Using predicted imports of 2019-nCoV cases to determine locations that may not be identifying all imported cases

Authors: De Salazar PM*, Niehus R*, Taylor A, Buckee C, Lipsitch M

*contributed equally

Affiliation: Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America

Correspondence: Pablo M. De Salazar. 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: pablom@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: mlispsitc@hsph.harvard.edu

Keywords: outbreak, coronavirus, travel, pneumonia

Abstract:

Cases from the ongoing outbreak of atypical pneumonia caused by the 2019 novel coronavirus (2019-nCoV) exported from mainland China can lead to self-sustained outbreaks in other populations. Internationally imported cases are currently being reported in several different locations [1]. Early detection of imported cases is critical for containment. Based on air travel volume estimates between Wuhan and locations in other countries and using generalized linear regression model we identify locations which may potentially have underdetected internationally imported cases.

Introduction:

A novel coronavirus 2019-nCoV has been identified with the first confirmed patient cases in December 2019 in the city of Wuhan, Hubei province, in China. Since then the number of confirmed cases has increased drastically. Model estimates suggested that by Jan 25, 2020 there were a total of over 75,000 infected cases and a doubling time of about 6 days [2]. By the end of January 2020, a city-wide quarantine was implemented for Wuhan and neighbouring cities. Nonetheless, the virus has managed to spread from Wuhan to other Chinese cities and also outside of China. Currently (as of Feb 4, 2020), internationally exported cases are being reported in 28 different countries.

Many of the imported cases have been linked to a recent travel history from Wuhan, suggesting that air travel volume may play an important role for the risk of cases being exported outside of China. To

prevent other cities and countries becoming epicentres of the 2019-nCoV epidemic, substantial targeted public health interventions are required. Here we use estimates of air travel volume from Wuhan to international destinations and a generalised linear regression model to predict imports of 2019-nCoV cases acress 26 locations. Using these predictions we can identify locations that might not be identifying imported cases.

Methods:

We regressed the cumulative number of confirmed and reported cases of 2019-nCoV in locations outside of mainland China on the estimated number of daily air flight passengers leaving from Wuhan airport by direct international flights (daily air travel volume) using a Poisson model. Data on cases were obtained by accessing Wuhan Coronavirus Global Cases [1]. Estimated air travel volume was obtained from [2]. Air travel volume for France and US was estimated using average daily passengers from [3] adjusted to have similar flight occupancy as the UK. Both 2019-nCoV cases and air travel volume were aggregated by location.

Across different locations, we assumed that the observed case counts follow a Poisson distribution, and that the expected case count is linearly proportional to the air travel volume:

$$C_i \sim Pois(\lambda_i), \qquad \text{(Equation 1)}$$
$$\lambda_i = \beta(x_i + \gamma) \quad for\ i = 1, ..., n,$$

where C_i denotes the observed case count in the *i*-th location, λ_i denotes the expected case count in the *i*-th location, β denotes the regression coefficient, x_i denotes the number of daily flight passengers from Wuhan to the *i*-th location (daily air travel volume of the *i*-th location), and γ denotes a small arbitrary value that ensures the expected case count is positive when air travel estimates are zero (Equation 1) [4]. In our case we set $\gamma = 1$, which is at most 2% of the non-zero air travel estimates.

We fit this model in R (version 3.6.1) [5] to compute $\hat{\beta}$, the maximum likelihood estimate (MLE) of the regression coefficient β, and thus the expected case count predictions for each location (Figure 1, solid line). We compute the 95% prediction interval (PI) bounds (Figure 1, dashed lines) as follows. First, we generate a bootstrapped dataset by sampling *n* observed case counts from $C_1,...,C_n$ with replacement. Second, we re-estimate β using the bootstrapped dataset and corresponding travel volumes. Third we simulate case counts for the *n* different locations under our model using this estimate of β. These three steps are repeated 50,000 times to generate 50,000 simulated case counts for each of the *n* locations from which the lower and upper PI bounds (2.5th and 97.5th percentiles) are computed. Figure 1 was generated with the 95% PI bounds smoothed using functions from the package ggplot2 [6]; see code available online https://github.com/c2-d2/cov19flightimport .

Results:

Figure 1

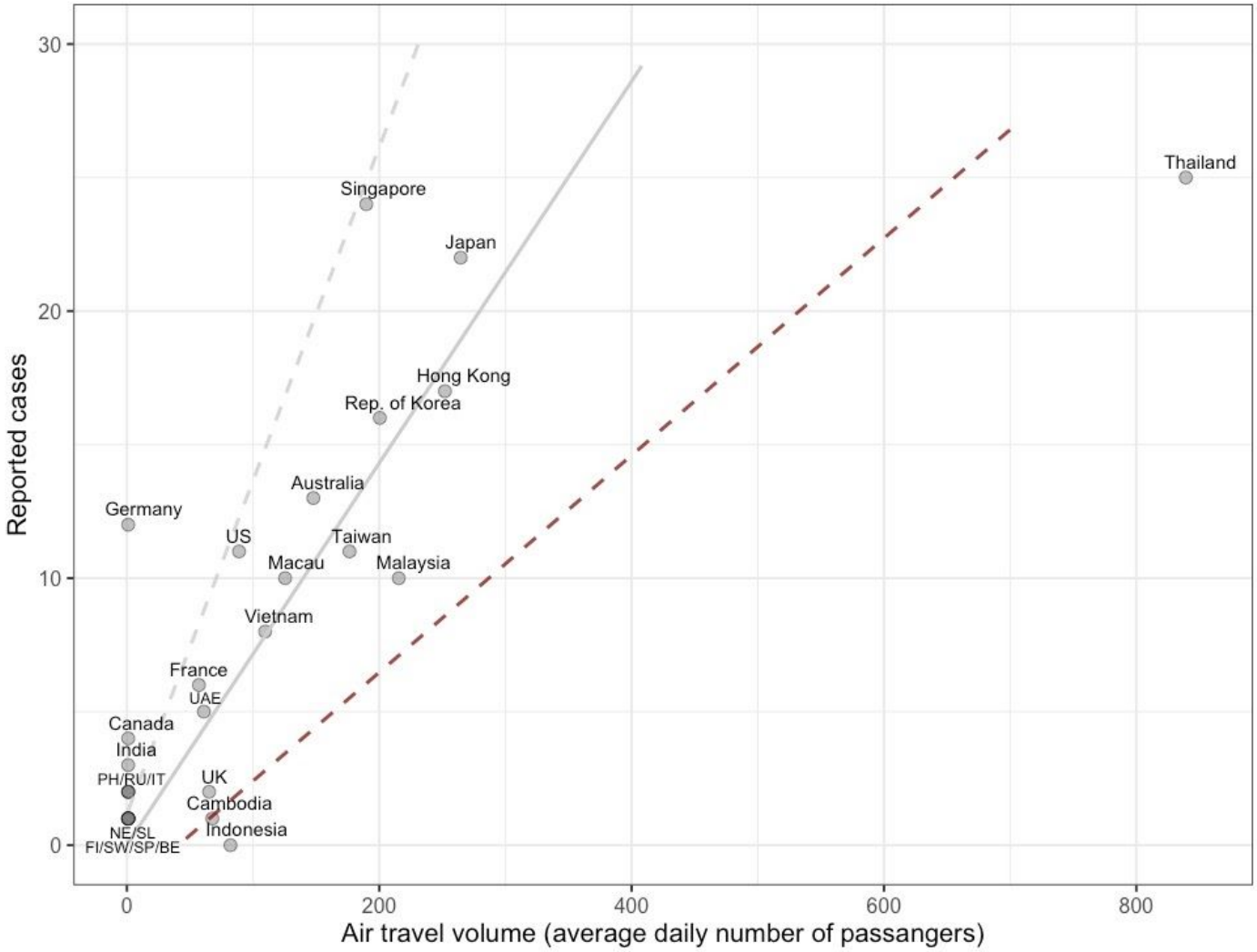

Figure 1 Reported case counts and air travel volume. The solid line shows the expected case counts based on our model, and the dashed lines show the 95% prediction interval bounds (grey dashed line for the upper bound, red for the lower bound). The Philippines (PH), Russia (RU) and Italy (IT) all report two cases each. Nepal (NE), Sri Lanka (SL), Finland (FI), Sweden (SW), Spain (SP) and Belgium (BE) all report one case each. Other abbreviations: United Arab Emirates (UAE), United States of America (US).

Discussion:

Location-aggregated reported case counts outside of mainland China show positive correlation with estimated Wuhan-to-location daily air travel volume. Locations whose case counts exceed the 95% prediction interval (PI) of our very simple model could be interpreted as having higher case-detection capacity and/or more connection with Wuhan than that captured by available daily air travel volume estimates, such as land transportation. Local within-location transmission also increases case counts, as is the case of Germany [7].

Locations with direct flights from Wuhan and reported case counts below the 95% PI may suggest the potential for undetected cases in these locations given the expected connection before travel control measures were implemented. In particular, Indonesia and Cambodia with direct flights from Wuhan during the outbreak have case counts on and under the lower bound 95% PI and have reported zero and one cases so far, while Thailand has the most reported cases but is still below the PI. We recommend that outbreak surveillance and control capacity should be rapidly strengthened in those locations lying below the 95% PI lower bound to ensure cases are detected if occurring and avoid emergence of self-sustained transmission.

References:

1. Wuhan Coronavirus Global Cases (by John Hopkins CSSE). https://systems.jhu.edu/ (2020).
2. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* (2020) doi:10.1016/S0140-6736(20)30260-9.
3. Huff, A., Allen, T., Whiting, K., Breit, N. & Arnold, B. FLIRT-ing with Zika: A Web Application to Predict the Movement of Infected Travelers Validated Against the Current Zika Virus Epidemic. *PLoS Currents* (2016) doi:10.1371/currents.outbreaks.711379ace737b7c04c89765342a9a8c9.
4. Gelman, A. & Hill, J. Generalised Linear Models. *Data Analysis Using Regression and Multilevel/Hierarchical Models* 110–116 doi:10.1017/cbo9780511790942.014.
5. R Development Core Team, R. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing* (2011).
6. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer, 2016).
7. European Centre for Disease Prevention and Control. Outbreak of acute respiratory syndrome associated with a novel coronavirus, China: first local transmission in the EU/EEA – third update. *ECDC: Stockholm* (2020).

Funding: This work was supported by Award Number U54GM088558 from the US National Institute Of General Medical Sciences. P.M.D was supported by the Fellowship Foundation Ramon Areces. A.R.T. and C.O.B. were 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.