Response to Revision Requests

Reviewer #1 Comments

I suggest the authors do a thorough search of the similar studies and summarize them in a table to highlight the similarities and differences among different studies.

*Thank you for your comments. We have organized and summarized a table of published or preprinted estimates of the Wuhan infections in our revision (**Table 2**). We have highlighted our estimated number of infections is on the higher side of the estimates but not out of proportion.*

Reviewer #2 Comments

1) The introduction sets the topic and explains the outbreak of Covid-19 and the measures implemented in Wuhan. However, it does not say how it contributes to the previous literature and the importance of this study. Covid-19 is still a new virus but there are already lots of paper analyzing it and there is previous literature in epidemics that can help you position your paper on the literature.

*We agree with your comments about positioning our paper in regard to other estimates. We discuss the value added of our manuscript in the revised introduction (lines 73-76). We have also included a summary of the literature to highlight how our manuscript differs from previous publications (**Table 2**).*

2) The objective of the paper needs to be clearly stated in the introduction.

We have added the objective of our paper to the introduction (lines 65-66).

3) The main variable is "total cases with travel history to China" – you need to define it properly in the text. Moreover, the data source on all the variables is needed. Provide a table with all the relevant information and cases.

*We have added definitions to the variables (lines 80-82, 91-94, 99-100). We have added **Table 1** with all the information used to graph figure 2 and 3. **Supplementary Table 1** has a list of data complied used to graph Figure 1A. **Supplementary Table 2** has a list of data and sources to plot Figure 1B.*

4) The figures have a very low quality.

We apologize when we observed that figures were low quality upon submission. We have high quality figures and hopefully they will appear so in our upload.

5) Equation 1 – this is not the best way to specify a model. You need to show the dependent variable and the covariates with their coefficients and the error term. Notation is important to follow the paper.

We have modified our expression of our Equation 1 to best address this issue (105-112).

6) The methods are not clear. They are misleading and they don't really match what you do in the results. It is important that you rewrite the methods section. The results are interesting, but they need to be well supported by the methods.

We have made large changes and additions to the methods. We hope this clarified how we obtained our results.

7) I would suggest that you say only once the software used (excel or R).

We have made this change and deleted redundant software references.

8) While the journal requires you to publish the dataset, I would suggest you create nice tables out of the two excels that you are uploading as supplementary material. They could be very informative. You need to show a descriptive table – number of cases, exported, travelers…etc.

*We agree and have added **Table 1** with the data sources listed to address this comment.*

9) It looks like from the data, that most of the infected cases traveled 4 days before the ban – however, you have a very small sample. Moreover, the incubation period is not known, so these people could have been infected long before traveling.

We agree the sample size from Japan, Singapore, and Korea is not large but we believe it is appropriate and sufficient for sampling purpose (40/~140). The incubation period of these infected travelers is not known and we presume that they were infected before leaving Wuhan. Therefore, we used data weeks after the travel ban to make sure we had the full dataset of exported infected cases. We also mention that Chinazzi's study using a larger exported dataset had a similar depiction of a critical period of ~5 days in their figure (Lines 185-189).

10) You analyze the exports from Wuhan, but since the outbreak and before the ban, people in China were moving around. Flights from other regions in China to different countries could also carry infected people. Could that be checked too?

We performed analysis on Wuhan travelers because it was a significant contributor to exported cases. Theoretically, we can do the same analysis on other cities if they had a large sample size of exported cases but it does not appear that other cities contributed significantly to the exportation of COVID-19. From the WHO report on February 14th, 2020 [5], there is graph of 200 COVID-19 cases outside China with the suspected method transmission. Exported cases from Hubei province account for over 80% of exported cases from China, so it can be assumed the city of Wuhan accounts for the majority of all exported cases from China.

11) In the discussion, you talk about a model, but the model is not really explained in the methods section.

We agree and have elaborated on the realistic scenario and model scenario in our methods (lines 119-125).

12) While the discussion is interesting, you need to write down the main limitations. You are looking at the correlation of two variables – so, you might have omitted variable bias. You don't know the incubation period. And other data limitations that you faced during the study.

We agree and have added many more limitations we faced in our discussions. Some of these include asymptomatic infected travelers, COVID-19 testing methodology, and underreporting (lines 200-203). We do not know the incubation period of the virus in the travelers although we cite literature regarding the COVID-19 incubation period (line 142). However, we do not use the incubation period in our methods except that we account for the fact that there is a delay between the travel ban and the detection of exported cases (lines 245-249).

13) The paper would benefit from a full rewording – in many cases it feels like the authors are listing concepts (mainly in the methods and results). The text should flow.

We agree and have made significant changes to the methods and results sections.

Editors Comments

1. Review of the literature: table with other papers examining similar issues. (R1#1, R2#1)

*We agree and have added **Table 2** to compare our results with other estimates of the Wuhan infected numbers.*

2. Data sources table and output table R2#3 R2#8

*We agree and added **Table 1** to address this. Additionally, we have made changes to Supplementary Tables to better display our data and sources.*

3. Clarify methods and models. R2#5, R2#6, R2#12

We agree and have made changes to our methods section to address this. We have added clarification to variables and our equation (lines 80-82, 91-94, 99-100, 105-112). We have added an explanation of the models depicted in the supplementary figure (lines 116-122).

4. Discussion: Please focus your discussion on your results and your limitations. One of them is the incubation period, which according to Bati et al (2020) the median incubation period was estimated to be 5.1 days.

We agree and have added more discussion on our results and of the study limitations. We included the incubation period as a reason our dataset may be more accurate in estimating the infected population in Wuhan at the time of the travel ban compared to previous studies which were published around the time of the travel ban (lines 240-249). We discuss the impact asymptomatic travelers may have on our study (lines 242-245)

5. Future lines of research: Please suggest, based on your findings and limitations, what/how should researchers continue investigating. For example, R2#11 is a relevant issue for further research.

We have suggested in our discussion that government, researchers, and healthcare workers should reassess the infectivity of COVID-19 in an unrestricted environment (lines 253-257). Furthermore, these estimates and research are also useful for future use in case of a new disease outbreak.