natureresearch

Peer Review File

Manuscript Title: Effect of non-pharmaceutical interventions to contain COVID-19 in China

Reviewer Comments & Author Rebuttals

.

Reviewer Reports on the Initial Version:

Referee #1 (Remarks to the Author):

This study describes a mathematical model used to interpret the course of the COVID-19 epidemic in China. The authors also use the model to explore the impact of a set of interventions and a conduct a what-if analysis of differences in when these were initiated.

This is clearly an important study providing information of use to other countries combatting COVID-19. I am very pleased that the authors pre-released their work so it is already publicly available.

The study is in part a fairly standard exercise in spatially structured SEIR modelling. But the data available from Baidu to estimate movement rates are truly remarkable. The city level and county level data are 6 years old (which is a slight worry, but far better than most) but there is also a real-time data set at the city level.

My major concerns are as follows.

- I could not find a clear statement of how the model was fitted/parameterized. The fits to the shape of the epidemic curve (Figure 2) are good (though not perfect) so I assume there must have been some kind of calibration. Or did they just input pre-existing parameter estimates and accept a need for re-scaling (especially in Wuhan, by a factor of 3 or so). What is their explanation for this re-scaling and, crucially, are they assuming that it is constant in time, which seems intuitively highly unlikely?

- My biggest concern is with their no intervention scenario. In Table 1 it is reported that in the absence of interventions the epidemic across China would have been 67.3 times larger (with an uncertainty range). That equates to 5.37M people, a tiny fraction of the population of China. With R0=2 we would expect (very crudely) the majority of the population to become infected. This is a huge discrepancy and needs a full explanation.

- I was initially concerned about the assumption that this is a SEIR process (complete, permanent immunity is by far the best case scenario). However, given my comment above I am not sure that herd immunity can play much of a role in the dynamics of the simulated epidemic.

- The study's conclusions about the benefits of earlier interventions follow directly from their analysis and there is nothing surprising there. But it is extremely important that the authors make clear that this conclusion applies only to the first wave and/or the interventions are permanent. The

authors seem to accept the latter point in the Discussion (LL220-223).
Once you start to consider the longer term dynamics the picture will change dramatically. On this point I think the authors do not properly explore the implications of their results.

- That said, there is merit in early intervention if the objective is eradiation of the virus at national level. However, the authors are (correctly in my view) cautious about eradication. So this does not seem to be their motivation for recommending early intervention and my preceding comment still applies.

Minor points.
L46 and L49. The authors are mostly admirably diligent about giving ranges to estimates, but not here.

LL57-60. May as well update these figures if the paper is to be published.

L100. The point about economic and social costs is tremendously important. But simply to put "high" seems inadequate. Do the authors have any further data on these costs?

Table/Figure numbering wayward in places.

L136. What was the interval between implementation and epidemic peak? That's a key output.

Figure 2. I am unclear about the nature of the case data. The data I have seen include a very prominent spike where there was an adjustment of case definition. That is not visible here. Indeed, the reported data look remarkably clean.

L220. "Some" is too vague. How many?

L249. The claim "was likely to be detected" runs counter to a preprint published by Imperial College which estimates very large numbers of undetected infections. This discrepancy is important and merits discussion.

LL416-421. How important is this rescaling to the results? It sounds like an ad hoc solution to a messy technical issue.

L439. I imagine that initiating the epidemic with I(0)=5 on 01/12/19 affects the results. How was this arrived at? If by trial and error to get an epidemic peak at the right time (which it does, Fig 2) then we should be told.

Referee #4 (Remarks to the Author):

I've reviewed the manuscript titled "Effectiveness of non-pharmaceutical interventions to contain COVID-19 in China," which is an interesting and timely account of this current public health crisis. The manuscript introduces a travel network-based stochastic SEIR metapopulation model across 340 prefecture-level cities in China. The aim is to estimate the effectiveness of three types of NPI interventions: inter-city travel restrictions, early identification and isolation of cases, and contact restrictions and social distancing measures. The authors address various types of what-if or counterfactual scenarios, such the effectiveness of each type of intervention on its own and the impact of intervention timing on the epidemic. The model is trained on two types of data: mobility data obtained from cell phones and epidemiological data obtained from various sources. The research has been carefully carried out, and the manuscript reads very well. The results are impressive, and will no doubt be useful as countries attempt different strategies to control and contain the epidemic. The figures are informative, and the methods section and the supplement provide useful technical details and additional results. The code used in the simulations is available on GitHub, which improves reproducibility of the work. In my view, this is an excellent manuscript.

In the following, I raise a few minor points the authors should address:

1. State the source of mobility data, i.e., cell phones, early in the manuscript; this will help readers better assess potential biases and confounders that may be present.

2. Is every cell phone owner visible in the data or only smartphone owners? According to Wikipedia, smartphone penetration in China is around 55%. Whether or not that's the most accurate number, it would be useful to mention the number in the manuscript. Are there significant differences between the characteristics (e.g., socioeconomic status) between smartphone owners and non-owners? If so, would you expect smartphone owners to be more likely to travel than non-owners? If so, this would be expected to bias mobility estimates upwards. Some discussion about this would be insightful.

3. Around line 255, the authors discuss the possibility of mobility being affected by unknown seasonal factors, such as temperature and absolute humidity. Is there any evidence in the data the authors use to suggest this is the case? I assume the data set doesn't cover long enough time period to make this assessment.

4. In the data sources section, around line 314, the authors discuss three different types of location / mobility data. Are the locations based on cell tower triangulation, Wi-Fi positioning, GPS, or some other approach to determining location? What is the average frequency of location ascertainment?

5. Around line 390, the authors discuss simulating the number of exposed individuals within a given patch on a given day through a random draw from the Poisson distribution for each infectious person. There's however a large body of work showing that the number of social contacts people have is not as homogeneous as the Poisson distribution assumes, but instead follows some type of fat-tailed distribution, i.e., a small number of people have a very large number of contacts. I assume that using a distribution with fat tails would increase variability and hence uncertainty in the simulation results. Are the results robust to the choice of using Poisson vs. some more heterogeneous distribution? (As a side note, it appears that the number of exposed individuals is drawn independently for each individual from a Poisson; if that's the case, note that the sum of

independent Poisson distributions is a Poisson distribution, which might simplify the simulation.)

6. Figure 1 might benefit from a point-wise mean or median curve as a summary of the overall effectiveness of travel restrictions between the two groups (restricted vs. unrestricted).

7. In Figure 3, can you please clarify (if this interpretation is correct) that panel A is empirical data and panel B is simulated data?

I have no major concerns about this manuscript, and I'm strongly in favor of its publication.

Author Rebuttals to Initial Comments:

Referee #1 (Remarks to the Author):

This study describes a mathematical model used to interpret the course of the COVID-19 epidemic in China. The authors also use the model to explore the impact of a set of interventions and a conduct a what-if analysis of differences in when these were initiated.

This is clearly an important study providing information of use to other countries combatting COVID-19. I am very pleased that the authors pre-released their work so it is already publicly available.

The study is in part a fairly standard exercise in spatially structured SEIR modelling. But the data available from Baidu to estimate movement rates are truly remarkable. The city level and county level data are 6 years old (which is a slight worry, but far better than most) but there is also a real-time data set at the city level.

My major concerns are as follows.

- I could not find a clear statement of how the model was fitted/parameterized. The fits to the shape of the epidemic curve (Figure 2) are good (though not perfect) so I assume there must have been some kind of calibration. Or did they just input pre-existing parameter estimates and accept a need for re-scaling (especially in Wuhan, by a factor of 3 or so). What is their explanation for this re-scaling and, crucially, are they assuming that it is constant in time, which seems intuitively highly unlikely?

Re: The parameterisation of the model is detailed in the Data analysis section of the Methods. As we only used the epidemiological parameters estimated for the early stage of the outbreak in Wuhan and NPI indicators derived from both population movement data and case reporting delays, technically we did not calibrate or fit the SEIR model by using a re-scaling factor to the reported epidemic curve. We simply calculated the correlation between estimates from the model and the daily number of reported cases using R-square and P values by region. In terms of spatial accuracy, we also calculated the sensitivity and specificity of this model to highlight how well the model can potential predict the geographical distribution of the affected area (those with at least one case), compared with the actual distribution of cities with reported COVID-19 cases. However, the parameters did affect model estimates, especially R0, so we conducted a series of analyses using various R0 values and varying other parameters, with the results provided in the Extended Data Figs 3 to 9.

- My biggest concern is with their no intervention scenario. In Table 1 it is reported that in the absence of interventions the epidemic across China would have been 67.3 times larger (with an uncertainty range). That equates to 5.37M people, a tiny fraction of the population of China. With R0=2 we would expect (very crudely) the majority of the population to become infected. This is a huge discrepancy and needs a full explanation.

Re: The 67.3 times larger number of cases under the no intervention scenario, presented in Table 1, is the estimate as of February 29, 2020. Certainly, the total number of infections without interventions over the longer term would reach a high fraction of the Chinese population. We have clarified this in the Results.

- I was initially concerned about the assumption that this is a SEIR process (complete, permanent immunity is by far the best case scenario). However, given my comment above I am not sure that herd immunity can play much of a role in the dynamics of the simulated epidemic.

Re: Yes. It is a potential limitation of the SEIR model if there isn't permanent immunity, although most infections from pathogens can potentially generate short-term or permanent immunity. Because our simulation runs were not extended beyond five months and the susceptible populations are still large enough, we did not include the addition of new susceptibles, or conversion of recovered people back to susceptibles. We added this information to the Methods section.

- The study's conclusions about the benefits of earlier interventions follow directly from their analysis and there is nothing surprising there. But it is extremely important that the authors make clear that this conclusion applies only to the first wave and/or the interventions are permanent. The authors seem to accept the latter point in the Discussion (LL220-223).

Once you start to consider the longer term dynamics the picture will change dramatically. On this point I think the authors do not properly explore the implications of their results.

- That said, there is merit in early intervention if the objective is eradiation of the virus at national level. However, the authors are (correctly in my view) cautious about eradication. So this does not seem to be their motivation for recommending early intervention and my preceding comment still applies.

Re: Thanks for these comments. As we only simulated the outbreak and NPIs for five months, this study focussed on the control for the first wave of the virus. However, continuing the implementation of interventions, e.g. early case identification and isolation, will also help to prevent and/or delay the arrival of the next wave, considering the increasing numbers of cases imported from other countries and the presence of asymptomatic or subclinical infections found in China. We have added this information to the Discussion.

Minor points.

L46 and L49. The authors are mostly admirably diligent about giving ranges to estimates, but not here.

Re: Thanks. Due to the word count limit, we removed this sentence and these estimates from the Summary.

LL57-60. May as well update these figures if the paper is to be published.

Re: Done.

L100. The point about economic and social costs is tremendously important. But simply to put "high" seems inadequate. Do the authors have any further data on these costs?

Re: It is important, but currently we haven't found any quantitative information on these costs in China.

Table/Figure numbering wayward in places.

Re: We have re-numbered the table and figures.

L136. What was the interval between implementation and epidemic peak? That's a key output.

Re: Agreed. We added the intervals between the implementation of NPIs and epidemic peak in the Extended Data Table 1 and the Results in main text.

Figure 2. I am unclear about the nature of the case data. The data I have seen include a very prominent spike where there was an adjustment of case definition. That is not visible here. Indeed, the reported data look remarkably clean.

Re: The epicurves were plotted based on the date of illness onset. Indeed the case definition has been adjusted and a large number of clinically diagnosed cases before laboratory confirmation have been retrospectively reported into the information system since 12 February. We also found that there was an abnormal increase of cases in Wuhan City and Hubei Province on a single date of February 1, 2020. We discussed this issue and underlying causes (e.g. reporting delay, system error, incorrect reporting of the onset date) with epidemiologists in China, and this spike might not represent the actual infection patterns, but exact reasons remain unclear. Therefore, we interpolated the number on February 1 by using the mean of the numbers of cases reported on January 31 and February 2 in the epicurves of Wuhan and Hubei Province. We have described this in the Methods section.

L220. "Some" is too vague. How many?

Re: We have rephased the sentence.

L249. The claim "was likely to be detected" runs counter to a preprint published by Imperial College which estimates very large numbers of undetected infections. This discrepancy is important and merits discussion.

Re: Based on Table 1, we also found that 38% of cases were undetected in Wuhan and 24% undetected in other provinces. We have removed this sentence from the Discussion.

LL416-421. How important is this rescaling to the results? It sounds like an ad hoc solution to a messy technical issue.

Re: We deliberately took this approach to facilitate movement in a completely data-driven way. Using other methods of deriving movement between areas may have introduced distributional assumptions on the movement probabilities that we wanted to avoid. Notably, because we rescaled using the total number of infected people in each patch, but retained the relative values of people moving between areas, the rescaling step only serves to revert movement patterns slightly back towards the mean. Therefore, this rescaling only reduces some stochasticity introduced by the movement function, and should not bias the movement patterns in any particular direction.

Because the error between the sum of people initially predicted to move and the actual number of infected people in each patch was generally small, this rescaling generally only affected numbers of movers during the initial stages of the epidemic; once the number of infected people surpassed 2500, the amount of overestimated/underestimated movement per day through this function was <5%. As noted, accounting for this over/underestimation does not affect the relative patterns, and should only serve to reduce the stochasticity introduced by movement.

L439. I imagine that initiating the epidemic with I(0)=5 on 01/12/19 affects the results. How was this arrived at? If by trial and error to get an epidemic peak at the right time (which it does, Fig 2) then we should be told.

Re: We *a priori* chose 5 individuals as the starting number of infected people to reduce the likelihood of the epidemic dying out from stochastic recovery events before the epidemic was able to spread to sufficient numbers of new people. Transmission parameters were selected from other studies using nearly identical model formulations that were fit using very similar data from across China (particularly Li et al 2020), while making similar assumptions about the date of initial infection (this was in particular additionally informed through the earliest case of illness onset being Dec 2, 2019) . As a result, while we did not need to use trial and error to obtain the correct initial conditions that yielded the correct timing of the epidemic peak, we also did not choose these values purely fortuitously, as the usage of a similar model structure to previous papers meant we could use parameter values and initial conditions in a literature-informed way that we could trust would provide reasonably accurate results about the actual epidemic progression.

Referee #4 (Remarks to the Author):

I've reviewed the manuscript titled "Effectiveness of non-pharmaceutical interventions to contain COVID-19 in China," which is an interesting and timely account of this current public health crisis. The manuscript introduces a travel network-based stochastic SEIR metapopulation model across 340 prefecture-level cities in China. The aim is to estimate the effectiveness of three types of NPI interventions: inter-city travel restrictions, early identification and isolation of cases, and contact restrictions and social distancing measures. The authors address various types of what-if or counterfactual scenarios, such the effectiveness of each type of intervention on its own and the impact of intervention timing on the epidemic. The model is trained on two types of data: mobility data obtained from cell phones and epidemiological data obtained from various sources. The research has been carefully carried out, and the manuscript reads very well. The results are impressive, and will no doubt be useful as countries attempt different strategies to control and contain the epidemic. The figures are informative, and the methods section and the supplement provide useful technical details and additional results. The code used in the simulations is available on GitHub, which improves reproducibility of the work. In my view, this is an excellent manuscript.

In the following, I raise a few minor points the authors should address:

1. State the source of mobility data, i.e., cell phones, early in the manuscript; this will help readers better assess potential biases and confounders that may be present.

Re: We now briefly describe the source of mobility data in the first paragraph of Results.

2. Is every cell phone owner visible in the data or only smartphone owners? According to Wikipedia, smartphone penetration in China is around 55%. Whether or not that's the most accurate number, it would be useful to mention the number in the manuscript. Are there significant differences between the characteristics (e.g., socioeconomic status) between smartphone owners and non-owners? If so, would you expect smartphone owners to be more likely to travel than non-owners? If so, this would be expected to bias mobility estimates upwards. Some discussion about this would be insightful.

Re: The mobility data only included the smartphone owners who used the apps using the Baidu location-based services. However, we don't have data on the number of smartphone owners in China and the differences between the characteristics of smartphone owners and non-owners. We have added smartphone penetration numbers to the manuscript and discussed the potential biases of smartphone coverage on our estimates in the limitations section.

3. Around line 255, the authors discuss the possibility of mobility being affected by unknown seasonal factors, such as temperature and absolute humidity. Is there any evidence in the data the authors use to suggest this is the case? I assume the data set doesn't cover long enough time period to make this assessment.

Re: We discussed that the decrease of COVID-19 in China might be also partially attributable to other factors, such as temperature and absolute humidity explored by other studies. But we didn't investigate this seasonal impact, as the long-term observation of these factors on COVID-19 are not available currently. Further investigations using data for a long enough time period are needed to measure the effects of these seasonal factors.

4. In the data sources section, around line 314, the authors discuss three different types of location / mobility data. Are the locations based on cell tower triangulation, Wi-Fi positioning, GPS, or some other approach to determining location? What is the average frequency of location ascertainment?

Re: According to Baidu (http://privacy.baidu.com/detail?id=288), the locations of smartphones are determined by the combination of data from GPS, IP address, Wi-Fi access points, and cell towers, and Baidu location-based service (LBS) provides over 7 billion positioning requests per day across China, but we don't have the data on average frequency of location ascertainment.

5. Around line 390, the authors discuss simulating the number of exposed individuals within a given patch on a given day through a random draw from the Poisson distribution for each infectious person. There's however a large body of work showing that the number of social contacts people have is not as homogeneous as the Poisson distribution assumes, but instead follows some type of fat-tailed distribution, i.e., a small number of people have a very large number of contacts. I assume that using a distribution with fat tails would increase variability and hence uncertainty in the simulation results. Are the results robust to the choice of using Poisson vs. some more heterogeneous distribution? (As a side note, it appears that the number of exposed individuals is drawn independently for each individual from a Poisson; if that's the case, note that the sum of independent Poisson distributions is a Poisson distribution, which might simplify the simulation.)

Re: We agree with this point that social networks have much longer tails than the Poisson distribution would imply. While we feel it is out of the scope of the manuscript as-is to compare the effect of different distributional assumptions, we believe the effect of this distributional assumption beyond the initial stages of the epidemic likely had a limited effect on variability in the results. This is partly because population sizes in each patch were fairly large (3.2M median population size); because each population was well-mixed, we believe these large population sizes would likely make the effect of few people having large numbers of contacts revert towards the mean.

6. Figure 1 might benefit from a point-wise mean or median curve as a summary of the overall effectiveness of travel restrictions between the two groups (restricted vs. unrestricted).

Re: Agreed. Figure 1a has been revised to present the overall movement patterns for all cities in mainland China, with the median and interquartile range of relative outbound flows shown. However, we kept Figure 1b to present patterns of movement for each city in Hubei Province, as Wuhan shows a different pattern under a longer travel restriction intervention.

7. In Figure 3, can you please clarify (if this interpretation is correct) that panel A is empirical data and panel B is simulated data?

Re: Yes. It is correct. Panel A is based on empirical data and panel B was plotted using simulated data. Figure 3 has been moved and integrated into Extended Data Fig. 2.

I have no major concerns about this manuscript, and I'm strongly in favor of its publication.

Re: Thanks.

Reviewer Reports on the First Revision:

Referee #1 (Remarks to the Author):

The authors have attempted to address the issues raised during the first round of reviews. With mixed success.

It is now even more obvious that this is a very ad hoc analysis indeed. While this may partly be down to the complex and messy data sets available I suspect it is also partly due to this work being done (quite understandably) in considerable haste.

How the central value R0=2.2 was chosen still isn't clear. The authors state, correctly, that their results are very sensitive to the value of R0 (Extended Data). So this is rather unsatisfactory.

That the 67.3x greater result applies only to 20/02/20 is now clear in the text but not in the Abstract.

The estimate of 38% undetected infections is still there. The authors approach to the discrepancy with other estimates (and with emerging serological data) is now simply to ignore it.

Regarding the re-scaling exercise reported in the Methods, I have read their explanation of this several times and I still do not understand what was done and why.

The authors appear now to be saying that their initial value of I(0)=5 was a lucky guess! This is most unsatisfactory.

Referee #4 (Remarks to the Author):

I have reviewed a revision of the manuscript titled “Effect of non-pharmaceutical interventions to contain COVID-19 in China.” In my review of the original submission, I raised only a few minor points of criticism. The authors have successfully addressed these points in their response letter and have added clarifications to the manuscript where needed. I'm strongly in favor of publication.

Author Rebuttals to First Revision:

Referee #1 (Remarks to the Author):

The authors have attempted to address the issues raised during the first round of reviews. With mixed success.

It is now even more obvious that this is a very ad hoc analysis indeed. While this may partly be down to the complex and messy data sets available I suspect it is also partly due to this work being done (quite understandably) in considerable haste.

How the central value R0=2.2 was chosen still isn't clear. The authors state, correctly, that their results are very sensitive to the value of R0 (Extended Data). So this is rather unsatisfactory.

Re: To parameterise the model and simulate the epidemic before widely implementing the NPIs as a baseline scenario, we used the R0 (2.2) estimated by using case data reported by January 22, 2020 at the early stage of the outbreak in Wuhan (Li Q, et al. NEJM 2020). Then using this R0, we could conduct before-and-after comparable analyses to quantify the relative effect of the three major groups of NPIs in China.

That the 67.3x greater result applies only to 20/02/20 is now clear in the text but not in the Abstract.

Re: The date has been added in the Abstract.

The estimate of 38% undetected infections is still there. The authors approach to the discrepancy with other estimates (and with emerging serological data) is now simply to ignore it.

Re: Compared with reported data in Wuhan, our analyses indeed found that 38% cases might be undetected. As discussed in the first point of limitation section in the paper, our simulations were based on parameters estimated for symptomatic cases in Wuhan, our estimates might have underestimated the total number of infections. Additionally, a sero-epidemiological survey has been conducted across China in April, but the data is not yet available.

Regarding the re-scaling exercise reported in the Methods, I have read their explanation of this several times and I still do not understand what was done and why.

Re: The Baidu data recorded the proportion of people who moved between cities across China per day. To recognize the inherent uncertainty and variance in these values, rather than directly implement the proportions to move people between places each day, we used these proportions as daily probabilities of each person moving to each city (including the number of stayers).

Including variance in this way, however, means that the net number of people for each patch each day who move or stay can exceed or be fewer than the total number of people in that patch. We only wanted to introduce stochasticity in the relative patterns of movement, however, not the total number of people who exist/are infected based on this movement process. While it is reasonable to assume there is uncertainty in the relative movement patterns reflected in the Baidu data, it is not reasonable to assume this uncertainty could somehow modify the total number of infected people. To ensure these Bernoulli trials only influenced relative patterns, we rescaled values such that the sum of all movers and non-movers equalled the number of people in the patch. Because this includes the predicted number of non-movers, this rescaling does not on average bias movement in any way (towards increased or decreased movement), and, at worst, reduces slightly the stochasticity introduced by this process, but even this stochasticity reduction becomes less important as the epidemic progresses.

We have clarified the text in the Methods (under "Movement" heading) to better reflect exactly why this rescaling step was needed, and why this rescaling should not introduce any bias.

The authors appear now to be saying that their initial value of I(0)=5 was a lucky guess! This is most unsatisfactory.
Re: The reviewer's initial comment regarded both the initial value and transmission parameters used, such as R0, recovery rate, exposure periods, etc. The transmission parameters were chosen without trial and error, but this is because they were fitted in previous studies that used nearly identical model formulations to nearly identical data, meaning we did not have to do comprehensive parameter estimation. The initial value of 5 was therefore chosen as a minimum value necessary to prevent many simulation runs from going to 0 due to stochastic recovery events. We have since performed some sensitivity analyses and found no difference with choosing 3, 5, or 8 initially infected people over 3 months of simulation; fewer than 3 tended to cause our model to go to 0 in most cases (having 3 initially infected caused it to happen roughly 50% of the time), and over 15 began generating significantly more cases than when 5 were initially infected.

We have added text justifying this initial condition in the Methods.

Referee #4 (Remarks to the Author):

I have reviewed a revision of the manuscript titled "Effect of non-pharmaceutical interventions to contain COVID-19 in China." In my review of the original submission, I raised only a few minor points of criticism. The authors have successfully addressed these points in their response letter and have added clarifications to the manuscript where needed. I'm strongly in favor of publication.

Re: Thanks.