Estimating the number of undetected COVID-19 cases among travellers from mainland China [version 1; peer review: awaiting peer review]


1MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, Imperial College London, London, UK
2Department of Statistics, University of Oxford, Oxford, UK

Abstract

Background: Since the start of the COVID-19 epidemic in late 2019, there have been more than 152 affected regions and countries with over 110,000 confirmed cases outside mainland China.

Methods: We analysed COVID-19 cases among travellers from mainland China to different regions and countries, comparing the region- and country-specific rates of detected and confirmed cases per flight volume to estimate the relative sensitivity of surveillance in different regions and countries.

Results: Although travel restrictions from Wuhan City and other cities across China may have reduced the absolute number of travellers to and from China, we estimated that more than two thirds (70%, 95% CI: 54% - 80%, compared to Singapore; 75%, 95% CI: 66% - 82%, compared to multiple countries) of cases exported from mainland China have remained undetected.

Conclusions: These undetected cases potentially resulted in multiple chains of human-to-human transmission outside mainland China.

Keywords
Epidemiology, COVID-19, novel coronavirus, SARS-CoV-2, international
Corresponding author: Christl A. Donnelly (c.donnelly@imperial.ac.uk)

Author roles: Bhatia S: Formal Analysis, Investigation, Writing – Review & Editing; Imai N: Data Curation, Investigation, Resources, Writing – Review & Editing; Cuomo-Dannenburg G: Data Curation, Investigation, Resources, Writing – Review & Editing; Baguelin M: Investigation, Writing – Review & Editing; Boonyasiri A: Investigation, Writing – Review & Editing; Cori A: Investigation, Writing – Review & Editing; Cucunubá Z: Investigation, Writing – Review & Editing; Dorigati I: Investigation, Writing – Review & Editing; FitzJohn R: Investigation, Writing – Review & Editing; Fu H: Investigation, Writing – Review & Editing; Gaythorpe K: Investigation, Writing – Review & Editing; Ghani A: Investigation, Writing – Review & Editing; Hamlet A: Investigation, Writing – Review & Editing; Hinsley W: Investigation, Writing – Review & Editing; Laydon D: Investigation, Writing – Review & Editing; Nedjati-Gilani G: Investigation, Writing – Review & Editing; Okell L: Investigation, Writing – Review & Editing; Riley S: Investigation, Writing – Review & Editing; Thompson H: Investigation, Writing – Review & Editing; van Elsland S: Investigation, Writing – Review & Editing; Volz E: Investigation, Writing – Review & Editing; Whittaker C: Investigation, Writing – Review & Editing; Xi X: Investigation, Writing – Review & Editing; Donnelly CA: Conceptualization, Formal Analysis, Investigation, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; Ferguson NM: Conceptualization, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust [200861]. This work was also supported by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement, also part of the EDCTP2 programme supported by the European Union [UK, Centre MR/R015600/1]; and the National Institute for Health Research (UK, for Health Protection Research Unit funding) [HPRU-2012–10080]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2020 Bhatia S et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Bhatia S, Imai N, Cuomo-Dannenburg G et al. Estimating the number of undetected COVID-19 cases among travellers from mainland China [version 1; peer review: awaiting peer review] Wellcome Open Research 2020, 5:143

https://doi.org/10.12688/wellcomeopenres.15805.1

First published: 15 Jun 2020, 5:143 https://doi.org/10.12688/wellcomeopenres.15805.1
Background

As of 18th March 2020, over 80,000 cases of COVID-19 (formerly 2019-nCoV) have been reported in China (with 3231 deaths), and more than 110,000 cases have been detected in 152 regions and countries outside mainland China (including Hong Kong SAR and Macau SAR). Several analyses have been undertaken to predict or estimate the risk of exported cases by country on the basis of flight connections between Wuhan City, China or mainland China as a whole and other regions and countries. In this analysis we built on published work to analyse COVID-19 cases reported and confirmed in different countries that were exported from mainland China, comparing the region- and country-specific rates of detected cases per flight volume to estimate the relative sensitivity of surveillance in different countries. We then estimate the number of COVID-19 cases exported from mainland China that have remained undetected worldwide.

Methods

Data sources

Air traffic volume. Air travel data for the months of January, February, and March 2016 were obtained from the International Air Travel Association (IATA), with the sum divided by three to get destination-region- and destination-country-specific monthly averages. These numbers were not scaled up to reflect recent growth in air travel because any constant scaling of the monthly averages would simply be absorbed into the λ estimate and not affect other results. Flows of passengers within mainland China were excluded from this analysis.

Number of cases detected outside mainland China. We collated data on 3276 cases in international travelers from media reports and provincial and national department of health press releases up until 27 February 2020. We defined a local transmission as any transmission that occurred outside mainland China (Hong Kong SAR and Macau SAR are considered outside mainland China for this analysis). We only consider cases that were not transmitted locally. That is, we only considered cases detected outside mainland China that had a travel history to China and arrived outside mainland China by air, excluding repatriation flights. Based upon these inclusion criteria, a total of 173 cases were included in our analysis. The earliest date of travel for the cases included in the analysis is 1 January 2020, and the latest date of travel is 25 February 2020.

Analysis

We assume that the number of exported cases in a country \( i \) is Poisson distributed with a mean that depends on the air traffic from Wuhan to \( i \), and the sensitivity of surveillance in \( i \) relative to a country \( j \), denoted by \( s_{ij} \). For each country \( i \), let \( X_i \) be the number of exported cases (a count) and let \( F_i \) be the volume of air traffic from Wuhan to country \( i \). We can then write a joint log likelihood for the data from countries \( i \) and \( j \):

\[
I = X_i \ln \left( s_{ij} \lambda_{ij} F_i \right) - s_{ij} \lambda_{ij} F_i + X_j \ln \left( \lambda_{ij} F_j \right) - \lambda_{ij} F_j
\]

ignoring additive constants. Thus, the maximum likelihood estimates for \( \lambda_{ij} \) and \( s_{ij} \) are:

\[
\hat{\lambda}_{ij} = \frac{X_i}{F_i} \quad \text{and} \quad \hat{s}_{ij} = \frac{X_i}{\lambda_{ij} F_i}
\]

The likelihood-based confidence intervals are obtained by calculating the maximum log likelihood (over values of \( \lambda \)) for each value of \( s_{ij} \). Then the 95% confidence interval includes all those values of \( s_{ij} \) such that \( 2 \left( \frac{I - I_{0.025}}{\text{SS}} \right) \leq 3.84 \) (the 95th centile of the chi-squared distribution with 1 degree of freedom). These calculations were all performed using R version 3.6.0.

The relative sensitivities can also be estimated relative to \( J \) countries simultaneously using a method similar to above but with the log likelihood:

\[
I = X_i \ln \left( s_{ij} \lambda_{ij} F_i \right) - s_{ij} \lambda_{ij} F_i + \sum_{j=1}^{J} X_j \ln \left( \lambda_j F_j \right) - \lambda_j F_j
\]

Expected values can then be calculated for every country \( i \) as simply \( \hat{j} \lambda_{ij} F_i \), and the expected value for all countries \( \lambda \sum_{j=1}^{J} F_j \), where \( N \) is the total number of countries with air traffic from Wuhan Tianhe International Airport (\( N = 119 \)).

Results

The number of exported cases by country was plotted as a function of the average monthly passenger volume originating from Wuhan Tianhe International Airport on international flights (Figure 1). This showed Singapore to be an outlier in terms of having relatively many exported cases compared to the measure of air traffic volume.

The relative sensitivity of surveillance in individual countries was estimated compared to Singapore, Finland, Nepal, Philippines, Sweden, India, Sri Lanka, and Canada were all found to have relative sensitivity estimates greater than 1 (i.e. more cases were detected per passenger flight than in Singapore). Thus, a second set of relative sensitivity estimates was obtained for all other individual countries compared simultaneously to Singapore, Finland, Nepal, Philippines, Sweden, India, Sri Lanka, and Canada.

The region- and country-specific expected numbers of exported COVID-19 cases were in several cases substantially higher than the numbers detected (Figure 2). The sum of the expected numbers of exported COVID-19 cases for all regions and countries other than mainland China was 576.8 (95% CI: 372.2 - 845.4), based on the analysis relative to Singapore only, and 704.4 (95% CI: 510.3 - 942.3), based on the analysis relative to Singapore, Finland, Nepal, Philippines, Sweden, India, Sri Lanka, and Canada. Given that 173 such cases were detected, these central estimates suggest that between 70% (95% CI: 54% - 80%, relative to Singapore only) and 75%
Table 1. Number of cases detected outside mainland China with travel history to China.

<table>
<thead>
<tr>
<th>Country</th>
<th>Travel History to Hubei</th>
<th>No Travel History to Hubei</th>
<th>Unknown Travel History within China</th>
<th>Total (cases with a travel history to China)</th>
<th>Travelled by air (not repatriation flight)</th>
<th>Travelled on repatriation flight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>Belgium</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Cambodia</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Canada</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Finland</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>France</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Germany</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Hong Kong SAR</td>
<td>12</td>
<td>3</td>
<td>0</td>
<td>15</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>India</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Italy</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Japan</td>
<td>24</td>
<td>4</td>
<td>0</td>
<td>28</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>Macau SAR</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Malaysia</td>
<td>9</td>
<td>4</td>
<td>0</td>
<td>13</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Nepal</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Philippines</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Singapore</td>
<td>21</td>
<td>3</td>
<td>0</td>
<td>24</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>South Korea</td>
<td>12</td>
<td>2</td>
<td>0</td>
<td>14</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Sweden</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Taiwan</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Thailand</td>
<td>14</td>
<td>3</td>
<td>5</td>
<td>22</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>United Arab Emirates</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>United States of America</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>13</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>Vietnam</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>162</td>
<td>26</td>
<td>16</td>
<td>204</td>
<td>173</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 1. Exported COVID-19 cases vs average air traffic from Wuhan Tianhe International Airport by destination. The number of exported COVID-19 cases detected by region and country plotted against the average monthly international air traffic volume from Wuhan Tianhe International Airport aggregated by destination country. The colour of the points denotes the continent of the destination country (Asia - orange, Europe - light blue, Africa - green, North America - dark blue, South America - pink, and Oceania - dark orange).
Figure 2. The expected and observed numbers of exported COVID-19 cases by country, relative to Singapore only. Values above the diagonal line indicate more cases were expected than were observed. The colour of the points denotes the continent of the destination country (Asia - orange, Europe - light blue, Africa - green, North America - dark blue, South America - pink, and Oceania - dark orange).

(95% CI: 66% - 82%, relative to Singapore, Finland, Nepal, Philippines, Sweden, India, Sri Lanka, and Canada) remained undetected.

Discussion
Consistent with similar analyses, we estimated that more than two thirds of COVID-19 cases exported from mainland China have remained undetected worldwide, potentially leaving sources of human-to-human transmission unchecked (70%, 95% CI: 54% - 80% and 75%, 95% CI: 66% - 82%, undetected, based on comparisons to Singapore only and to Singapore, Finland, Nepal, Philippines, Sweden, India, Sri Lanka, and Canada, respectively). Undoubtedly, the exported cases vary in the severity of their clinical symptoms, making some cases more difficult to detect than others. However, some countries have detected significantly fewer than would have been expected based on the volume of flight passengers arriving from Wuhan City, China. These undetected cases potentially resulted in multiple chains of human-to-human transmission outside mainland China.

Data availability
Source data
The air travel data used in this analysis can be purchased from International Air Transport Association (IATA) via the following link: https://www.iata.org/en/services/statistics/air-transport-stats/.

Underlying data

This project contains the following underlying data:
- exported_cases.csv (information on the date of report, country of report and travel history of 3,276 cases outside mainland China)

Extended data

This project contains the following extended data:
- data_processing.R (R code to post-process international case data)

Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

References