Peer Review File
Article information: https://dx.doi.org/10.21037/jtd-23-234

Reviewer A

Professor Huang and colleagues developed a dynamic model to predict the incidence of COVID-19 by using multi-source data during the "Zero-Covid" period in China. The topic is interesting and the paper is well written. I suggest to publish the paper in the journal after some modifications.

Comment 1: Please provide the details of the values of parameters in the models, such as incubation period of different VOCs.
Reply 1: We thank the reviewer for asking this question. We have added the discussion about the values of parameters in the models as required, which can be found in the sub-section of Modified SEIR model in line 166. We agree that the values of parameters should be mentioned in the paper, and we have added the details of the values of parameters in the text.
Changes in the text: "The range of each parameter in the model are shown in the table 1:
Table 1 Range of the parameters

Parameters	**Range**	**Initial Values**
α (protection rate)	-1 ~ 1	0.06
β (transmission rate)	0 ~ 5	1.0
γ (inverse of latent time)	0 ~ 1	0.2
δ (transition rate of infected people get quarantined)	0 ~ 1	0.1
λ (recovery rate)	0 ~ 1	0.1
κ (death rate)	0 ~ 1	0.001

"

Comment 2: It would be better to discuss the differences by using the reported date and the illness onset date of COVID-19 cases during the model fitting process.
Reply 2: We appreciate the kind reminder from the reviewer. It is a very good suggestion, and we actually planned to simulate the results by using data from these two different dates in the first place. Unfortunately, the data sources are limited and we have searched online trying to obtain the related data, but nothing was found. It asks for the cooperations with some departments. Therefore, we have not figured out a way to get these data, but it will definitely be one of the key problems we need to solve in the short future. Nevertheless, we have added the corresponding discussion in the **Discussion and Conclusion** section in line 397 and line 403.

Changes in the text: "…Moreover, the details of the epidemic data need to be improved. For instance, the reported date and illness onset date of the patients should be obtained and added into model to simulate the results accordingly, thus, compare the difference between these two simulation results to reveal more transmission mechanisms of COVID-19."; "…At the same time, we will also seek for cooperation and other ways to obtain more relevant data to perfect our model"

Comment 3: During the zero-covid period, the shape of the epidemic curve depends on the strength of PHSMs. It is suggest to discuss these affects during the modelling.
Reply 3: We thank the reviewer's suggestion. We agree that the shape of the curve relies on the strength of public health and social measures (PHSMs). As the epidemic curves (Figure 4) suggested that cities with more strict control measures had lower daily cases and shorter duration of the outbreaks. We have added the corresponding discussion and explanation in a new sub-section called "Parameter for control measurements" under the "Materials and Methods" section in line 187, as well as the sub-sections of "Case study in China" in line 325.
Changes in the text: "In the GPCP system, different parameterization schemes are employed into the model. As for MDEP method, the most important one is the parameterization for control measurements. Since Chinese government applied strict control measures in terms of contain the outbreaks in a timely manner. Therefore, to achieve a better performance of the prediction results, we established a parameterization scheme to depict the different effects of different strengths of control measures. There are three coefficients are included in the model, namely government response time (Days_con), initial exposed cases (E0), and the attenuation rate. In this case, the infection rate β will be calibrated as $\beta = \beta_0 \cdot \text{attenuation rate}^t \ (t > Days\_con)$, where β_0 is the base infection rate. With the strict control measures applied, the value of Days_con will be smaller, and attenuation rate will be smaller. As a result, the duration of an outbreak will be shorter and the peak value and cumulative cases will be smaller" and "…When taking the population and control measures into consideration, we can observe that cities like Beijing, Lanzhou and Urumqi took relatively strict control measures compared with the other three cities, and they all have shown lower peak values and shorter durations."

Reviewer B

The manuscript focuses on the accurate prediction of the transmission of the infectious illness COVID-19 using the Multi-source dynamic ensemble prediction (MDEP) approach, which integrates the initial prediction result and subsequent dynamic prediction results to generate the final prediction result. This manuscript is fairly clearly written and the results are convincing. However, it has several weaknesses that need to be addressed to move toward publication.

Comment 1: what is the protection rate at line 144? Please indicate the exact data utilized.
Reply 1: We appreciate the question proposed by the reviewer. The protection rate (α) represents the proportion of susceptible people who will be highly immune to COVID-19 due to the awareness of self-protection, such as wearing face masks and keep social distance. When $\alpha > 0$, people will enter into protected phase from susceptible population, while if $\alpha < 0$, people will return to susceptible group. We have added relevant information in the text, which can be found in line 161, in order to specify the meanings of each parameter.
Changes in the text: "…Particularly, the protection rate (α) represents the proportion of susceptible people who will be highly immune to COVID-19 due to the good awareness of self-protection, such as wearing face masks and keeping social distance. It should be noted that all the above-mentioned parameters are greater than zero, except the protection rate. When $\alpha > 0$, people will enter into protected phase from susceptible population, while if $\alpha < 0$, people will return to susceptible group (role of seasonality…)."

Comment 2: The data in Result 1 should be shown in detail as much as possible. For example, in line 217, where the authors said that "The simulated curves cannot reflect the real trends and show very poor accuracy rates in daily cases and cumulative cases", it is proposed to show an error value.
Reply 2: We thank the kind reminder from the reviewer. We have added the plot of error value of each simulation result in the text, which can be found in line 249
Changes in the text: "…The calculated error values ($E = N_{pdaily/pcumu\_i} - N_{rdaily/rcumu\_i}$, where $i = 1,2,3 \dots$ and $N_{pdaily/pcumu\_i}$ represents the prediction results of daily and cumulative cases, while $N_{rdaily/rcumu\_i}$ represents the real daily and cumulative cases), which are shown in Figure 3, indicated that the error values of daily and cumulative cases in Beijing and Lanzhou fluctuated greatly, ranging from -250 to 25 and -2000 to 500, respectively."

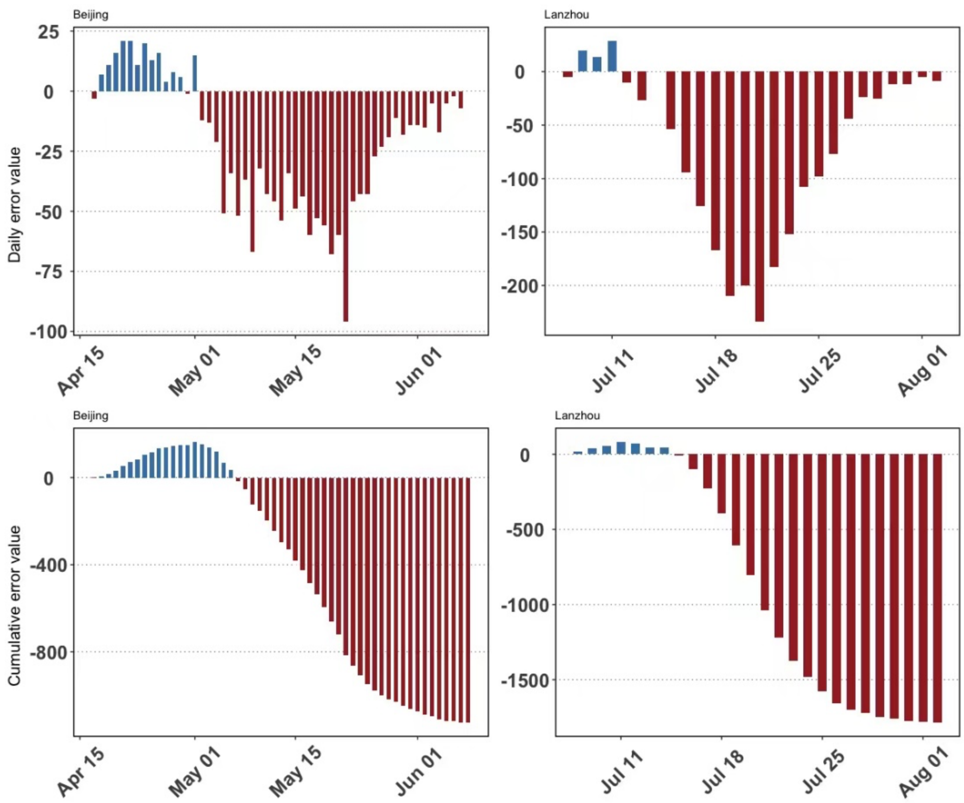

Fig. 3. The error values of simulated results using one-way epidemiological model, where the upper panel are the daily error values of Beijing and Lanzhou, respectively, and the bottom panel are the cumulative error values of Beijing and Lanzhou, accordingly.

Comment 3: What is the X-axis title in Figure 5, where it is suggested to improve?

Reply 3: We thank the kind reminder from the reviewer. The label of X-axis of Figure 6 (a) is “confirmed cases”, which has been specified in the corresponding place. We have added the title for X-axis as “City” in Figure 6 (b)

Changes in the text: In figure 6 (b), we have added the label “City” under X-axis