Rotation Report

Santiago M. Castro Dau

8/11/2021

Introduction

The effective reproductive number (R_e) is a crucial indicator of the transmission dynamics of an epidemic. Its importance has been reaffirmed during the ongoing SARS-CoV-2 pandemic in which it has been repeatedly used to inform public health policy. Huisman and collaborators have proposed a robust method to estimate R_e using confirmation, hospitalization or death incidence data which is currently featured by Switzerland's Federal Office of Public Health in response to the SARS-CoV-2 pandemic (Huisman et al. 2020). Moreover, Huisman and collaborators applied their method on SARS-CoV-2 case data from 170 countries and made the results available in an online dashboard. However, for all but 3 of these 170 countries certain assumptions were made that might affect the accuracy of the estimates. To be precise, the delay distributions between onset of symptoms and observation was taken to be "gamma-distributed, with parameters taken from the literature" (Huisman et al. 2020, 14). Ideally these distributions would be derived empirically from publicly available line list data but unfortunately this type of information is rarely readily available.

The objective of this project was to find and explore line list data from different countries to build delay distributions. As a result of said search we present here line list data for 12 countries and their corresponding delay distributions. For each country we explore both the number of observations and the change in the shape of the distributions through time. We also investigate the overall differences between the distributions of the different countries. Finally we use these distributions to estimate the effective reproductive number using the method described in Huisman et al. (2020), and contrast the results against the R_e estimates using the distributions taken from the literature.

Before we dive into the data let us give a quick overview of where these delay distributions come into play in the method, and why they are important. Huisman's method relies on EpiEstim, a method for R_e estimation using infection incidence time series (Cori et al. 2013). However, in an epidemic one is rarely able to observe the infection events directly, and therefore it is very hard to obtain these infection incidence time series. On the other hand, observing events that naturally follow infection like positive test confirmations, hospitalizations or deaths, is much easier. Huisman's method estimates the infection incidence time series by deconvolving the observable events (confirmations, hospitalizations or deaths) time series using observation-type and time specific delay distributions. The output (an estimate for the infection incidence time series) is then used as an input for EpiEstim which in turn outputs an R_e time series (Cori et al. 2013).

Evidence was found that model misspecification (e.g. having the "real" delay distributions be different from the assumed ones) is only grave when the difference in mean is larger than 2 days (Huisman et al. 2020, 5). Likewise, allowing for the delay distributions to vary through time only had a strong effect on the estimates with large changes in the mean (Huisman et al. 2020, 5). Consequently, in order to know how much we can trust the estimates under the assumed distributions, it is important to explore how different they are from the empirical ones.

The data

The Global Health Data Set

Originally, the goal was to obtain the data from the Global Health (GH) data set, a collection of over 40 million SARS-CoV-2 cases from over 100 countries (Health 2021). Sadly more than half of the observations

in this data set were not annotated with the onset of symptoms date (from now on referred to as "onset"), making them useless for our purposes. The remaining observations that were annotated came from 37 countries, but only 9 of them had over 90 annotated observations.

Furthermore, upon inspecting the sources that the GH data set drew from, it became apparent that for some countries (e.g., USA and Germany) the data was incomplete, meaning that although in the original source there was data for every day since the beginning of the pandemic, the GH data set was missing data from the most recent months.

Moreover, the examination of the original sources revealed some inconsistencies between data sets. These can be partly explained by the fact that the GH data set is composed of multiple sources for each country (newspaper articles, sub-national health authorities, etc.), but other differences were found to be irreconcilable. In the case of Mexico for example, all of the observations in the GH data have a confirmation date, but in the original data set confirmation date is not even field. It is possible that there might have been some translation mistakes since the original data sets are in the language of their country of origin, and as a result incomplete or inaccurate data was logged into the GH data set (e.g. symptom onset date, or the date in which the observation was logged in to the data set was taken to be the confirmation date). Overall this and other inconsistencies weakened our trust in the GH data set. Nevertheless the GH data set proved to be very useful because it provided a list of countries and sources for which a centralized and publicly available line list data set was available.

Independent Line List Data Sets

Drawing inspiration from Sorci, Faivre, and Morand (2020), a research project which analyzed the correlation between political, economical and health related variables and SARS-CoV-2 Case Fatality Rate (CFR), a list of 21 countries was devised and for each country an extensive search for line list data was carried out. This list included the countries for which we knew there was line list data (thanks to the GH data set) but also incorporated additional countries such that the countries selected would represent a diverse sample in terms of variables that were found to be correlated with CFR (Sorci, Faivre, and Morand 2020). Taking this diversity into account is important because countries with a similar CFR profile might have similar delay distributions; therefore if we only consider a homogeneous set of countries we might misleadingly conclude that delay distributions across countries do not vary significantly.

The typical search consisted of (but was not limited to) surfing the web in search of the public institution responsible for recording and/or reporting the SARS-CoV-2 cases and checking whether they had a publicly available line list data set. Additionally an extensive search for literature that could provide line list data was performed.

Through this systematic approach we found 7 countries which currently record and make available to the public a line list data set (Disease Control and Prevention 2021; Secretaria de Salud 2021; Instituto Nacional de Salud 2021; Ministerio de Salud 2021; Coordenação-Geral do Programa Nacional de Imunizações 2021; Department of Health 2021; Robert Koch-Instituts 2021). In addition, a publication that provides line list data for 5 countries in the first few months of the epidemic was found (Berry et al. 2020). It is with these 8 data sets that we construct the following analyses in which we explore the completeness of the data and their corresponding delay distributions.

The countries for which we found data were Argentina, Brazil, Colombia, Germany, Hong Kong, Japan, Mainland China, Mexico, Philippines, South Korea, Singapore, and the United States of America. The CFR values and profiles for these countries are shown in figures 2 and 3. A thorough explanation of the variables in the CFR profiles can be found in the supplementary materials.

Five out of the 12 counties are in the Americas, 6 are in Asia and only 1 is in Europe (Figure 1). These 12 countries cover a wide range of CFR values and have very distinct CFR profiles (Figure 2 and 3), however there is no obvious correlation between the CFR profiles and the observed distributions (Figure 5 and Tables 1-3). We would have expected for example, that countries with a high number of hospital beds per capita (one of the variables considered in the CFR profile) might have better infrastructure to treat and test for cases thereby making the distributions of onset-confirmation, and onset-hospitalization narrower and

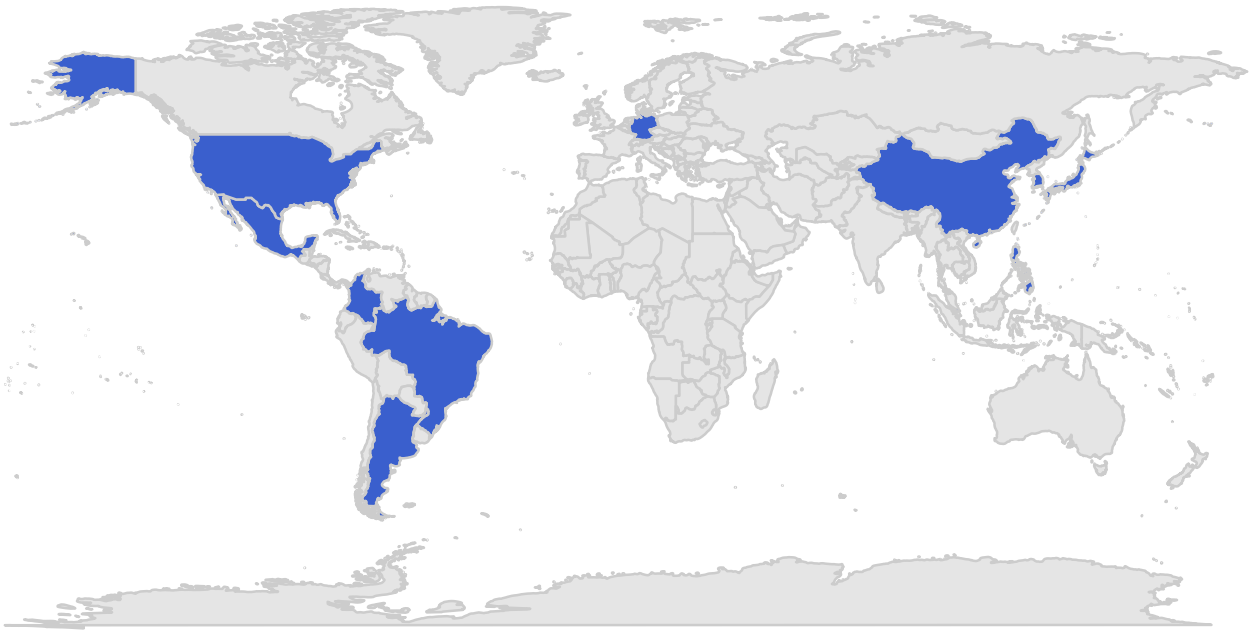

Figure 1: Map of countries for which we have line list data.

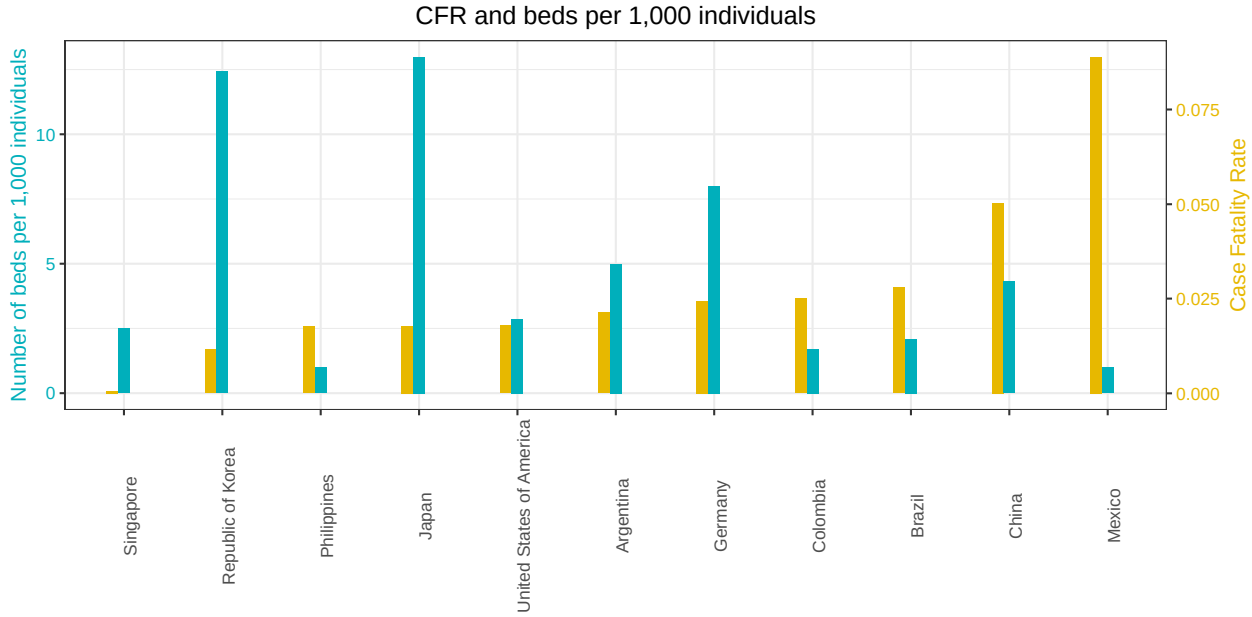

Figure 2: Case fatality rate and beds per 1,000 individuals. Hong Kong is not shown as the variables were only available for China.

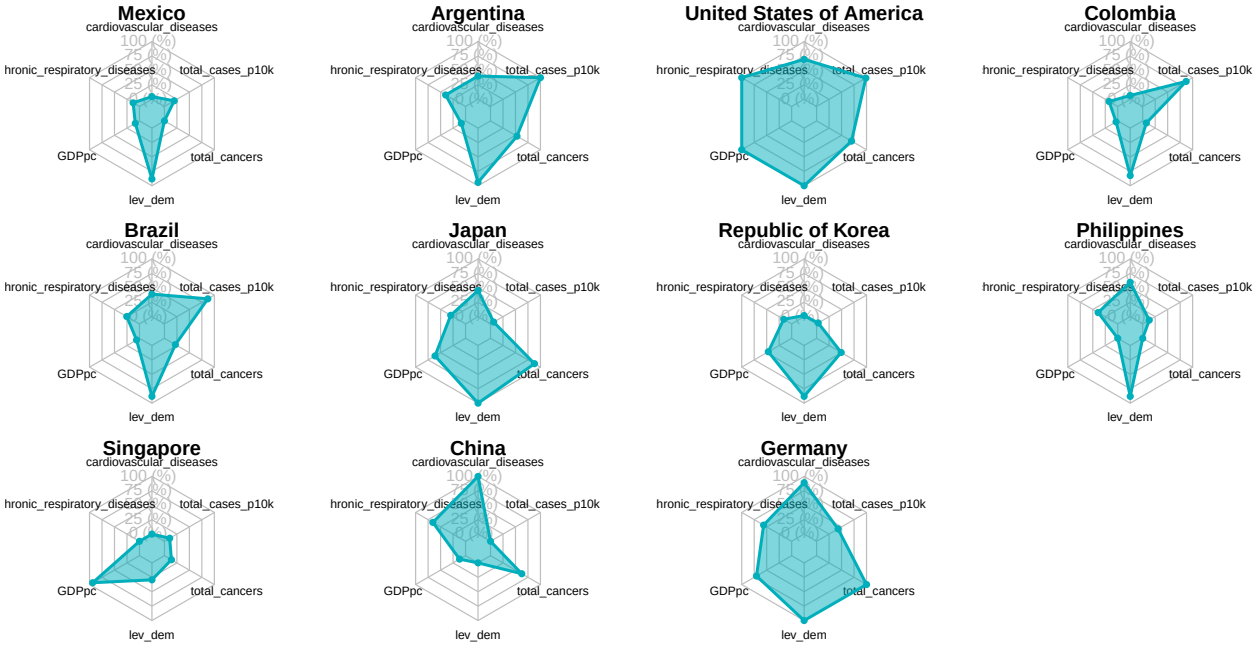

Figure 3: CFR profiles for 11 of the 12 countries. Hong Kong is not shown as the variables were only available for China. Explanation of the variables is found in the supplementary materials.

more skewed to the left. Or we would have expected that countries with high incidence of risk factors like chronic respiratory diseases (another variable considered in the CFR profile) to have shorter death-onset distributions. However just by visually inspecting the CFR profiles and the distributions, these expected trends are not apparent, but also they cannot be discarded as the number of countries considered is small and the relationships were not formally tested (Tables 1-3).

Data exploration

Data completeness

The first thing that we explored in these data sets was their completeness relative to the case incidence of each individual country. In the following figure we show for each country the number of new cases reported and the number of annotated observations for each month since the beginning of the pandemic. New reported cases were retrieved from the OWID data set (Hannah Ritchie and Roser 2020).

We can see that Japan, China, Singapore, South Korea and The Philippines only have annotated observations for the first few months of the pandemic, since they correspond to the data obtained from Berry et al. (2020). We can also see the different types of annotations that each data set provides. For countries like Argentina we have death, hospitalization and confirmation dates available as opposed to countries like the USA where only the confirmation date is provided. One can also see this in figure 5.

Figure 4 allows us to explore whether the amount of annotated observations is proportional to the number of new cases or whether there is a limit to the amount of observations that can be reported. For the countries for which we have pervasive annotation (data covering all months of the epidemic, e.g. Argentina, Brazil, Colombia, Germany, Hong Kong, Mexico and the United States of America) it seems like the former is more likely. It is worth noting that because some of these data sets are updated at the end of the month or on a weekly/bi-weekly basis, the number of annotations in the last month (July 2021) appears to be truncated in comparison to previous months.

To further explore the completeness of the data we sought to explore the fraction that the annotated observations per month represent relative to the number of reported cases per month. We thought this to be

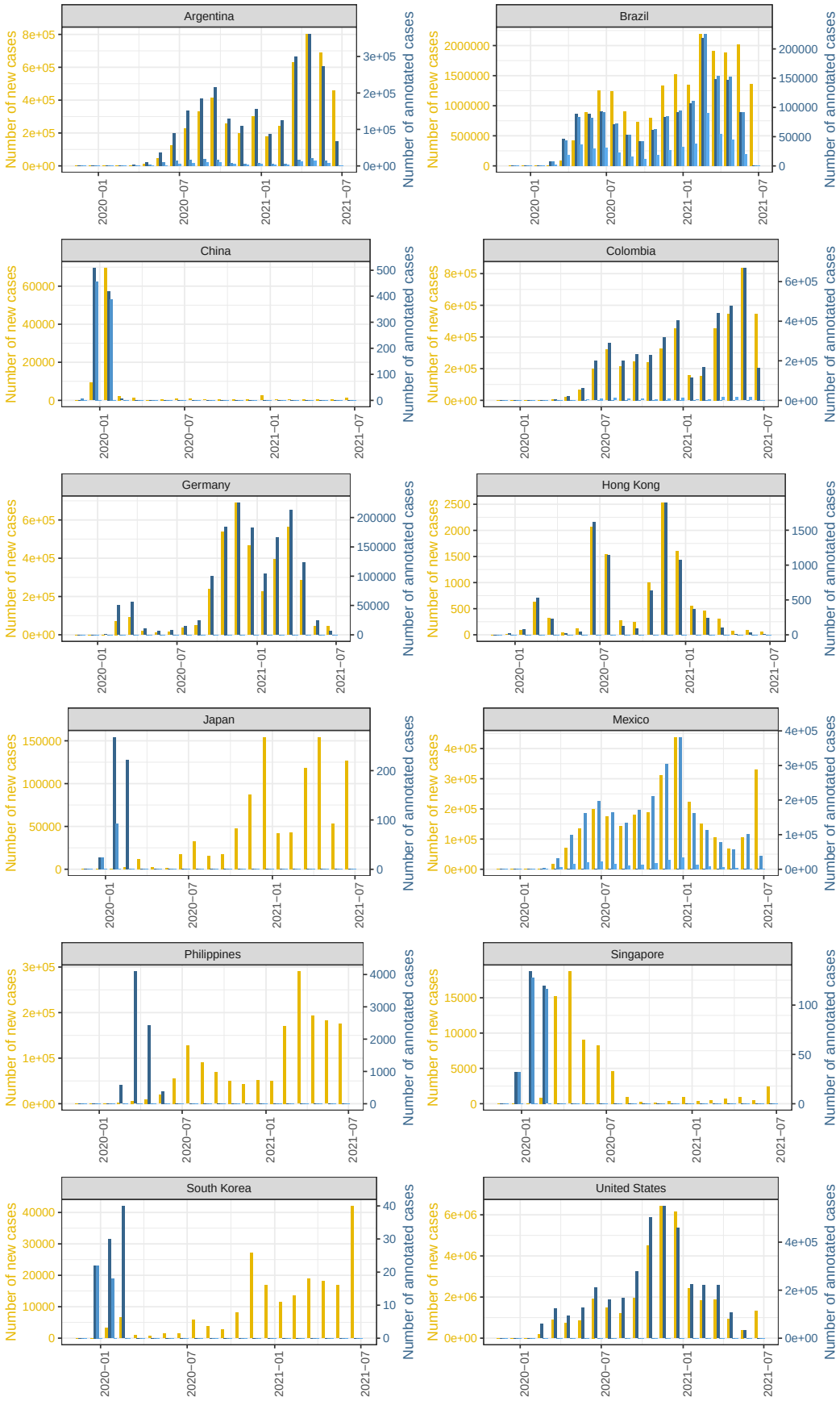

Figure 4: Yellow bars correspond to new cases while blue bars correspond to different types of annotated observations. The darkest blue corresponds to the number of observations annotated with both onset and confirmation date. The second darkest blue corresponds to the number of observations annotated with both onset and hospitalization date. Finally the lightest blue corresponds to the number of observations annotated with death and onset date.

an important dimension of the data because it could also uncover if there is a limit to the extent to which annotations can be reported imposed by the intensity of the incidence (Supplementary Figure 1). For most countries for which we have complete data, the fraction of annotated cases started very high at the beginning of the epidemic but has now decreased and stayed constant, with the exception of countries like Colombia which have kept a high percentage of annotated reporting (Supplementary Figure 1). It is also interesting to see how for some months in some countries there is a larger number of annotated observations than newly reported cases (Supplementary Figure 1). We observe this for the first few months of the epidemic, when the number of new cases was still very low for all countries, hence the magnitude of the difference between new cases and annotated cases is not so large. There could be many reasons for this phenomenon, including the under-reporting of cases.

Country wise comparison of the delay distributions

Next we show the overall delay distributions for each country (the distributions considering observations from all time points). Often there are outliers in the delay distributions, therefore to improve the clarity of the figures we filtered out delays that were smaller or equal to 0 or larger/equal to 60 days. We also show the statistics of these filtered distributions (Tables 1-3).

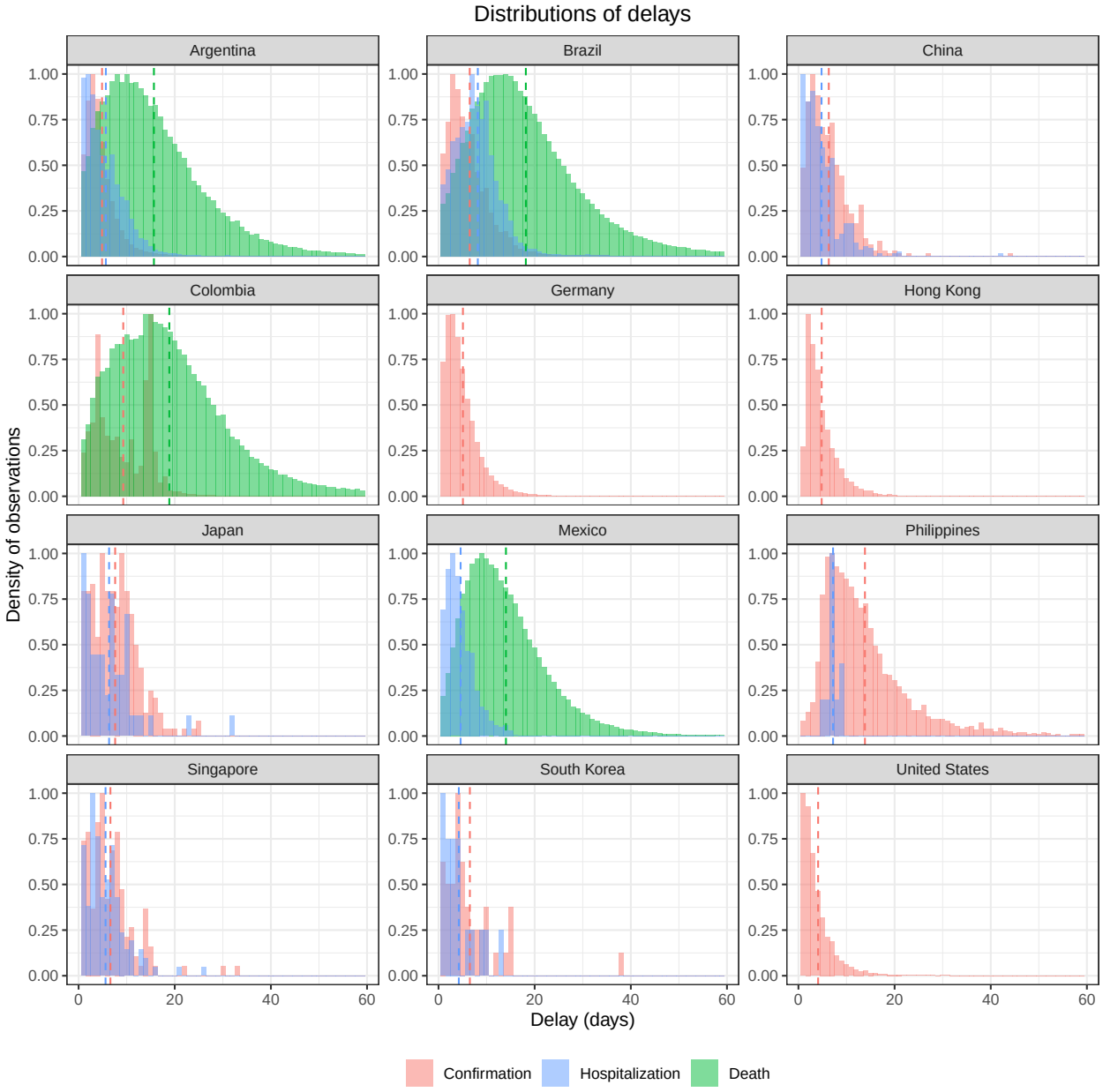

Figure 5: Distribution of delays for every country. All delays are with respect to onset date. Dashed lines indicate the mean of the distribution.

One can easily see the number of observations that inform these distributions which range from several million (i.e., Argentina, Brazil, Colombia, Germany, USA, Mexico), to less than one hundred (i.e., South Korea). Again this mostly reflects the fact that only 7 out of these 12 countries have data sets that are continuously updated and made available to the public (namely Argentina, Brazil, Colombia, Germany, USA, Mexico and Hong Kong).

Table 1: Statistics for the onset-confirmation delay distributions.

Country	Number of observations	Mean	Median	Standard Deviation
Argentina	2144902	4.9	4.0	4.7
Brazil	1294856	6.5	5.0	5.2
China	922	6.3	5.0	4.5
Colombia	3842626	9.3	8.0	5.6
Germany	1397268	5.1	4.0	4.3
Hong Kong	8161	4.8	4.0	3.6
Japan	508	7.6	7.0	4.7
Philippines	7287	13.8	12.0	9.1
Singapore	284	6.6	5.5	5.0
South Korea	88	6.5	5.0	6.2
United States	2441933	4.1	3.0	4.6

Table 2: Statistics for the onset-hospitalization delay distributions.

Country	Number of observations	Mean	Median	Standard Deviation
Argentina	141276	5.7	4	4.9
Brazil	1334758	8.2	8	5.2
China	622	4.8	4	4.1
Japan	110	6.3	5	5.6
Mexico	2104612	4.6	4	3.1
Philippines	10	7.2	7	1.2
Singapore	248	5.6	5	4.0
South Korea	36	4.2	3	3.4

Table 3: Statistics for the onset-death delay distributions.

Country	Number of observations	Mean	Median	Standard Deviation
Argentina	87930	15.7	14	10.4
Brazil	476807	18.2	16	11.0
Colombia	117199	18.9	17	11.4
Mexico	212021	14.0	12	8.7

Huisman et al. (2020) assumed onset-confirmation, onset-hospitalization and onset-death distributions that have means of 5.5, 5.1 and 15 days, and standard deviations of around 4, 4 and 7 respectively (Linton et al. 2020; Bi et al. 2020; Pellis et al. 2020). By looking at the tables 1 through 3 we can already see that most countries stay within a 2 days range of the assumed mean and also come close to the assumed standard deviations. However there are some countries which show differences in mean larger than 2 days and also significant differences in the standard deviations.

Colombia, Japan and The Philippines' onset-confirmation delay distributions exhibit such divergences. In the case of Colombia it is unclear whether we should trust this divergence since the distribution exhibits a second mode that seems unnaturally large and is perhaps the product of some artifact. In the case of Japan the small number of observations available makes it hard to draw any decisive conclusions about the observed difference in means. In the case of the Philippines it is suspicious to see such a large difference in mean and spread when compared to the other countries, but there is no special reason to doubt the veracity of the data.

Brazil and The Philippines' hospitalization-onset distributions exhibit differences in means larger than 2 days. However it's hard to trust this difference in the case of the Philippines since the distribution is only made up of 10 observations. Brazil and Mexico's death-onset distributions show significant differences in mean with respect to the assumed one. Furthermore all death-onset distributions show a larger standard deviation than the assumed one, therefore taking into account the empirical delay distributions in the R_e estimation might have a significant effect on the estimates.

Variation of the delay distributions through time

Next we show the same distributions but now breaking them down by month. For clarity we filter out delays of 30 days or larger for the confirmation and hospitalization distributions (Figure 6 and Supplementary Figure 2). For the death delay distribution we filtered out delays of 60 day or larger (Supplementary Figure 3).

An important observation from these figures (Figure 6, Supplementary Figure 2 and 3) is that the delay distributions are different in the first few months of the epidemic. Perhaps this is due to the fact that in the first few months there were less cases, and hence the resulting distributions are more variable when compared to later ones which are informed by more observations.

It is worth mentioning again that the reason why distribution in the month of July of 2021 appears to be so irregular is the data sets were downloaded some time in the middle of the month and are therefore incomplete.

Interestingly we can see the emergence of the secondary mode in the Colombian onset-confirmation distributions around July of 2020. Again it remains unclear whether this reflects the underlying distribution of delays or whether it is an artifact.

After an unstable period (1 to 5 months) most distributions settle down into a more stable form which does not fluctuate much in terms of mean and standard deviation. This is more easily seen in the next figure which shows the evolution of the mean and standard deviation of these distributions through time (Figure 7). As before, we only consider delays larger than 0 and smaller than 60 days.

For the onset-confirmation distributions we see that the mean tends to stay constant through time with the exception of Hong Kong which exhibits a more erratic behavior (Figure 7). By looking at the traces of the standard deviations for the onset-confirmation distributions for this country we can see that the spread tends to vary more than in other countries (Figure 7). Because of this we expect that using the empirical delay distributions to estimate R_e for these countries will have a greater impact in the estimates.

In the onset-death and onset-hospitalization delay distributions the means are notoriously more stable than in the onset-confirmation distributions. This trend is also observed in the traces for the standard deviation. It is also interesting to note how for the onset-death distributions all country's exhibit a slight increase in mean over time, perhaps reflecting the accumulated expertise of the medical staff in treating SARS-CoV-2 patients.

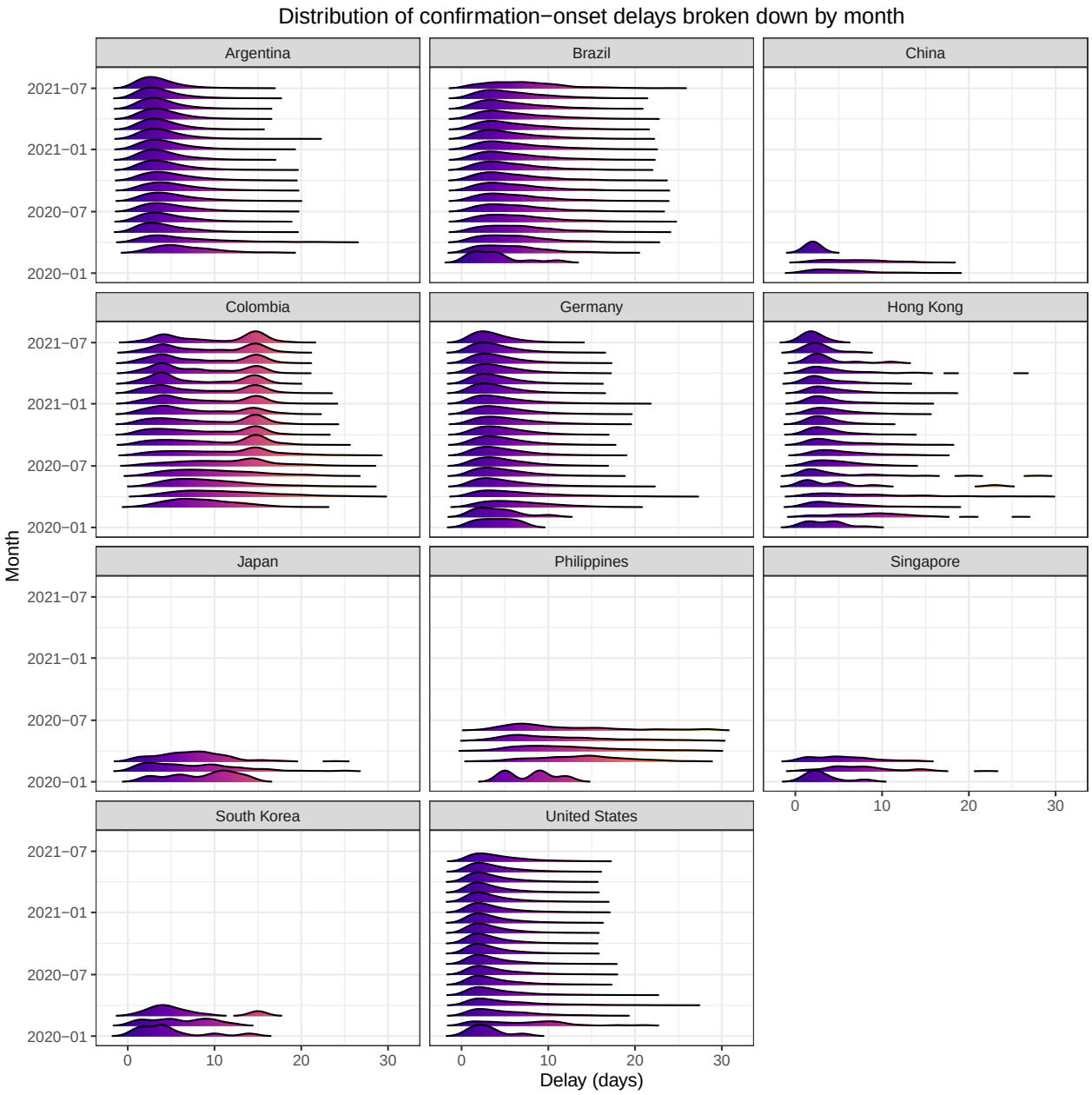

Figure 6: Distribution of confirmation-onset delays for every country for which there is data. The band width used for the density estimation is 1.

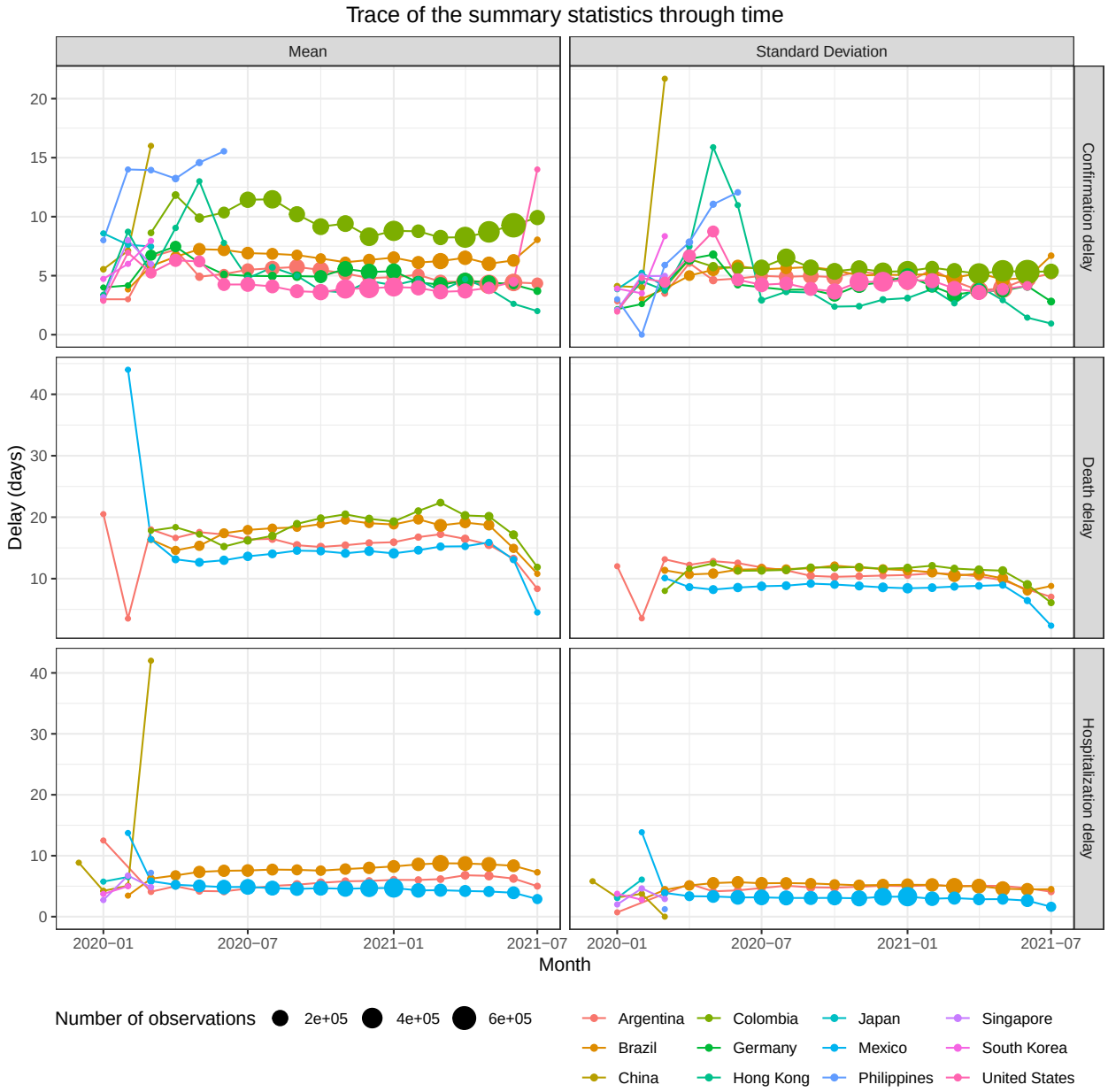

Figure 7: Trace of the summary statistics for the different delay distributions. The size of the dots represents the number of observations available for that month and country.

Estimating R_e

Finally in order to test the effect of using the empirical distributions that we found we estimated R_e using the method described in Huisman et al. (2020). For each country we do this using both the empirical delay distributions and the fixed distributions and analyze the observed differences.

The fixed distributions (from now on referred to as fixed from the literature) are parameterized with the means and standard deviations mentioned in the previous section (Linton et al. 2020; Bi et al. 2020; Pellis et al. 2020). For the empirical distributions we consider both a gamma distribution parameterized by the mean and standard deviation of the empirical delay distribution (from now on referred to as fixed empirical) and the empirical distribution themselves (from now on referred to as variable empirical).

This analysis was only performed for countries for which we have pervasive reporting, namely we excluded China, Japan, Philippines, Singapore and South Korea, because they only have data for the first few months of the pandemic. Incidentally we will only estimate R_e using confirmation and death incidence data. We leave out hospitalization data because there is no readily available hospitalization incidence data for the countries for which we have pervasive annotated observations (Mexico, Colombia, Brazil and Argentina). Again to improve the quality of the delay distributions we filtered out delays that were smaller or equal to 0 or larger/equal to 60 days.

Onset-Confirmation R_e estimates

USA, Argentina, Brazil, Hong Kong and Germany show onset-confirmation traces that are almost identical regardless of whether we use the variable empirical distributions or the fixed from the literature (Figure 8). This is expected as their onset-confirmation distributions are very stable through time and resemble the fixed from the literature delay distribution (they are within a 2 day window of the fixed from the literature mean, which is 5.5). Perhaps the most notable differences between the traces are at the beginning of the epidemic when the empirical delay distribution was much more unstable.

Interestingly, Hong Kong's variable empirical R_e trace shows larger divergences even though its empirical delay distributions also resemble the assumed fixed delays distribution. This is perhaps due to the variation in the construction of the delay distribution through time. Since Hong Kong has relatively few annotated observations compared to the rest of the countries that we consider here the number of observations that inform the distributions is small for each time window considered, making the distribution vary significantly between them (Table 1).

Colombia's variable empirical trace and fixed from the literature trace show significant divergences, which was expected as both underlying distributions are very different. Furthermore the variable empirical trace can hardly be approximated with a gamma distribution (Figure 5). On the other hand the fixed empirical trace follows very closely the fixed from the literature trace, especially following the same pattern but with a slight delay. This similarity is perhaps due to the fact that these distributions are both gamma distributions (even if they differ significantly in mean and variance) as opposed to the empirical distribution which is bimodal.

It is also interesting how both Brazil and Colombia show very wide confidence intervals in the variable empirical traces. One probable explanation is that this uncertainty is caused by the larger spread in the empirical delay distribution.

Onset-Death R_e estimates

The differences between the fixed empirical and the variable empirical traces might be explained by the variance induced by a low number of observations informing the distributions at each time window. This would explain why the variable empirical traces for the onset-death R_e estimates are so noisy, since the onset-death distributions only have a couple hundred thousand observations per country as opposed to some of the onset-confirmation distributions which have millions.

As with the onset-confirmation traces, the fixed empirical and the fixed from the literature traces follow the same pattern but are slightly out of phase; shifted forward or backwards depending on whether the mean of

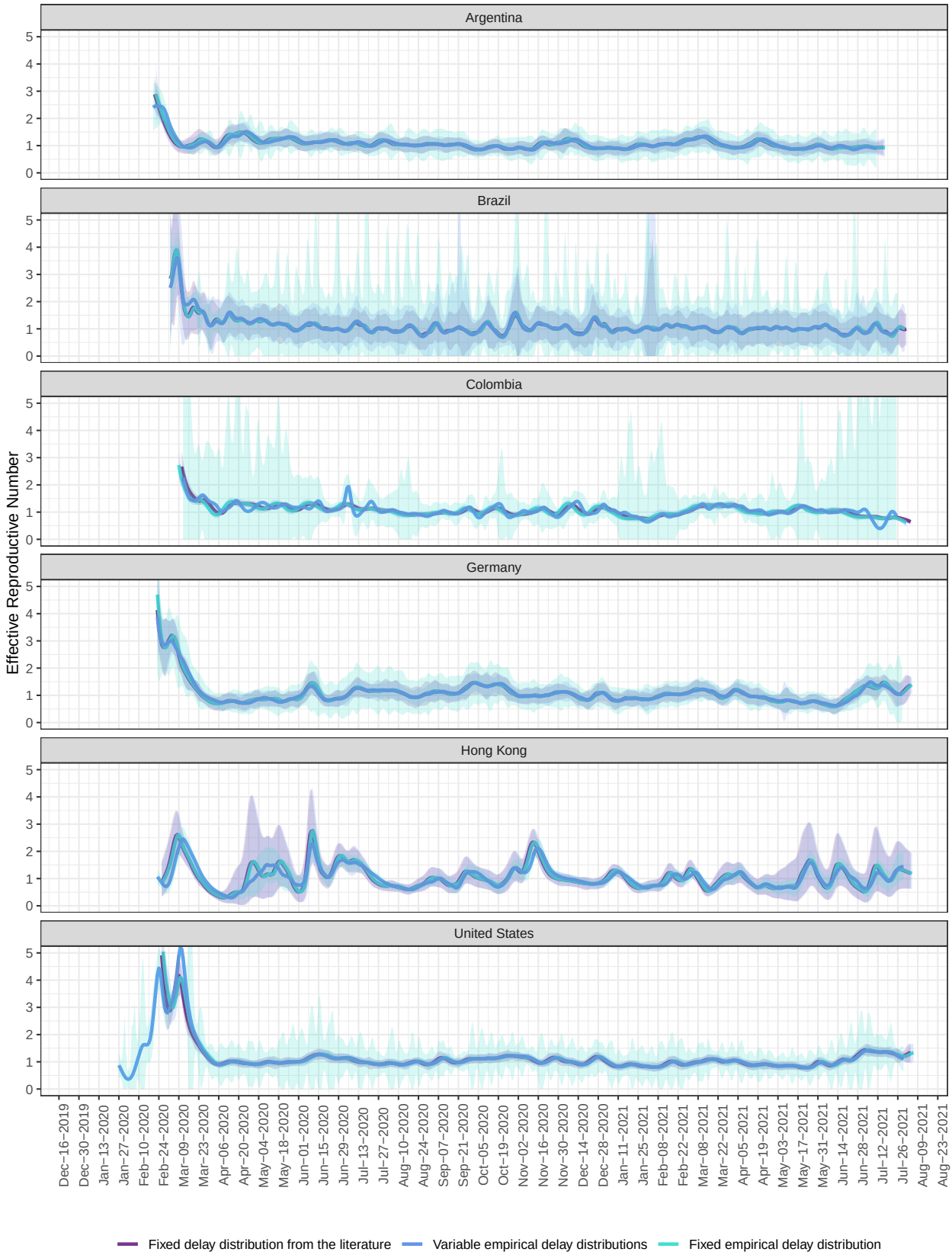

Figure 8: Effective reproductive number estimates using empirical and fixed onset-confirmation delay distributions.

Figure 9: Effective reproductive number estimates using empirical and fixed onset-death delay distributions.

the empirical distribution is larger or smaller than the mean of the assumed distribution.

Interestingly Mexico seems to be the exception with periods where the magnitude of the R_e estimates vary significantly between the fixed empirical and the fixed form the literature traces (Figure 9). These periods seem to coincide with steep changes in the value of R_e which hints at the possibility of this being an artifact of the smoothing step in the method (Huisman et al. 2020, 22).

Conclusions

Out of the 12 countries for which we found reliable line list data, there were 7 for which there is pervasive reporting of annotated cases. We found that the reporting of the annotated cases seems to be proportional to the number of new cases. For the 12 countries we constructed type specific delay distributions and did an analysis of the difference between these and the distributions reported in the literature (Linton et al. 2020; Bi et al. 2020; Pellis et al. 2020). We found that there were noticeable differences, which implied that using the empirical delay distributions might have a significant effect on the R_e estimates of the effective reproductive number. We also investigated whether these distributions have evolved throughout the course of the pandemic and found that after the first few months all distributions adopt a stable form.

We made use of the onset-confirmation and onset-death delay distributions that we constructed to estimate R_e for 7 out of the 12 countries and found that even when the difference between the mean of the empirical and the assumed distribution is larger than 2 days the estimates are quite similar. We also found that the R_e estimates derived from using variable empirical delay distributions makes the estimates very noisy unless there are many available annotated observations.

In the case of Colombia's onset-confirmation distribution, there are many available observations at every time window to construct the time varying delay distributions but the estimates become noisy because the empirical distribution is bi-modal, which presumably makes the resulting distributions subject to noise from sampling more observations from one mode or the other. The fact that we don't observe this noise when using a fixed delay distribution parametrized by the mean and variance of the empirical onset-confirmation distribution, further supports this hypothesis.

Moreover it is hypothesized that abrupt changes in the value of R_e might make the signal more sensitive to changes in the delay distribution, like we can see for Mexico's fixed empirical and fixed from the literature R_e estimates (Figure 9). Overall these findings show the importance of exploring the distributions of different countries in order to verify the assumed delay distributions used in the method proposed in Huisman et al. (2020).

Supplementary Material

The CFR profile was composed of the following variables:

- `total_cases_p10k` and `CFR` are the the number of accumulated cases over time per 10,000 inhabitants in each country and the Case Fatality Rate (number of deaths by COVID divided by the number of cases) respectively (Hannah Ritchie and Roser 2020).
- `cardiovascular_diseases`, `chronic_respiratory_diseases` and `total_cancers` values are DALYs per 100,000 population in 2019 (Disease Collaborative Network 2020). Disability Adjusted Life Years (DALYs) are measuring lost health and are a standardized metric that allow for direct comparisons of disease burdens of different diseases across countries, between different populations, and over time. Conceptually, one DALY is the equivalent of losing one year in good health because of either premature death or disease or disability. One DALY represents one lost year of healthy life. This variable was found to be positively correlated with CFR (Sorci, Faivre, and Morand 2020).
- `GDPpc` in US dollars in 2020. This variable was found to be positively correlated with CFR (Sorci, Faivre, and Morand 2020; Bank 2021).
- `lev_dem` each country receives a Polity IV score, which goes from -10 (full autocracy) to 10 (full democracy) 2015. This variable was found to be positively correlated with CFR (Sorci, Faivre, and Morand 2020; Roser 2013).
- `year` indicates the year for which `beds_per_1000`, the number of hospital beds per 1000 individuals corresponds to. This variable was found to be negatively correlated with CFR (Sorci, Faivre, and Morand 2020; Bank 2021).

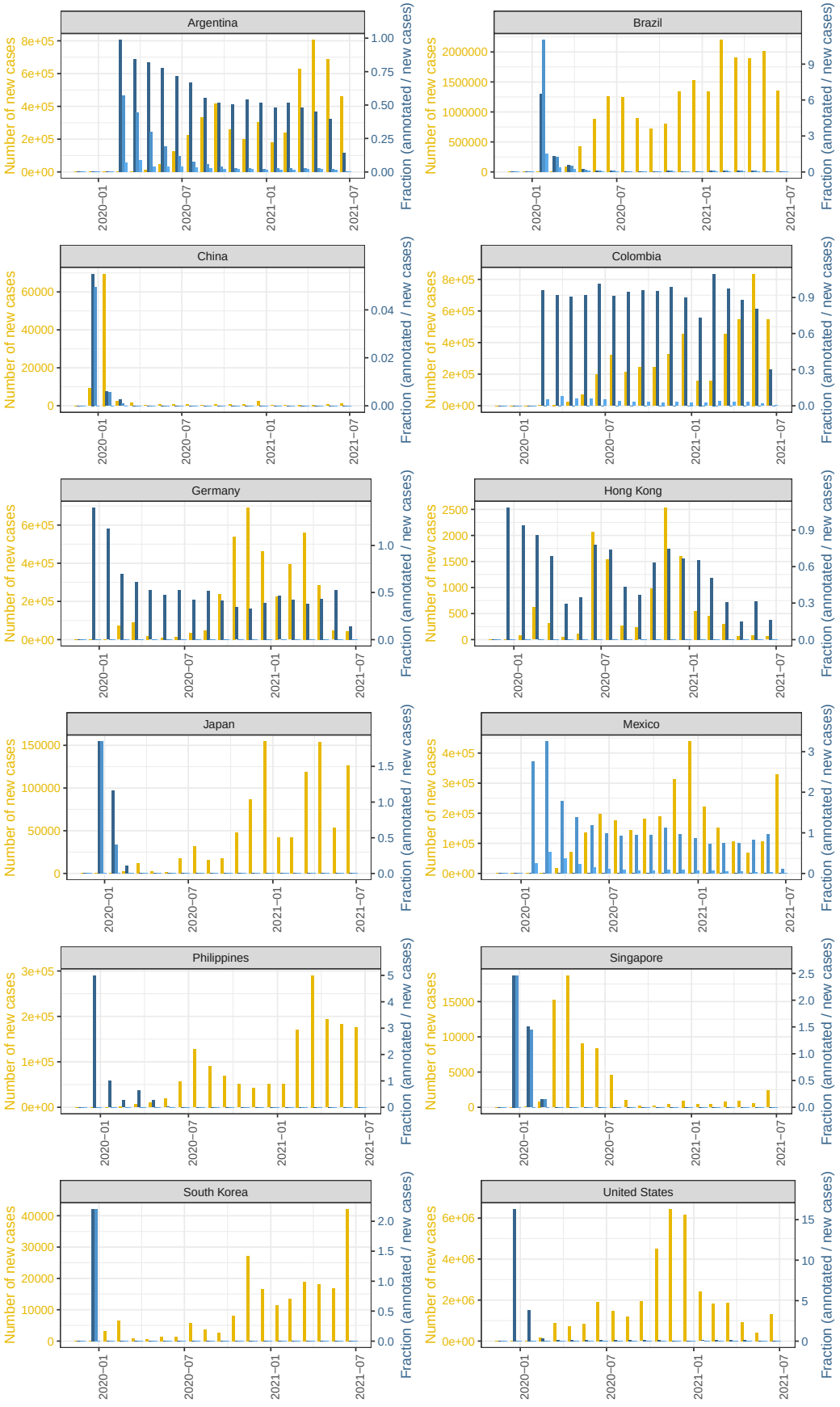

Figure 10: / Supplementary Figure 1. Same coloring scheme as in figure 4.

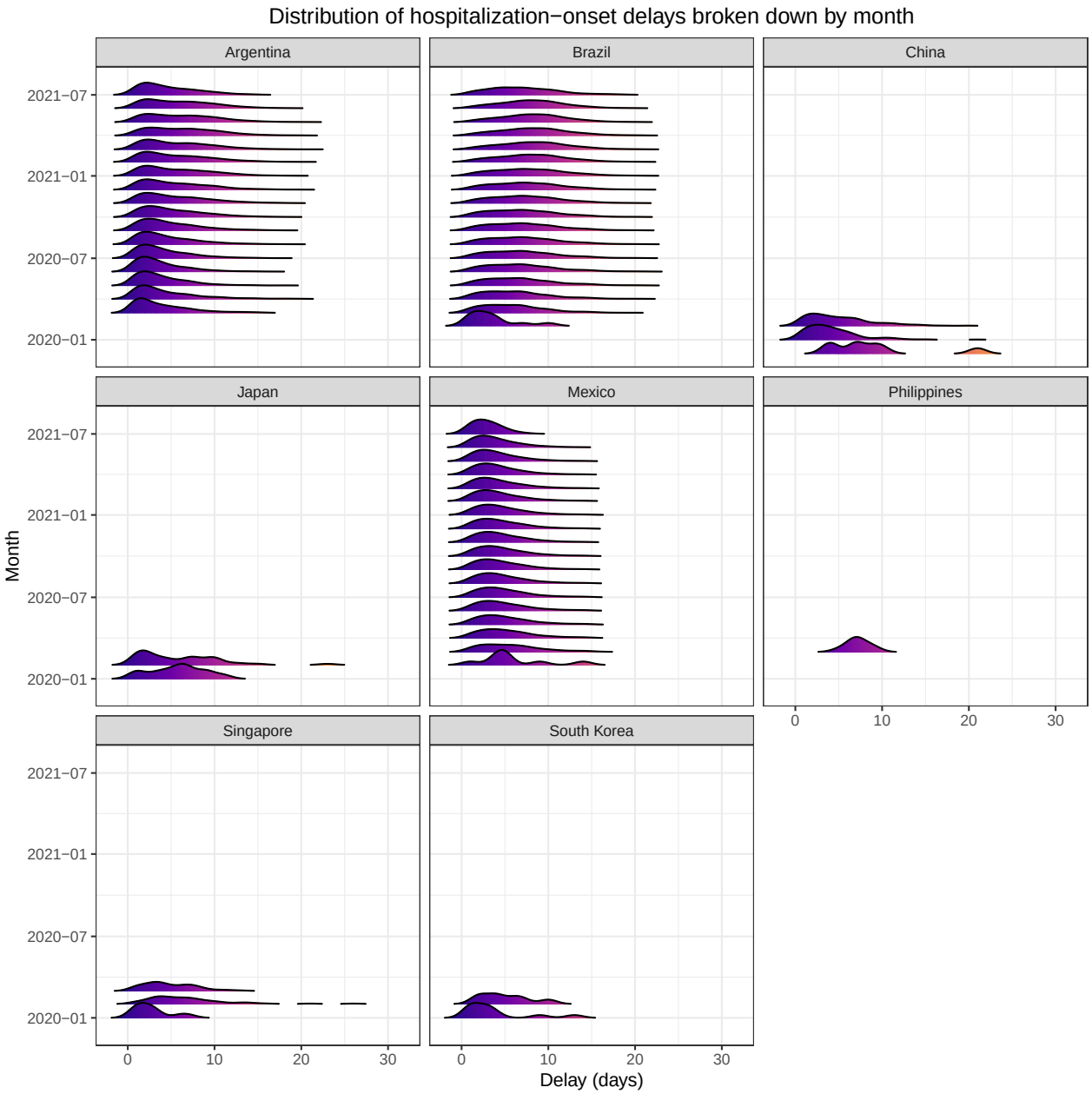

Figure 11: / Supplementary Figure 2. Distribution of hospitalization-onset delays for every country for which the is data.

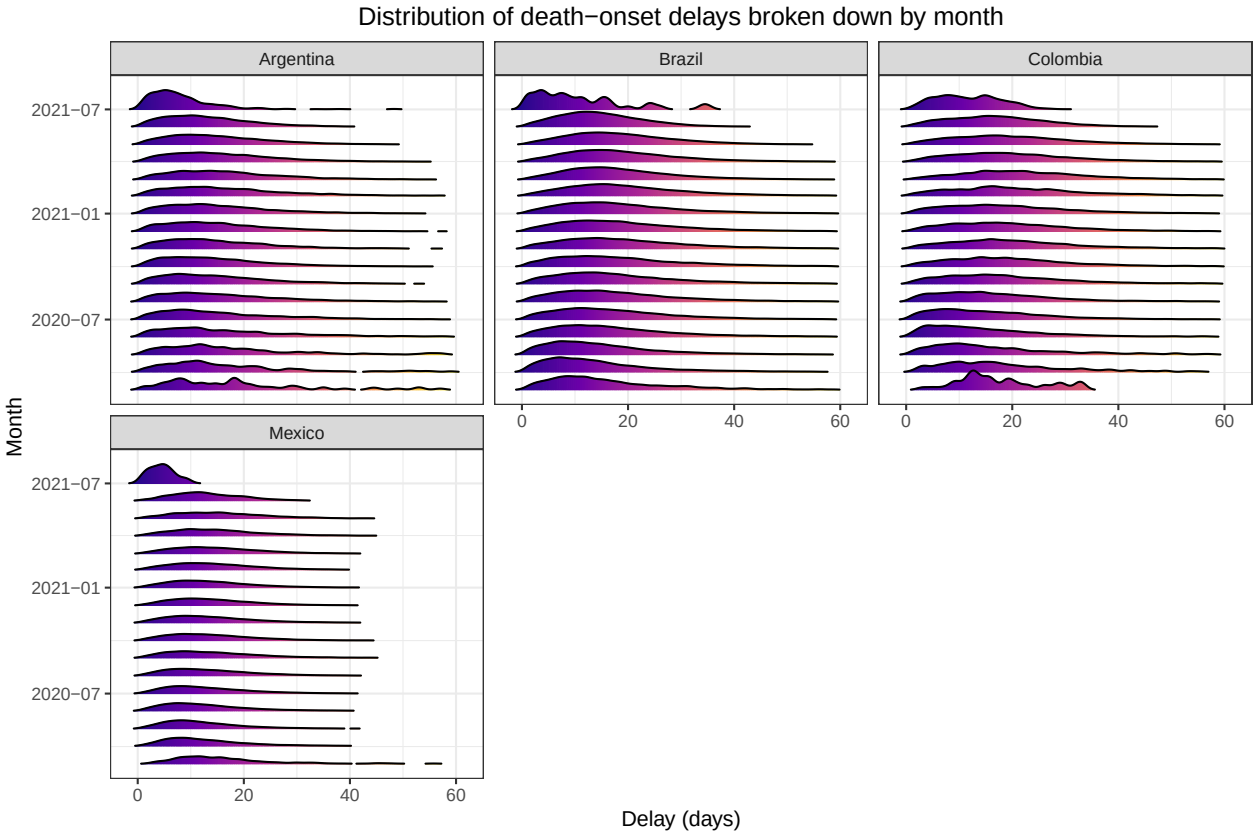

Figure 12: / Supplementary Figure 3. Distribution of death-onset delays for every country for which the is data.

Sources

Bank, The World. 2021. "GDP Per Capita and Number of Beds Per 1000 Individuals." Accesed on July 2021, https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.PCAP.CD&country=, https://databank.worldbank.org/reports.aspx?source=2&series=SH.MED.BEDS.ZS&country=.

Berry, Isha, Jean-Paul R. Soucy, Ashleigh Tuite, and David Fisman. 2020. "Open Access Epidemiologic Data and an Interactive Dashboard to Monitor the Covid-19 Outbreak in Canada." *CMAJ* 192 (15): E420–E420. https://doi.org/10.1503/cmaj.75262.

Bi, Qifang, Yongsheng Wu, Shujiang Mei, Chenfei Ye, Xuan Zou, Zhen Zhang, Xiaojian Liu, et al. 2020. "Epidemiology and Transmission of Covid-19 in 391 Cases and 1286 of Their Close Contacts in Shenzhen, China: A Retrospective Cohort Study." *The Lancet Infectious Diseases* 20 (8): 911–19. https://doi.org/https://doi.org/10.1016/S1473-3099(20)30287-5.

Coordenação-Geral do Programa Nacional de Imunizações, Brazil. 2021. "SRAG 2021 - Banco de Dados de Síndrome Respiratória Aguda Grave - Incluindo Dados Da Covid-19." *Publicly Avilable Data Set*. Accessed: 2021-07-15, https://opendatasus.saude.gov.br/dataset.

Cori, Anne, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. 2013. "A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics." *American Journal of Epidemiology* 178 (9): 1505–12. https://doi.org/10.1093/aje/kwt133.

Department of Health, Hong Kong. 2021. "Details of Probable/Confirmed Cases of Covid-19 Infection in Hong Kong (English)." *Publicly Avilable Data Set*. Accessed: 2021-07-15, https://data.gov.hk/en-data/dataset/hk-dh-chpsebcddr-novel-infectious-agent/resource/e04656ce-530a-45ab-84d1-45eaec43f743.

Disease Collaborative Network, Global Burden of. 2020. "Global Burden of Disease Study 2019 (Gbd 2019) Results." *Institute for Health Metrics and Evaluation (IHME)*. Accesed on July 2021, http://ghdx.healthdata.org/gbd-results-tool.

Disease Control, Centers for, and USA Prevention. 2021. "COVID-19 Case Surveillance Public Use Data with Geography." *Publicly Avilable Data Set*. Accessed: 2021-07-15, https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data-with-Ge/n8mc-b4w4.

Hannah Ritchie, Diana Beltekian, Esteban Ortiz-Ospina, and Max Roser. 2020. "Coronavirus Pandemic (Covid-19)." *Our World in Data*.

Health, Team Global. 2021. "Line List Epidemiological Data from the COVID-19 Outbreak." Accessed on 2021-06-24, https://global.health.

Huisman, Jana, Jérémie Scire, Daniel Angst, Richard Neher, Sebastian Bonhoeffer, and Tanja Stadler. 2020. "Estimation and Worldwide Monitoring of the Effective Reproductive Number of Sars-Cov-2," November. https://doi.org/10.1101/2020.11.26.20239368.

Instituto Nacional de Salud, Colombia. 2021. "Casos Positivos de Covid-19 En Colombia." *Publicly Avilable Data Set*. Accessed: 2021-07-15, https://www.datos.gov.co/Salud-y-Protecci-n-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr.

Linton, Natalie M., Tetsuro Kobayashi, Yichi Yang, Katsuma Hayashi, Andrei R. Akhmetzhanov, Sung-Mok Jung, Baoyin Yuan, Ryo Kinoshita, and Hiroshi Nishiura. 2020. "Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data." *Journal of Clinical Medicine* 9 (2): 538. https://doi.org/10.3390/jcm9020538.

Ministerio de Salud, Argentina. 2021. "Casos Covid-19." *Publicly Avilable Data Set*. Accessed: 2021-07-15, http://datos.salud.gob.ar/dataset/covid-19-casos-registrados-en-la-republica-argentina/archivo/fd657d02-a33a-498b-a91b-2ef1a68b8d16.

Pellis, Lorenzo, Francesca Scarabel, Helena B. Stage, Christopher E. Overton, Lauren H. K. Chappell, Katrina A. Lythgoe, Elizabeth Fearon, et al. 2020. "Challenges in Control of Covid-19: Short Doubling Times and Long Delay to Effect of Interventions." *medRxiv*. https://doi.org/10.1101/2020.04.12.20059972.

Robert Koch-Instituts, Germany. 2021. "CSV Mit Den Aktuellen Covid-19 Infektionen Pro Tag (Zeitreihe)." *Publicly Avilable Data Set*. Accessed: 2021-07-15, https://www.arcgis.com/home/item.html?id=f10774f1c63e40168479a1feb6c7ca74.

Roser, Max. 2013. "Democracy." *Our World in Data*.

Secretaria de Salud, Mexico. 2021. "Información Referente a Casos Covid-19 En México." *Publicly Avilable Data Set*. Accessed: 2021-07-15, https://datos.gob.mx/busca/dataset/informacion-referente-a-casos-covid-19-en-mexico.

Sorci, Gabriele, Bruno Faivre, and Serge Morand. 2020. "Explaining Among-Country Variation in Covid-19 Case Fatality Rate." *Scientific Reports* 10 (1): 18909. https://doi.org/10.1038/s41598-020-75848-2.