RESEARCH ARTICLE

REVISED Estimating the number of undetected COVID-19 cases among travellers from mainland China [version 2; peer review: 3 approved]

Sangeeta Bhatia[1], Natsuko Imai[1], Gina Cuomo-Dannenburg[1], Marc Baguelin[1], Adhiratha Boonyasiri[1], Anne Cori[1], Zulma Cucunubá[1], Ilaria Dorigatti[1], Rich FitzJohn[1], Han Fu[1], Katy Gaythorpe[1], Azra Ghani[1], Arran Hamlet[1], Wes Hinsley[1], Daniel Laydon[1], Gemma Nedjati-Gilani[1], Lucy Okell[1], Steven Riley[1], Hayley Thompson[1], Sabine van Elsland[1], Erik Volz[1], Haowei Wang[1], Yuanrong Wang[1], Charles Whittaker[1], Xiaoyue Xi[1], Christl A. Donnelly[1,2], Neil M. Ferguson[1]

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

v2 **First published:** 15 Jun 2020, **5**:143
https://doi.org/10.12688/wellcomeopenres.15805.1
Second version: 13 Sep 2021, **5**:143
https://doi.org/10.12688/wellcomeopenres.15805.2
Latest published: 06 Dec 2021, **5**:143
https://doi.org/10.12688/wellcomeopenres.15805.3

Abstract

Background: As of August 2021, every region of the world has been affected by the COVID-19 pandemic, with more than 196,000,000 cases worldwide.
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 up to 70% (95% CI: 54% - 80%) of imported cases could remain undetected relative to the sensitivity of surveillance in Singapore. The percentage of undetected imported cases rises to 75% (95% CI 66% - 82%) when comparing to the surveillance sensitivity in multiple countries.
Conclusions: Our analysis shows that a large number of COVID-19 cases remain undetected across the world. These undetected cases potentially resulted in multiple chains of human-to-human transmission outside mainland China.

Open Peer Review

Reviewer Status ✓ ✓ ✓

	Invited Reviewers		
	1	**2**	**3**
version 3 (revision) 06 Dec 2021			
version 2 (revision) 13 Sep 2021	✓ report	✓ report	✓ report
version 1 15 Jun 2020	? report	? report	? report

1. **Jeremy Bingham**, Stellenbosch University, Stellenbosch, South Africa
Juliet R.C. Pulliam, Stellenbosch University, Stellenbosch, South Africa

Keywords
Epidemiology, COVID-19, novel coronavirus, SARS-CoV-2, international

This article is included in the Coronavirus (COVID-19) collection.

2. **Sebastian Funk**, London School of Hygiene & Tropical Medicine, London, UK

3. **Hannah E. Clapham**, National University of Singapore and National University Health System, Singapore, Singapore

Any reports and responses or comments on the article can be found at the end of the article.

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; **Dorigatti 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; **Wang H**: Investigation, Writing – Review & Editing; **Wang Y**: 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.

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 2; peer review: 3 approved]** Wellcome Open Research 2021, **5**:143 https://doi.org/10.12688/wellcomeopenres.15805.2

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

REVISED **Amendments from Version 1**

The revised version addresses the concerns raised by the reviewers. We have updated the language to clarify the findings and use improved notation to explain the methodology. We have also updated the code to include an example file needed to run the analysis. The code has also been commented in depth so that readers can rerun the analysis.

Any further responses from the reviewers can be found at the end of the article

Background

As of August 2021, over 196,000,000 cases of COVID-19 have been reported across the world with over 4,000,000 deaths[1]. 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[2–8]. Salazar *et al.*[4], for instance, fit the number of reported cases in high surveillance countries and report that countries in Southeast Asia such as Indonesia and Thailand had reported fewer imported cases than expected despite a high volume of air travel with China. In this analysis we built on published work[4] 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- (Hong Kong SAR and Macau SAR) and destination-country-specific monthly averages. The data from 2016 were the most recent data to which we had access. 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 estimates of model parameters (see Analysis) 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[1,9]. Media reports on new cases of COVID-19 were followed daily from 15th January 2020 to 27th February 2020. Where possible, the details reported in the news were validated against official sources. Relevant websites such as ministries of health or local news media were identified through web searches. Reports in languages other than English were translated into English using translation services available online (e.g. Google translate). 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 (Table 1). Everyone we classified as a "case detected overseas" had the mode of travel either explicitly mentioned as air, or implied as the most probable mode of travel from mainland China to the destination (e.g. from China to Italy). Where multiple modes of travel are possible e.g. from mainland China to Hong Kong, we have only classified individuals as cases detected overseas where the mode of travel was explicitly mentioned as air. In most instances, all or most of the passengers on repatriation flights had been tested for the presence of SARS-CoV2. The cases detected through surveillance of repatriation flights are therefore not representative of the general sensitivity of surveillance in a country. We have therefore excluded these from the analysis.

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 observed 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:

$$l = X_i \ln\left(s_{ij}\lambda_j F_i\right) - s_{ij}\lambda_j F_i + X_j \ln\left(\lambda_j F_j\right) - \lambda_j F_j$$

ignoring additive constants. Thus, the maximum likelihood estimates for λ_j and s_{ij} are:

$$\hat{\lambda}_j = \frac{X_j}{F_j} \text{ and } s_{ij} = \frac{X_i F_j}{X_j F_i}$$

The likelihood-based confidence intervals are obtained by calculating the maximum log likelihood (over values of λ_j) for each value of s_{ij}. Then the 95% confidence interval includes all those values of s_{ij} such that $2\,(l_{\hat{s}ij} - l_{sij}) \leq 3.84$ (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:

[1] This has been updated since the analysis presented here was released as a public report by the Imperial College London Coronavirus Response Team on available 22nd February 2020. This report is available at https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-6-international-surveillance/. See https://doi.org/10.5281/zenodo.3736643.

Table 1. Number of cases detected outside mainland China with travel history to China.

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

$$l = X_i \ln(s_{iJ}\lambda_J F_i) - s_{iJ}\lambda_J F_i + \sum_{j=1}^{J}(X_j \ln(\lambda_j F_j) - \lambda_i F_j)$$

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

Results

The observed 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[9]). This showed Singapore to be an outlier in terms of having relatively many observed exported cases compared to the measure of air traffic volume.

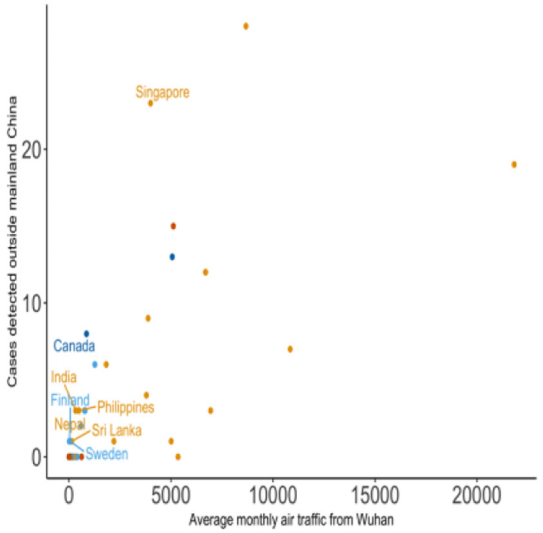

Figure 1. Exported COVID-19 cases vs average air traffic from Wuhan Tianhe International Airport by destination. The number of exported COVID-19 cases detected by region and country plotted against the average monthly international air traffic volume from Wuhan Tianhe International Airport aggregated 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).

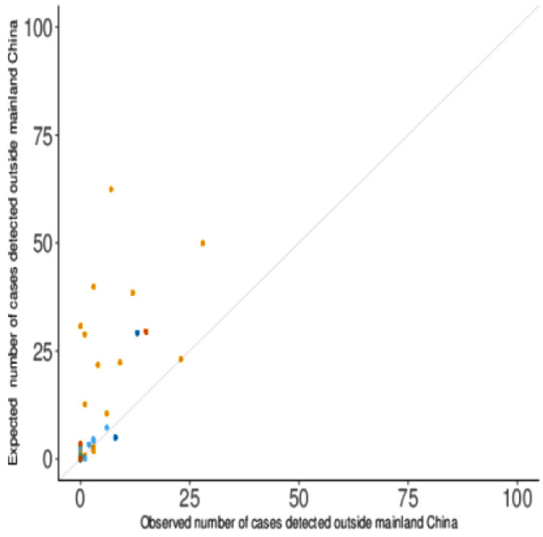

Figure 2. The expected and observed numbers of exported COVID-19 cases by country, with surveillance sensitivity relative to Singapore only. Values above the diagonal line indicate more cases were expected than were observed. The colour of the points denotes the continent of the destination country (Asia - orange, Europe - light blue, Africa - green, North America - dark blue, South America - pink, and Oceania - dark orange).

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[9]). 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% (95% CI: 66% - 82%, relative to Singapore, Finland, Nepal, Philippines, Sweden, India, Sri Lanka, and Canada) remained undetected.

Discussion

Consistent with similar analyses[4,10], we estimated that more than two thirds of COVID-19 cases exported from Wuhan 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).

A limitation of our study is that we do not take into account the changes in air travel due to the travel advisories and restrictions imposed by various governments (though only those in force before 27 February 2020 would be relevant), which may have changed the volume of passengers flying into particular countries. Further, in using the data from 2016, we assume that the passenger volumes in early 2020 into each country is scaled by a constant factor. Access to more recent data on the changes in the number of passengers would likely improve the estimates of the sensitivity of surveillance presented here. For countries/regions that are connected to Wuhan using multiple modes of transport such as train links and water routes e.g., Hong Kong, surveillance is likely to have been enhanced at ports of entry other than airports. If so, the estimate of the sensitivity of surveillance as estimated here would therefore likely present an underestimate for these regions.

During the period of this study, Wuhan was the epicenter of the outbreak. Hence, it was reasonable to assume that a case detected outside China with travel history to Hubei province in this period is likely to be an imported case. However, epidemiological investigations are critical to ascertain the origin of a case. Timely public release of the results of such investigations could help public health professionals better assess the spread of the disease.

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

Zenodo: mrc-ide/COVID19_surveillance_sensitivity: Data and code used for submission. http://doi.org/10.5281/zenodo.3736643[9].

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

Zenodo: mrc-ide/COVID19_surveillance_sensitivity: Data and code used for submission. http://doi.org/10.5281/zenodo.3736643[9].

This project contains the following extended data:

- data_processing.R (R code to post-process international case data)

References

1. **Coronavirus disease 2019 (COVID-19) Situation Report -208**. Reference Source
2. Bogoch II, Watts A, Thomas-Bachli A, *et al.*: **Potential for global spread of a novel coronavirus from China.** *J Travel Med.* 2020. **27**(2): taaa011. PubMed Abstract | Publisher Full Text | Free Full Text
3. 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.** *Lancet.* 2020; **395**(10225): 689–697. PubMed Abstract | Publisher Full Text | Free Full Text
4. Pablo M, De Salazar PM, Niehus R, *et al.*: **Using predicted imports of 2019-nCoV cases to determine locations that may not be identifying all imported cases.** *medRxiv.* 2020. Publisher Full Text
5. Boldog P, Tekeli T, Vizi Z, *et al.*: **Risk Assessment of Novel Coronavirus COVID-19 Outbreaks Outside China.** *medRxiv.* 2020. Publisher Full Text
6. Gilbert M, Pullano G, Pinotti F, *et al.*: **Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study.** *Lancet.* 2020; **395**(10227): 871–877. PubMed Abstract | Publisher Full Text | Free Full Text
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. Publisher Full Text
8. Menkir TF, Chin T, Hay JA, *et al.*: **Estimating internationally imported cases during the early COVID-19 pandemic.** *Nat Commun.* 2021; **12**(1): 311. PubMed Abstract | Publisher Full Text | Free Full Text
9. Cuomo-Dannenburg G: **mrc-ide/COVID19_surveillance_sensitivity: Data and code used for submission (Version v1.0).** *Zenodo.* 2020. http://www.doi.org/10.5281/zenodo.3736643
10. Niehus R, De Salazar PM, Taylor AR, *et al.*: **Estimating underdetection of internationally imported COVID-19 cases.** *medRxiv.* 2020. Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔ ✔

Version 2

Reviewer Report 08 November 2021

https://doi.org/10.21956/wellcomeopenres.18948.r45821

✔ **Juliet R.C. Pulliam**
South African DSI-NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, Stellenbosch, South Africa
Jeremy Bingham
South African DSI-NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, Stellenbosch, South Africa

The authors have addressed most of our concerns, except that the article still links to the old version of the code. We note that there does appear to be updated code available via the GitHub repository* that is linked from the Zenodo page, but some of the changes referred to in the response (eg updates to the README) do not seem to be available. We recommend that the authors re-check that all updates have been pushed to the repository and request that a direct link to the updated code be provided.

* https://github.com/mrc-ide/COVID19_surveillance_sensitivity

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: infectious disease epidemiology and modelling

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 19 Nov 2021

Christl Donnelly, Imperial College London, London, UK

We thank the reviewer for highlighting that the Zenodo link pointed to an older version of the code. We have now added a link to the latest version of the repository. We have also added additional instructions to the README file to help the readers replicate the analysis.

Competing Interests: No competing interests were disclosed.

Reviewer Report 06 October 2021

https://doi.org/10.21956/wellcomeopenres.18948.r45820

Sebastian Funk
Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, UK

Thanks for the revisions, which I think have improved the article.

As a last point, can I just ask to clarify that you can't add a column with the source (website?) for the data table. Was there no records kept of this? I couldn't find the answer to this question in the response.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: infectious disease dynamics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 14 Oct 2021

Christl Donnelly, Imperial College London, London, UK

Thanks for your review of our manuscript.

We have now uploaded to GitHub a list of the many sources consulted to identify COVID-19 cases among travellers.

list_of_sources_consulted.csv
COVID19_surveillance_sensitivity/list_of_sources_consulted.csv at bf210752a493a61bc230cfd21d10337c6de70fb4 · mrc-ide/COVID19_surveillance_sensitivity · GitHub

Competing Interests: No competing interests were disclosed.

Reviewer Report 27 September 2021

https://doi.org/10.21956/wellcomeopenres.18948.r45822

Hannah E. Clapham
Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore

No further comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Infectious disease epidemiology and dynamics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 13 August 2020

https://doi.org/10.21956/wellcomeopenres.17332.r39749

Hannah E. Clapham
Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore

This is a well-done analysis that estimates the number of unreported cases that were imported from China early on in the pandemic. It does this using data on air travel volume between China and other countries, and the reported numbers of cases in the places that reported the highest numbers of cases given their air traffic volume. This analysis was highly relevant early in the pandemic as infections spread from China.

I have a few comments on points for clarification below.

Abstract:

The statement of the results about 70/75%... in the abstract is confusing. Suggest rephrasing.
Conclusion: I wonder if this is a conclusion from the paper. I would suggest the addition here that the analysis leads to estimates that there were many unreported imported infections, and that potentially lead to transmission.

Main text:
Background: It would be helpful to have a statement about what the previous analysis in reference 4 did/showed.

Methods: Please add more detail on how airports in China were used in the flow calculation, and also on the definition of a destination region. In the results section initially Wuhan is focused on but the general conclusions seem to be from all of mainland China. Please clarify throughout.
Please add more detail from where the collated data on imported cases was obtained.
Were all excluded cases excluded because they were defined as local or due to missing information on this?
Was data available on which location was travelled from within China for the imported cases? If not, how was this dealt with in the analysis?

Results:
Figure 2 legend, are the numbers shown relative to Singapore numbers, or is the analysis done relative to Singapore and then the estimates of imported cases shown? At the moment, the legend reads as the former, but my understanding of the analysis is that is it the latter. Please clarify.

Discussion: Please add on limitations of the analysis, in particular how this relates to the available data including classification as imported vs local, and that this data needed to be publicly available.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Infectious disease epidemiology and dynamics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 12 Aug 2021

Christl Donnelly, Imperial College London, London, UK

1. The statement in the abstract has been rephrased for clarity.
2. The conclusion in the abstract has been edited as suggested.
3. We have added a sentence about the analysis and conclusions of Reference 4.
4. Destination-regions refer to Hong Kong SAR and Macau SAR, this clarification has been added to the text.
5. We have now edited the Results and the Discussion sections to emphasise that we are estimating the number of exported cases from Wuhan (rather than China).
6. We have expanded the section on data collation to clarify the points raised by the reviewer. The inclusion criteria have been elaborated to clarify when a case was excluded.
7. As the reviewer has rightly noted, the numbers in figure 2 are the estimates of imported cases. The legend has been edited to reflect this.
8. We have expanded the discussion to highlight the limitations and potential biases of the analysis.

Competing Interests: No competing interests were disclosed.

Reviewer Report 31 July 2020

https://doi.org/10.21956/wellcomeopenres.17332.r39133

?

Sebastian Funk
Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, UK

This manuscript aims to quantify the amount of underdetection of cases outside of China during the early phase of the COVID-19 pandemic. It assumes that the numbers of cases are centred around the product of the number of flights from Wuhan, an estimated per-country parameter \lambda, and an estimated relative per-country surveillance capacity.

Overall the methodology is sensible and the findings of interest at the time.

Comments:

- The manuscript is missing detail on how the 3276 cases were collated. Was this done systematically and if so how? Were non-English reports collated and was there any attempt to correct for language biases? Have you got references for the cases and, if yes, could they be added, e.g., to the csv file?

- The index could be more consistent, e.g. it seems \lambda=X_j/F_j should have an index j, and s should have indices i and j (and no e).

- It would be interesting to see the results here compared to Golding *et al.*, https://www.medrxiv.org/content/10.1101/2020.07.07.20148460v1, which had the same aim but used a different methodology/data - also, to compare to other estimates of underdetection in the relevant countries.[1]

- A thorough discussion of potential biases/limitations is missing.

References

1. Golding N, Russell T, Abbott S, Hellewell J, et al.: Reconstructing the global dynamics of under-ascertained COVID-19 cases and infections. *medRxiv*. 2020. Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: infectious disease dynamics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have

significant reservations, as outlined above.

Author Response 12 Aug 2021

Christl Donnelly, Imperial College London, London, UK

1. The section on data collation has been expanded to clarify the points raised by the reviewer.
2. Thanks, we have edited the indices.
3. A direct comparison with the results of Golding et al is difficult because they present country-specific ascertainment estimates while the goal of our analysis was to estimate the number of undetected COVID-19 cases globally and we have therefore not provided per-country estimates of surveillance sensitivity.
4. We have expanded the discussion to highlight the limitations and potential biases of the analysis.

Competing Interests: No competing interests were disclosed.

Reviewer Report 08 July 2020

https://doi.org/10.21956/wellcomeopenres.17332.r39131

?

Jeremy Bingham
South African DSI-NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, Stellenbosch, South Africa

Juliet R.C. Pulliam
South African DSI-NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, Stellenbosch, South Africa

Summary: Bhatia et al. estimate the numbers of undetected COVID-19 cases exported from Mainland China based on historical air traffic patterns and public data on COVID-19 cases imported to other countries that originated in China. The observed number of cases imported to each country is assumed to be Poisson distributed, and the expected number of imported cases relative to Singapore (used as a reference country because of its high ratio of detected cases per flight volume) is derived using standard maximum likelihood estimation methods. A similar procedure is used to compare the expected number of imported cases relative to a set of reference countries which had estimated sensitivities of surveillance higher than that of Singapore. The main finding is that at least 2/3 of cases exported from China remained undetected worldwide.

Issues and comments:

Major
The code provided does not run out-of-the-box. In data_processing.R `exported_cases_paper_test.csv` should be `exported_cases.csv` and the line defining the total number of exported cases should be moved after the definition of `exported`.

Furthermore, only the data processing script for cases in international travelers is provided, not the code for data analysis or visualization of results.

While we understand that the authors cannot share the IATA data, they should *at least* provide results sufficient to reproduce Figure 2, and preferably all code for analysis and visualization such that anyone with access to the data could reproduce the results.

Minor
There is a disconnect between the use of passenger flow data from only WTIA and analysis of cases thought to have acquired infection in mainland China, regardless of whether they travelled through Wuhan or Hubei; some explanation of the decision to use these inconsistent definitions is warranted.

Please clarify why repatriation flights have been excluded when selecting cases for analysis.

At the beginning of the analysis section, "We assume that the number of exported cases in a country *i*..." should read "We assume that the observed number of exported cases in a country *i*...". This distinction should be clarified throughout.
Some mention could be made of the reason for using 2016 flight data rather than more recent data.

Some mention of potential biases and pitfalls in methods and the data would be appropriate.

There are a small number of relevant, recently published works not mentioned - e.g. https://www.medrxiv.org/content/10.1101/2020.03.23.20038331v2[1] and https://doi.org/10.1016/S0140-6736(20)30411-6[2].

The background in the abstract is very out of date; suggest updating numbers and adding a date marker ('as of...' or similar)

Additional labels for the figures would be useful (beyond the select countries labeled in Figure 1). Alternatively, per-country estimates could be made available as a table.

In the methods section ('Data Sources') the parameter lambda is referred to before being introduced.

In the discussion "Consistent with similar *analyses*" has only one citation.

References
1. Menkir T, Chin T, Hay J, Surface E, et al.: Estimating the number of undetected COVID-19 cases exported internationally from all of China. *medRxiv*. 2020. Publisher Full Text

2. Gilbert M, Pullano G, Pinotti F, Valdano E, et al.: Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. *The Lancet*. 2020; **395** (10227): 871-877 Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: infectious disease epidemiology and modelling

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Author Response 12 Aug 2021

Christl Donnelly, Imperial College London, London, UK

1. We apologise for the omission of relevant code. We have now added all relevant code so that someone with relevant data can reproduce the results. We have also included a file with dummy data on travel volume to help the readers. The README file has also been updated to include instructions on running the code.
2. Everyone we classified as cases detected overseas had travel by air either explicitly mentioned, or implied as the most probable mode of travel from mainland China to the destination (e.g. from China to Italy). Where multiple modes of travel are possible e.g. from mainland China to Hong Kong, we have only classified individuals as cases detected overseas where the mode of travel was explicitly mentioned as air. The text has been updated to clarify this.
3. In most instances, all passengers on repatriation flights had been tested for the presence of SARS-CoV2. The cases detected through surveillance of repatriation flights are therefore not representative of the typical sensitivity of surveillance in a

country. We have therefore excluded these from the analysis. The text has been updated to clarify this.

4. This sentence has now been edited and the distinction has been emphasised in the rest of the text.
5. The data from 2016 were the most recent data to which we had access when undertaking our analysis. Further, any constant scaling of the volume of passengers would not affect the estimates of model parameters (lambda and s_e). This has been emphasised in the text.
6. We have included the limitations of the method and data sources in the discussion.
7. Thanks for highlighting these relevant references. We have now included reference to these in the section Background.
8. The abstract and the reference were updated as of August 2021.
9. The reference to lambda has been removed from the methods section and a reference added to the appropriate section.
10. We have added reference to other studies conducted at the time which provide estimate surveillance sensitivity globally.

Competing Interests: No competing interests were disclosed.