Peer Review File

Article Information: http://dx.doi.org/10.21037/atm-20-5248

Reviewer A

Comment 1:

The manuscript deals with the estimation of hospital bed shortage for patients with mild, severe and critical COVID-19 infections, considering the eects of underreporting and diagnosis delay. To this end, they establish an SEIR-based compartmental model to simulate the inter-city transmission of SARS-CoV-2 viruses, driven by the population inow/outow between the epicenter Wuhan and 50 other cities in mainland China. The authors consider different scenarios with an earlier or later lockdown and different decreases of tra_x001D_c volume.

The authors also studied the possibility of a second wave of the epidemic, using different scenarios for the speed of reducing the control measures in Wuhan. The simulations suggest that a second wave in Wuhan could have occurred in May if social distancing measures had been lifted at the beginning of April, however, as this second wave is estimated smaller than the first one, the hospital bed shortage is supposedly much smaller than in the case of the first wave.

I think that the paper is interesting and deals with an important issue. The manuscript is well written and well structured. It seems to me that a very thorough work was done in the evaluation of data using posts on the Weibo platform.

Unfortunately, there seem to be some incorrect characters (probably due to PDF conversion) in model (1) and its description which hinder the understanding of the notations and the structure of the model. Probably because of this reason, I cannot understand the last three lines of (1) either. It is also not clear to me how q appears as a multiplier in the equation for R0 while one divides by the same q in the equation for D0 . In general, a bit more detailed description of the model would be welcome.

Before I can suggest the paper for publication, I would be happy to see the mathematical part without the bad characters and thus clearly understand the model.

Response:

We thank the reviewer for positive comments and insightful suggestions. We now add more details of the model in the main text. We revise the formula in model (1) and add more descriptions of the model structure accordingly. The last three lines of the model (1) describe the inflow and outflow of population across the Susceptible, Exposed and Infectious stages in each city. $\vec{J}$ is the city-specific transmission rate to adjust for newly imported cases from other cities. In equation for $\vec{J}$, δ is the transmission rate change during the transportation process. The multiplier q in the previous version of the equation for $\vec{D}'$ is a typo, which should be the denominator instead. We revise this in the revised manuscript. We also use $\vec{N}'$(the daily

population change of each city) to replace $\vec{S}_{in}, \vec{E}_{in}, \vec{I}_{in}$ to simplify the model (P.2 line.39 in the supplementary material).

The following sentence has been added to the main text to clarify the above point (P.4 line.92-96):

"We built a susceptible-exposed-infected-recovered (SEIR) model to simulate the inter-city transmission of SARS-CoV-2 viruses, considering of underreport and test delay in the early stage of the epidemic. We also incorporated into the model the implementation of individual and governmental control measures, the population inflow/outflow between the epicenter Wuhan and 50 other cities in mainland China, and a higher transmission risk in public transportation."

Reviewer B

Major comments

Comment 1:

The assumption that the reporting rate increased from 5% to 100% linearly is far too arbitrary, and this choice absolutely determines the results. First, assuming that the reporting rate of infections can be as high as 100% is far too optimistic and disregards asymptomatic and mild cases. Second, the authors could come up with a better idea of how to estimate the reporting rate dynamically (e.g. by using exported cases as in Verity et al, Lancet or the age structure of cases and deaths, see Hauser et al, Plos Medicine). Another solution would be to introduce uncertainty on this trend (with multiple values and shapes for the increase in reporting rate), and propagate this uncertainty in the results (and not just add sensitivity analyses in the appendix). The results as they are presented now are too certain (uncertainty intervals are extremely narrow considering the data available) and are entirely driven by this very strong assumption.

There is a complete disregard for uncertainty in most parameters values used in the compartmental model. The authors need to include uncertainty on all parameter values (reporting rates as stated above, but also delays, R_0, transmission rates, exposure, and case fatality ratio) as they did for p, delta and I0, and propagate this uncertainty in all results (not just in the appendix).

Response:

We agree with the reviewer that the 100% reporting rate might not hold during the early phase of the COVID-19 pandemic. Although it is impossible to know the true reporting rate, previous studies have provided the reasonable range, as the reviewer pointed out. We now use the reporting rate estimated by other studies (1-5) and the public available data. Specifically, we assumed five reporting rates at different date points prior to 18 February 2020. We believe that the date points are sufficient to describe the nonlinear increasing trend of reporting rates due to gradually increased testing capacities. Between each time point, we simply assume a linear increase in reporting rates. As suggested by the reviewer, we also incorporate the uncertainty in other factors (such as delay of report, R_0, transmission rate, incubation period, infectious period, and others) to give wider confidence intervals in our model

estimates.
We have updated the results in Appendix Table 1 and 4. The confidence intervals are also added to Figure 1, 2 and Appendix Figure 4-6.
We added the following sentences about the assumptions of reporting rates to the main text (P.4 line.101-105):
"Underreport of COVID-19 cases in the early pandemic has been widely reported but it is nearly impossible to get the true reporting rates. We therefore estimated the reporting rates according to previous studies (12-16) and the public available data: 1.8% on January 3, 3.0% on January 18, 14.0 % on January 23, 34.0% on February 8 and 35.3% on February 18 (Table 1). The daily reporting rates between these dates were interpolated by a linear regression."

Comment 2:
The model should be stratified by age, which is not only fundamental for accurate predictions (as shown in Pellis et al, Systematic selection between age and household structure for models aimed at emerging epidemic predictions, Nature Com 2020) but is also very important considering the age-dependency of the probabilities of complication, hospitalisation and death.
Response
According to the guidelines by the National Health Commissions of China, the admission policy was mainly determined by severity of COVID-19 infections. Moreover, all the age groups are susceptible to the novel SARS-CoV-2 virus due to the lack of the preexisting immunity. Hence, like most SEIR models developed by other researchers (6-8), our SEIR model did not consider age structure, but estimated the admission risks by mild, severe and critical infections.

Comment 3:
The proposed use of social media only to estimate the lack of hospital beds for non-COVID patients is a good idea but absolutely not validated in any way. I don't find the arguments from the authors convincing in that regard. The authors should discuss thoroughly all limitations of this approach, and refer to previous works using social media to estimate hospitalisation needs.
The use of social media to estimate non-COVID hospital bed shortage is unconvincing, and could only be acceptable if framed as a secondary objective of the study, insisting on its exploratory nature.
Response:
We agree with the reviewer that using social media to estimate the hospital bed needs of non-COVID patients could an exploratory approach rather than a confirmatory one, although this approach could be the best we can do at this stage to address this important research question. Nevertheless, we have discussed the limitation of using social media posts to estimate the bed shortage of non-COVID patients in our previous manuscript. We now explicitly point out its exploratory nature in Discussion, and make a more conservative conclusion about the non-COVID in the revised manuscript.

We added more information in Methods (P.5 line.116-118):
"*Previous work has demonstrated the feasibility of using social media posts and news to monitor and assess unexpected disease outbreaks (17,18). Here we proposed a simple model utilizing social media posts to estimate hospital bed shortages of non-COVID-19 patients.*"
The following sentences have been added to Discussion (P.10 line.279-283):
"...It is of note that our estimates of hospital bed shortages for non-COVID-19 patients might need a cautious interpretation. We assumed that the probability of sending posts to seek help via social media in non-COVID-19 patients was the same as that in COVID-19 cases, which might not hold due to different risk perceptions and media attention to the COVID-19 outbreak and chronic diseases."
The following sentence is added to Abstract – Conclusions:
"The healthcare needs of non-COVID-19 patients in the pandemic warrant more investigations."

Comment 4:
The simulation of scenarios with different timings and magnitude for the implementation of control measures is interesting, but is far too overconfident. The authors should revise this part and include stochastic uncertainty and parameter uncertainty to obtain acceptable prediction intervals on their results.
Response:
We thank the reviewer for positive comments and insightful suggestions. As suggested, we incorporate the stochastic and parameter uncertainties into the model and update the results with wider and more reasonable prediction intervals.
We have updated the results in Table 2 and Appendix Figure 4.

Comment 5:
The assessment of the second wave is also interesting, but I have the same strong concerns about the inclusion of uncertainty in the predictions. It has been shown particularly important to account for stochastic uncertainty (eg, using stochastic models) when assessing the risk of emergence or reemergence, or more generally when modelling the spread of an infectious disease at low levels of incidence.
Response:
We thank the reviewer for positive comments and insightful suggestions. As suggested, we incorporate the stochastic uncertainty into the model and update the results with wider and more reasonable prediction intervals.
We have modified the results in Appendix Figure 5.

Comment 6:
It is unclear why the implementation of traffic restrictions in Wuhan would lead to more cases in Wuhan soon after the restriction.
Response:
We speculate that there could have been two reasons for immediate surge of cases in Wuhan after the lockdown. First, as reported in news, many people rushed to the

designated hospitals and waited for a few hours to get throat swabs for lab tests. This could have increased the transmissions within the city. Second, all the infected cases in latent period were forced to stay in the city due to traffic restrictions, leading to more sources of transmission chains.

We added the above speculations in Discussion (P9, line 249-253):

"This immediate surge could have been caused by two reasons. First, many people rushed to the designated hospitals and waited for a few hours to get throat swabs for lab tests. This could have increased the transmissions within the city. Second, all the infected cases in their latent period were forced to stay in the city due to traffic restrictions, leading to more sources of transmission."

Minor comments

Comment 7:

Abstract, background: please be more specific about the objectives of the study, which should have more justification than "few studies have investigated".

Response:

We added the following sentences to Abstract, as suggested by the reviewer:

"The global pandemic of COVID-19 first emerged in Wuhan, China since December 2019 and the lockdown of Wuhan city for 76 days has successfully contained the first wave. However, to date few studies have evaluated the hospital bed shortage for COVID-19 and non-COVID-19 patients during the lockdown, although such data are important for better preparedness of the second wave."

Comment 8:

Abstract, methods (and elsewhere): COVID-19 is the disease that is caused by SARS-CoV-2. It is thus improper to write "COVID-19 infection". It should be replaced by "SARS-CoV-2 infection" or just "COVID-19" if the authors are talking about the consequences of infection and not the virus.

Response:

We have revised these parts accordingly.

Comment 9:

Introduction, paragraph 1: The authors should use another reference than a news site in Chinese to support the claim that 3 trillion trips occurred during that period.

Response:

We have replaced the reference with a journal article (9).

Comment 10:

Introduction, paragraph 2: The statement that "reported cases rapidly increased in the early stage of the lockdown" is imprecise, as reported cases already started increasing before January 24.

Response:

We have revised the text (P3, line 60-61) into "*The rapidly increased cases soon overburdened the healthcare system in Wuhan in the early stage of the outbreak…*".

Comment 11:
Introduction, paragraph 2: The statement that "mental stress exacerbates underlying conditions" is too unprecise (what kind of conditions?) and should be supported by references.
Response:
We clarified this point by revising the sentence into (P3, line 65-66) "*Moreover, it was found that the lockdown could have increased anxiety and stress of local residents, thereby increasing their susceptibility to infection (5,6).*" Two references have been added to support this statement.

Comment 12:
Introduction, paragraph 2: The statement that "In addition, cases exported from Wuhan spread fast in other cities in late January, suggesting that when the lockdown policy was enforced, the best timing may have already passed." is unclear. Are you talking about cities in China or abroad? Are you talking about the lockdown in Wuhan or other areas? You seem to disregard the incubation period, it is also possible that lockdown was efficient but that cases exported before the lockdown in Wuhan led to reported cases in the following weeks.
Response:
We are sorry for the confusion caused. The statement has been revised into (P3, line 66-68):
"In addition, many cases exported from Wuhan had been reported in other cities in China and overseas weeks before the lockdown, suggesting the need to evaluate the optimal timing of lockdown implementation."

Comment 13:
Introduction, paragraph 3: "Non-COVID-19 patients means those who had acute or chronic diseases unrelated to COVID-19 infection." This definition should come earlier.
Response:
As suggested by the reviewer, we have moved the definition of Non-COVID-19 patients to paragraph 2 (P3, line 63-65).

Comment 14:
Methods, Mathematical modeling: The description of the model in the main text is lacking, at least a summary of the structure and main assumptions should be given here.
Response:
We added more sentences to the main text (P4, line 92-101)
"We built a susceptible-exposed-infected-recovered (SEIR) model to simulate the inter-city transmission of SARS-CoV-2 viruses, considering of under-reporting and test delay in the early stage of the epidemic. We also incorporated into the model the implementation of individual and governmental control measures, the population

inflow/outflow between the epicenter Wuhan and 50 other cities in mainland China, and a higher transmission risk in public transportation"
"We assumed that the lockdown and traffic restrictions had reduced 99% of the population flow from and within Wuhan since January 23. The daily number of diagnosed cases of SARS-CoV-2 infection in China was assumed to follow a Poisson distribution."

Comment 15:
Methods, Simulating scenarios: The data and code should be made public (eg on github) and not only available upon request.
Response:
We have uploaded the code and data to Github. A statement has been added in the main text (P.11 line153-154):
"The datasets and codes used in this study can be found at https://github.com/Larryzza/The-shortage-of-hospital-beds-for-COVID-19-Wuhan."

Comment 16:
Results, paragraph 1: The statement that "the model fits well" cannot be made without including uncertainty intervals in figure 1. Also the authors should add uncertainty intervals to all results.
Response:
We added confidence intervals to all the results in Figure 1.

Comment 17:
Appendix Table 2 and 3: not shown properly.
Response:
We have modified Table 2, 3 and Figure 1 accordingly.