A data-driven assessment of early travel restrictions related to the spreading of the novel COVID-19 within mainland China

Alberto Aleta [a,1,*], Qitong Hu [b,c,1], Jiachen Ye [b,c], Peng Ji [b,c], Yamir Moreno [d,e,a]

[a] ISI Foundation, Via Chisola 5, 10126 Torino, Italy
[b] Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, China
[c] Research Institute of Intelligent and Complex Systems, Fudan University, Shanghai 200433, China
[d] Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain
[e] Department of Theoretical Physics, University of Zaragoza, 50018 Zaragoza, Spain

ARTICLE INFO

ABSTRACT

Two months after it was firstly reported, the novel coronavirus disease COVID-19 spread worldwide. However, the vast majority of reported infections until February occurred in China. To assess the effect of early travel restrictions adopted by the health authorities in China, we have implemented an epidemic metapopulation model that is fed with mobility data corresponding to 2019 and 2020. This allows to compare two radically different scenarios, one with no travel restrictions and another in which mobility is reduced by a travel ban. Our findings indicate that *i)* travel restrictions might be an effective measure in the short term, however, *ii)* they are ineffective when it comes to completely eliminate the disease. The latter is due to the impossibility of removing the risk of seeding the disease to other regions. Furthermore, our study highlights the importance of developing more realistic models of behavioral changes when a disease outbreak is unfolding.

1. Introduction

In Dec. 31st, 2019, Chinese authorities reported an outbreak of a novel coronavirus disease, called COVID-19 by the World Health Organization. Due to the proximity of the Spring Festival, the Chinese Government implemented quarantine in Wuhan, where the outbreak started, as well as in several nearby cities since Jan. 23rd, 2020. As of Feb. 16th the outbreak had already infected 68,584 individuals in China of which 1666 died [1]. Much is still unknown about the characteristics of this pathogen. For instance, it remains unclear the animal source of this zoonotic disease, being bats and pangolins currently the two most likely sources [2–4]. It has also been proposed that more than half of the cases might have gone undetected by routinely screening passengers, due to the special characteristics of this disease [5], which makes it possible that infected individuals are asymptomatic while infectious.

Several studies predict a much larger number of infections than the actual number reported by the authorities, claiming that only between 10% and 20% of the cases have been detected and re-

ported [6–9]. The reasons for such deviations between models and actual count of cases are diverse, for instance, the fact that the symptoms could be mild and similar to other flue-like diseases for some people, may induce infected individuals not to seek medical care [10]. On the other hand, on Feb. 13th, 14,840 new cases were reported [11], in contrast to 2022 cases counted during the previous day [12]. The reason was that previously to that day, only those cases that had been laboratory-confirmed were being recognized as so, whereas from that date onwards, also the clinically diagnosed cases were accounted for. Therefore, the model-based prediction of the number of infected individuals can plausibly be larger than the official reports.

From a theoretical and computational point of view, there are groups that have proposed new epidemic models to properly account for the special characteristics of this disease [10,13,14]. However, our knowledge of the dynamics of the disease is too limited to be constrained to use such sophisticated models. In fact, in some of these works, the models are fitted to reproduce exactly the reported number of infected individuals, which, as noted before, can be counterproductive given that the actual number of infected individuals in the population is surely higher than those detected either by clinical diagnosis or in the laboratory. Lastly, there has also been intense research aimed at computing the probability that the outbreak extends beyond Wuhan to other cities in China, as well as

---

* Corresponding author.
  E-mail addresses: albertoaleta@gmail.com (A. Aleta), pengji@fudan.edu.cn (P. Ji), yamir.moreno@gmail.com (Y. Moreno).
  [1] These two authors contributed equally.

to the rest of the world [6,7,15–18]. These works use historical data in order to produce a risk assessment and obtain probabilities that the disease is imported in other populations. Likewise, modeling efforts have been directed towards gauging the effect of Wuhan's quarantine on the spreading of the epidemic all over the country (which so far has determined to be a delay of 3 days in the arrival of the peak [7]) and worldwide.

Given the unprecedented characteristics of this outbreak, we adopt a slightly different approach and study a data-driven metapopulation model that makes use of the actual flows of the population to properly measure the early effects of the travel reductions in China. Specifically, we built a basic metapopulation model [19,20] of 31 regions in China (except HongKong, Macau and Taiwan). Inside each population, we considered that the individuals in the population interact following a homogeneous mixing scheme. This approach is similar to the ones proposed in [7] and [15]. However, in our case, we perform a data-driven simulation with the actual flows of individuals that have taken place during the period of study. That is to say, we do not rely on transportation data (which is most of the time given as the maximum flow capacity between subpopulations), but on a real dataset gathered for this occasion (see Materials and Methods). The results align with previous findings and indicate that travel restrictions do not have a significant impact in containing the expansion of the disease, though reducing the flow of individuals could lead to a delay in the importation of new cases in other subpopulations. Importantly, we show that using the real mobility patterns of the population is a key factor to understand the spreading of the disease. In this sense, our conclusions point out that despite the many advances in disease modeling during the last two decades, there still remain many open challenges, most of them related to how to sensibly incorporate human behavioral changes during the unfolding of an outbreak.

2. Model

We implement a stochastic SEIR-metapopulation model to simulate the spread of the epidemic across mainland China. The model can be divided in two discrete processes: the disease dynamics governed by a SEIR compartmental model; and the mobility of individuals. In a SEIR compartmental model individuals are assigned into each group according to their health status: susceptible (S) if they are susceptible to the disease; exposed (E) if they have been infected but cannot infect; infectious (I) once the incubation period is over and the individuals can infect others; and removed (R) when they are either recovered or deceased. Within each subpopulation (henceforth, region), the transition between compartments results from the following rules, iterated at each time step, corresponding to 1 day:

- **S → E:** Susceptible individuals in region i might get infected with probability $P(S \to I) = 1 - (1 - R_0/(T_I N_i))_i^{I_i}$, where R_0 is the reproduction number, T_I the mean infectious time, N_i the number of individuals in region i and I_i the number of infected individuals in the region.
- **E → I:** Exposed individuals enter the infected compartment with a rate inversely proportional to the mean latent period, T_E.
- **I → R:** Infected individuals enter the removed compartment with a rate inversely proportional to the mean infectious period, T_I.

We define the generation time as $T_g = T_E + T_I$. On the other hand, the mobility of individuals is implemented through a data-driven approach. We have obtained the number of individuals leaving each region each day, N_i^o, as well as the probability, p_{ij}, of going from each region i to region j (see Materials and Methods

for a thorough description of the data). Hence, at each time step, we select N_i^o individuals at random from within each region - excluding infectious individuals in i, which are supposed to be under quarantine or hospitalized - , and distribute them across the country according to such probabilities. Note that this implies that the disease will be propagated to other regions by exposed individuals. We also implemented a randomized version of the model for the mobility, in which the fraction of the population traveling from each region at each time step is 0.008 (estimated from the average fraction of individuals traveling during the first week of January 2020) and their destinations are chosen randomly with probability proportional to the population of the destination region.

Furthermore, since we have data of N_i^o from 2019 and 2020, we simulated the spreading of the disease in both years. In this way, we are able to asses the impact of travel reductions due to mobility restrictions across regions without making any - however sensible - assumptions about possible changes in individuals' mobility. Nonetheless, this period of the year has some peculiarities due to the Spring Festival, an event that completely modifies the travel patterns of the population. In 2020, the Spring Festival was celebrated on Jan. 25th, while in 2019 it took place on Feb. 5th. For this reason, we align both simulations so that "day 0" will correspond in both years to the day of the Spring Festival (Jan. 25th and Feb. 5th respectively). Lastly, the simulation will run in both cases to 13 days after the Festival, which corresponds to Feb. 5th in 2020 and Feb. 16th in 2019, so that exogenous effects such as the strong social distancing measures enforced by the government or the new way of defining a positive case do not enter into play.

3. Results and discussions

Fig. 1 clearly shows that the changes induced by the epidemic in the overall flow and movement of the population throughout China are not restricted just to the region in which the city of Wuhan is located (Hubei). In 2019, a large number of individuals moved right before the Spring Festival. Then, it reached a minimum at the Spring Festival and, later on, a large number of individuals moved again, see Fig. 1A. On the other hand, in 2020, the situation was similar only until Jan. 23rd, when the travel restrictions in Wuhan were implemented. Two features are worth highlighting. First, when the travel restrictions were implemented, the flow of population had peaked already just before the Spring festival. This is key to understand why, despite the reduction in mobility, the daily number of new cases continued to increase for weeks. Secondly, most of the second wave observed in 2019 - which corresponds to travel back to the original region - had not taken place yet (see Fig. 1B), and thus, there was still a high risk that a subsequent large outbreak could occur. The existence of recurrent local outbreaks is a feature that some models have anticipated [20]. Further indications that the population has not yet reached the original distribution by region are provided in Fig. 1C and D, where we show the evolution of the number of individuals living in each region (initial data obtained from the Chinese Yearbook, see Materials and Methods). Clearly, while in 2019 most of the population returned to their regions, in 2020 this had not taken place yet.

To parameterize the metapopulation epidemiological model we follow Chinazzi et al. [7] and set a generation time of 7.5 days and a reproduction number R_0 equal to 2.4. The latent period is set to 3 days [21], which is compatible with pre-symptomatic transmission of the disease, since the incubation period is between 5 and 6 days [7]. The outbreak is seeded by introducing 40 exposed individuals on Dec. 1st 2019 (Dec. 12th 2018 for 2019) [7]. Then, the simulations run for 66 days in both cases and we extract the cumulative number of infected cases in each region as a function of time. Note that as there were no travel restrictions in 2019, one can see the results obtained with the 2019 data as the more plau-
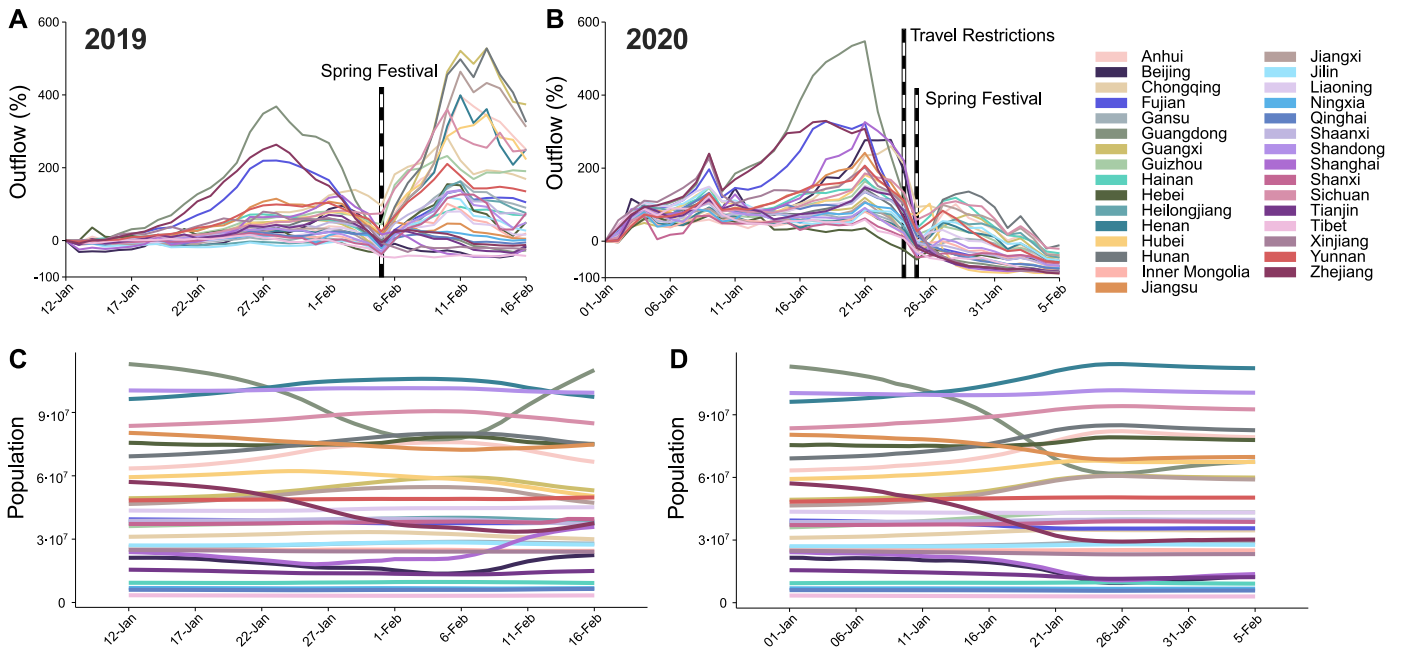
Fig. 1. Movement patterns of individuals across Mainland China in 2019 and 2020. Panels A and B represent the percentage of individuals traveling from each region compared to such fraction on Jan. 1st in 2020 (Jan. 12th in 2019). Additionally, panels C and D show the population of each region assuming that the number of individuals traveling from each region is given by the flow data and that their destinations are randomly selected according to the mixing patterns p_{ij}.
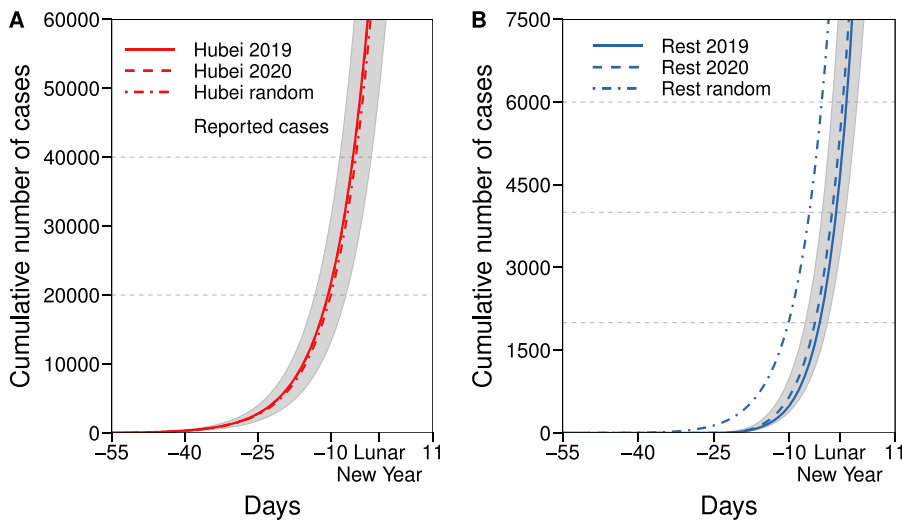

Fig. 2. Predicted cumulative number of cases: in Hubei (panel A) and the rest of mainland China (panel B) using mobility data of 2019 (solid lines, scenario equivalent to no travel restrictions), of 2020 (dashed lines, with travel restrictions considered) and under the assumption of random movement of the population (dashed-dotted lines). Dots represent the actual total value in Hubei reported by the authorities.

sible outcome for the 2020 outbreak without travel restrictions. In other words, by comparing 2019 with 2020, we can factor out the impact of the early travel reduction in the city of Wuhan and the subsequent changes in the mobility patterns of the population.

Fig. 2 shows the cumulative number of infected individuals for Hubei and for the rest of mainland China. The large majority of cases, in all situations considered, are contained within Hubei province. The results in Fig. 2B show that travel restrictions have a small impact in the temporal evolution of the disease in the rest of the country(compare 2019 with 2020). As it can also be deduced from the trend of the curves in Fig. 2B, there is no indication that the growth in the number of cases will evolve following a different functional form. This contrasts with the results that would be obtained if individuals moved at random (with a probability proportional to the size of the subpopulations).

Fig. 2 A also shows that there is a large difference between the cumulative number of cases predicted in 2020 and the reported one. Although, as previously discussed, this has been reported by several other studies, to ensure that the methodology is correct, and to further analyze the effect of the mobility patterns, in Fig. 3 we show the correlation between the real values of infected individuals and the simulated ones for 2020. We obtain a Pearson correlation of 0.80 implying that the assumptions behind the model, albeit simplistic, can correctly describe the basic dynamics of the epidemic. Furthermore, we also see that the data-driven model is able to predict better the dynamics than its random counterpart. In Fig. S1, we perform a sensitivity analysis on the different parameter choices. For sets of parameters that induce a slower spreading (such as smaller R_0 or larger T_g) the disease spreads outside Hubei only after the restrictions have been
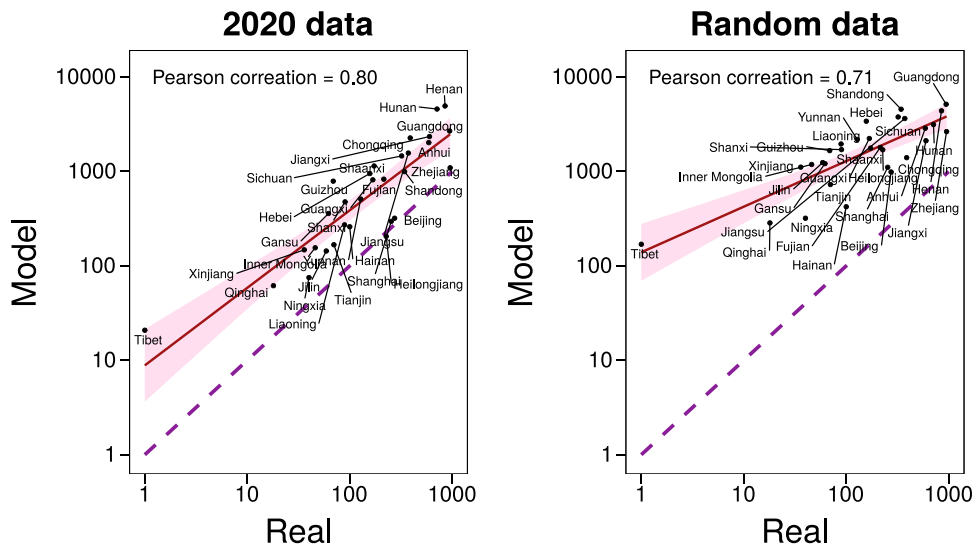
Fig. 3. Correlation between model and reality. Predicted cumulative number of cases in each region, except Hubei, compared to the real number reported by the authorities by Feb. 5th. Left: estimations obtained using 2020 data. Right: estimations obtained using random mobility data. In both cases the dashed line represents the identity line.
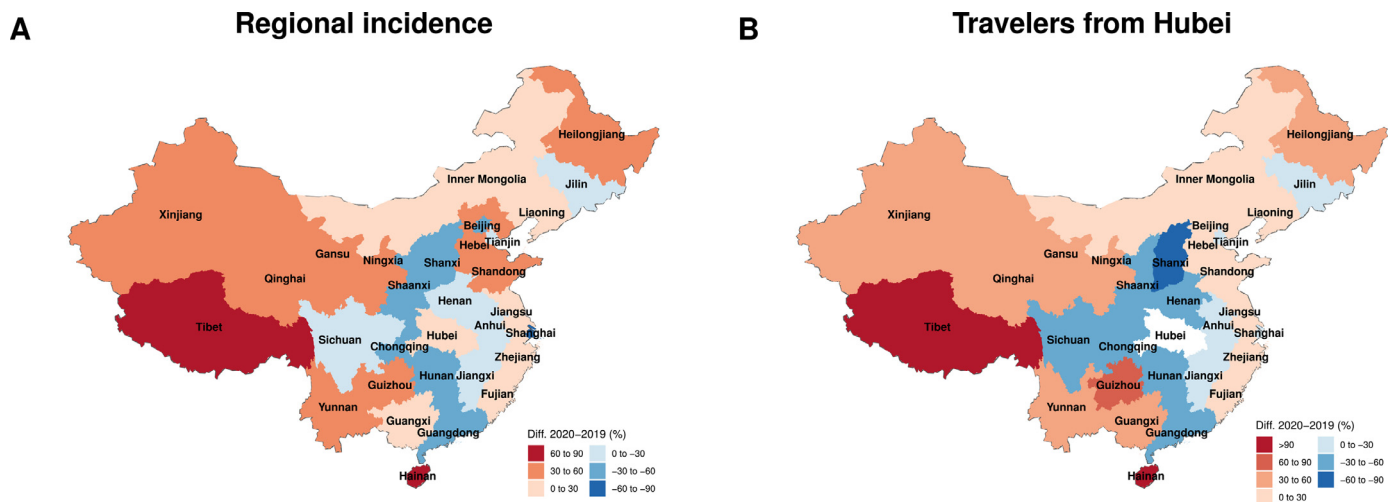

Fig. 4. Expansion of the disease. A: relative difference of the incidence in 2020 and 2019 in each region by Feb. 5th (Feb. 16th in 2019). B: relative difference in the total number of travelers that went from Hubei to any other region for the period under consideration.

implemented. This, in turn, enlarges the differences between 2019 and 2020. Conversely, for sets of parameters that produce a faster spreading, the disease spreads outside Hubei before the restrictions are implemented, producing a larger incidence in those regions. In either case, the correlation between the model predictions and the observed values is high, signaling that the real mobility patterns of the population are a key element to understand the spreading of the disease.

Lastly, we have also studied the difference in the predicted number of cases in each region between 2019 and 2020 at the end of the simulation period. Results are reported in Table 1, where we show the mean number of infected individuals per region. As it can be seen in the table, the estimated total number of cases is quite similar in both cases. However, the distribution across the country is fairly different. Indeed, while there are many areas where the number of infected individuals would be larger in 2019, there are others where it would be smaller.

To elucidate the reasons behind this behavior, in Fig. 4A we show the relative difference in the incidence per region, comparing 2019 and 2020. We observe that regions near Hubei tend to have smaller incidences in the simulation with data from 2020.

Conversely, regions further away from Hubei tend to have larger incidences in 2020. We can compare this distribution with Fig. 4B, where we show the total number of individuals who travel from Hubei to any other region from Jan. 1st to Feb. 5th in 2020 with the ones that did the same from Jan. 12th to Feb. 16th in 2019. We can see that regions closer to Hubei had a smaller number of visitors in 2020, while the ones further away from Hubei had more visitors. This explains the observed differences in the simulations and highlights the importance of using real and updated data to properly account for the behavior of individuals.

Summarizing, we have studied a data-driven metapopulation model that allows assessing the effect of early travel restrictions in Wuhan and Hubei province. Even if our modeling framework is simpler than other more sophisticated implementations of the disease dynamics, our results are in line with several available studies in that: *i)* travel restrictions have limited efficacy unless they are applied very early, and *ii)* reducing the travel does not appear to have a long term impact on the spreading of the disease if not accompanied by other measures. We also note the effect of travel restrictions might be underestimated because the large majority of people had already moved before these mobility restrictions were

Table 1

Reported value of the total number of infected individuals by Feb. 5th per region compared to the values predicted by the model using data from 2019 (Feb. 16th) and 2020 (Feb. 5th). The mean value of the prediction, obtained from 10^4 independent runs of the model, alongside the 95% confidence interval is shown.

| Region	Reported value	2019 simulated no travel restriction	2020 simulated travel restriction
Henan	815	6887 [4095-10,203]	4938 [2796-7633]
Hubei	19,665	348,540 [214,850-504,160]	361,876 [221,094-526,075]
Hunan	711	9647 [5839-14,095]	4580 [2589-7013]
Guangdong	944	5765 [3464-8528]	2681 [1482-4183]
Jiangxi	600	2531 [1403-3927]	2342 [1217-3746]
Chongqing	389	4909 [2933-7228]	2263 [1211-3606]
Anhui	591	2148 [1164-3396]	2021 [1030-3294]
Jiangsu	373	1443 [732-2346]	1566 [775-2608]
Zhejiang	954	1003 [523-1643]	1093 [542-1835]
Shaanxi	173	1809 [1008-2811]	1143 [520-2206]
Sichuan	321	2061 [1127-3230]	1460 [706-2472]
Beijing	274	482 [257-778]	320 [143-564]
Shanghai	254	805 [453-1259]	296 [131-529]
Shandong	343	731 [305-1429]	996 [415-1871]
Fujian	215	716 [308-1134]	828 [354-1526]
Guizhou	69	494 [184-1007]	789 [305-1495]
Hebei	157	711 [301-1337]	950 [417-1713]
Guangxi	168	647 [282-1224]	815 [341-1525]
Yunnan	128	373 [128-811]	510 [164-1091]
Shanxi	90	734 [383-1244]	476 [154-1005]
Hainan	100	155 [30-426]	260 [69-606]
Liaoning	89	217 [63-571]	272 [67-668]
Gansu	62	238 [67-605]	359 [92-832]
Tianjin	70	183 [68-408]	167 [35-441]
Heilongjiang	227	144 [30-445]	205 [40-549]
Xinjiang	36	103 [18-352]	148 [15-443]
Inner Mongolia	46	134 [39-373]	155 [23-451]
Jilin	59	150 [44-406]	144 [21-435]
Qinghai	18	43 [6-177]	62 [1-286]
Ningxia	40	53 [7-221]	75 [2-317]
Tibet	1	11 [0-56]	21 [0-165]
Total	27,982	393,869 [242,817-569,435]	393,812 [241,466-572,552]
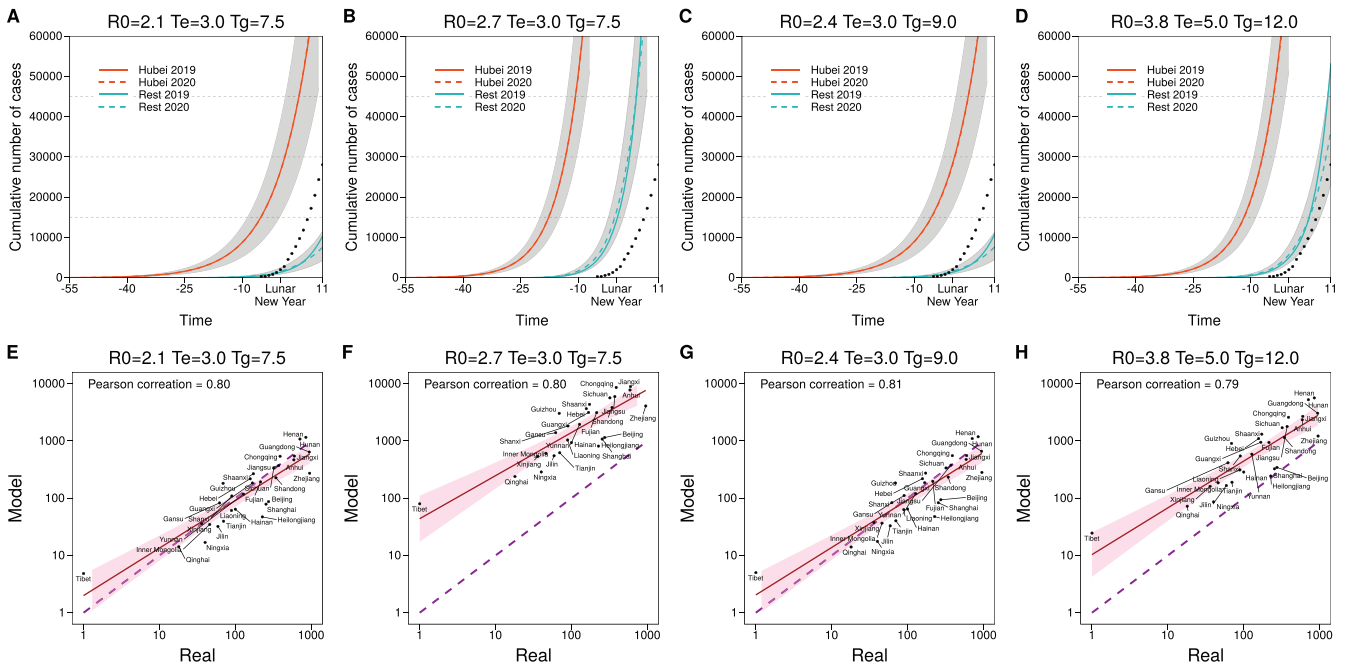
Fig. S1. (Top Panels) Predicted cumulative number of cases in Hubei and the rest of mainland China using mobility data of 2019 (solid lines, scenario equivalent to no travel restrictions), and 2020 (dashed lines, with travel restrictions considered) for several parameter values of the epidemic model as indicated. Dots represent the actual total value reported by the authorities. (Bottom Panels) Predicted cumulative number of cases in each region, except Hubei, compared to the real number reported by the authorities by Feb. 5th, 2020 for several disease parameters as indicated.

implemented (as it can be seen in Fig. 1). Our study is limited in several aspects that can constitute future research goals. First, the geographic resolution allowed by the mobility data used here is low. Considering large regions have the undesired effect that one can not add structure to the population and therefore the dynamics within each subpopulation is constrained by the homogeneous mixing hypothesis. This limitation could be overcome if less granular spatial and temporal data becomes available. Secondly, and perhaps more important as it currently represents a scientific challenge, we have assumed that the transmissibility does not change during the whole simulation period. This implies that changes in behavioral patterns of the population are not fully accounted for nor they can be completely disentangled from those associated with travel restrictions. Understanding how to deal with such behavioral changes is key for the development of more realistic descriptions of the large-scale spreading of diseases. Finally, another critical feature of current models that needs to be improved in future research is the use of disease parameters - notably R_0 - that are constant both in time and across populations.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix. Materials and Methods

Data description

Resident population. The statistics about the distribution of the population across 31 regions in China were obtained from the "China Statistical Yearbook 2019" [22]. This annual publication reflects comprehensively the economic and social development of China.

Population flow. We obtained migration data from Baidu Qianxi, an open platform based on Baidu Location Based Services (LBS) that provides information about the population flow within China [23]. The dataset comprises two types of data: the Baidu migration index and the Baidu migration ratio. The former is a number proportional to the number of individuals leaving each region. To obtain the proportionality constant, we averaged the amount of individuals leaving Wuhan between Jan. 1st and 10th and compared it to the value of 502,013 estimated by Wu et al. [15]. The second part of the dataset contains the fraction of individuals going from region i to region j, p_{ij}. As such, it is possible to estimate the number of individuals going from region i to j by multiplying the total outflow of the region by the corresponding p_{ij}. The outflow of the individuals is available for the years 2019 and 2020, while the p_{ij} values only for 2020. Nevertheless, as it can be seen in Fig. 1 in

the main text, it represents a good proxy of the situation in 2019. Indeed, even though some populations were closed after Jan. 23rd, 2020, using these values for 2019 correctly describes the return of the population that had left for the Spring Festival.

Infected individuals. The number of infected individuals as a function of time, as well as their distribution across regions on Feb. 5th was obtained from the WHO reports [1].

Sensitivity analysis.

To gauge the effect of the chosen parameterization of the model, we have repeated the analysis with several values reported in the literature [7,24], Fig. S1. As expected, if the incubation period is larger or the basic reproduction number is smaller, the overall number of cases decreases and the disease extends outside Hubei only after the restrictions were in place. Conversely, when using large values of R_0 or longer generation times, the amount of infected individuals increases and the spreading outside Hubei takes place before the restrictions are implemented. In any case, the mobility model is quite independent from the dynamics of the disease, since the correlation is almost always the same in all cases considered.

References

[1] Tech. rep., World Health Organization; 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200216-sitrep-27-covid-19.pdf.

[2] Randhawa GS, Soltysiak MPM, Roz HE, de Souza CPE, Hill KA, et al. PLOS ONE 2020. doi:10.1371/journal.pone.0232391.

[3] Wahba L, Jain N, Fire AZ, Shoura MJ, Artiles KL, McCoy MJ, Jeong DE. bioRxiv 2020. doi:10.1101/2020.02.08.939660.

[4] Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang CL, et al. Nature 2020. doi:10.1038/s41586-020-2012-7.

[5] Gostic K, Gomez ACR, Mummah RO, Kucharski AJ, Lloyd-Smith JO. eLife 2020. doi:10.7554/eLife.55570.

[6] Du Z, Wang L, Cauchemez S, Xu X, Wang X, Cowling BJ, Meyers L. Emerging Infectious Diseases 2020. doi:10.3201/eid2605.200146.

[7] Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, Piontti APy, Mu K, Rossi L, Sun K, et al. Science 2020. doi:10.1126/science.aba9757.

[8] Cao Z, Zhang Q, Lu X, Pfeiffer D, Wang L, Song H, Pei T, Jia Z, Zeng DD. medRxiv 2020. doi:10.1101/2020.02.07.20021071.

[9] Zhou Y, Dong J. medRxiv 2020. doi:10.1101/2020.02.10.20021774.

[10] Zhu X, Zhang A, Xu S, Jia P, Tan X, Tian J, Wei T, Quan Z, Yu J. medRxiv 2020. doi:10.1101/2020.02.09.20021360.

[11] 2020 World Health Organization Tech. rep. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200213-sitrep-24-covid-19.pdf.

[12] 2020 World Health Organization Tech. rep. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200212-sitrep-23-ncov.pdf.

[13] Li X, Zhao X, Sun Y. medRxiv 2020. doi:10.1101/2020.02.09.20021477.

[14] Xiong H, Yan H. medRxiv 2020. doi:10.1101/2020.02.10.20021519.

[15] Wu JT, Leung K, Leung GM. Lancet 2020. doi:10.1016/S0140-6736(20)30260-9.

[16] Tian H, Li Y, Liu Y, Kraemer MUG, Chen B, Cai J, Li B, Xu B, Yang Q, Yang P, et al. medRxiv 2020. doi:10.1101/2020.01.30.20019844.

[17] Yuan H-Y, Hossain MP, Tsegaye MM, Zhu X, Jia P, Wen T-H, Pfeiffer D. medRxiv 2020. doi:10.1101/2020.02.01.20019984.

[18] Boldog P, Tekeli T, Vizi Z, Denes A, Bartha F, Rost G. J. Clin. Med. 2020. doi:10.1101/2020.02.04.20020503.

[19] Colizza V, Vespignani A. J Theor Biol 2008. doi:10.1016/j.jtbi.2007.11.028.

[20] Aleta A, Hisi ANS, Meloni S, Poletto C, Colizza V, Moreno Y. R Soc Open Sci 2017. doi:10.1098/rsos.160914.

[21] Guan W-j, Ni Z-y, Hu Y, Liang W-h, Ou C-q, He J-x, Liu L, Shan H, Lei C-l, Hui DS, et al. N Engl J Med 2020. doi:10.1056/NEJMoa2002032.

[22] China Statistical Yearbook 2019. Compiled by National Bureau of Statistics of China. Accesed February 24, 2020. http://www.stats.gov.cn/tjsj/ndsj/2019/indexeh.htm.

[23] Baidu Qianxi open platform. Accessed February 24, 2020. https://qianxi.baidu.com.

[24] Yang Y, Lu Q, Liu M, Wang Y, Zhang A, Jalali N, Dean N, Longini I, Halloran ME, Xu B, et al. medRxiv 2020. doi:10.1101/2020.02.10.20021675.