

Dating the emergence of the first case of COVID-19 with flights: a retrospective modelling study

Liwei Yang (✉ loveblairsky@gmail.com)

Bayview Secondary School

Jing Yan Zhu

University of Western Ontario

Yihe Zhang

University of Toronto

Wenjie Zhou

Bayview Secondary School

Yicheng Si

Worcester Polytechnic Institute

Research Article

Keywords: Dynamic SEIR model, novel coronavirus, origin-tracing, retrospective modelling, index case dating

Posted Date: November 3rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-92854/v2>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Title: Dating the emergence of the first case of COVID-19 with flights: a retrospective modelling study

Authors: Liwei Yang¹, Jing Yan Zhu², Yihe Zhang³, Wenjie Zhou¹, Yicheng Si⁴

Affiliations:

1. Bayview Secondary School
2. University of Western Ontario
3. University of Toronto
4. Worcester Polytechnic Institute

Keywords: Dynamic SEIR model, novel coronavirus, origin-tracing, retrospective modelling, index case dating

One Sentence Summary: We estimated the date of the first COVID-19 case.

Abstract: Commercial flights contributed to the early-stage international transmission of SARS-CoV-2. Understanding the effect of international and inter-state flights on 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. We constructed a geographical-structured epidemiology model, integrating 2541 province-level units, 250 country-level units, and 26,094,036 flight plans to evaluate the possible date of the emergence of the first case. Using the model, we estimated the number of cumulative deaths and the date of first death caused by COVID-19 in different countries. The pattern of the three parameters we evaluated suggests a high likelihood of the emergence of the first case of COVID-19 be around September 15 and September 22, 2019.

Covering Letter:

Dear *Nature* Editors,

We are a group of high school and freshmen undergrad students from different parts of the world who are passionate about the topic of epidemiology and want to contribute to the mitigation of COVID-19. We researched together during this difficult time with tools that are available and we worked cooperatively to tackle this research even though we are separated by the pandemic and physically in different parts of the world. We submitted a research report entitled “Dating the emergence of the first case of COVID-19 with flights: a retrospective modelling study” for consideration for publication in *Science*. It describes the construction of a geographical-structured epidemiology model and estimates potential dates of the emergence of the first COVID-19 case using this model.

Science, as one of the most prestigious scientific journals in the world, has published some key papers in this field. In July 2020 Britton and his colleagues created a mathematical model to assess the effect of population heterogeneity on the threshold of shielding immunity (herd immunity). We advanced further using a similar model. Inspired by the game *Plague Inc.*, we constructed a model integrating 2154 province-level regions and 21,081,805 commercial flights from September 1, 2019 to June 1, 2020 to date the first case of COVID-19. We used three parameters regarding the deviation from the actual scenario to evaluate our results, finding all of these parameters to indicate a peak of the degree of fit when the emergence was on September 22, 2020.

Our study provides new insights into the early-stage transmission of the COVID-19 and the potential date of the origin of the virus. We believe our model and findings can be used for both origin-tracing of viruses and decisions regarding virus mitigation - especially border restrictions.

Sincerely and respectfully,

The students

October 22, 2020

Dating the emergence of the first case of COVID-19 with flights: a retrospective modelling study

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has impacted multiple countries after first detected in Hubei, China.^{1,2} Evidence proved that the virus can spread through aerosols.^{1,2} Flights were determined to be a crucial contributor to case exportations.^{3,4} Some countries have enforced strict border closures to mitigate the spread of this deadly virus.⁵ Although multiple non-pharmaceutical interventions on the COVID-19 pandemic were introduced, they were not sufficient to contain the spreading of the virus, on both local and global scales.^{6,7}

Multiple modelling studies were conducted to assess the local COVID-19 transmission in the UK, Singapore, and Canada.^{7,8,9} These studies provided insight into the virus transmission and mitigation based on a static population. Results given by these studies can reflect the country-level transmission dynamics and potential epidemic sizes without the influx and outflux of the cases.^{5,10}

International flights contributed to the early-stage spread of the virus,¹ but none of these previous mathematical models included this crucial factor. In 2018, 1,811,324,000 passengers traveled internationally through air.¹² Since commercial flights is one of the main contributors to the importation of COVID-19 cases,^{11,13} we focused on the effect of flights on the transmission of SARS-CoV-2 to better describe the global transmission dynamic.

We present our result of a geographical-structured epidemiology model for the COVID-19 on a global scale. We estimated the number of cumulative deaths caused by COVID-19 by May 1, 2020 and the date of the first death caused by COVID-19 with different settings of the emergence of the first case. By comparing our results with actual scenario, we estimated the most likely date of the emergence of the first case of COVID-19.

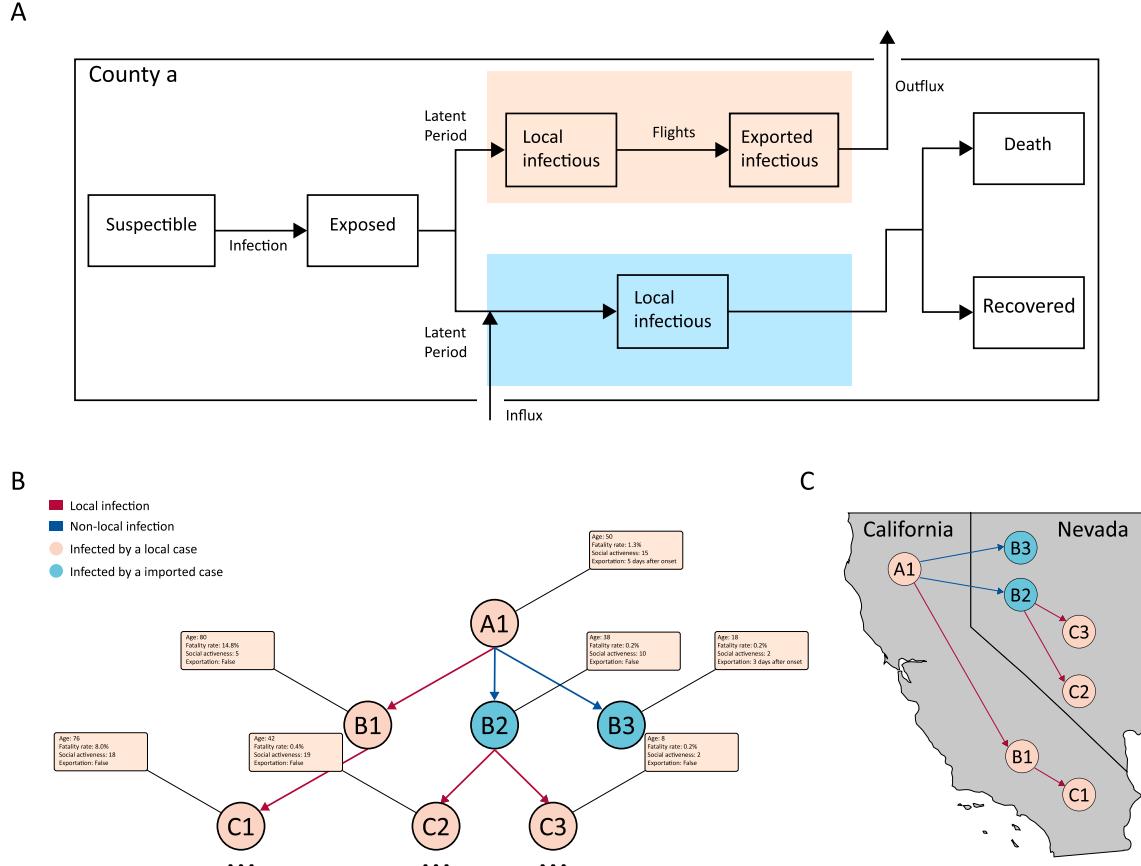


Figure 1: Description of the transmission model.

(A) Individuals in the stochastic compartmental model are classified into susceptible, exposed, infectious, and removed states. The patient will be determined to become whether an exported case or a local case by the stochastic algorithm. (B) Flow chart of local and non-local transmission of COVID-19. Each circle represents a patient. The blue arrow represents a non-local infection, whereas the red arrow represents a local infection. The tag indicates the epidemiological information about the individual. (C) Sample map for case exportation from California, the United States to Nevada, the United States.

Our model described the transmission scenario of COVID-19 with a latent period and an infectious period.¹⁴ 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.^{7,15,16} 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 1).

We completed 50 simulations for each scenario with the date of emergence of the first case of 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 2.79.¹⁷ The fatality rate of each individual in the model was determined by a function based on age. Our data about fatality rate was acquired from a publication by the Chinese CDC (supplementary material).¹⁸

The imported cases will preserve its data (time since onset, age, and social activity pattern) and start transmission accordingly in the new area (Figure 1). Since the average fatality rate of COVID-19 patients demonstrates a strong correlation with the age,^{19,20,21} we stratified the population into 10-year age bands according to the country.^{22,23} The age of each individual will be assigned according to the age stratification.

Our stratification was completed on the country level; thus, we did not consider the difference of population age distribution between different provinces in the same country.²⁸

Due to the reason that the first case of COVID-19 was reported in Hubei, 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.^{24,25}

We introduced two types of lockdown into the model: local level lockdown, and country-level lockdown. The lockdown in a certain province or country will be triggered when a certain number of cumulative deaths caused by COVID-19 is surpassed, 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 provinces, altering the relative frequency of contacts to our best estimate of a reasonable reduction in contact rates under lockdown.⁷

According to ISO-3166 and IATA standard, we introduced 250 country-level regions, then further divided them into 2154 province-level regions (Figure 1).^{26,27} The population of each region was collected through census data (supplementary material).²⁸

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

We assumed the population in each province to be distributed evenly; the population of each province is recorded and integrated into the model (supplementary material).²⁹

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).^{30,31} We picked the global average passenger load factor in 2019 (82.6%) to be the uniform passenger load factor for our model.³²

For a region with a population of N , a number of flights taken off on a certain day n , numbers of passengers of each flight of α_i , and the passenger load factor of μ (82.6%), the possibility of each individual to be onboard is:

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

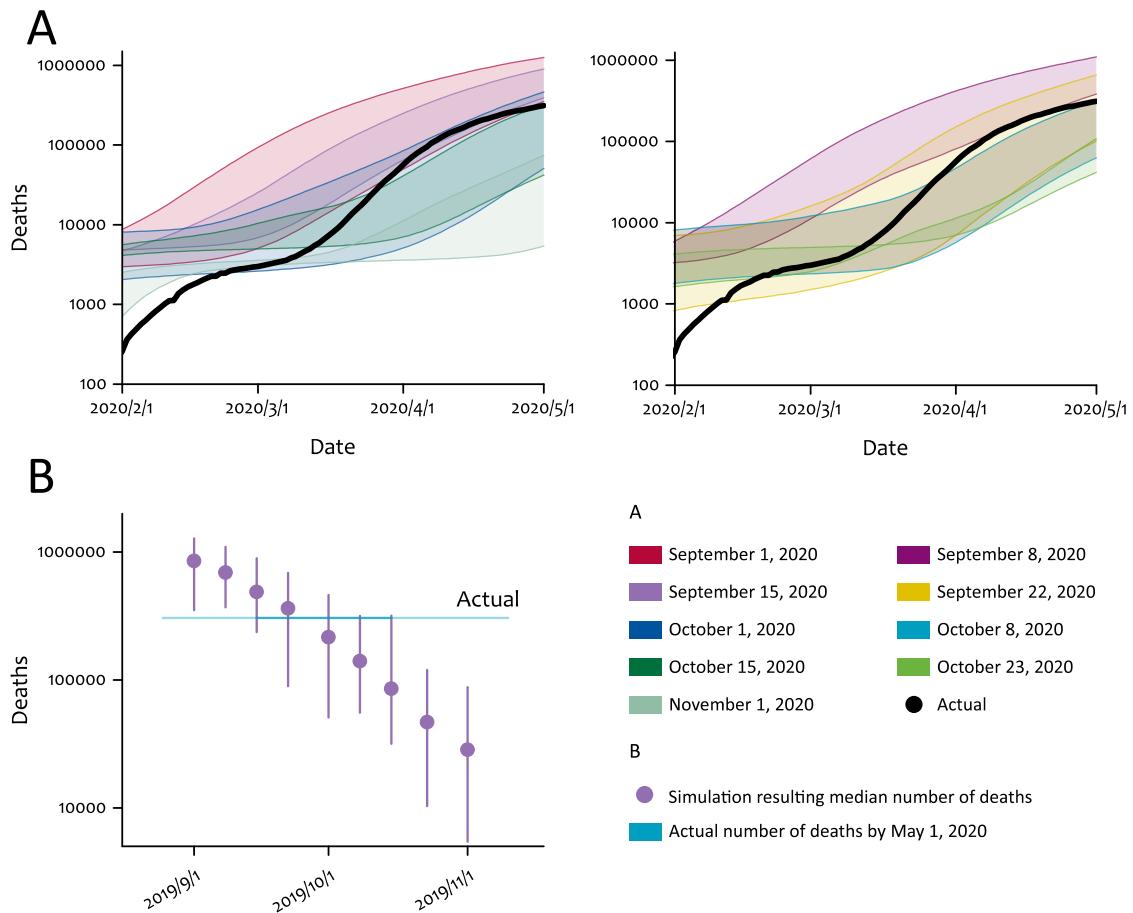


Figure 2: Number of cumulative deaths worldwide by May 1, 2020.

(A) Number of cumulative deaths caused by COVID-19 from February 1, 2020 to May 1, 2020. The colored curves represent the 97.5th percentile and 2.5th percentile of the number of cumulative deaths from each scenario; the colored areas represent 95% prediction interval. The bold line is the actual number of cumulative deaths worldwide. Red, purple, light purple, yellow, blue, sky blue, pine green, green, and light green represents the emergence of the first case of COVID-19: 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 cumulative 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 cumulative deaths on May 1, 2020; the vertical droplines represent 95% prediction interval. The x-axis represents the date of the emergence of the first case. The figures are on a logarithmic scale.

We modelled 9 different dates of the emergence of the first case, 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 was 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 2). The deviation of the median number of cumulative deaths (363746) from the actual scenario (305875) reaches the minimum when the emergence of the first case of COVID-19 was on September 22, 2019.

Overall, the number of cumulative deaths caused by COVID-19 on May 1, 2020 worldwide demonstrates a pattern of exponential decay as the emergence of the first case defers.³³ When the date of the emergence of the first case is as early as September 8, 2019, the actual number of cumulative deaths is lower than the 2.5th percentile; when the date is set as late as than October 22, 2019, the actual number is higher than the 97.5th percentile (Figure 2).

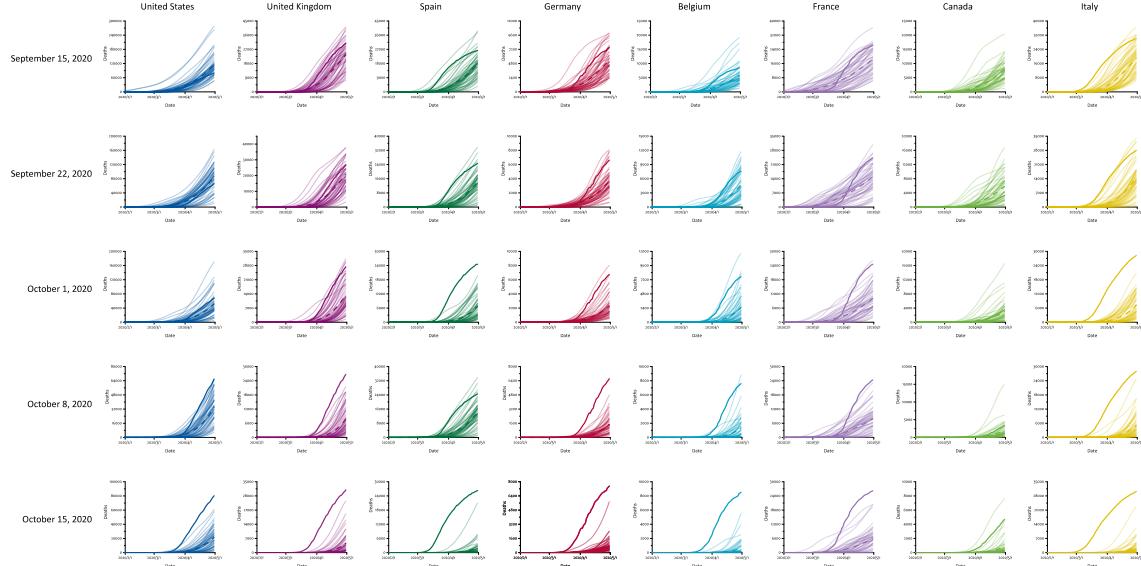


Figure 3: Detailed number of cumulative deaths in the eight major epicentres.

Number of cumulative deaths resulted from 50 realisations from five scenarios with different dates of the emergence of the first case in the eight major epicentres (United States, United Kingdom, Spain, Germany, Belgium, France, Canada, and Italy). The bold lines represent the actual number of cumulative deaths; the dash line is the trial resulting in a median number of cumulative deaths; the transparent line represents one realisation. Scenarios where the actual number of cumulative deaths is not within the 2.5th to 97.5th percentile of the simulation result (September 1, 2019, September 8, 2019, October 23, 2019, November 1, 2019) were excluded for clarity issues. The actual number of cumulative deaths is not within the 95% prediction interval if it is not inside the interval between the trial with the greatest number of cumulative deaths and the trial with the lowest one.

The median number of cumulative deaths in the United States, United Kingdom, Spain, Germany, Belgium, France, Canada, and Italy, the eight major epicentres, demonstrated the lowest deviation from the actual scenario (according to the percentage deviation from the actual number) on September 15, September 22, September 15, September 15, September 15, September 15, October 1, and September 15, respectively (Figure 3). Obviously, as the emergence of the first case of COVID-19 defers, the number of countries that do not include the actual scenario within its 95% prediction interval increases. The actual numbers of deaths are included inside the 95% prediction interval of the simulation result in all of the eight epicentres when the emergence was on September 15, 2019 and September 22, 2019. Seven of eight countries failed to include their actual numbers inside the 95% prediction interval when the emergence was on October 15, 2019 (Figure 3).

Overall, the average date which resulted in the lowest deviation amongst the eight epicentres is between Septembers 15, 2019 and Septembers 22 2019. Among these eight major epicentres, the average date which results in the lowest deviation in North American countries (the United States, and Canada with an average of September 22) is overall later than that of European countries (United Kingdom, Spain, Germany, Belgium, France, and Italy with an average on September 16).

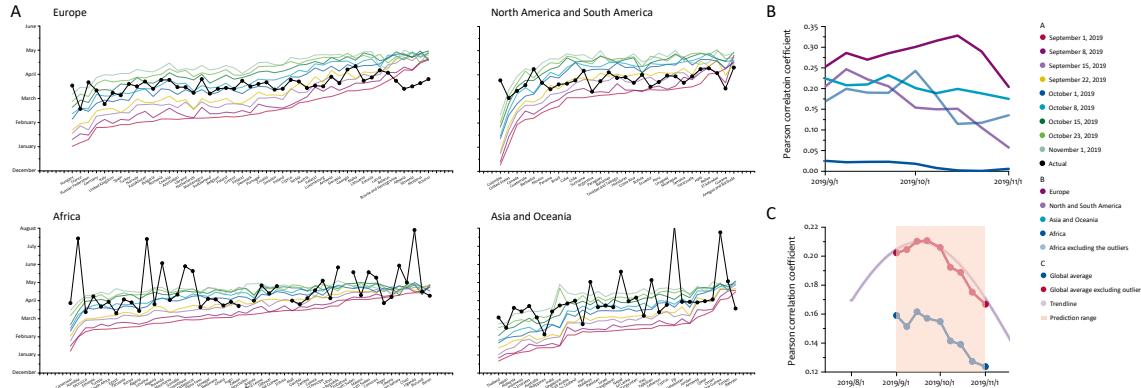


Figure 4: Dates of the first death in 155 countries.

(A) Dates of the first death caused by COVID-19 in 155 countries around the globe (that have COVID-19 death cases); the countries are categorized into Europe, North America and South America, Africa, and Asia and Oceania based on their continents. The points on the bold line represent the actual date of the first death in the countries. Each y-value on the colored line represents the date of the first death in these countries from the trial that resulted the median number of cumulative deaths from each scenario with different dates of the emergence of the first case. The countries are sorted by the date of the first death from the simulation when September 1 is date of the emergence of the first case. Red, purple, light purple, yellow, blue, sky blue, pine green, green, and light green represents the emergence of the first case of COVID-19 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 date from the simulation and the actual date in four geographical divisions. The transparent green line represents the correlation coefficient in Africa after removing the outliers (Namibia, Lesotho, and Uganda). (C) The global correlation coefficient before and after removing the outliers. The x-axis represents the date of the emergence of the first case of COVID-19. The light purple curve represents the trendline.

The deviation of the date of the first death of the simulation result from the actual scenario demonstrated different patterns depending on the geographical divisions. Overall, Europe, North America, and South America possess a higher Pearson 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 (Figure 4). A maximum of the global average coefficient can be observed on September 15 (Figure 4).

There are some outliers from our simulation which are mainly African and Oceania countries. The simulation result of Namibia, Lesotho, and Uganda, for example, which the date of the first death caused by COVID-19 is 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 4). When the 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 4). As a result of that, the global average increased after the outliers were removed, with a more obvious maximum on September 22, 2019 (Figure 4). The global average demonstrated a normal 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 4).

Equation for the trendline (x represents the number of days after September 22):

$$f(x) = 0.21e^{-0.0001x^2}$$

Using this model, we evaluated three different parameters relating to the deaths caused by COVID-19 worldwide. We found that the number of global deaths by May 1, 2020, the number of cumulative deaths in the eight major epicentres on May 1, 2020, and the deviation of the first death in 155 different countries,

all demonstrated the smallest deviation when the emergence of the first case of COVID-19 is in the week between September 15, 2019 and September 22, 2019.

We observed an exponential decay pattern of the number of cumulative deaths caused by COVID-19 as the emergence of the first case defers (Figure 2). We found that the date where the number of cumulative deaths from the simulation shows the smallest deviation from the actual scenario is September 22, 2019 (Figure 2).

The pattern of the deviation of the date of the first death caused by COVID-19 in the four different geographical divisions varies to a great extent (Figure 4). The global Pearson correlation coefficient (after removing the outliers) demonstrates a clear normal distribution pattern when the date of the emergence of the first case changes (Figure 4).

We observed the inconsistency between different 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, 2019, whereas the counterparts for the other continents are before October 1 (Figure 4). 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.³⁴ The outliers contributed to a great reduction of the overall correlation coefficient.

Our study is subject to several limitations. First, the census data we acquired is not up to date, thus the transmission scenario in our simulation may be different in some areas. This may also have an impact on our age stratification, affecting the fatality rate in some countries.¹⁸ We did not include inter-state and international travels through ground vehicles, thus may underestimate the influx and outflux of cases in some countries where the trains and buses play a dominant role in transportation.

Our study provides insights into the early stage transmission of the SARS-CoV-2 virus with the presence of flights. Our simulation results suggest a high likelihood that the date of onset of the first case is 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). Since the global Pearson 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 date of emergence of the first case of COVID-19 is December 1, 2019. This may indicate some untested cases at the early stage of the outbreak. Our estimates of the underlying effect of the flight on this transmission dynamic will help health officials to re-evaluate the early-stage transmission trajectory and policies regarding border closures.

References and Notes

- [1] WHO. Pneumonia of unknown cause – China World Health Organization. World Health Organization; 2020. <https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/> (accessed September 15, 2020).
- [2] 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.
- [3] 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.
- [4] 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.
- [5] Mangili A, Gendreau MA. Transmission of infectious diseases during commercial air travel. *The Lancet*. 2005;365(9463):989–96.
- [6] Diseases TLI. Air travel in the time of COVID-19. *The Lancet Infectious Diseases*. 2020;20(9):993.
- [7] 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).
- [8] 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.
- [9] 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.
- [10] 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).
- [11] Nakamura H, Managi S. Airport risk of importation and exportation of the COVID-19 pandemic. *Transport Policy*. 2020;96:40–7.
- [12] Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*. 2020;
- [13] Devi S. Travel restrictions hampering COVID-19 response. *The Lancet*. 2020;395(10233):1331–2.
- [14] IATA. Slower but Steady Growth in 2019. <https://www.iata.org/en/pressroom/pr/2020-02-06-01/> (accessed September 15, 2020).
- [15] IATA. World Air Transport Statistics. <https://www.iata.org/en/publications/store/world-air-transport-statistics/> (accessed September 15, 2020).
- [16] 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.
- [17] 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;
- [18] Glauser W. Communication, transparency key as Canada faces new coronavirus threat. *Canadian Medical Association Journal*. 2020;192(7).
- [19] 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).
- [20] 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.
- [21] 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.
- [22] Normile D. Source of Beijing's big new COVID-19 outbreak is still a mystery. *Science*. 2020;
- [23] 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. 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] Ghisolfi S, Almås I, Sandefur JC, Carnap TV, Heitner J, Bold T. Predicted COVID-19 fatality rates based on age, sex, comorbidities and health system capacity. *BMJ Global Health*. 2020;5(9).
- [28] Kang S-J, Jung SI. Age-Related Morbidity and Mortality among Patients with COVID-19. *Infection & Chemotherapy*. 2020;52(2):154.
- [29] United Nations. World Population Prospects - Population Division [Internet]. United Nations. United Nations; [cited 2020Oct13]. Available from: <https://population.un.org/wpp/DataQuery/>
- [30] CIA. The World Fact Book. Central Intelligence Agency. Central Intelligence Agency; 2018. <https://www.cia.gov/library/publications/the-world-factbook/fields/341.html> (accessed September 15, 2020).
- [31] 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;10.3760/cma.j.issn.0254-6450.2020.02.003.
- [32] IATA. Airline and Location Code Search. <https://www.iata.org/en/publications/directories/code-search/> (accessed September 15, 2020).
- [33] ISO. ISO 3166 - Country Codes. 2020. <https://www.iso.org/iso-3166-country-codes.html> (accessed September 15, 2020).
- [34] United Nations. World Population Prospects - Population Division. <https://population.un.org/wpp/DataQuery/> (accessed September 15, 2020).

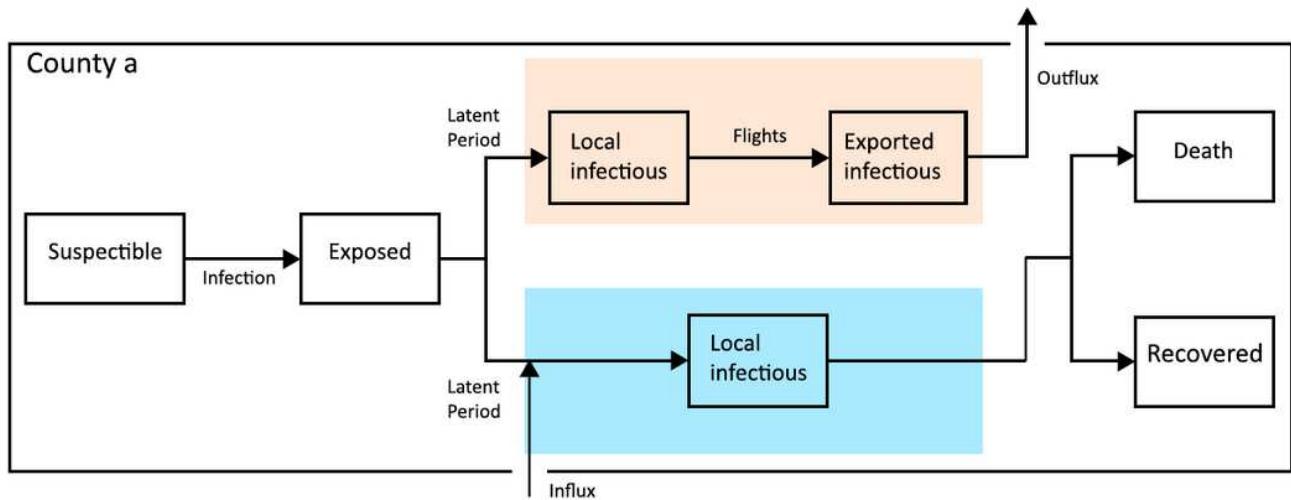
Data and materials availability

All the data and code are available can be accessed through internet
(https://github.com/blaireyung/Global_Flight_Covid-19)

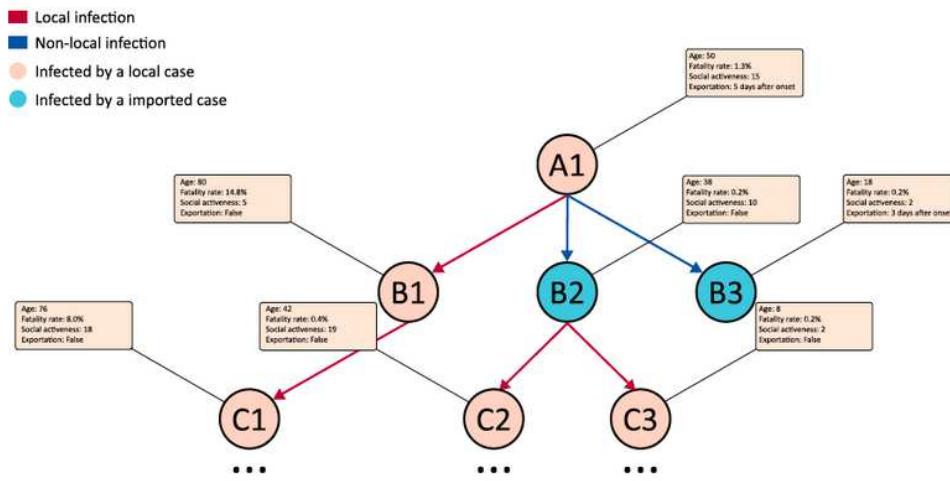
Most parts of our material are available excluding the majority of the flight data (part of the data was shared as a sample).

Figures

A



B



C

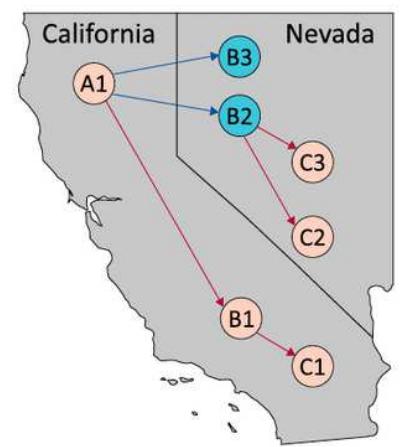


Figure 1

Description of the transmission model. (A) Individuals in the stochastic compartmental model are classified into susceptible, exposed, infectious, and removed states. The patient will be determined to become whether an exported case or a local case by the stochastic algorithm. (B) Flow chart of local and non-local transmission of COVID-19. Each circle represents a patient. The blue arrow represents a non-local infection, whereas the red arrow represents a local infection. The tag indicates the epidemiological information about the individual. (C) Sample map for case exportation from California, the United States to Nevada, the United States.

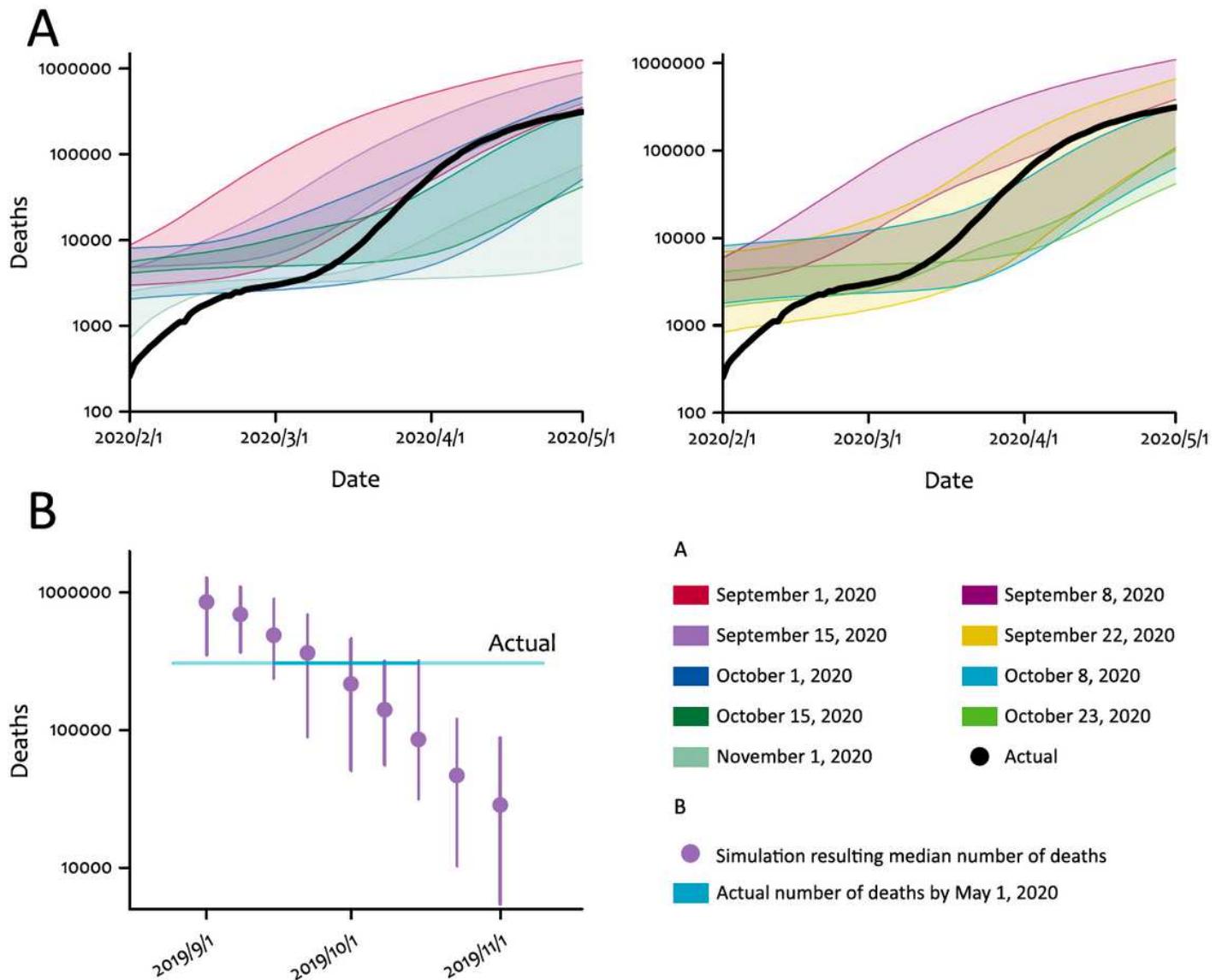


Figure 2

Number of cumulative deaths worldwide by May 1, 2020. (A) Number of cumulative deaths caused by COVID-19 from February 1, 2020 to May 1, 2020. The colored curves represent the 97.5th percentile and 2.5th percentile of the number of cumulative deaths from each scenario; the colored areas represent 95% prediction interval. The bold line is the actual number of cumulative deaths worldwide. Red, purple, light purple, yellow, blue, sky blue, pine green, green, and light green represents the emergence of the first case of COVID-19: 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 cumulative 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 cumulative deaths on May 1, 2020; the vertical droplines represent 95% prediction interval. The x-axis represents the date of the emergence of the first case. The figures are on a logarithmic scale.

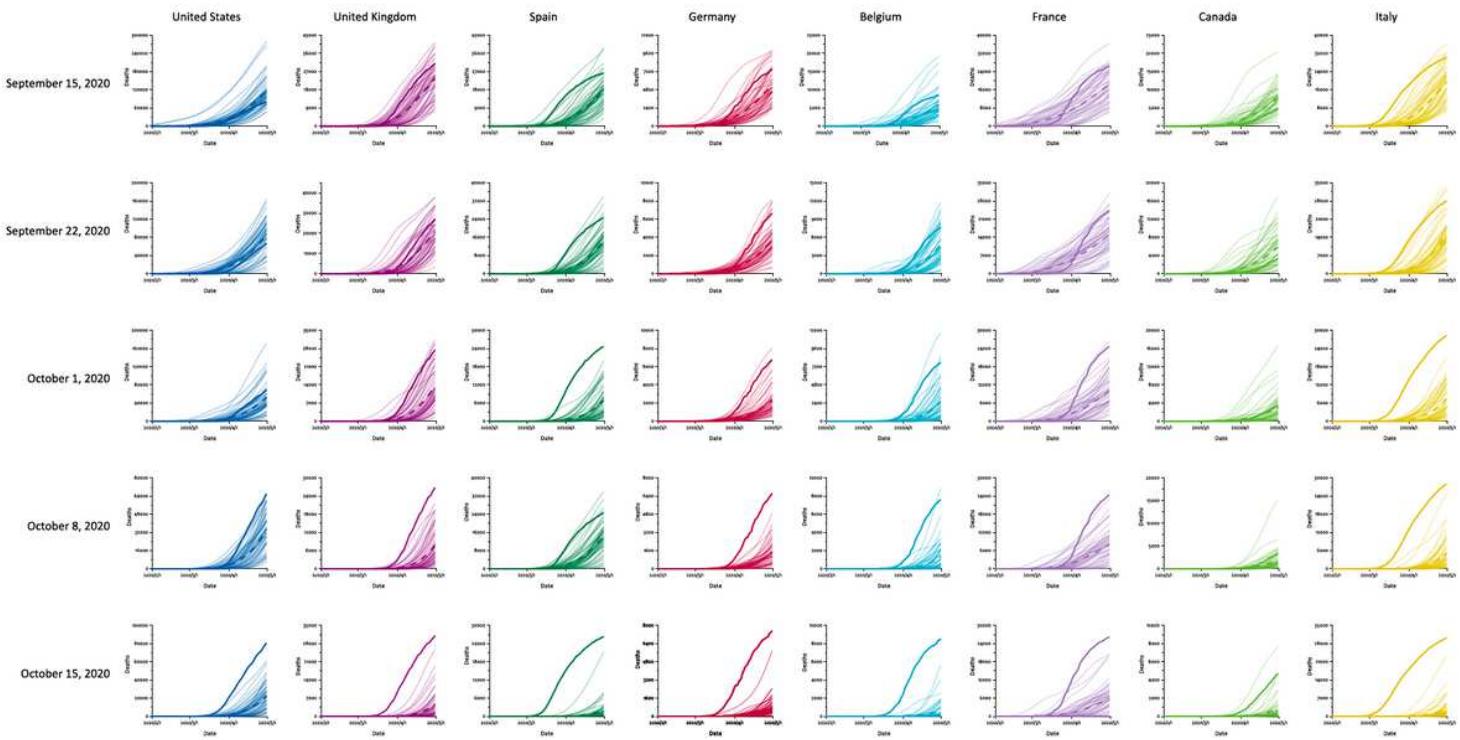


Figure 3

Detailed number of cumulative deaths in the eight major epicentres. Number of cumulative deaths resulted from 50 realisations from five scenarios with different dates of the emergence of the first case in the eight major epicentres (United States, United Kingdom, Spain, Germany, Belgium, France, Canada, and Italy). The bold lines represent the actual number of cumulative deaths; the dash line is the trial resulting in a median number of cumulative deaths; the transparent line represents one realisation. Scenarios where the actual number of cumulative deaths is not within the 2.5th to 97.5th percentile of the simulation result (September 1, 2019, September 8, 2019, October 23, 2019, November 1, 2019) were excluded for clarity issues. The actual number of cumulative deaths is not within the 95% prediction interval if it is not inside the interval between the trial with the greatest number of cumulative deaths and the trial with the lowest one.

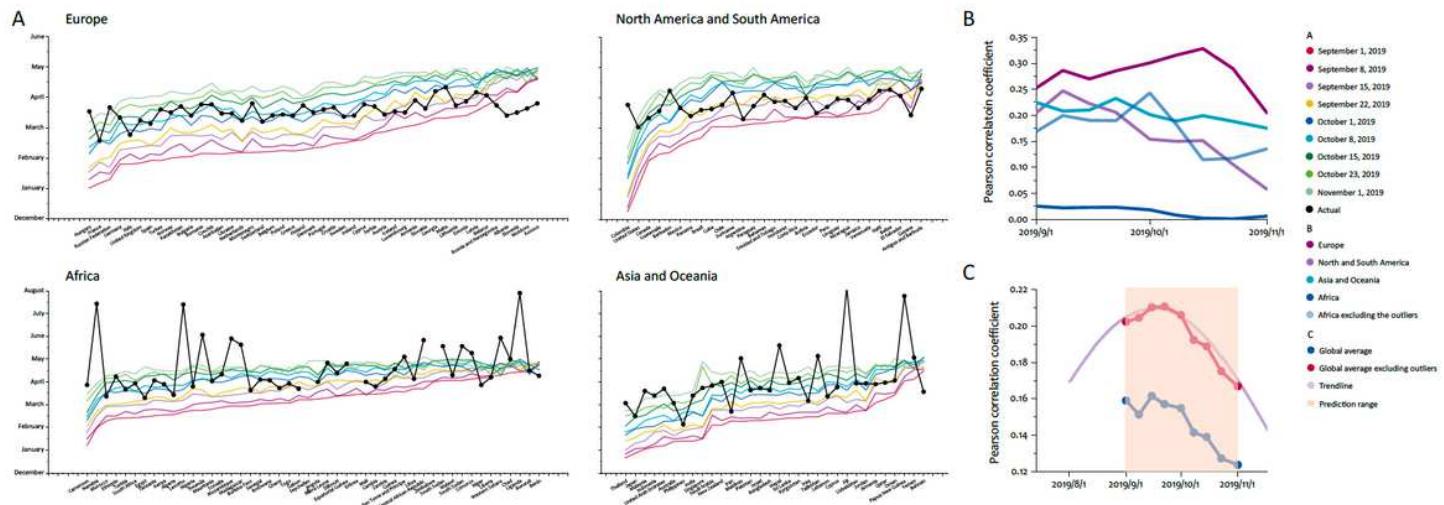


Figure 4

Dates of the first death in 155 countries. (A) Dates of the first death caused by COVID-19 in 155 countries around the globe (that have COVID-19 death cases); the countries are categorized into Europe, North America and South America, Africa, and Asia and Oceania based on their continents. The points on the bold line represent the actual date of the first death in the countries. Each y-value on the colored line represents the date of the first death in these countries from the trial that resulted the median number of cumulative deaths from each scenario with different dates of the emergence of the first case. The countries are sorted by the date of the first death from the simulation when September 1 is date of the emergence of the first case. Red, purple, light purple, yellow, blue, sky blue, pine green, green, and light green represents the emergence of the first case of COVID-19 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 date from the simulation and the actual date in four geographical divisions. The transparent green line represents the correlation coefficient in Africa after removing the outliers (Namibia, Lesotho, and Uganda). (C) The global correlation coefficient before and after removing the outliers. The x-axis represents the date of the emergence of the first case of COVID-19. The light purple curve represents the trendline.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AIR.csv](#)
- [Popfile.csv](#)
- [WPP2019TotalPopulationBySex.csv](#)
- [Path2.csv](#)
- [SupplementaryMaterials.pdf](#)