Reviewer's report

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

Version: 2 Date: 9 May 2014

Reviewer: Anne AC Cori

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?

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”.

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).

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.

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?

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.

6. Results about the boarding school in 1977/78: could the authors comment on possible reason for such high R estimates in that case?

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.

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.

Minor issues not for publication

1. Line 119 “of R reproduction number” should be “of R” or “of the reproduction number”

2. Line 185 “because one” should be “because of one”

3. Line 255 I suggest removing the sentence on confined settings, which is then repeated and discussed on line 339.

4. Line 330: “estimation of” should be “estimation of R”

5. In all Table notes: “The generation time” should be “The mean generation time”

6. In all Figure legends: why do you have “Confidence Limit” and “Lower CI”? What is the difference? Please clarify.
7. Table 6: why is there no R estimate presented at all for study [131]?

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests