference between the two quantities.

Reference:


10. P8, l1: Please refer to FIG S3 and additional information for explanation about how accuracy was defined and measured in this context.

Response: We referred to Fig S3 in the last line, first paragraph of the results section, which reads:

“For matched cases, we found discrepancies in dates of hospitalization, death and discharge when comparing to the China CDC line list (Additional file 3: Figure S3).”

Accuracy was defined in the title of Figure S3, which reads:

“Accuracy was defined to be exact for age, sex, province and severe cases. A two day discrepancy was allowed for onset, hospitalization, death and discharge dates.”

11. P8, l15: it should read “…cumulative number of deaths by the cumulative number of hospitalized cases…”.

Response: We apologized for the missing word. It is now added in the revised manuscript.
12. P12, l18: Add Ref.#27 as this also investigated severity based on case fatality ratios.

Response: We agree with the reviewer. Ref #27 (now ref #20) is now added.

13. Please check and update reference list (e.g. missing information Ref. #9)

Response: The reference list is now updated.

14. FIG2: Please add HFR1 and HFR2 in the label of the y-axis.

Response: “HFR1,” and “HFR2” are now added in the y-axis of Figure 2.