The influence of passenger air traffic on the spread of COVID-19 in the world

Yves Morel Sokadjo [a],[*], Mintodê Nicodème Atchadé [b]

[a] *Université d'Abomey-Calavi/International Chair in Mathematical Physics and Applications (ICMPA: UNESCO-Chair), 072 BP 50 Cotonou, Benin*
[b] *National Higher School of Mathematics Genius and Modelization, National University of Sciences, Technologies, Engineering and Mathematics, Abomey, Benin*

ARTICLE INFO

ABSTRACT

Countries in the world are suffering from COVID-19 and would like to control it. Thus, some authorities voted for new policies and even stopped passenger air traffic. Those decisions were not uniform, and this study focuses on how passenger air traffic might influence the spread of COVID-19 in the world. We used data sets of cases from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University and air transport (passengers carried) from the World Bank. Besides, we computed Poisson, QuasiPoisson, Negative binomial, zero-inflated Poisson, and zero-inflated negative binomial models with cross-validation to make sure that our findings are robust. Actually, when passenger air traffic increases by one unit, the number of cases increases by one new infection.

1. Introduction

COVID-19 is a deadly virus that is at the base of a pandemic in the world since December 2019. Actually, in the beginning, 27 infections have been detected in China from Wuhan City (Ayittey et al., 2020). On 30/01/2020, the World Health Organization (WHO) declared COVID-19 as a pandemic and alarmed about the public health emergency (World Health Organization et al., 2020). This situation speeds up many countries to step into the research on how that virus can spread and be controlled. Experts also found that the virus could be transmitted person-to-person (Chan et al., 2020). Consequently, face mask wearing and social distancing (Alice et al., n.d.), home quarantine and confinement and sanitary cordon (Pan et al., 2020), contact tracing (Hellewell et al., 2020), travel restrictions (Chinazzi et al., 2020), and many other decisions have been set to decrease and control the spread of the pandemic.

The spread of the first COVID-19 cases in the world has been imported from other countries (Wu et al., 2020; Hellewell et al., 2020; Yuan et al., 2020). Besides, it has been found that human mobility can influence significantly the COVID-19 epidemic in China (Kraemer et al., 2020; Zhao et al., 2020). Especially, we think that the influence of means of transportation on the spread of COVID-19 in the world is quite possible. At the beginning of the pandemic, the airport of Wuhan did not stop any passenger flights (Bogoch et al., 2020b; Wilson and Chen, 2020). Taking into account the global passenger air traffic network, it is easy to discover how unknown contagious diseases can be spread in the world (Wilson and Chen, 2020). Actually, air transport contributed to the spread of many diseases in the world (Tatem et al., 2006; Colizza et al., 2006; MacFadden et al., 2015). Moreover, some researchers also discovered that airline travel could influence the spread of viruses such as influenza (Grais et al., 2003; Brownstein et al., 2006; Hosseini et al., 2010), Ebola (Bogoch et al., 2015), Zika (Bogoch et al., 2016), and Dengue (Tian et al., 2017). All those works support our hypothesis to think that the number of passengers from air traffic is related to the spread of the pandemic. Even while China was overcoming the peak, new infections were noted in other regions of the world with significant increases (Liu et al., 2020). Besides, there is also a recent similar work (Lau et al., 2020) that focused on the association between air traffic and the coronavirus outbreak,but the methods used were just about linear correlation. Many infected cases were imported from China to the other continents (Kim et al., 2020; Rodriguez-Morales et al., 2020). And we should recall that numerous countries did not strictly stop passenger air traffic while the virus was still infecting people.

Considering the aforementioned points, we also notice that there is no scientific body of knowledge (with robust modeling process and cross-validation) about the plausible influence of air transport on the spread of COVID-19. Besides, we also hypothesized in this study that passenger air traffic has a significant influence on the number of COVID-19 infections in the world. This results of this paper are going to be technical adding information with strong pieces of evidence to control the spread of the pandemic in the world.

---

* Corresponding author.
E-mail address: yvesmorel17@yahoo.fr. (Y.M. Sokadjo).

2. Sample description and methodology

2.1. Sample description

In this work, we used a data set of two variables such as aggregated global cases (GC) and worldwide recent data about passenger air traffic (PAT). The variables were accessed on 12/03/2020 and are from the Center for Systems Science and Engineering (CSSE) of Johns Hopkins University (JHU) (https://github.com/CSSEGISandData/COVID-19) and the latter has been downloaded from (World Bank, 2018). GC is a count of the total confirmed cases in a country per day, and PAT count of the total passenger air traffic in 2018 (it is the very recent data set available) per country. The data cover the period from 23/01/2020 to 13/07/2020.

The quality and accuracy of WHO data sets about COVID-19 are based on the developed technical guidance by topic (World Health Organization, 2020a) related to the virus. Every country is to respect all those processes, otherwise their data provided are not validated. Moreover, a confirmed case is a person that got laboratory confirmation of COVID-19 infection, whatever the symptoms and clinical signs. The data sets about passenger air traffic are international and domestic aircraft passengers registered in countries and aggregated by the World Bank.

2.2. Methodology

The goal is to propose a model that uses predictors to approach the dependent variable and check whether there is significant influence or not. The variable of interest in this work is a counting variable (with zeros at some dates), that is, the total of confirmed cases GC, denoted by (y_i). Actually, to model counting variables, researchers use many approaches such as Poisson model (PM), quasi-Poisson model (QPM), Negative binomial model (NBM), zero-inflated models (ZIM) (Lambert, 1992), and Hurdle models (HM) (Mullahy, 1986). HM and ZIM both provide computation of a Bernoulli and Poisson probabilities and give flexibility to model a zero outcome. In addition, we are suspecting overdispersion because at the beginning of the pandemic there was great variability (disproportion among countries) in terms of total cases. Consequently, in the case of the current work, we will compute all the models without Hurdle ones because it is similar to the ZIM. In addition, in the context of COVID-19, we think that the zero case means the absence of the pandemic or the non detection in a given country. With Hurdle models, one of the hypotheses stipulates that zeros should be from only one point, and this does not fit our study.

2.2.1. Poisson and quasi-Poisson regression

In Poisson regression model, the variance is assumed to be equal to the mean. Besides, in quasi-Poisson model, the variance is hypothesized to be the dispersion parameter times the mean (i.e., the variance is proportional to the mean). Be N_i, X_i, μ_i that are, respectively, the number of total cases, the passenger air traffic, and the mean number of total cases in a given country i, we have:

$$\log(\mu_i) = E\,(N = N_i | x_i, x_i) + \varepsilon_i \tag{1}$$

We recall that $E(y_i | x_i, x_i) = \alpha + \beta\, x_i$. Then,

$$\begin{aligned}\log(\mu_i) &= \alpha + \beta x_i + \varepsilon_i \\ \mu_i &= e^{\alpha + \beta x_i + \varepsilon_i}\end{aligned} \tag{2}$$

$$Pr(N = N_i | x_i, x_i) = e^{-\mu_i} \frac{\mu_i^{N_i}}{N_i!} \tag{3}$$

With that general modeling, we then have the following specificities:

$$PM : Var(N_i) = \mu_i \tag{4}$$

$$QPM : Var(N_i) = \theta \mu_i \tag{5}$$

Such that θ is the dispersion parameter, α is the intercept, β is the influence of X on N, and $bJ\varepsilon_i$ represents the error term. Let us note that PM is QPM with $\theta = 1$. When $\theta > 1$, we state that there is over-dispersion relative to Poisson.

2.2.2. Negative binomial regression

We use the negative binomial regression model when we suspect overdispersion in the modeling. Actually, it is a generalization of PM that uses the assumption that the variance is a quadratic function of the mean. In other words, when the parameter λ of the PM follows the gamma distribution of parameter (r, ρ), we use NBM meaning a Poisson-gamma mixture. Thus, we have:

$$\mu_i = e^{\alpha + \beta x_i + \varepsilon_i} \tag{6}$$

$$\Pr(N = N_i | x_i, x_i) = \frac{\Gamma(N_i + \rho)}{\Gamma(N_i + 1)\Gamma(\rho)} \left(\frac{r}{r + \rho^{-1}\mu_i}\right)^\rho \left(\frac{\rho^{-1}\mu_i}{r + \rho^{-1}\mu_i}\right)^{N_i} \tag{7}$$

Such that Γ is the gamma function.

2.2.3. Zero-inflated regressions

In this work, we can notice that there is an excess of zeros cases in the beginning of the pandemic. It is a type of regression that researchers use to handle zero-inflated count data. ZIM use the mixing of a process (logistic regression in this study) that generates zeros and another one (PM, QPM, NBM) that models counts.

$$Pr(N = N_i | x_i, x_i) = \begin{cases} \pi_i + (1 - \pi_i)g(N_i | x_i) & \text{when } N_i = 0 \\ (1 - \pi_i)g(N_i | x_i) & \text{when } N_i > 0 \end{cases} \tag{8}$$

With π_i the probability to get zero in a country i, we have the following possible models:

$$ZIM = \begin{cases} \text{Zero-inflated Poisson} & \text{when } g(N_i | x_i) \text{ is (3)} \\ \text{Zero-inflated Negative Binomial} & \text{when } g(N_i | x_i) \text{ is (7)} \end{cases} \tag{9}$$

2.3. Analysis process

The checking of the influence of PAT on the spread of COVID-19 is a very crucial point that requires a very strong method before any final conclusion. Then, taking into account that we have daily case count per country and only a data of PAT representing the air traffic per country, we decided to check the significance influence of PAT on GC with every day of the data set. It means that we computed to check our findings 174 times for each model. In addition, to be sure about the goodness of fit of our models, we implemented cross-validation method. The summary of the process is on Fig. 1.

3. Results and discussion

3.1. Preliminary checking

The variable of interest is the number of cases per country in the world. The daily evolution of that pandemic is on Fig. 2.

The hypothesis in this work is that Air traffic has influenced the spread of COVID-19. Thus, we computed and illustrated the Pearson correlation between those two variables (daily cases and Passenger Air Traffic) on the study period (from 22/01/2020 to 13/07/2020). It is illustrated on Fig. 3.

The minimum value as Pearson correlation coefficient is 0.05 (from 01/03/2020 to 04/03/2020) and the maximum 0.94 (from 07/05/2020 to 25/05/2020) with a mean of 0.66. These results motivated us to deepen our analysis with the models for better understanding of the plausible daily influence of air traffic on COVID-19 spread.
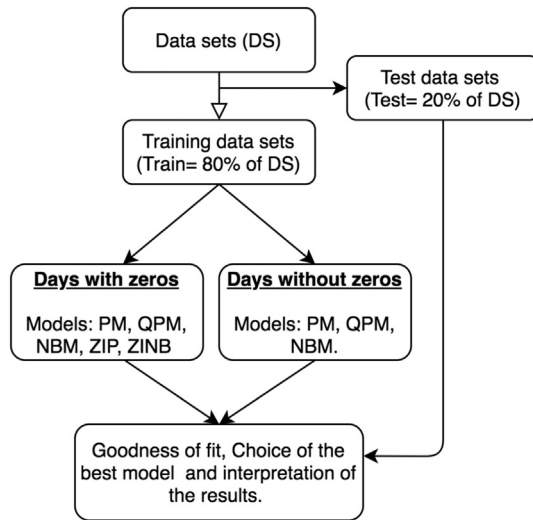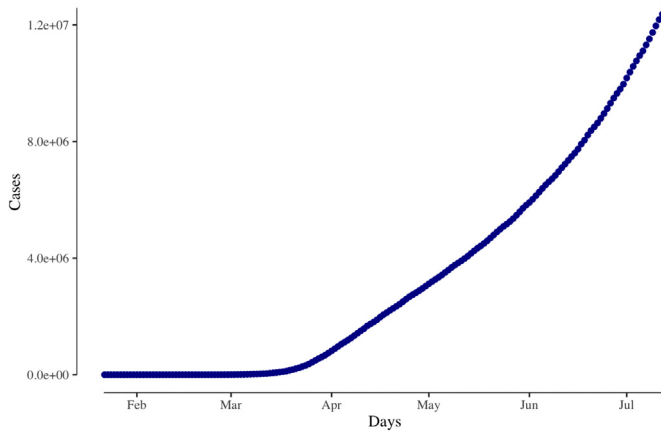
Fig. 1. Study analysis process.

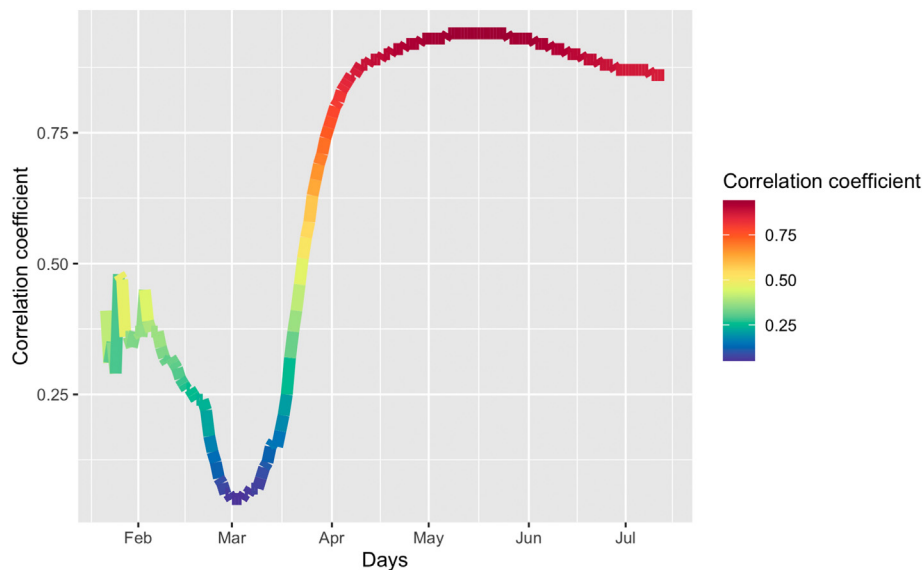**Fig. 2.** Daily number of registered cases of COVID-19.

3.2. Modeling

Let us recall that from 22/01/2020 to 09/04/2020 we have zeros cases in some countries. After that date, every country in our data set was with at least one case. These situations of many zeros (128 on 22/01/2020 and 1 on 09/04/2020) motivate us also to use zero-inflated models on the period with zeros. To interpret any model result, we did:

- Firstly, after the initial computation of the models, we have noticed that PM, QPM, and NBM regressions converge (i.e. the algorithm gets the coefficient estimates such that the values become approximately the same with iteration), however, ZIP and ZINB did not (only a part of ZIP from 24/03/2020 to 09/04/2020 converged). According to our main objective, that is, to check the influence of air traffic on the spread of COVID-19, consequently, we cannot consider a model that does not converge because it does not allow to output the p-value to interpret the influence.
- Secondly, we noticed that during the study period, all the four retained models (PM, QPM, and NBM about every day and ZIP only from 24/03/2020 to 09/04/2020) have p-values that are all smaller than 0.05 except QPM from 22/02/2020 to 17/03/2020. All the coefficients are near zero and consequently their exponential gives one. It means that our finding is robust whatever the model. When the air traffic increases of one unit, the number of cases in a country increases of one too.

Though all the retained models (with significant p-value) give the same results, we can however have an idea on the variability among coefficients per model on Fig. 4.

Actually, we can notice that in the beginning, NBM tends to estimate higher coefficients and it is a result of higher influence when we use that model. After getting estimates with the train data sets, we computed the forecasts with the test data and with the root mean square error (RMSE), we output the best model. Moreover, in Table 1, the column of best models considering p-values is the one taking into account the process to leave firstly models that did not converge before any choice. Meanwhile, the column of best models without considering p-values focuses on only best fitting without firstly checking any p-values but only the smallest RMSE in general after cross-validation. Especially, from Table 1, we can notice that regardless of p-values, the best models when we have an excess of zero are the zero-inflated ones. And this is helpful because we were expecting the ZIM for the period with excess of zeros and due to convergence we did not choose them. In other words, if there was convergence, the best models for the period with zeros should be the ZIM.
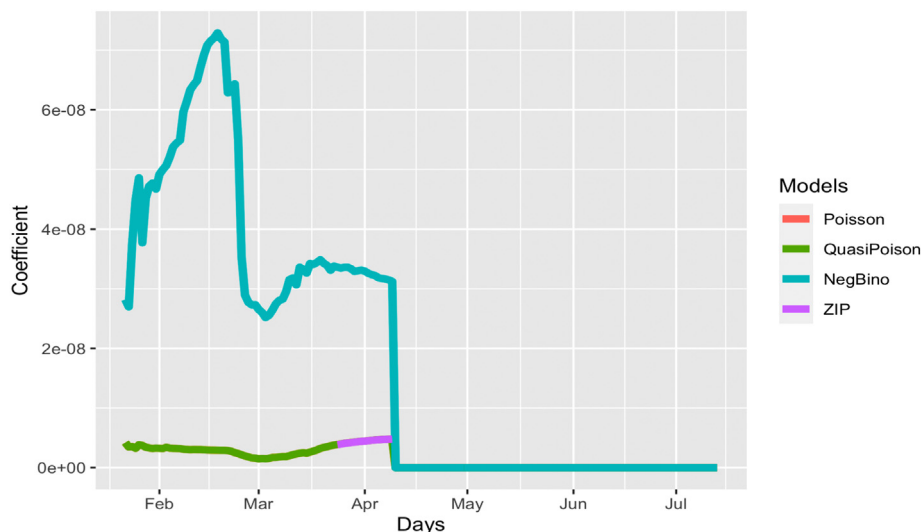

Fig. 3. Pearson correlation coefficient between Passenger Air Traffic and COVID-19 cases.

Fig. 4. Passenger Air Traffic coefficient in the models.

3.3. Discussion

The principal target of this study is to check whether passenger air traffic had a significant influence on the number of COVID-19 infections with a robust process. In the results, we noticed that the higher passenger air traffic, the higher the number of infections. Our finding is quite similar to the one in the recent work (Lau et al., 2020) related to air traffic and the pandemic. Wuhan City is the town in which COVID-19 was firstly mentioned (Ayittey et al., 2020) and many other countries' first cases were imported. A recent study (Bogoch et al., 2020a) from Wuhan in China, found that commercial air travel can influence the international dissemination of pneumonia of unknown aetiology. Even in the following work (Wilson and Chen, 2020), the authors mentioned that travellers can give wings to COVID-19 too. In the beginning of the pandemic, all the countries did not rigorously stop passenger air traffic, and it explains the finding of this work because it allowed imported transmissions to trigger the local transmission. Many works (Grais et al., 2003; Brownstein et al., 2006; Hosseini et al., 2010; Bogoch et al., 2015, 2016; Tian et al., 2017) have also mentioned that air traffic can facilitate a virus spread as in this study. When we also take into account spatial parameters with passenger mobility, we can understand that when a passenger traveled to a country, he is able to infect or to be infected. And once back to his country, the incidence of COVID-19 increases.

In addition, the spatial concerns in this work are considered thanks to each country of the data set. Furthermore, the conclusion in this work is from the comparison and variability among all the countries affected by the pandemic. For instance, though Africa is predicted to be significantly affected by the pandemic (Lone and Ahmad, 2020), according to our data, Africa is with the smallest passenger air traffic, and this also explains the small number of cases compared to other regions. Actually, this work outputs an explanation of how passenger air traffic contributes to the relative number of infections in each country. To control the pandemic, authorities should strictly stop Passenger Air Traffic to their countries, excepted emergencies. Once internal transmissions have been controlled, they can progressively open their air traffic and highly strengthen the screening tests for passengers. Passenger air traffic should be considered when there are contagious public health issues. Additionally, this work is the first one that computed 5 different models on 174 days with cross-validation to check how robust is the finding of Passenger Air Traffic on the spread of COVID-19 and proposed an appropriate model. This work is quite important for researchers in the domains of applied statistics, transportation, and public health.

However, only QPM does not confirm the influence of air traffic on the number of cases of COVID-19 from 22/02/2020 to 17/03/2020. We can recall that WHO declared COVID-19 outbreak as a pandemic on 11/03/2020 (World Health Organization, 2020b). Consequently, the absence of influence in that period can be explained by the fact that countries were facing the pandemic crisis and might change significantly their daily air traffic. Even on Fig. 3, we can notice that it is in the same period there are smaller correlation coefficients between the number of cases and air traffic passengers. This study has some limitations that are mostly related to the cross-sectional PAT data set. We think that daily time series data of PAT and COVID-19 cases will give stronger pieces of evidence. We also lack information about means of transportation of COVID-19 imported cases. Consequently, a research of those information (daily PAT time series and imported cases' means of transportation) will provide a deeper insight for the current work.

CRediT authorship contribution statement

Yves Morel Sokadjo: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Mintodê Nicodème Atchadé:** Conceptualization, Formal analysis, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing - review & editing.

Declaration of competing interest

None.

Table 1
Best models (with smallest RMSE).

Periods	Best models considering p-values	Best models without considering p-values
22/01/2020–23/01/2020	PM and QPM	ZINB
24/01/2020–28/01/2020	PM and QPM	ZIP
29/01/2020–21/02/2020	PM and QPM	ZINB
22/02/2020–24/02/2020	NBM	ZINB
25/02/2020–02/03/2020	NBM	ZIP
03/03/2020	NBM	ZINB
04/03/2020–23/03/2020	NBM	ZIP
24/03/2020–09/04/2020	ZIP	ZIP
10/04/2020–13/07/2020	NBM	NBM

Innocent Koovi, an independent researcher, who helped us in the proof-reading of this paper.

References

Alice, C., François, Q., Yann, C., 2020. Beware of virus! wearing a face mask against covid-19 results in a reduction of social distancing.

Ayittey, F.K., Dzuvor, C., Ayittey, M.K., Chiwero, N.B., Habib, A., 2020. Updates on Wuhan 2019 Novel Coronavirus Epidemic.

Bogoch, I.I., Creatore, M.I., Cetron, M.S., Brownstein, J.S., Pesik, N., Miniota, J., Tam, T., Hu, W., Nicolucci, A., Ahmed, S., Yoon, J.W., Berry, I., Hay, S.I., Anema, A., Tatem, A.J., MacFadden, D., German, M., Khan, K., 2015. Assessment of the potential for international dissemination of ebola virus via commercial air travel during the 2014 West African outbreak. Lancet 385, 29–35.

Bogoch, I., Brady, O., Kraemer, M., German, M., Creatore, M., Kulkarni, M., Brownstein, J., Mekaru, S., Hay, S., Groot, E., Watts, A., 2016. Anticipating the international spread of zika virus from Brazil. Lancet 387, 335–336.

Bogoch, I.I., Watts, A., Thomas-Bachli, A., Huber, C., Kraemer, M.U., Khan, K., 2020a. Pneumonia of unknown etiology in wuhan, china: Potential for international spread via commercial air travel. J. Travel Med. 27 (2), taaa008 Oxford University Press.

Bogoch, I.I., Watts, A., Thomas-Bachli, A., Huber, C., Kraemer, M.U., Khan, K., 2020b. Potential for global spread of a novel coronavirus from China. Journal of Travel Medicine 27, taaa011.

Brownstein, J.S., Wolfe, C.J., Mandl, K.D., 2006. Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. PLoS Med. 3, e401.

Chan, J.F.-W., Yuan, S., Kok, K.-H., To, K.K.-W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C.C.-Y., Poon, R.W.-S., et al., 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 395, 514–523.

Chinazzi, M., Davis, J.T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., y Piontti, A.P., Mu, K., Rossi, L., Sun, K., et al., 2020. The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. Science 368, 395–400.

Colizza, V., Barrat, A., Barthélemy, M., Vespignani, A., 2006. The role of the airline transportation network in the prediction and predictability of global epidemics. Proc. Natl. Acad. Sci. 103, 2015–2020.

Grais, R.F., Ellis, J.H., Glass, G.E., 2003. Assessing the impact of airline travel on the geographic spread of pandemic influenza. Eur. J. Epidemiol. 18, 1065–1072.

Hellewell, J., Abbott, S., Gimma, A., Bosse, N.I., Jarvis, C.I., Russell, T.W., Munday, J.D., Kucharski, A.J., Edmunds, W.J., Sun, F., et al., 2020. Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts. The Lancet Global Health.

Hosseini, P., Sokolow, S.H., Vandegrift, K.J., Kilpatrick, A.M., Daszak, P., 2010. Predictive power of air travel and socio-economic data for early pandemic spread. PLoS One 5, 1–8.

Kim, J., Choe, P., Oh, Y., Oh, K., Kim, J., Park, S., Park, J., Na, H., Oh, M.-d, 2020. A first case of 2019 novel coronavirus pneumonia imported into Korea from Wuhan, China: implication for infection prevention and control measures. J. Korean Med. Sci. 35.

Kraemer, M.U., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D.M., Du Plessis, L., Faria, N.R., Li, R., Hanage, W.P., et al., 2020. The effect of human mobility and control measures on the covid-19 epidemic in China. Science 368, 493–497.

Lambert, D., 1992. Zero-inflated poisson regression, with an application to defects in manufacturing. Technometrics 34 (1), 14.

Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Ichii, H., Zacharski, M., Bania, J., Khosrawipour, T., 2020. The association between international and domestic air traffic and the coronavirus (covid-19) outbreak. J. Microbiol. Immunol. Infect. 53, 467–472.

Liu, Q., Liu, Z., Zhu, J., Zhu, Y., Li, D., Gao, Z., Zhou, L., Yang, J., Wang, Q., 2020. Assessing the global tendency of covid-19 outbreak. medRxiv Cold Spring Harbor Laboratory Press.

Lone, S.A., Ahmad, A., 2020. Covid-19 pandemic – an African perspective. Emerging Microbes & Infections 9, 1300–1308 (PMID: 32458760).

MacFadden, D.R., Bogoch, I.I., Brownstein, J.S., Daneman, N., Fisman, D., German, M., Khan, K., 2015. A passage from India: association between air traffic and reported cases of New Delhi metallo-beta-lactamase 1 from 2007 to 2012. Travel Med. Infect. Dis. 13, 295–299.

Mullahy, J., 1986. Specification and testing of some modified count data models. J. Econ. 33, 341–365.

Pan, A., Liu, L., Wang, C., Guo, H., Hao, X., Wang, Q., Huang, J., He, N., Yu, H., Lin, X., et al., 2020. Association of public health interventions with the epidemiology of the covid-19 outbreak in Wuhan, China. Jama 323, 1915–1923.

Rodriguez-Morales, A.J., Gallego, V., Escalera-Antezana, J.P., Méndez, C.A., Zambrano, L.I., Franco-Paredes, C., Suárez, J.A., Rodriguez-Enciso, H.D., Balbin-Ramon, G.J., Savio-Larriera, E., Risquez, A., Cimerman, S., 2020. Covid-19 in Latin America: the implications of the first confirmed case in Brazil. Travel Med. Infect. Dis. 35, 101613.

Tatem, A.J., Rogers, D.J., Hay, S.I., 2006. Global transport networks and infectious disease spread. Adv. Parasitol. 62, 293–343.

Tian, H., Sun, Z., Faria, N.R., Yang, J., Cazelles, B., Huang, S., Xu, B., Yang, Q., Pybus, O.G., Xu, B., 2017. Increasing airline travel may facilitate co-circulation of multiple dengue virus serotypes in asia. PLoS Negl. Trop. Dis. 11, 1–15.

Wilson, M.E., Chen, L.H., 2020. Travellers Give Wings to Novel Coronavirus (2019-ncov).

World Bank, 2018. Air Transport, Passengers Carried.

World Health Organization, et al., 2020. Statement on the Second Meeting of the International Health Regulations. Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-nCoV). [Internet].

World Health Organization, 2020a. All Technical Guidance by Topic. https://www.who.int/emergencies/diseases/novel-coronaviovid-19/novel-coronavirus-2019-ncov. (Accessed 22 July 2020). .

Wu, J., Liu, J., Zhao, X., Liu, C., Wang, W., Wang, D., Xu, W., Zhang, C., Yu, J., Jiang, B., et al., 2020. Clinical characteristics of imported cases of covid-19 in Jiangsu province: a multi-center descriptive study. Clin. Infect. Dis. 10.

Yuan, J., Li, M., Lv, G., Lu, Z.K., 2020. Monitoring transmissibility and mortality of covid-19 in europe. Int. J. Infect. Dis.Cold Spring Harbor Laboratory Press.

Zhao, S., Zhuang, Z., Ran, J., Lin, J., Yang, G., Yang, L., He, D., 2020. The association between domestic train transportation and novel coronavirus (2019-ncov) outbreak in China from 2019 to 2020: a data-driven correlational report. Travel Med. Infect. Dis. 33, 101568.