medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Estimating the number of undetected COVID-19 cases exported internationally from all of China

Authors: Tigist F. Menkir*[1], Taylor Chin*[1], James Hay[1], Erik D. Surface[1], Pablo M. De Salazar[1], Caroline O. Buckee[1], Alexander Watts[4], Kamran Khan[2,3,4], Michael Mina[1], Marc Lipsitch[†1], Rene Niehus[†1]

*shared first authorship

[†]shared last authorship

Affiliation:

[1]Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

[2]Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada.

[3]Department of Medicine, Division of Infectious Diseases, University of Toronto, Canada.

[4]BlueDot, Toronto, Canada

Correspondence:

Tigist F. Menkir, Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, Massachusetts, United States of America. E-mail: tmenkir@hsph.harvard.edu.

Taylor Chin, Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health,677 Huntington Ave, Boston, Massachusetts, United States of America. E-mail: tchin@hsph.harvard.edu.

Marc Lipsitch, Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, Massachusetts, United States of America. E-mail: mlipsitc@hsph.harvard.edu

Keywords: outbreak; Wuhan, China, underdetection; prevalence; COVID-19; flight data; travel ban; Africa

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

During the early phase of the COVID-19 pandemic, when SARS-CoV-2 was chiefly reported in the city of Wuhan, cases exported to other locations were largely predicted using flight travel data from Wuhan. However, given Wuhan's connectivity to other cities in mainland China prior to the lockdown, there has likely been a substantial risk of exportation of cases from other Chinese cities. It is likely that many of these exportations remained undetected because early international case definitions for COVID-19 required a recent travel history from Wuhan. Here, we combine estimates of prevalence in 18 Chinese cities with estimates of flight volume, accounting for the effects of travel bans and the timing of Lunar New Year, to approximate the number of cases exported from cities outside of Wuhan from early December 2019 to late February 2020. We predict that for every one case from Wuhan exported internationally, there were approximately 2.9 cases from large Chinese cities exported internationally that likely remained undetected. Additionally, we predict the number of exported cases in six destinations for which predictions on exported cases have yet to be made, surveillance has likely been low, and where health care systems will likely face issues in managing current or potential outbreaks. We observe heterogeneities in exported case counts across these destinations. The predicted number of cases exported to Egypt and South Africa exceeds the predicted number of cases exported to Mauritania. These trends may anticipate differences in the timing and emergence of local transmission in these countries. Our findings highlight the importance of setting accurate travel history requirements for case definition guidelines in the initial phase of an epidemic, and actively updating these guidelines as the epidemic advances.

Introduction

In late December 2019, researchers identified a new coronavirus disease, later named Coronavirus Disease 2019 (COVID-19), in Wuhan, Hubei Province, China.[1] Rigorous measures to curtail the spread of COVID-19 including travel restrictions and school and workplace closures, have largely controlled the outbreak in mainland China.[2] However, international exportation of COVID-19 cases before the outbreak was contained in mainland China has been critical in the global spread of COVID-19,[3] which has now become a pandemic. As of 18 March 2020, 191,127 confirmed cases of COVID-19 have been registered worldwide, with 81,116 detected in mainland China and the remainder detected internationally in 156 locations.[4]

Because the majority of cases in the early phase of the epidemic were reported in Wuhan, early COVID-19 case definitions and clinical guidelines required individuals suspected of infection to have had a recent travel history from Wuhan.[3,5] Due to the high travel volume within China in January for the Lunar New Year Spring Festival, it has been suggested that a significant number of COVID-19 cases were introduced to other large cities in China before travel restrictions were instituted on 23 January 2020.[6,7] The restrictive surveillance case definition used internationally in January suggests that cases of COVID-19 remain potentially undetected in many countries.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The potential for introduction of COVID-19 from China to African countries was significant considering their highly interconnected commercial relationship and daily flight volume to and from China. Before 1 March 2020, Ethiopian Airlines – a major airline carrier in Africa – continued operating flights from China to the continent.[8] More than 600 confirmed cases have been reported in 34 African countries as of 19 March 2020.[9] Coverage of COVID-19 diagnostic and control interventions is expanding yet still limited, and many of these nations may struggle to adapt to increased demand.[10,11] In light of these issues, assessing the risk of COVID-19 introduction across countries in Africa is a major public health challenge.

Existing approaches have used historical air connectivity data to estimate risk of importation to large cities in China[6,12] and locations outside of China.[7,13,14] Gilbert *et al.* (2020), meanwhile, combined travel data with incidence data to estimate the risk of importation from all Chinese provinces, excluding Hubei, to all African countries.[15] This analysis, however, used historical flight data from January 2019, which may not accurately reflect current travel trends, given that the Lunar New Year occurred in January this year and travel restrictions were in place starting late January. In addition, by using the reported number of cases and the province's population size to estimate incidence in each province, the analysis did not take into account the reliability of data on reporting rates.

Recent work using flight data has discussed the marked variation in detection capacity for imported cases across locations and has identified locations that may have undetected imported cases.[16] Further work estimated that only a minority of cases exported from Wuhan were detected in their destination locations.[3] These predictions are limited given that 1) a substantial number of cases have been reported outside of Hubei province and 2) Wuhan Tianhe International Airport contributes only a fraction of all air-travel volume from mainland China.

Here, we estimate the number of internationally exported cases from mainland China to 16 destinations, and identify countries in Africa with the highest number of potentially underdetected imported cases. We first estimate daily flight volume from 18 cities in China to 16 international destinations from December 2019 to February 2020. Importantly, our estimates of flight volume take into account increased travel in early January due to the Lunar New Year holiday and reduced travel in late January due to travel restrictions. We combine our flight volume estimates with a measure of prevalence in each province in China to quantify the ratio of the number of internationally exported cases from Wuhan to the number of internationally exported cases from other large cities in China.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Methods

1. Estimating number of airplane passengers from 18 Chinese cities to international destinations

Data

We define three groups of locations for our analyses: 1) Chinese cities as likely sources of exported cases; 2) international locations with high surveillance capacity and high air travel connectivity to Wuhan (used for model calibration); and 3) African locations as destinations (used for model prediction). For all flight data, in addition to Wuhan, we include as origin locations (N_0 =17) the 17 Chinese cities that were previously identified by Lai *et al.* (2020)[7] as high-risk cities for importation of COVID-19 from Wuhan: Beijing, Shanghai, Guangzhou, Zhengzhou, Tianjin, Hangzhou, Jiaxing, Changsha, Nanjing, Nanchang, Shenzhen, Chongqing, Chengdu, Hefei, Fuzhou, Xi'an, and Donngguan.

For destinations outside of China, we consider a selection of locations with both high air travel connectivity to Wuhan and high surveillance capacity (thus 'high surveillance locations') for model validation. We assess surveillance capacity using the Global Health Security (GHS) Index, in particular its components of "early detection and reporting epidemics of potential international concern", published in 2019[17]. We thus select locations with the highest connectivity to Wuhan as estimated by Lai *et al.* (2020)[7], and within the top 5% percentile of the GHS index rank. We additionally include Singapore as it has demonstrated a strong capacity to identify, trace and document COVID-19 cases[16,18], despite having a relatively low GHS index. We incorporate N_D = 16 total destination locations, which include Singapore, U.S., Australia, Canada, Korea, UK (and Northern Ireland), Netherlands, Sweden, Germany, and Spain, for model validation. The cities we select for model prediction are Mauritius, Johannesburg, Nairobi, Addis Ababa, Cairo, and Casablanca, which are the 6 top destination cities in Africa in terms of air-travel volume from 18 high-risk cities in mainland China.[7]

We use data from BlueDot[19] on the monthly number of confirmed passengers on flights (direct and indirect) for each of the *N* origin-destination pairs, henceforth referred to as air-travel volume, from December 2018 to February 2019. In addition to air-travel volume, we use data from Cirium[20] on the number of daily departed (and landed) direct passenger flights for each of the *N* origin-destination pairs, henceforth referred to as 'flight departures', for the period December 2018 to February 2019, and December 2019 to February 2020.

At the time of writing, data on air-travel volume were not available for the pre-lockdown epidemic period and total epidemic period, which we considered to be from 8 December 2019 (date of epidemic seeding in Wuhan) to 23 January 2020 and from 8 December 2019 to 28 February 2020, respectively. Historical air-travel volume data are likely not representative of the epidemic time period for two reasons: 1) Lunar New Year was earlier than in preceding years (25 January 2020), and 2) large-scale travel bans and flight cancellations took place in late

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

January 2020. We therefore combined monthly historical air-travel volume data with recent daily flight departure data to estimate daily air-travel volume out of cities in China to our destinations of interest during the total epidemic period, as described in the Supplement.

2. Estimating daily prevalence of COVID-19 in 18 Chinese cities

Daily incidence of new COVID-19 cases per day in each of the Chinese provinces included in our analysis is sourced from Hay *et al.* (forthcoming), which combined confirmed case count data by province, Baidu mobility data[21,22] from Wuhan to all other provinces, and delay distributions from infection onset to reporting, in order to map estimated incidence in Wuhan to other provinces. This method assumes logistic growth of cumulative infection incidence in Wuhan and both importation of infected individuals and limited local transmission in all other provinces.

We assume that the prevalence of infected individuals relevant to exportation included individuals that were infected and not yet symptomatic, and symptomatic but not yet confirmed. Number of prevalent cases per day from daily incidence for province *l* at time *t* is estimated by summing over individuals who are infected but not yet symptomatic and individuals who are symptomatic but not yet confirmed as:

$$Prev_l(t) = \sum_{x=0}^{t} I_l(x)(1 - F(t-x)) + \sum_{x=0}^{t} E_l(x)(1 - G(t-x)),$$

where $I_l(x)$ is the incidence of infections on day x in province l; $E_l(x)$ is the incidence of symptom onsets on day x in province l; $F(t-x)$ is cumulative density of the incubation period distribution (assumed to follow a log normal distribution), giving the probability that an individual infected on day x becomes symptomatic within $t-x$ days post infection; and $G(t-x)$ is the cumulative density of the confirmation delay distribution (assumed to follow a gamma distribution), giving the probability that an individual with symptom onset on day x is confirmed within $t-x$ days post symptom onset. For each province, we calculate the median number of prevalent cases by day using 1000 posterior samples of new infections per day estimated by Hay *et al.* (forthcoming).

Finally, we convert province-level daily number of prevalent cases estimates to city-level daily number of prevalent cases estimates for the 18 cities of interest by allocating total prevalent cases in a city's province entirely to that city. The estimated number of prevalent cases in cities in our dataset which share a province are set to equal the quotient of the total number of prevalent cases in the province and the number of cities within it. In so doing, we assume that individuals in a province are equally likely to go to the specific airports in our analysis. The final measure of prevalence (as a proportion) is standardized by population size and divided by a non-Wuhan ascertainment rate of 10.5%, which is equivalent to the estimated ascertainment rate in Wuhan up to 30 January 2020. We note that this ascertainment rate is likely an under-estimate for the majority of China, and prevalence estimates under this assumption therefore represent an upper bound.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

3. Estimating number of exported cases to international destinations

3.1 Model training: Associating flight volume of infected passengers from Wuhan to observed number of Wuhan-origin cases in validation set locations

We first fit a model to the number of imported COVID-19 cases from Wuhan observed in the high surveillance locations to determine the relationship between prevalence, air-travel volume and imported case counts. We use this model fit to subsequently make predictions using data from Wuhan and the remaining cities in China. The number of observed cumulative cases imported from Wuhan to destination *j* is denoted as y_j. Further, $y_j^* = 2.5y_j$ denotes the number of cases that each destination location *j,* excluding Singapore, could have detected with a surveillance capacity of Singapore[3] (for Singapore $y^*_{j=Singapore} = y_{j=Singapore}$). We assume that across the high surveillance destinations (U.S., Australia, Canada, Korea, UK and Northern Ireland, Netherlands, Sweden, Germany, Spain, and Singapore), this number follows a Poisson distribution, as follows:

$$y_j^* \sim Poisson(\alpha C_{w,j})$$

$$C_{w,j} = \sum_t prev_{w,t} * v_{t,w,j} \,,$$

where $C_{w,j}$ represents the force of exportation from Wuhan to each destination *j* , which is calculated as the product of COVID-19 prevalence in Wuhan ($prev_{w,t}$) and volume of passengers from Wuhan to destination *j* ($v_{t,w,j}$) on day *t*, summed over all days in the pre-lockdown epidemic period. We fit this model in R (version 3.6.1)[23].

3.2 Model application: Predicting exported case counts to subset of African countries

The force of exportation of COVID-19 from all selected cities in China to destination *j* (one of Mauritius, Johannesburg, Nairobi, Addis Ababa, Cairo, and Casablanca) is computed as:

$$C_j = \alpha \sum_i^{N_0+1} \sum_t prev_{i,t} * v_{t,i,j}$$

where $prev_{i,t}$ is the prevalence of COVID-19 in Chinese city *i* at time *t*, and $v_{t,i,j}$ is the total volume of passengers across flights from each origin city *i* to each of the destinations *j* at time *t*. The product of daily passenger volume ($v_{t,i,j}$) and COVID-19 prevalence in city *i* ($prev_{i,t}$) is summed over all days of the total epidemic time period, and over all N_0 Chinese cities and Wuhan. We use this force of importation to make predictions for all six African locations. Further, we compute the 95% prediction interval (PI) bounds under our model fit to the high surveillance locations. To do so, we first generate bootstrapped datasets by sampling six locations with replacement among the six African countries. Second, we re-estimate α using this bootstrapped dataset. Third, we simulate case counts for all six African prediction locations under our model using the estimate of α based on the bootstrapped dataset and C_j as described above. These three steps are repeated 50,000 times to generate for each of the

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

African locations 50,000 simulated case counts from which the lower and upper PI bounds (2.5th and 97.5th percentiles) are computed.

3.3 Ratio of force of exportation from Wuhan versus from the rest of China

To estimate the ratio of expected exported cases from Chinese cities without Wuhan versus expected exported cases from Wuhan, we computed both the force of exportation from Wuhan ($C_{w,j}$, as defined above) and the force of exportation from the 17 Chinese cities, excluding Wuhan, as follows:

$$C_{\overline{w},j} = \sum_{i}^{N_0} \sum_{t} prev_{i,t} * v_{t,i,j} ,$$

where the product of daily passenger volume ($v_{t,i,j}$) and COVID-19 prevalence in city *i* ($prev_{i,t}$) is summed over all days of the total epidemic time period, and over the N_0 Chinese cities (excluding Wuhan). The ratio of $C_{\overline{w},j}$ and $C_{w,j}$ gives the ratio of expected cases exported from outside Wuhan and cases exported from Wuhan. We compute this ratio for three different sets of locations (*j*∈{high surveillance locations}, *j*∈{African locations}, *j*∈{all locations}), according to:

$$R = \frac{\sum_j C_{\overline{w},j}}{\sum_j C_{w,j}} .$$

Results

1. Estimating number of airplane passengers from 18 Chinese cities to international destinations

Supplementary Figure 1 shows the estimated daily air-travel volume from each of the 18 airports in the 18 Chinese cities of our analysis to the 16 international destinations of interest. The trends in Supplementary Figure 1 illustrate the necessity of adjusting historical flight data to reflect 1) the earlier Lunar New Year holiday in 2020 and 2) the widespread travel restrictions in place by late January 2020. Air-travel volume increased in January during the Chunyun holiday – the 40-day holiday period surrounding Lunar New Year that began on 10 January 2020 this year. The number of flight departures from 10 January 2020 to 22 January 2020 was approximately 8% higher than what we would expect based on data from the same time period in 2019. The sharp decline in air-travel volume after 23 January 2020 is a result of the travel restrictions and flight cancellations that occurred starting late January; the number of flight departures 23 January 2020 to 28 February 2020 is approximately 60% less than that of the same time period in 2019.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2. Estimating daily prevalence of COVID-19 in 18 Chinese cities

Supplementary Figure 2 presents daily prevalence estimates in 18 Chinese cities from 1 December 2019 to 28 February 2020. Prevalence is estimated to have peaked in all cities between 24 January 2020 and 28 January 2020. Wuhan has the largest estimated daily prevalence in the time period, at approximately 1.7×10^{-2} on 27 January 2020, representing the proportion of the population relevant to exportation (that is, infected and not yet symptomatic, and symptomatic but not yet confirmed). Jiaxing has the second largest estimated daily prevalence after Wuhan, with a value of 8.4×10^{-3} on 28 January 2020. Shanghai and Tianjin have the lowest daily peak prevalences in the time period, with values of 1.0×10^{-4} and 8.7×10^{-5}, respectively.

3. Estimating number of imported cases to international destinations

3.1 Model training: Associating flight volume of infected passengers from Wuhan to observed number of Wuhan-origin cases in validation set locations

From our poisson regression, we estimate the scaling factor to be 0.876. The scaling factor, α, is interpreted as the ratio of the number of observed cases, corrected for low surveillance relative to Singapore, exported from Wuhan to our list of high-surveillance locations to the flight volume of infected travelers from Wuhan to those locations. By assumption, the scaling factor then represents the ratio of the number of actual (but not yet observed, due to lack of testing) cases exported from other Chinese cities outside of Wuhan to African countries and the flight volume of infected travelers from other cities in China to these African locations.

3.2 Model application: Predicting exported case counts to subset of African countries

Our model estimates that South Africa and Egypt received the most imported cases (3.0 and 2.5, respectively) from the 18 Chinese cities in our analysis over the period 8 December 2019 to 28 February 2020. We estimate the next highest import count in Kenya (1.3), followed by Morocco (1.0), Ethiopia (0.8), and Mauritania (0.4). Figure 1A shows the total estimated number of exported cases to the 6 African countries. Supplementary Table 2 provides bounds to our estimates based on the bootstrap procedure. The relatively low number of mean cases estimated to be imported to each of the African countries in our analysis during the total epidemic period may in part be due to the sharp reduction in air-travel volume starting late January to these countries (Supplementary Figure 3). Figure 1B illustrates the weekly predicted number of imported cases in each of the 6 African countries over time. The peak in weekly number of imported cases coincides for all countries on the week of 19 January 2020, with similar trajectories of imported cases from the 18 Chinese cities over the study period.

Figure 2 presents the ranking of the 18 Chinese cities in our analysis, ordered by their contribution to the total exported cases from all origin cities in China, from 8 December 2019 to 20 February 2020 for each of the 6 African countries. The African countries are ordered based

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

on the number of imported cases over the total epidemic period. Beijing, Wuhan, Guangzhou, Chengdu, and Jiaxing, rank among the top three cities in terms of fraction of exported cases across the 6 African countries in our analysis. In addition to these cities, Shanghai is the leading contributor of exported cases in Mauritania. Nanchang consistently contributes the lowest share of exported cases across all countries. This figure highlights the variability in Chinese cities' contribution to the number of imported cases in African countries.

3.3 Ratio of force of exportation from Wuhan versus from the rest of China

We estimate that for every case from Wuhan exported to the high-surveillance international destinations, there were approximately 2.9 cases from 17 other large Chinese cities exported internationally that were likely undetected, due to case definitions requiring a travel history from Wuhan only. For the subset of African countries this figure was higher at 5.1, suggesting discordances across destination regions in the relative role of Wuhan in driving total imported cases from China.

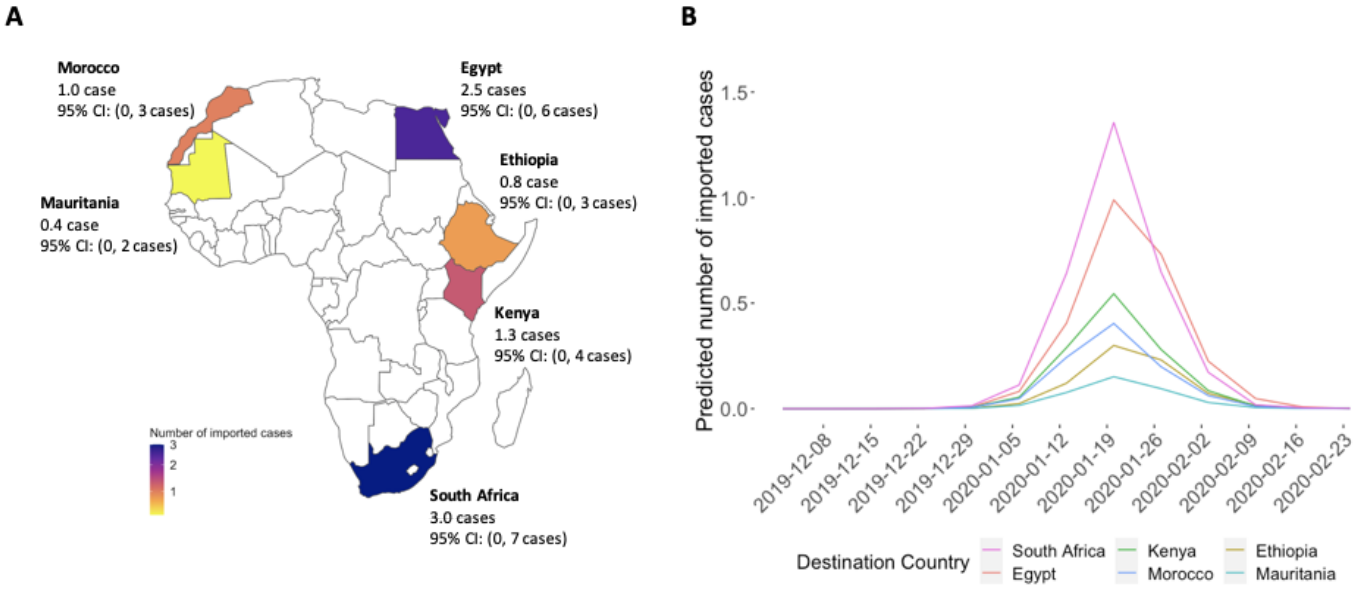


Figure 1. A. Map of 6 African countries in analysis (Ethiopia, South Africa, Egypt, Mauritania, Morocco, and Kenya) and the predicted number of imported cases (with 95% Confidence Intervals) from 18 Chinese cities during the total epidemic period (8 December 2019 to 28 February 2020). B. Weekly number of predicted imported cases from 18 Chinese cities in analysis to 6 countries in Africa during the total epidemic period.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

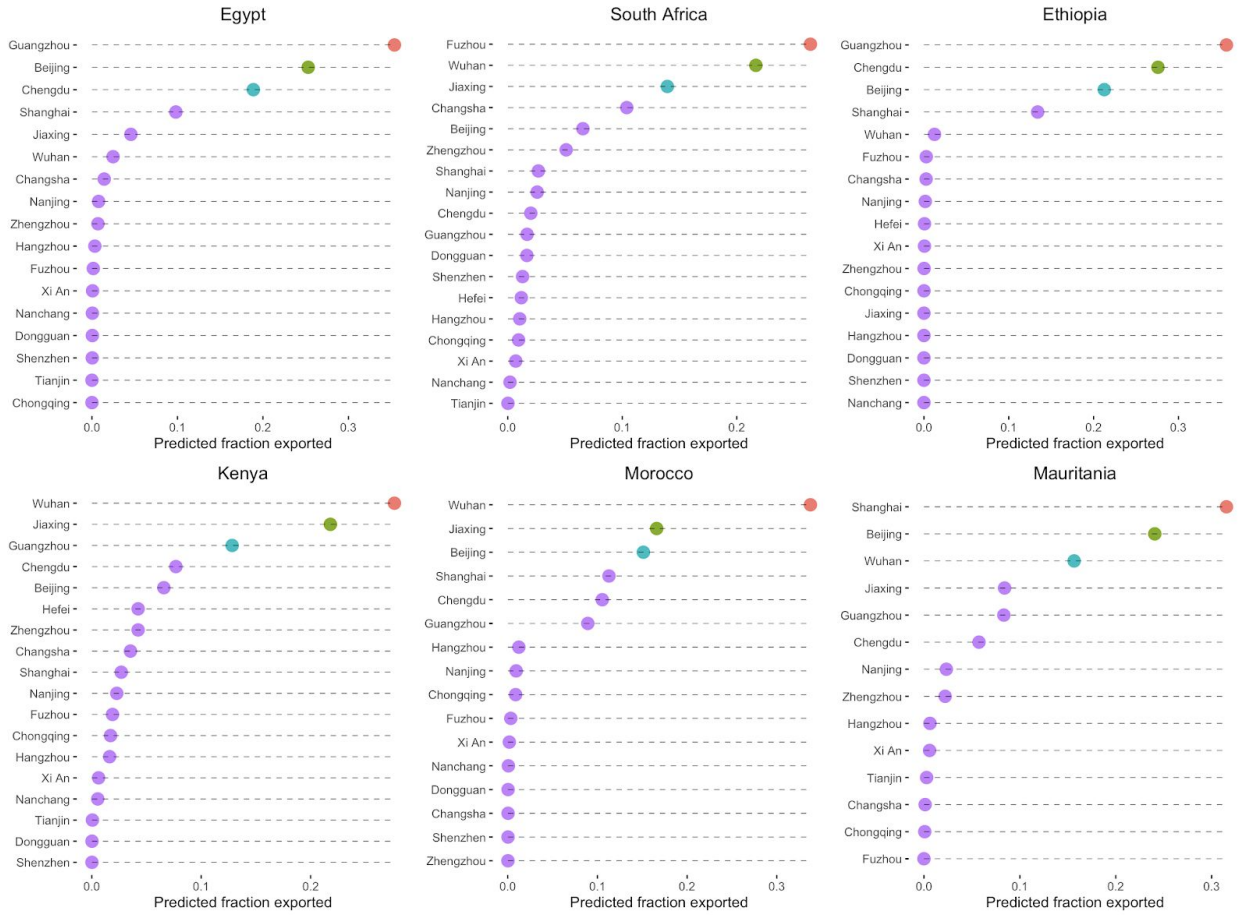


Figure 2. Rank of 18 Chinese cities by fraction of all predicted exported COVID-19 cases to each of the 6 countries in Africa included in our analysis. Countries are ranked from left to right by the total number of imported cases from 18 Chinese cities from 8 December 2019 to 28 February 2020. The red dot indicates the top Chinese city for the country in terms of predicted fraction of all exported cases over the time period; green indicates the second top Chinese city; blue indicates the third top Chinese city; and purple indicates all other cities.

Discussion

This study aimed to make predictions about exported COVID-19 cases from all of China, which differs from previous predictions in two fundamental ways: 1) instead of estimating risk of importation,[15] our model predicts actual number of cases, and importantly does so for countries on the African continent; and 2) instead of accounting only for travellers from Wuhan,[16] we account for travellers from all major Chinese airports as a potential source population. Using prevalence estimates of COVID-19 across all Chinese provinces and estimates of passenger flight volume during the ongoing epidemic, our model predicts that until the end of February (28 February 2020), for every internationally exported case from Wuhan to a high-surveillance location, approximately 2.9 cases were exported from other cities in China. For the African countries studied here taken together, the ratio was 5.1. We further predict that before the end

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

of February a total of 9 COVID-19 cases from all of China were imported to the six African destinations. Of those, the highest number was likely exported to Johannesburg, South Africa (3.0; 95% CI: 0, 7 cases) and Cairo, Egypt (2.5; 95% CI: 0, 6 cases), followed by Nairobi, Kenya (1.3; 95% CI: 0, 4 cases), Addis Ababa, Ethiopia (0.8; 95% CI: 0, 3 cases), Casablanca, Morocco (1.0; 95% CI: 0, 3 cases) and Mauritius, Mauritania (0.4; 95% CI: 0, 2 cases) (Figure 1A). The majority of those 9 cases (approximately 71%) have likely been exported between 12 January 2020 and 2 February 2020 (Figure 1B). The probability that a single undetected introduction leads to sustained local transmission is a relevant question that remains unknown. Our findings provide an initial framework for approaching these questions.

Recent work has suggested that about 62% of internationally exported COVID-19 cases have remained undetected, even in high surveillance locations.[3] Here, we suggest that underdetection might have been substantially larger because surveillance had focused on exported cases from Wuhan. Our predictions suggest that for each case exported from Wuhan, 2.9 cases from the rest of China were exported. This would mean that for each imported and then detected case, 1.8 cases remained undetected due to imperfect detection[3], and 7.1 additional cases remained undetected due to surveillance focusing on exportation from Wuhan. As a result, despite concerted efforts to halt case importation through airports, we expect that the majority of cases evaded surveillance and may have in part caused the observed unexpectedly high levels of local transmission. These findings underscore the importance of appropriate case definitions in efforts to contain an epidemic.

At the time of writing, 18 March 2020, Egypt reports 166 confirmed cases, South Africa 62, Morocco 38, Kenya 3, Ethiopia 5, and Mauritania 1.[4] Importantly, by the end of February (February 28), none of these locations had confirmed any cases, apart from Egypt with one confirmed case on 14 February 2020.[24] The other locations detected their first cases on March 2 for Morocco,[25] March 5 for South Africa,[26] March 13 Ethiopia[27] and Kenya[28], and March 14 Mauritania.[29] If our predictions are accurate, then undetected imports of cases already occurred a month before the first cases were actually confirmed, indicating a serious risk that these imports have initiated unobserved local transmissions.

Our findings may only partially explain imported case counts in Ethiopia, South Africa, Egypt, Mauritania, Morocco, and Kenya, as other modes of transport, such as marine travel, or trips from cities outside of China, could be more significant sources of transmission to these nations. Furthermore, since 28 February 2020, patterns of travel are likely to have changed appreciably, due to recently implemented travel restrictions and airline cancellations, and growing widespread concern, among other factors. The significant decline in flights to African countries from all the 18 Chinese cities included in our analysis in late January is shown in Supplementary Figure 3. It is thus essential to regularly update the methods we proposed here with the most recent connectivity data and to consider a broader range of origin cities and forms of travel.

We provide a framework for incorporating available travel data, accounting for flight trends in the present climate, and back-calculating prevalence estimates to construct, validate, and test a

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

measure of the number of exported cases to a subset of African destination nations. Through this analysis, we quantified the number of (potentially untested and unknown) cases from a subset of cities in China to select African countries with high air traffic to and from China from 8 December 2019 to 28 February 2020. This methodology can be applied to other outbreak settings to estimate the number of exported cases potentially missed due to early restrictive case definitions. Our findings are relevant to understanding time-to-sustained local transmission for outbreaks in locations with limited air-travel connectivity to Wuhan, Hubei, and forecasting future transmission and potential sources of introductions or reintroductions.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Taylor DB. A Timeline of the Coronavirus [Internet]. N. Y. Times. 2020 [cited 2020 Feb 21];Available from: https://www.nytimes.com/2020/02/13/world/coronavirus-timeline.html
2. Lai S, Ruktanonchai NW, Zhou L, et al. Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak in China. medRxiv 2020;2020.03.03.20029843.
3. Niehus R, Salazar PMD, Taylor A, Lipsitch M. Quantifying bias of COVID-19 prevalence and severity estimates in Wuhan, China that depend on reported cases in international travelers. medRxiv 2020;2020.02.13.20022707.
4. World Health Organization. Coronavirus disease 2019 (COVID-19): Situation Report – 58 [Internet]. 2020. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200318-sitrep-58-covid-19.pdf?sfvrsn=20876712_2
5. Clinical management of severe acute respiratory infection when novel coronavirus (2019-nCoV) infection is suspected: Interim guidance [Internet]. World Health Organization; 2020. Available from: https://www.who.int/docs/default-source/coronaviruse/clinical-management-of-novel-cov.pdf
6. Du Z, Wang L, Cauchemez S, et al. Risk for Transportation of 2019 Novel Coronavirus (COVID-19) from Wuhan to Cities in China. medRxiv 2020;2020.01.28.20019299.
7. Lai S, Bogoch I, Ruktanonchai N, et al. Assessing spread risk of Wuhan novel coronavirus within and beyond China, January-April 2020: a travel network-based modelling study. medRxiv 2020;2020.02.04.20020479.
8. Coronavirus – Airlines Suspend Flights. Afr Res Bull Econ Financ Tech Ser 2020;57(1):22876A-22876A.
9. WHO, Regional Office for Africa. More than 600 confirmed cases of COVID-19 in Africa [Internet]. WHO Reg. Off. Afr. 2020 [cited 2020 Mar 19];Available from: https://www.afro.who.int/news/more-600-confirmed-cases-covid-19-africa
10. WHO, Regional Office for Africa. COVID-19 Situation update for the WHO African Region: External Situation Report 2 [Internet]. WHO, Regional Office for Africa; 2020. Available from: https://apps.who.int/iris/bitstream/handle/10665/331425/SITREP_COVID-19_WHOAFRO_20200311-eng.pdf
11. WHO, Regional Office for Africa. More than 15 countries in Africa report COVID-19 cases [Internet]. WHO Reg. Off. Afr. 2020 [cited 2020 Mar 16];Available from: https://www.afro.who.int/news/more-15-countries-africa-report-covid-19-cases
12. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet 2020;395(10225):689–97.
13. Adiga A, Venkatramanan S, Schlitt J, et al. Evaluating the impact of international airline suspensions on the early global spread of COVID-19. medRxiv 2020;2020.02.20.20025882.
14. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MUG, Khan K. Potential for global spread of a novel coronavirus from China. J Travel Med [Internet] [cited 2020 Mar 2];Available from: https://academic.oup.com/jtm/advance-article/doi/10.1093/jtm/taaa011/5716260
15. Gilbert M, Pullano G, Pinotti F, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. The Lancet [Internet] 2020 [cited 2020 Feb 27];0(0). Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30411-6/abstract

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

16. Salazar PMD, Niehus R, Taylor A, Buckee CO, Lipsitch M. Using predicted imports of 2019-nCoV cases to determine locations that may not be identifying all imported cases. medRxiv 2020;2020.02.04.20020495.
17. John Hopkins Bloomberg School of Public Health. Global health Security Index [Internet]. [cited 2020 Mar 8];Available from: https://www.ghsindex.org
18. Pung R, Chiew CJ, Young BE, et al. Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. The Lancet [Internet] 2020 [cited 2020 Mar 23];0(0). Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30528-6/abstract
19. BlueDot. BlueDot [Internet]. [cited 2020 Mar 10];Available from: https://bluedot.global
20. Cirium homepage [Internet]. [cited 2020 Mar 10];Available from: https://www.cirium.com
21. 全国迁徙详情——百度地图迁徙大数据 [Internet]. Baidu. [cited 2020 Mar 22];Available from: https://qianxi.baidu.com/
22. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. The Novel Coronavirus, 2019-nCoV, is Highly Contagious and More Infectious Than Initially Estimated. medRxiv 2020;2020.02.07.20021154.
23. {R Development Core Team, R}. R: A Language and Environment for Statistical Computing. Available from: http://dx.doi.org/10.1007/978-3-540-74686-7
24. World Health Organization, Regional Office for the Eastern Mediterranean. Update on COVID-19 in the Eastern Mediterranean Region [Internet]. World Health Organ. 2020 [cited 2020 Mar 20];Available from: http://www.emro.who.int/media/news/update-on-covid-19-in-the-eastern-mediterranean-region.html
25. COVID-19 Information: Morocco has three confirmed cases of COVID-19 within its borders [Internet]. CNBC Afr. 2020 [cited 2020 Mar 20];Available from: https://www.cnbcafrica.com/apo/2020/03/11/covid-19-information-morocco-has-three-confirmed-cases-of-covid-19-within-its-borders/
26. FIRST CASE OF COVID-19 CORONAVIRUS REPORTED IN SA [Internet]. Natl. Inst. Commun. Dis. 2020 [cited 2020 Mar 20];Available from: http://www.nicd.ac.za/first-case-of-covid-19-coronavirus-reported-in-sa/
27. Ethiopia confirms 9th #COVID19 patient, advises travelers from affected countries to 14-day self-quarantine [Internet]. Addis Stand. 2020 [cited 2020 Mar 20];Available from: https://addisstandard.com/news-ethiopia-confirms-9th-cocid19-patient-advises-travelers-from-affected-countries-to-14-day-self-quarantine/
28. Ombok E. Kenya Reports First Case of Coronavirus [Internet]. Bloomberg. 2020 [cited 2020 Mar 20];Available from: https://www.bloomberg.com/news/articles/2020-03-13/kenyan-health-secretary-reports-first-case-of-coronavirus
29. Burkhardt P. First Coronavirus Cases Reported in Countries Across Africa [Internet]. Bloomberg. 2020 [cited 2020 Mar 20];Available from: https://www.bloomberg.com/news/articles/2020-03-14/first-coronavirus-cases-reported-in-countries-across-africa

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Information

Methods

Adjustment of historical flight data

We generate the imputed daily air-travel volume for the epidemic period, taking into account yearly increases in flight numbers and the impact of travel bans, using two data sources: monthly air-travel volume data from BlueDot[18] from 1 December 2018 to 28 February 2019; and Cirium[20] data on daily landed flight departures from 1 December 2018 to 28 February 2019 and from 1 December 2019 to 28 February 2020.

First, we use Cirium[20] data on landed flight departures to estimate the expected change in air-travel volume from December 2019 to February 2020 relative to December 2018 to February 2019, γ_i, for each origin airport *i* as a result of yearly increase in air travel as well as travel bans implemented in late January 2020 as the following:

$$\gamma_i = \sum_{d=Dec2019}^{Feb2020} \sum_{j=1}^{N_D} f_{d,i} \Big/ \sum_{d=Dec2018}^{Feb2019} \sum_{j=1}^{N_D} f_{d,i} ,$$

where the number of flight departures on day *d*, $f_{d,i}$, is summed from 1 December 2019 to 28 February 2020 and summed over all destination airports *j.* This value is divided by the number of landed flight departures from the same period last year, 1 December 2018 to 28 February 2019, summed over all destination airports *j*. We sum over all destination airports *j* instead of calculating this for each origin-destination connection due to sparse data limitations.

The total number of passengers traveling from origin *i* to destination *j* from 1 December 2018 to 28 February 2019, $V_{18-19,i,j}$, is then multiplied by the adjustment factor, γ_i, to yield an estimate of the total three-month flight volume for each origin-destination connection, with both increased annual travel and travel bans in late January taken into account, $V'_{19-20,i,j}$:

$$V'_{19-20,i,j} = \gamma_i * V_{18-19,i,j}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Note that, by using data on the number of landed direct flights to adjust air-travel volume, we make the assumption that the number of direct flight departures is proportional to the number of passengers from each origin-destination connection.

To estimate the daily distribution of passengers from December 2019 and February 2020, we distribute the adjusted total air-travel volume, $V'_{19-20,i,j}$, across the three-month period into daily air-travel volume for each day *d*, using the proportion of daily landed flight departures out of the total number of daily landed flights in the three month period.

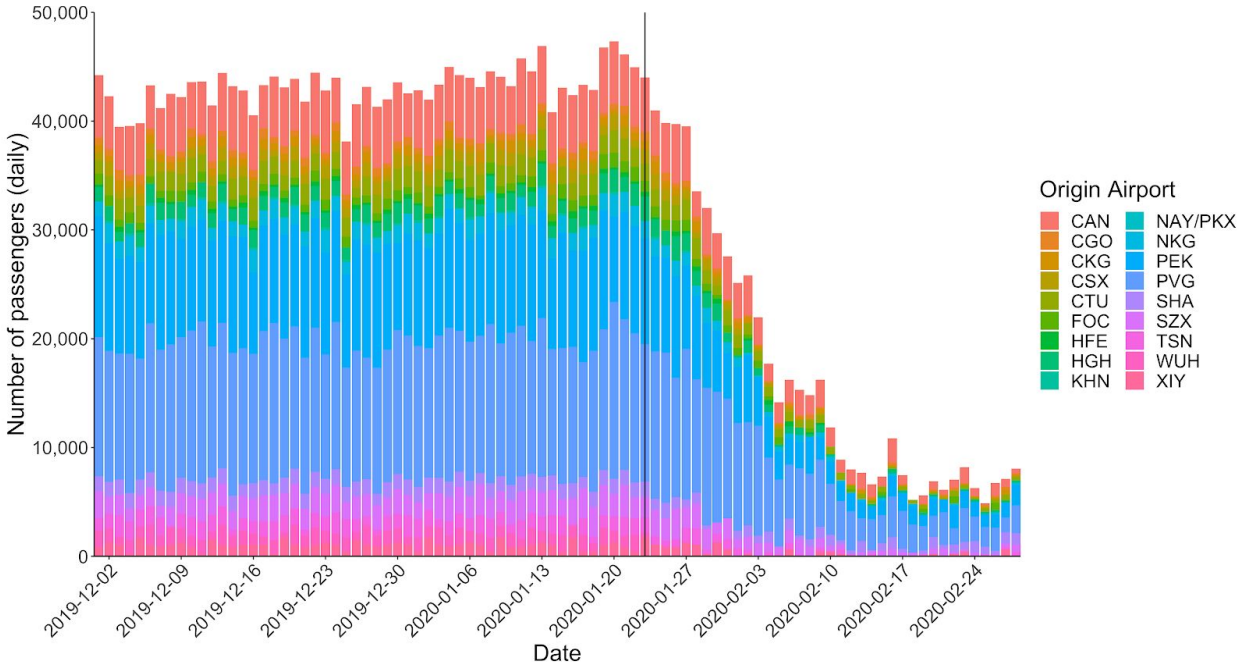
$$v'_{d,i,j} = V'_{19-20,i,j}\left(\sum_{j=1}^{N_D} f_{d,i} \Big/ \sum_{d=Dec2019}^{Feb2020} \sum_{j=1}^{N_D} f_{d,i}\right) ,$$

where $v'_{d,i,j}$ is the imputed air-travel volume from origin *i* to destination *j* on day *d*, and $f_{d,i}$ is the number of landed flight departures on day *d* from origin *i.* We calculate the proportion of daily flight departures for each origin *i* summed over all destinations *j* instead of calculating this for each origin-destination connection again due to sparse data limitations.

Since residents of the cities Shenzhen and Dongguan both use the Shenzhen Bao'an International Airport, we divide the air-travel volume equally between these two cities. Similarly, the air-travel volume for Hangzhou Xiaoshan International Airport airport is divided equally for the cities Jiaxing and Hangzhou. Importantly, Cirium data only documents the daily number of direct flights, but we use this data source 1) to estimate the expected flight volume reduction on all direct and indirect flights, and 2) to distribute a total three-month volume into daily volume. We therefore make the assumption that the number of direct flights is proportional to number of direct and indirect flights.

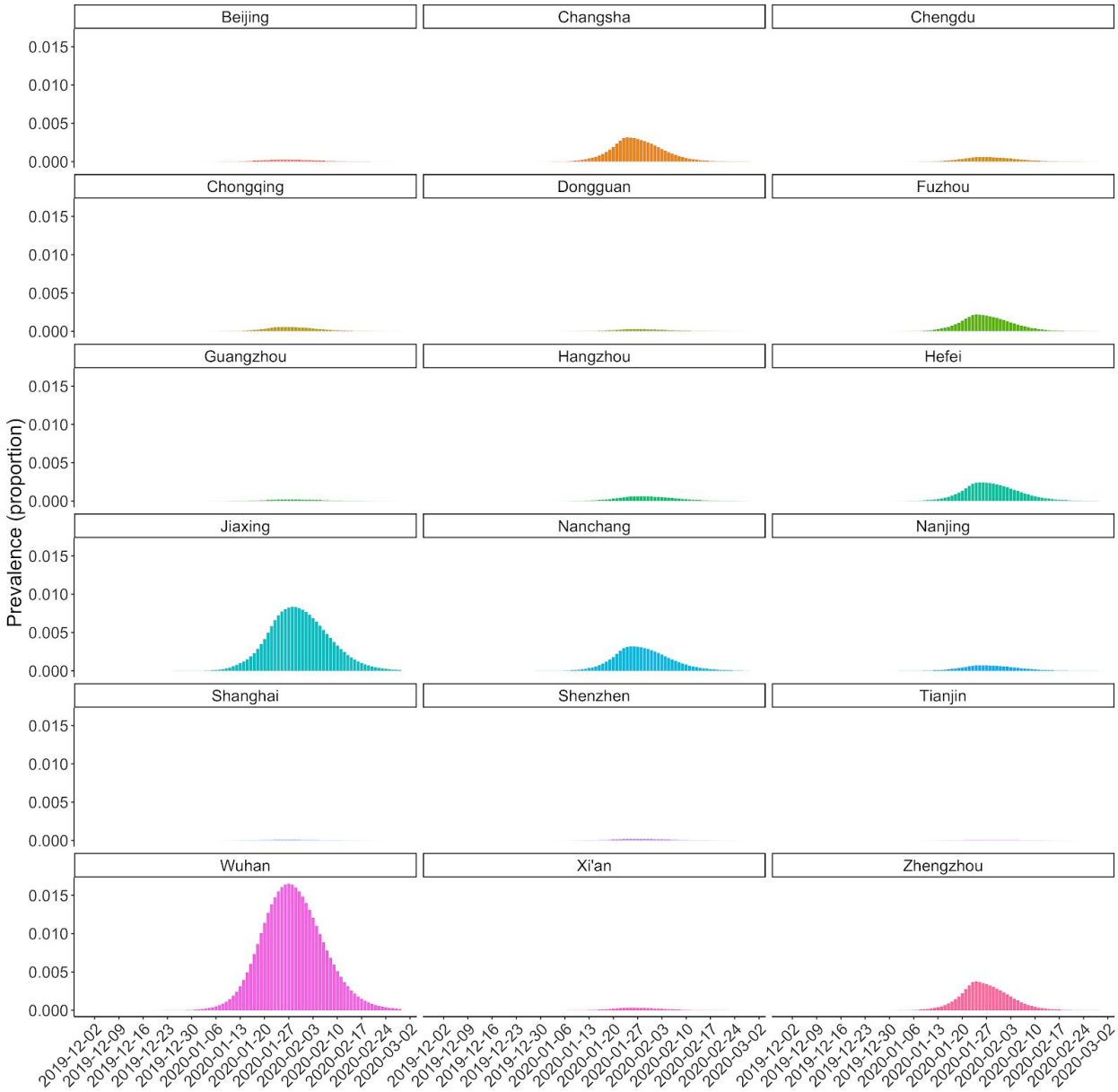
medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Results



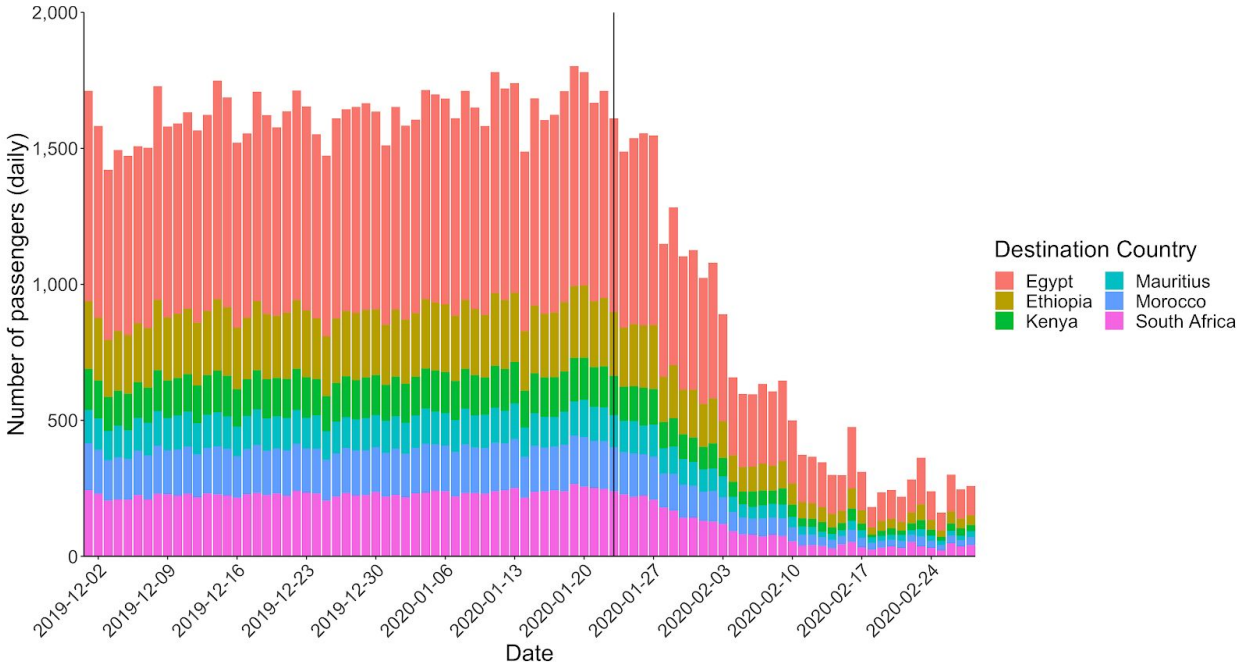
Supplementary Figure 1. Estimated daily flight volume from 18 airports in 18 Chinese cities to 16 international destinations from 1 December 2019 to 28 February 2020

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



Supplementary Figure 2. Estimated daily prevalence of COVID-19 in 18 Chinese cities from 1 December 2019 to 28 February 2020

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



Supplementary Figure 3. Estimated daily flight volume from 18 airports in 18 Chinese cities to 6 African locations from 1 December 2019 to 28 February 2020

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Destination Country	Lower Bound	Mean predicted number	Upper bound
Egypt	0	2.5	6
Ethiopia	0	0.8	3
Kenya	0	1.3	4
Mauritania	0	0.4	2
Morocco	0	1.0	3
South Africa	0	3.0	7

Supplementary Table 1: Predicted exported cases by destination country in Africa, with 95% CI (Lower Bound, Upper Bound)

Contributors

TFM, TC, ML, RN and JH designed the study. TFM, TC and JH performed the analyses. TFM, TC and RN wrote the manuscript. All authors reviewed and approved the final manuscript.

Role of the funding source

The funding bodies of this study had no role in the study design, data analysis and interpretation, or writing of the manuscript. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

Marc Lipsitch has received consulting fees from Merck. All authors declare no competing interests.

Data availability

Transformed flight data, derived prevalence data, and all code used in these analyses are fully available online at https://github.com/c2-d2/COVID_allchina_export

Acknowledgments

We thank Rebecca Kahn and Christine Tedijanto for their valuable input and feedback. ML,

medRxiv preprint doi: https://doi.org/10.1101/2020.03.23.20038331; this version posted March 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

TFM, TC and RN were supported by Award Number U54GM088558 from the US National Institute Of General Medical Sciences. PMD was supported by the Fellowship Foundation Ramon Areces. COB was supported by a NIGMS Maximizing Investigator's Research Award (MIRA) R35GM124715-02.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute Of General Medical Sciences or the National Institutes of Health.