Author's response to reviews

Title: Estimates of the Reproduction Number for Seasonal, Pandemic, and Zoonotic Influenza: A systematic review of the literature.

Authors:

Matthew Biggerstaff (mbiggerstaff@cdc.gov)
Simon Cauchemez (simon.cauchemez@pasteur.fr)
Carrie Reed (creed1@cdc.gov)
Manoj Gambhir (manoj.gambhir@monash.edu)
Lyn Finelli (lfinelli@cdc.gov)

Version: 3
Date: 6 August 2014

Author's response to reviews: see over
August 7, 2014

Dr. Philippa Harris
Executive Editor, *BMC Infectious Diseases*

Dear Dr. Harris;

My co-authors and I, revising in response to reviewer comments, are re-submitting the above referenced article for publication in *BMC Infectious Diseases*.

We thank the reviewers for their thoughtful and constructive comments. Our point-by-point response to each of their comments is provided below. Additionally, we have included a completed PRISMA checklist as an additional file and a completed copy of the PRISMA flowchart as Figure 1 in the manuscript. We have also ensured the revised manuscript conforms to the journal style.

**Reviewer 1**

Reviewer's report:
In this manuscript, the authors review all published estimates of the reproduction number $R$ (both the basic reproduction number $R_0$ and the effective reproduction number $R_E$) for seasonal, pandemic and zoonotic human influenza. The authors show that there is surprisingly little variability in the estimates of $R$ for seasonal influenza over different seasons and geographical locations, despite them being based on various data sources and estimation methods. Estimates of $R$ for pandemic influenza are consistently higher and much more variable across pandemics but also within a pandemic. Estimates for zoonotic influenza are, on the other hand, much smaller but still quite variable. This manuscript, which is well presented, constitutes an interesting database of estimates of $R$ for influenza, together with details on the type of data used, the geographical location, and assumed mean generation time. I have a few comments, detailed below, that I think deserve to be addressed.

**Major Revisions**

1. My main worry about this study regards the completeness of the review. Indeed I was surprised (and a bit worried) to see no $R$ estimates for the 1889 pandemic, despite some articles (which seem to fall in the search criteria) reporting some estimates (e.g. Valleron et al. PNAS, 2011). Could the authors clarify why these articles are absent i.e. was it a selection criterion? If so why is that?

   **Response:** Our selection criteria did capture the Valleron 1889 reproductive number PNAS article (listed as reference # 112 in the revised manuscript; located in Table 5). We included its reproductive number estimate in the seasonal table because there is no confirmation available that the 1889 event met the definition as a true pandemic, and it is not considered one of the four pandemics that occurred in the modern era.

**Minor revisions**

1. In the introduction the authors state that “the value of $R$ characterizes the final number infected”; I suggest adding “in absence of intervention”.

   **Response:** We have added the suggested phrase (line 97 in the track-changes version).

2. In the results on the 1968 pandemic, the authors claim that “the value of $R$ increased across the waves”. However this is not based on a statistical test and is largely led by the results of a single study (Jackson et al. AJE, 2010). Thus I think the authors should moderate their statement. On another note, it seems that estimates of the first wave are for $R_0$ and estimates for the second wave are for $R_E$, which should be discussed (see also my general point about $R_0$ versus $R_E$ below).
Response: We agree and have moderated the language about the reproductive number increasing between the 1968 waves (line 195 in the track-changes version).

3. In the results for zoonotic influenza, I noticed that some of the estimates were very close to zero, and some above one, but none really in-between. Could the authors discuss why this is? Also I am wondering whether reporting the median of these values is the best choice. 
Response: We agree and have removed the report of the median value for zoonotic influenza. We prefer not to add a discussion about why the reproductive values are either close to zero or close to one because it would be purely speculative on our part and not based on the results of this literature review or other outbreak investigations (line 255 in the track-changes version).

4. In the discussion, authors suggest that improvements in surveillance systems would result in increased R estimates whereas disruptions to surveillance systems would result in decreased R estimates. Can they explain why?
Response: We have added an explanation to this section that explains how changes in case ascertainment can affect estimates of the reproductive number (lines 317-325 in the track-changes version).

5. In the discussion, it is said that “this study found consistently higher reproduction numbers for confined settings” which is not true for the 1968 pandemic.
Response: We agree and have modified our statement to indicate that this finding was generally found but was not consistent (lines 362–364 in the track-changes version).

6. Results about the boarding school in 1977/78: could the authors comment on possible reason for such high R estimates in that case?
Response: We have added some possible reasons for the high R value found during the 1978 H1N1 boarding school outbreak (lines 372–378 in the track-changes version).

Discretionary revisions
I understand this review already represents a lot of work, which is why I am suggesting the two following additions as “discretionary revisions” only, but I feel quite strongly they would be incredibly helpful to get a full critical view on the estimates presented.
1. First, estimates of R_0 and R_E are presented, and while clearly distinguished in the results, they are pooled to provide overall median estimates. Comparison between these medians across settings (e.g. pandemic/seasonal) or between pandemic waves is then made without any distinction between R_0 and R_E, which I think should be at least acknowledged and discussed.
Response: Thank you for this suggestion. When writing the results section, we considered dividing the results out by effective and basic reproductive numbers. However, we were already presenting the results by pandemic/season, wave, and whether the estimate was made in a confined setting. We felt having to present two R values for each of these stratification variables could be overwhelming to the reader without adding a lot of practical information. Additionally, some papers were not precise in their use of the terms basic and effective reproductive numbers. We felt some of the reproductive numbers presented as basic were more likely effective reproductive numbers since the authors did not state their immunity assumptions or made erroneous assumptions about the lack of population preexisting immunity (especially for the 2009 pandemic). We have added a limitations section to the paper and indicate in it that we did not differentiate to between basic and effective reproductive numbers for the median estimates (lines 387–395 in the track-changes version).

2. Second, the authors mention that the estimates reviewed were obtained using various data sources as well as various methods. Although the type of data used is recorded, the type of method is not. I think it would really be an improvement to add this as it would allow, when comparing estimates, to fully understand what can have led to differences.
Response: We appreciate this suggestion and considered adding the methodology to the review. However, some of the studies included in the analysis did not clearly explain the method used to estimate the reproductive number. We also believe adding this component will not provide much explanation for the values of R reported, do not meet the main objective of this review, and inclusion of this type of detail along with the necessary discussion will add considerable length to the manuscript.

Minor issues not for publication
1. Line 119 “of R reproduction number” should be “of R” or “of the reproduction number”
Response: We have made the suggested change (line 126 in the track-changes version).

2. Line 185 “because one” should be “because of one”
Response: We have made the suggested change (line 199 in the track-changes version).

3. Line 255 I suggest removing the sentence on confined settings, which is then repeated and discussed on line 339.
Response: We have made the suggested change (line 270 in the track-changes version).

4. Line 330: “estimation of” should be “estimation of R”
Response: We have made the suggested change (line 353 in the track-changes version).

5. In all Table notes: “The generation time” should be “The mean generation time”
Response: We have made the suggested change.

6. In all Figure legends: why do you have “Confidence Limit” and “Lower CI”? What is the difference? Please clarify.
Response: This was an error on our part. We have removed Lower CI from the legend.

Reviewer 2
Reviewer's report:
Discretionary revisions
This is a nice manuscript and provides a wealth of information for those interested in the transmission of influenza.
Might be worth thinking about an R0 versus Tg x-y plot to show the degree of correlation.
Response: Thank you for this suggestion. We created the graph below for the 2009 pandemic as an example of what this type of figure would look like. We don’t feel like this graph adds a lot of new, digestible information for the reader, even though there is a significant but low correlation between generation time and reproductive number for the 2009 pandemic. Given that, we would prefer not to include this figure in the revised manuscript especially with the number of tables and figures already included in this manuscript.
Is it worth adding in a simple statistical model to create a “corrected R0”. For the strains for which there are many estimates, would it be possible to standardize R0 by Tg using simple linear regression. This is a kind of poor-man's model reconciliation, but it might be interesting.

Response: Thank you for this interesting suggestion. We regressed GT against the reported reproductive numbers for the 2009 pandemic below as an example. While interesting, we do not believe the findings will add much to the overall objective of our systematic review and prefer not to include this in the revised manuscript.

. regress Median GT

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 66</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>.652479821</td>
<td>1</td>
<td>.652479821</td>
<td>F( 1, 64) = 4.46</td>
</tr>
<tr>
<td>Residual</td>
<td>9.3642018</td>
<td>64</td>
<td>.146319065</td>
<td>Prob &gt; F = 0.0386</td>
</tr>
<tr>
<td>Total</td>
<td>10.0169</td>
<td>65</td>
<td>.154106154</td>
<td>R-squared = 0.0651</td>
</tr>
</tbody>
</table>

Adj R-squared = 0.0505
Root MSE = 0.38252

| Median | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|--------|-------|-----------|---|-----|----------------------|
| GT     | .1680839 | .0795963 | 2.11 | 0.039 | .009072 .3270958 |
| _cons  | 1.07992 | .2289222 | 4.72 | 0.000 | .6225961 .537244 |

line 80 : should be “typical infectious individual” not “typical case”. The two aren’t usually the same and in some circumstances the difference might be important.

Response: We have made the suggested edits.

93 : R isn’t easily correlated with attack rate other than in homogeneously mixed epidemics. Consider lower than expected attack rates in 2009 (driven by children)

Response: We have clarified the statements (line 98 in the track-changes version).

121 : slight change to wording on the serial intervals and Tg? Its not an incorrect statement because the expectation of the serial interval distribution is the same as the expectation of the generation time distribution. However, serial interval is defined from symptoms to symptoms and generation time from infection to infection. Please just reword somehow to reflect this!

Response: We have made the suggested changes to the introduction (lines 97–105 in the track-changes version) where we originally discuss serial intervals and generation times.
201: similarity of median estimates between 2009 waves is very interesting. I’d promote that to the abstract and the expense of other results.

Response: We have made the suggested change to the abstract (lines 59–60 in the track-changes version).

282: I’d argue that R is not an intrinsic property and that the results here emphasize that finding. We’d expect it to be different in different populations and when observed at different scales. At best, we’d hope that for very large similar populations, it might be the same absolute values. It probably has better use as a comparator from year to year for the same population.

Response: We have reworded this part of the discussion section to reflect the reviewer’s comments above in lines 296–315 in the track-changes version.

348: one could comment on the dangers of great analysis on a small dataset for the 77 British boarding school.

Response: We have added this comment as one explanation for the large reproductive number reported from the British boarding schools (lines 372–378 in the track-changes version).

Tables: can they be made available as a csv? Its great work and others might use it.

Response: We have provided the tables as .csv files for inclusion as additional materials on the website.