Dating the emergence of the first case of COVID-19 in Hubei, China with commercial flights: a retrospective modelling study

Abstract

Background

Commercial flights contributed to the early-stage international transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Understand the effect of international and inter-state flights on the virus transmission is important to evaluate the initial response of the outbreak. This study investigated the likely date of the emergence of the first COVID-19 case in Wuhan, China.

Methods

We constructed a geographical-structured model, including 9122 county-level geographical units in 250 different regions or countries, and 26,094,036 flight plans. Using the model, we estimated the date of the number of deaths and the date of first death caused by COVID-19 in 155 different countries. We set a certain trigger for country and county level lockdown, and a built-in flight randomizer, to assess the different evidence that can suggest the possibility of different dates to be the emergence of the first COVID-19 case.

Findings

We found the median number of global deaths caused by COVID-19 from 50 trails of our simulation is between 853901 and 28432 when the emergence of the first case of COVID-19 case is within the time interval between September 1, 2019 and November 1, 2020. Overall, the average and median $R^2$ of 155 countries decrease as the date of the emergence postpone; however, a small increase of $R^2$ is observed when the date emergence of the first case of COVID-19 case is September 22. The deviation of the simulation result of the deaths from the actual scenario in the eight major epicenters demonstrates a negative trend as the first case emerges later. A similar pattern can be observed on the date of the first death in these eight countries, which can be applied to a global scale.

Interpretations

We found that our simulation results can only include the actual number of global deaths caused by COVID-19 on May 1, 2020 inside the 95%PI when the date of the emergence of the first case is between September 15, 2019 and October 15, 2020; The lowest deviation can be observed on September 22. We also found a decrease of $R^2$ when we deploy later emergence of the first case; unexpectedly, an abnormal increase of $R^2$ on September 22 is observed. When focusing on the eight major epicenters, we found that the number of deaths in these eight countries can only be all included inside the 95%PI when the emergence is on September 15 and September 22. The deviation of the first deaths demonstrates different patterns depending on the continent, whereas the global average demonstrates a pattern of normal distribution with a maximum of 0.2107 on September 22 after the removal of outliers. This could suggest a high likelihood of the emergence of the first case of COVID-19 be around September 15, 2020 and September 22, 2020.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has impacted multiple countries after first detected in Wuhan China. Evidence suggests that the virus can spread through aerosols. Commercial flights were determined to be a crucial factor in importing cases and resulting in local infection. Some countries have enforced strict border closures to mitigate the spread of this deadly virus. Although multiple non-pharmaceutical interventions on the COVID-19 pandemic were introduced, they were not sufficient to contain the spreading of the virus, both locally and globally.

Multiple modelling studies have been conducted to assess the local COVID-19 transmission in the UK, Singapore, Canada, and many other countries. These studies provided insight into the virus transmission and mitigation based on a local and static population. Results given by these studies reflect the country-level transmission dynamics and potential epidemic sizes without the influx and outflux of the cases.
International flights contributed to the early stage spread of the virus, but none of the mathematical models introduced before included this crucial factor. Due to this reason, the simulation results from these models were limited to the local level. To better describe the global transmission dynamic and investigate the early transmission trajectory of the virus, we construct a computer model to evaluate the potential date of the emergence of the first case of COVID-19 and give an assessment of the effect of the mobility of the population on the exportation of cases.

**Methods**

**Geographical structure**

![Geographical Division](image)

**Figure 1: Sample geographical division of the model**

The figure shows a sample geographical division implemented in the model at a (A) global level (B) country level (C) province level. The data is picked from a representative simulation.

According to ISO-3166 and IATA standard, we broke down the world into 250 country-level regions, then further divided 2154 district-level regions (Figure 1). The population of each region was collected through census data, and it is shown in the supplementary material (supplementary material).

A dedicated SEIR model was constructed for each region, and will only be activated when a case is imported from another province. The patients could be traveling from a province to another and initiated the transmission in the destination.

We assumed the population to be distributed evenly without heterogeneity; the population of each county is recorded and integrated into the model (Supplementary Material).

**Flights**

We integrated the flight plans from September 1st, 2019 to June 1st, 2020 (26,094,036 Flights), to help to describe a more dynamic transmission scenario. Due to the lockdown and travel ban in some areas, only 21,081,805 are included (Supplementary material). Flights are assumed to be the only way to travel between states and countries, the exchange of population who is not infected is excluded.

We picked the global average passenger load factor in 2019 (82.6%) to be the uniform passenger load factor for our model, regardless of the actual load factor for a certain flight.
In 2018, 1,811,324,000 passengers traveled internationally through air. Since commercial flights is one of the main contributors to the importation of COVID-19 cases, we focused on the effect of commercial flights on the transmission of SARS-Cov-2 and therefore investigate the date of the emergence of the first case.

Several studies have discussed the case of onboard transmission, and they discovered a low probability of in-flight transmission. Thus, we excluded the in-flight transmission, assuming the transmission scenario to be only local human-to-human infections.

**Lockdown**

We introduced two types of lockdown into the model: local level lockdown, and country-level lockdown. The lockdown in a certain county or country will be triggered when a certain number of deaths caused by COVID-19 is passed, causing the frequency of social activities to drop; the frequency of social activities will be reduced to no lower than 15% of its original level. Lockdowns were assumed to gradually decrease the number of contacts between the individuals in these counties, altering the relative number of contacts to our best estimate of a reasonable reduction in contact rates under lockdown.

The local level lockdown will be enforced only if the local threshold is surpassed, it will not affect the flight plans and the lockdown in other parts of the country.

The country-level lockdown will be triggered when the death poll of any county inside the country reached a threshold, it will cause the whole country to lockdown regardless of the severity of other counties (i.e. The whole China was locked down due to the high death poll in Hubei) (Supplementary Material).

**Initial Epicentre**

Since the origin of the virus is still unknown, this study will not address the location where the virus emerged, but rather focus on the date of the onset of the first case of COVID-19. Due to the reason that the first COVID-19 case was reported in Wuhan, China in December, 2019, we assumed the location of the first case of COVID-19 to be in Hubei, China in between the time interval from September 1, 2019 to November 1, 2019. The lockdown in Hubei provinces is assumed to be a country-level lockdown.

**Transmission**

![Figure 2: Description of the transmission model](image)

Individuals in the stochastic compartmental model are classified into susceptible, exposed, infectious, and removed states. The cases will be determined to be an exported case or a local case by the stochastic algorithm.

Although the virus could spread through imported goods and animals, our model includes only human-to-human transmission. We assumed all individuals at the start as susceptible without immunity. The individual will enter the exposed state when contracted with an infected case. After an average latent period
of 4.5 days, the person will remain infectious for 7 days; the infectious period includes 2 days of pre-
clinical period and 5 days of clinical period.\textsuperscript{6,23,24} The individual in the exposed and infected stage will be
determined to become whether a local case or an exported case according to a stochastic algorithm (Figure
2).
We completed 50 simulations for each scenario with the date of emergence of the first case of COVID-19
to be September 1, 2019, September 8, 2019, September 15, 2019, September 22, 2019, October 1, 2019,
October 8, 2019, October 15, 2019, October 23, and November 1, 2019. The basic reproduction number \( R_0 \)
of the virus is set to be 2.79.\textsuperscript{25} The imported cases will preserve its data (time since onset, age, and social activity pattern) and start
transmission accordingly in the new area. Since the average mortality rate of COVID-19 patients
demonstrates a strong correlation with the age,\textsuperscript{26} we stratified the population in some countries into 10-year
age bands.\textsuperscript{27} Our stratification was completed on the country level; thus, we did not consider the difference
of population age distribution between different provinces and counties in the same country.\textsuperscript{13} The mortality rate of each individual in the model was determined based on a function based on age. The
mortality rate for a certain ten-year age interval was assumed to be the same; the data was acquired from a
publication by the Chinese CDC (supplementary material).\textsuperscript{28}
Results

Figure 3: Number of deaths worldwide by May 1, 2020.

(A) Number of deaths from February 1, 2020 to May 1, 2020. The colored curves represent the maximum and minimum deaths resulted from the simulation; the colored areas represent 95% PI. The bold line is the actual number of deaths worldwide. Red, purple, light purple, yellow, blue, sky blue, dark green, green, and light green represents the emergence of the first case of COVID-19 case in Wuhan on September 1, 2019, September 8, 2019, September 15, 2019, September 22, 2019, October 1, 2019, October 8, 2019, October 15, 2019, October 23, and November 1, 2019, respectively. (B) The number of deaths resulted from the simulation when changing the date of the emergence of the first case of COVID-19. The horizontal blue line represents the actual number of deaths on May 1, 2020; the vertical droplines represent 95% PI. The x-axis represents the date of the emergence of the first case of COVID-19. The figures are in logarithmic scale.

We modelled 9 different dates of the emergence of the first case of COVID-19, resulting in median numbers of deaths of 853901, 691434, 487912, 363746, 215706, 140413, 85357, 46789, and 28432 when the emergence of the first case of COVID-19 case is on September 1, 2019, September 8, 2019, September 15, 2019, September 22, 2019, October 1, 2019, October 8, 2019, October 15, 2019, October 23, and November 1, 2019, respectively (Figure 3 b). The deviation of the number of the deaths of the median number of the result (363746) from the actual scenario (305875) reaches the minimum when the emergence of the first case of COVID-19 case was on September 22, 2019.

Overall, the number of deaths caused by COVID-19 by May 1, 2020 worldwide demonstrates an exponential decrease pattern as later dates of the emergence of the first case of COVID-19 were deployed. When the date of the emergence of the first case of COVID-19 as early as September 8, 2020, the actual number of
deaths is lower than the 95% prediction interval; when the date is set as late as than October 22, 2019, the actual number is higher than the 95% prediction interval (Figure 3).

Figure 4: Impact of the date of the emergence of the first case of COVID-19 in Wuhan.

(A) Number of deaths caused by COVID-19 in eight major epicenters with the greatest numbers of deaths (United States, United Kingdom, Spain, Germany, Belgium, France, Canada, and Italy), from February 1, 2020 to May 1, 2020 when the first human COVID-19 case emerged on September 1, 2019, October 1,2019, and November 1, 2019. 50 realisations are shown in the figure with light colors; the bold dashed line is the simulation resulted in a median number of deaths, with the bold solid curve showing the actual number of deaths. (B) The distribution of coefficient of determinations of the actual scenarios and the median projection results in 200 modelled countries. The red, purple, light purple, yellow, blue, light blue, dark green, green, and light green represents the emergence of the first case of COVID-19 case in Wuhan on September 1, 2019, September 8, 2019, September 15, 2019, September 22, 2019, October 1,2019, October 8, 2019, October 15, 2019, October 23, and November 1, 2019, respectively. (C) The effect of changing the date of the emergence of the first case of COVID-19 in Wuhan on the worldwide average (red) and median (blue) coefficient of determination in the modelled countries. The shadow area represents low degree of fit, the x-axis represents the date of the emergence of the first case of COVID-19.
We picked three different scenarios for the date of the emergence of the first case (September 1, 2019, October 1, 2019, and November 1, 2019), and then picked eight epicenters with the greatest number of deaths on May 1, 2020 (United States, United Kingdom, Spain, Germany, Belgium, France, Canada, and Italy) to represent the globe. When the emergence of the first case of COVID-19 was on September 1, 2020, four countries demonstrated a high degree of fit: United Kingdom, Spain, Belgium, and Italy have coefficient $R^2$ of determination of 0.9157, 0.9731, 0.9577, and 0.9654 (Figure 4 a). The coefficient of determination of these four countries decreases to 0.8627, 0.6261, 0.7768, and 0.5999, respectively (Figure 4 a). United States, France, and Canada, by contrast, increases from 0.9282, 0.8946, and 0.8934 to 0.9739, 0.9901, and 0.9708. Germany, however, did not demonstrate a great change in the coefficient of determination (From 0.9623 to 0.9722) (Figure 4 a). When the date was set to be November 1, 2019 all of the eight countries demonstrate an extremely low degree of fit, with an average coefficient of determination of 0.5845.

A similar pattern can be observed on the global scale (Figure 4 b, c). It is clear that the worldwide $R^2$ decreased as the emergence of the first case of COVID-19 is postponed: the $R^2$ distribution shifts left, and the worldwide average and median decreases (Figure 4 b). The global median increased from 0.8156 on September 15, 2019 to 0.8340 on September 22, 2019 before it drops sharply to 0.7540. When the date of the emergence of the first case of COVID-19 is later than October 23, 2019, the worldwide median and average $R^2$ dropped to lower than 0.7, indicating a low degree of fit (Figure 4 c).

![Figure 5: Number of deaths in eight major epicenters](image)

Number of deaths resulted from 50 realisations from 5 different scenarios for the date of the emergence of the first case (September 15, 2019, September 22, 2019, October 1, 2019, October 8, 2019, and October 15, 2019) in the eight major epicenters with the greatest number of cases on May 1, 2020 (United States, United Kingdom, Spain, Germany, Belgium, France, Canada, and Italy). The bold line represents the actual number of deaths; the dash line is the trail resulting in a median number of deaths. Cases that failed include the actual scenario in the 95%PI (September 1, 2019, September 8, 2019, October 23, 2019, November 1, 2019) are excluded.
The median of United States, United Kingdom, Spain, Germany, Belgium, France, Canada, and Italy, the eight major epicenters, demonstrated the lowest deviation from the actual scenario on September 15, September 22, September 15, September 15, September 15, September 15, October 1, and September 15 according to the percentage deviation from the actual number (Figure 5). Obviously, as the emergence of the first case of COVID-19 was set later, the number of countries that do not include the actual scenario within its 95%PI increases. When the emergence happened on September 15, 2020 and September 22, 2020, the simulation included all the actual numbers of the eight countries inside the 95%PI; seven of eight countries are excluded from 95%PI when the emergence is on October 15, 2020 (Figure 5).

Overall, the date of the emergence of the first case in which the eight epicentres demonstrate the lowest deviation from the actual scenario converged towards Septembers 15, 2020. The date of the emergence of the first case countries which results in the lowest deviation of countries in North America are overall later than that of those countries in Europe.

Figure 6: Date of the first death in 155 countries.

(A) Dates of the first death caused by COVID-19 in 155 countries around the globe; the countries are categorized into Europe, North America and South America, Africa, and Asia and Oceania based on their continents. The bold line with points represents the actual date of the first death in the countries. The colored line represents the date of the first death in the countries from the trail that resulted the median number of deaths. The countries are sorted according to the average date of the first death. Red, purple, light purple, yellow, blue, sky blue, dark green, green, and light green represents the emergence of the first case of COVID-19 case in Wuhan on September 1, 2019, September 8, 2019, September 15, 2019, September 22, 2019, October 1,2019, October 8, 2019, October 15, 2019, October 23, and November 1, 2019, respectively. (B) The average Pearson product-moment correlation coefficient between the actual date of first death in the countries and the simulated dates in the four geographical divisions. The transparent blue line represents the correlation coefficient after removing the three outliers (Namibia, Lesotho, and Uganda). (C) The global correlation coefficient before and after removing the three outliers (Namibia, Lesotho, and Uganda). The x-axis represents the date of the emergence of the first case of COVID-19.

The deviation of the median simulation result from the actual scenario demonstrated different degrees of fit depending on the continent. Overall, Europe, North America, and South America possess a higher Pearson product moment correlation coefficient than Asia, Africa, and Oceania.

For Africa, the coefficient remains at a low level and decreases to nearly 0 after October 1. The coefficients of Europe, North America and South America, and Asia and Oceania, however, are within an interval between 0.15 and 0.35 before October 15, after which they drop significantly (Figure 6 b). A maximum of the global average coefficient can be observed on September 15 (Figure 6 c).
The simulation with the emergence of the first case of COVID-19 date resulting in the lowest deviation of the first death from the actual date of the first death in the eight major epicenters which possesses the lowest deviation is later than the dates which they have the highest coefficient of determination. The dates with the lowest deviation of the United States, United Kingdom, Spain, Germany, Belgium, France, Canada, and Italy are October 23, October 8, October 8, October 15, October 8, October 1, October 8, and October 1, respectively. The similar pattern does apply to other countries.

There are some outliers from our simulation which are mainly African countries. The simulation result of Namibia, Lesotho, and Uganda, for example, which the date of the first death caused by COVID-19 happened after July, demonstrates significant deviation (the scenario that results in the latest date of the first death) from the actual scenario: 105 days, 87 days, and 89 days (Figure 6 a). When the three outliers are excluded, the correlation coefficient of Africa was significantly boosted, with a peak of 0.2427 (originally 0.0182) on October 1, 2019 (Figure 6 b). As a result of that, the global average increased after the outliers were removed, with a more obvious maximum on September 22, 2019 (Figure 6 c). The global average demonstrated a Gaussian distribution with an obvious central peak on September 22, 2019 after the removal of outliers; however, the pattern cannot be observed before the removal (Figure 6 c).

Equation for the Gaussian distribution:

\[ f(x) = 0.21 e^{-0.0001x^2} \]

Discussion

Using this model, we evaluated different parameters regarding the deaths caused by COVID-19 worldwide. We found that the number of deaths worldwide by May 1, 2020, the number of deaths in the eight major epicenters on May 1, 2020, and the deviation of the first death in 155 different countries, demonstrates a convergence pattern of the maximum value when the emergence of the first case of COVID-19 is in the week between September 15, 2020 and September 22, 2020.

We observed an exponential decay pattern of the number of deaths caused by COVID-19 when the emergence of the first case delays (Figure 3). We found that the date in which our simulation result is the closest to the actual scenario to be September 22, 2019 (Figure 3). Unexpectedly, the date of the emergence of the first case resulting in the highest degree of fit in the eight major epicenters differed to some extent depending on the continent (Figure 5). When the date is later than October 8, 2019, although the number of deaths worldwide can still fit the actual number of deaths into its 95%PI, the number of deaths in some countries (ie. Italy) resulted from the simulation is far less than the actual number (Figure 3, Figure 5).

We found the overall deviation of the number of deaths by May 1, 2020 from our simulation to be smaller when the emergence of the first case to be in September, 2019 than in October 2019. However, the deviation of the date of the first deaths in Europe demonstrates a minimum on October 15, 2019. The pattern of the deviation in the four different geographical divisions (Europe, North America and South America, Africa, and Asia and Oceania) varies to a great extent (Figure 6). Unexpectedly, the global average Pearson product moment correlation coefficient demonstrates a normal distribution pattern when the date of the emergence of the first case changes (Figure 6).

We observed the inconsistency between geographical divisions. For instance, the maximum degree of fit of the date of the first death in Europe is demonstrated when the emergence of the first case happened on October 15, 2020, whereas the counterparts for the other continents converge before October 1 (Figure 6). This may due to the different passenger load factors of the flights and the effect of ground vehicles on the overall population mobility and thus affect the transmission dynamics. The three outliers contributed to a great reduction of the overall correlation coefficient. We believe a higher degree of fit can be realized if some limitations to the countries for inclusion are added.
Our study is subject to several limitations. First, the census data we acquired is not completed at the same time, thus the transmission scenario in our simulation may be different in some areas. This may also have an impact on our age stratification, altering the mortality rate in some countries.

We did not include inter-state and international travels through ground vehicles, thus may underestimate the exportation in some countries where the trains and buses play dominant role in transportation. In terms of transmission, we did not include asymptomatic cases, which may change the transmission dynamics and therefore influence our modelling results from various aspects. 31,32 The passenger load factor of commercial fights has a great impact on the transmission scenarios; however, we failed to acquire the exact passenger load factor for every flight record but rather used a global average.15 This may lead to an overestimation of the passengers of some flights with lower demands.

Another assumption we made is that Wuhan, China is the location where the virus emerged. We only discussed the scenario which that Wuhan is the sole epicenter at early stage of the virus transmission, whereas the actual scenario could be more complicated.33

Our study provides insights into the early stage transmission of the SARS-CoV-2 virus with the presence commercial flights. Our simulation results suggest a high degree of fit and likelihood of the date of onset of the first case to be in between the time interval between September 15, 2019 and October 1, 2019 with the peak on September 22, 2019. The most likely date of the emergence of the first case we found is over 70 days earlier than the onset date of the first case identified (December 1, 2019).33 Since the global Pearson product moment correlation coefficient demonstrates normal distribution as the onset date of the first case changes, we estimated that the coefficient would be as low as 0.1179 if the onset date of the first case of COVID-19 is December 1, 2019. This may indicate some untested cases in Wuhan at the early stage of the outbreak.34,35 Nevertheless, since mutations will affect the infectivity of the virus, the actual origin of the virus may be far different from our estimation.36

References

1 Yu IT, Li Y, Wong TW, Tam W, Chan AT, Lee JH, et al. Evidence of Airborne Transmission of the Severe Acute Respiratory Syndrome Virus. New England Journal of Medicine. 2004;350(17):1731–9.

2 Hamner L, Dubbel P, Capron I, Ross A, Jordan A, Lee J, et al. High SARS-CoV-2 Attack Rate Following Exposure at a Choir Practice — Skagit County, Washington, March 2020. MMWR Morbidity and Mortality Weekly Report. 2020;69(19):606–10.

3 Silverstein WK, Stroud L, Cleghorn GE, Leis JA. First imported case of 2019 novel coronavirus in Canada, presenting as mild pneumonia. The Lancet. 2020;395(10225):734.

4 Mangili A, Gendreau MA. Transmission of infectious diseases during commercial air travel. The Lancet. 2005;365(9463):989–96.

5 Diseases TLI. Air travel in the time of COVID-19. The Lancet Infectious Diseases. 2020;20(9):993.

6 Davies NG, Kucharski AJ, Eggo RM, Gimma A, Edmunds WJ, Jombart T, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. The Lancet Public Health. 2020;5(7).

7 Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature. 2020;584(7820):257–61.
Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, et al. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. The Lancet Infectious Diseases. 2020;20(6):678–88.

Tuite AR, Fisman DN, Greer AL. Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada. Canadian Medical Association Journal. 2020;192(19).

Nakamura H, Managi S. Airport risk of importation and exportation of the COVID-19 pandemic. Transport Policy. 2020;96:40–7.

IATA. Airline and Location Code Search. https://www.iata.org/en/publications/directories/code-search/ (accessed September 15, 2020).

ISO. ISO 3166 - Country Codes. 2020. https://www.iso.org/iso-3166-country-codes.html (accessed September 15, 2020).

United Nations. World Population Prospects - Population Division. https://population.un.org/wpp/DataQuery/ (accessed September 15, 2020).

IATA. Slower but Steady Growth in 2019. https://www.iata.org/en/pressroom/pr/2020-02-06-01/ (accessed September 15, 2020).

IATA. World Air Transport Statistics. https://www.iata.org/en/publications/store/world-air-transport-statistics/ (accessed September 15, 2020).

Zhang L, Yang H, Wang K, Zhan Y, Bian L. Measuring imported case risk of COVID-19 from inbound international flights --- A case study on China. Journal of Air Transport Management. 2020;89:101918.

Nir-Paz R, Grotto I, Strolov I, Salmon A, Mandelboim M, Mendelson E, et al. Absence of in-flight transmission of SARS-CoV-2 likely due to use of face masks on board. Journal of Travel Medicine. 2020;

Glauser W. Communication, transparency key as Canada faces new coronavirus threat. Canadian Medical Association Journal. 2020;192(7).

Hoehl S, Karaca O, Kohmer N, Westhaus S, Graf J, Goetsch U, et al. Assessment of SARS-CoV-2 Transmission on an International Flight and Among a Tourist Group. JAMA Network Open. 2020;3(8).

Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. New England Journal of Medicine. 2020;382(13):1199–207.

Leung K, Wu JT, Liu D, Leung GM. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. The Lancet. 2020;395(10233):1382–93.

Normile D. Source of Beijing’s big new COVID-19 outbreak is still a mystery. Science. 2020;

Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. 2020;
24 Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet Infectious Diseases. 2020;20(6):669–77.

25 Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. Journal of Travel Medicine. 2020;27(2).

26 CDC. COVID-19 Hospitalization and Death by Age. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-age.html (accessed September 15, 2020).

27 United Nations. World Population Prospects - Population Division [Internet]. United Nations. United Nations; [cited 2020Oct13]. Available from: https://population.un.org/wpp/DataQuery/.

28 中国疾病预防控制中心新型冠状病毒肺炎应急响应机制流行病学组. 新型冠状病毒肺炎流行病学特征分析 [J/OL]. 中华流行病学杂志,2020,41 (2020-02-17).http://rs.yiigle.com/yufabiao/1181998.htm. DOI: 10.3760/cma.j.issn.0254-6450.2020.02.003.

29 WHO. Coronavirus disease 2019 (COVID-19) situation report—102. May 1, 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200501-covid-19-sitrep.pdf?sfvrsn=742f4a18_4 (accessed September 15, 2020).

30 Gushulak BD, Macpherson DW. Population Mobility and Infectious Diseases: The Diminishing Impact of Classical Infectious Diseases and New Approaches for the 21st Century. Clinical Infectious Diseases. 2000;31(3):776–80.

31 Yang R, Gui X, Xiong Y. Comparison of Clinical Characteristics of Patients with Asymptomatic vs Symptomatic Coronavirus Disease 2019 in Wuhan, China. JAMA Network Open. 2020;3(5).

32 Gao Z, Xu Y, Sun C, Wang X, Guo Y, Qiu S, et al. A systematic review of asymptomatic infections with COVID-19. Journal of Microbiology, Immunology and Infection. 2020;

33 Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet. 2020;395(10223):497–506.

34 Wu JT, Leung K, Bushman M, Kishore N, Niehus R, Salazar PMD, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nature Medicine. 2020;26(4):506–10.

35 Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science. 2020;368(6490):489–93.

36 Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell. 2020;182(4).
Supplementary Material

Geographical-structured model

Figure 1: Sample geographical division

We included all county/province level that has airports. Overall, there are 250 countries/regions and 2154 province level districts in our model.1,2

Transmission model

We constructed a conventional SEIR model for each county-level unit. The overall model from a macro scale can be written as:

\[
\begin{align*}
\frac{dS(t)}{dt} &= -\frac{\beta SI}{N} \\
\frac{dE(t)}{dt} &= \frac{\beta SI}{N} - \sigma E \\
\frac{dI(t)}{dt} &= \sigma E - (\delta + \mathcal{D})I \\
\frac{dR(t)}{dt} &= (\delta + \mathcal{D})I \\
\frac{dD(t)}{dt} &= \mathcal{D}I \\
\frac{dr(t)}{dt} &= \delta I
\end{align*}
\]

The flowchart for an individual patient is:
Figure 2: Description of the transmission model

Individuals in the stochastic compartmental model are classified into susceptible, exposed, infectious, and removed states. The cases will be determined to be an exported case or a local case by the stochastic algorithm.

**Stochastic algorithm for flight assign**

We integrated 21,081,805 flights from September 1, 2019 to June 1, 2020 to describe the mobility of the population. To utilize this day, we used a stochastic algorithm to determine the number of patients on board.

For a county with the population of $N$, the number of fights taken off of on a certain day $n$, passenger of each flight of $\alpha_n$, passenger load factor of $\mu$, the possibility of an individual onboard is:

$$P(\text{onboard}) = \frac{1}{N} \sum_{0}^{n} \alpha_n$$

The expectancy of the number of exposed/infectious individuals onboard ($E_O$ and $I_O$) is:

$$I_O + E_O = \sum_{0}^{n} \alpha_n \frac{(E + I)}{N}$$

However, to reduce the amount of computational power required for the model, we simplified this algorithm to a Gaussian distribution function. Therefore, the number of exposed/infectious passengers is:

$$I_O + E_O = (1 + 0 \cdot 3e^{-\eta^2})(\sum_{0}^{n} \alpha_n \frac{(E + I)}{N})$$

After determining the number of patients that will be onboard, randomly pick certain patients to be exported from the province. The algorithm will then assign them a flight and a destination.
We picked 10 flights on January 11, 2020 from Shanghai, China (ISO-3166: CN-SH) to demonstrate our algorithm (Table 1).

Table 1: Sample flight record on January 11, 2020

| Flight number | Departure | Arrival | Departure time | Arrival time | Aircraft | Seats | Passenger code |
|---------------|-----------|---------|----------------|--------------|----------|-------|----------------|
| MU735 PVG     | SYD       |         | 2020-01-10 00:36:00 | 2020-01-10 11:12:00 | B787     | 285   | μ (0-285)      |
| K6861 PVG     | REP       |         | 2020-01-10 01:05:00 | 2020-01-10 05:09:00 | A320     | 180   | μ (286-465)    |
| SQ825 PVG     | SIN       |         | 2020-01-10 00:41:00 | 2020-01-10 05:47:00 | B777     | 264   | μ (466-729)    |
| LH727 PVG     | MUC       |         | 2020-01-10 00:58:00 | 2020-01-10 12:04:00 | A380     | 509   | μ (730-1238)   |
| MU219 PVG     | FRA       |         | 2020-01-10 00:50:00 | 2020-01-10 12:11:00 | A350     | 288   | μ(1239-1526)   |
| MU771 PVG     | AMS       |         | 2020-01-10 01:17:00 | 2020-01-10 12:30:00 | A350     | 288   | μ (1527-1814)  |
| MU553 PVG     | CDG       |         | 2020-01-10 01:03:00 | 2020-01-10 12:59:00 | B777     | 316   | μ(1815-2130)   |
| MU779 PVG     | AKL       |         | 2020-01-10 00:47:00 | 2020-01-10 12:41:00 | B787     | 285   | μ(2133-2415)   |
| MU739 PVG     | MEL       |         | 2020-01-10 00:38:00 | 2020-01-10 11:03:00 | A330     | 234   | μ(2416-2649)   |
| UL867 PVG     | CMB       |         | 2020-01-10 01:09:00 | 2020-01-10 07:58:00 | A330     | 297   | μ(2650-2946)   |

Overall, there are 2946 seats available ($S_a$) to choose from. However, there may be some seats that do not have passengers; therefore, we will multiply $S_a$ by the passenger load factor ($\mu = 82.6\%$) to find the number of passengers on board ($S_o$), which $S_o$ equals:

$$S_o = \mu \cdot S_a = 0 \cdot 82.6 \cdot S_a$$

In this case, $S_o$ is 2,433. The algorithm will then generate a random integer from 0 to $S_o$ (2433) for each passenger, this number is the passenger code for them (Table 1). The passenger code is corresponding to one flight, and this is the flight which the passenger is going to take.

Key parameters in the model

The key epidemiology parameters in our model are acquired from published studies.\(^4\,5\,6\) We used a serial interval of 7 days based on published studies. We also used the mean of the latent period $4 \cdot 5$ days.\(^4\) The patient will be determined whether to demise or recover after a median of $18 \cdot 8$ days (95% credible interval [CrI] 15·7–49·7).\(^4\)

The mortality rate of the model for each individual was determined based on their age according to the table (Table 2).

Table 2: Mortality rate of COVID-19 patient by age in China

| Age   | Mortality rate |
|-------|----------------|
| ≤20   | 0.2%           |
| 20−   | 0.2%           |
| 30−   | 1.2%           |
| 40−   | 0.4%           |
| 50−   | 1.3%           |
| 60−   | 3.6%           |
| 70−   | 8.0%           |
| ≥80   | 14.8%          |
References

1 ISO. ISO 3166 - Country Codes. 2020. https://www.iso.org/iso-3166-country-codes.html (accessed September 15, 2020).

2 IATA. World Air Transport Statistics. https://www.iata.org/en/publications/store/world-air-transport-statistics/ (accessed September 15, 2020).

3 IATA. Slower but Steady Growth in 2019. https://www.iata.org/en/pressroom/pr/2020-02-06-01/ (accessed September 15, 2020).

4 Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet Infectious Diseases. 2020;20(6):669–77.

5 Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. 2020;

6 Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. Journal of Travel Medicine. 2020;27(2).
A

B

Simulation resulting median number of deaths

October 1, 2020

October 8, 2020

October 15, 2020

October 23, 2020

November 1, 2020

Actual number of deaths by May 1, 2020
