medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Patterns of the COVID19 epidemic spread around the world: exponential vs power laws

Natalia L. Komarova[1] **and Dominik Wodarz**[2]

[1] Department of Mathematics, University of California Irvine, Irvine, CA 92697
[2] Department of Population Health and Disease Prevention, Program in Public Health, Susan and Henry Samueli College of Health Science, University of California Irvine, Irvine, CA 92697

Abstract. We have analyzed the COVID19 epidemic data of more than 174 countries (excluding China) in the period between January 22 and March 28, 2020. We found that some countries (such as the US, the UK, and Canada) follow an exponential epidemic growth, while others (like Italy and several other European countries) show a power law like growth. At the same time, regardless of the best fitting law, most countries can be shown to follow a trajectory similar to that of Italy, but with varying degrees of delay. We found that countries with "younger" epidemics tend to exhibit more exponential like behavior, while countries that are closer behind Italy tend to follow a power law growth. We hypothesize that there is a universal growth pattern of this infection that starts off as exponential and subsequently becomes more power law like. Although it cannot be excluded that this growth pattern is a consequence of social distancing measures, an alternative explanation is that it is an intrinsic epidemic growth law, dictated by a spatially distributed community structure, where the growth in individual highly mixed communities is exponential but the longer term, local geographical spread (in the absence of global mixing) results in a power-law. This is supported by computer simulations of a metapopulation model that gives rise to predictions about the growth dynamics that are consistent with correlations found in the epidemiological data. Therefore, seeing a deviation from straight exponential growth may not be a consequence of working non-pharmaceutical interventions (except for, perhaps, restricting the air travel). Instead, this is a normal course of raging infection spread. On the practical side, this cautions us against overly optimistic interpretations of the countries epidemic development and emphasizes the need to continue improving the compliance with social distancing behavior recommendations.

1 Introduction

An outbreak of a novel coronavirus, named COVID19, was reported in December 2019 in Wuhan, China, and has been the source of significant morbidity and mortality due to progressive pneumonia [1, 2]. It has since spread around the world and become a pandemic, with large infection burdens reported in Europe and the United States. Disease severity and mortality seem age-dependent, with a higher chance of respiratory complications and death among older people [3], although morbidity and mortality also occur

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

to some extent in younger segments of the population. The overall mortality rate further depends on the prevalence of the infection in the population: Once the number of infected individuals has reached sufficiently large sizes, health care systems become overwhelmed and lose the capacity to treat all patients in need [4]. In particular, the number of respirator units poses an important limitation, which has resulted in difficult triaging policies in some countries, most notably Italy and Spain. Non-pharmaceutical intervention methods, such as social distancing, have become an important tool in the fight against coronavirus spread [5]. The idea is to slow down the rate at which the virus spreads through human populations, and thus to maintain the number of infected individuals that require treatment below a threshold level, thus avoiding a situation in which hospitals become overwhelmed. Hence, wide-spread closures of schools and businesses, and stay-home orders, have been implemented across Asia, America, Europe, and other affected regions.

A number of studies have been performed that investigated the kinetics of coronavirus spread through human populations. The pattern of virus spread was found to be approximately exponential, at least during the early stages of virus spread [6, 7], and importantly, the basic reproductive ratio of the virus has been estimated [8, 9, 10, 11]. The implementation of social distancing measures are expected to have a significant impact on the virus spread kinetics [5], which has been observed in some Asian countries, such as China [12]. While the effect of containment efforts have been very pronounced in these countries, they might be more subtle in other countries across the globe, especially during the initial phase following their implementation. To better assess observed virus spread trajectories following the implementation of social distancing measures, a more detailed understanding of the basic kinetics of virus spread in different countries is required. While the virus spread pattern is often quoted to be exponential [6, 7], power law patterns have been reported in data from China [13].

In this study, we compare the per capita virus spread kinetics observed for many countries around the globe, in order to obtain a better understanding of similarities and differences. Interestingly, we find that as per capita infections grow towards larger numbers, the growth pattern deviates from exponential and is better described by a power law. This can be hypothesized to be due to social distancing measures [5], or potentially to the build-up of immunity in the population [14]. We show, however, that the longer term per capita infection levels over time across a wide array of countries can be remarkably similar, and follow the same power law trajectory as seen in countries that are most strongly affected, such as Italy. This indicates that the long-term dynamics of COVID19 spread might be intrinsically governed by a power law, even in the absence of non-pharmaceutical interventions. We interpret these findings with computer simulations of a metapopulation model, which can account for an initial exponential spread phase, followed by a longer-term power law behavior, by assuming that the infection spreads in a well-mixed manner inside local demes, while also spreading spatially across different demes. We relate model predictions to epidemiological correlations found in the data. These insights into the spread kinetics of COVID19 might be useful for assessing the impact of non-pharmaceutical intervention methods in different countries.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2 Data sources

The data of confirmed COVID19 cases over time have been obtained from the data repository maintained by Johns Hopkins University Center for Systems Science and Engineering (CSSE) (`https://datahub.io/core/covid-19#data-cli`). As of March 28, 174 countries were represented in the database, as well as the cases on "Diamond Princess" (which were not used in the analysis). We only included the total counts for each country, even though information on the different provinces was available for several countries. The number of confirmed cases has been recorded since January 22, 2020, and has been updated daily.

To compare the time course of COVID19 cases across different locations, the per capita incidence was calculated, normalizing the numbers by the total population size of the country. The information on the population size and the area of different countries was taken from Wolfram Mathematica's database, "CountryData".

3 Comparing per capita infection levels over time in different countries

3.1 Per capita case numbers and time lags

Here, we present the comparison of the kinetics according to which cumulative COVID19 cases grow over time in different countries around the world. Figure 1(a) presents the raw data showing total case counts for a select number of key countries. Figure 1(b) shows the corresponding per million case counts.

A complication for comparing the growth dynamics is that the timing of the onset of community spread varies across locations. The growth curve of confirmed cases was therefore shifted in time to make them comparable, according to the following procedure. The cumulative confirmed COVID19 case counts in Italy were chosen to be the example against which the growth curves in all other countries were compared, due to Italy being a current epicenter of the outbreak. The (normalized, cases per million) infection growth curves of the other locations were shifted in time such that the difference between all data points of the country under consideration and Italy was minimized. The shift that minimized this Euclidean distance between the curves was assumed to indicate the number of days that the country lags behind Italy. Some examples of such results are presented in figure 2. We note that this assumes that all the countries test for COVID19 at comparable levels, which is an over-simplification. If a country tests less than Italy, it will lag behind Italy to a lesser extent. Conversely, if a country tests more than Italy,

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

it is predicted to be further behind Italy. Reliable data on the total number of tests in different countries are currently not available to our knowledge.

In this way, we obtained a time course of confirmed COVID19 cases that are temporally synchronized with Italy, which allows for a more straightforward comparison of the kinetics.

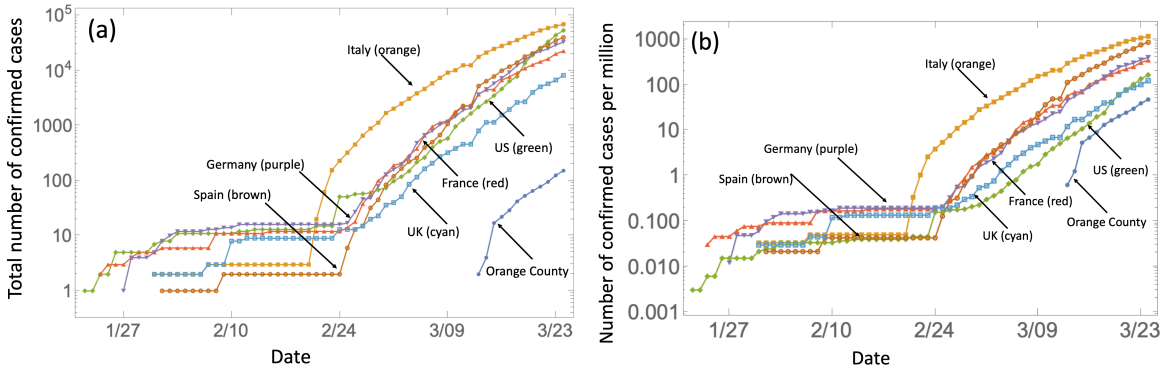


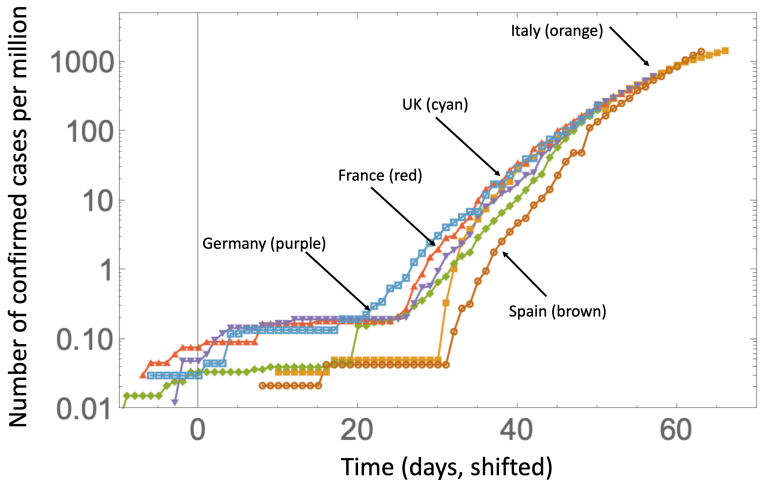
Figure 1: Example of the data. The number of confirmed cases is plotted as a function of time for 6 countries and Orange County: (a) the raw counts, (b) cases per million. The numbers of confirmed COVID19 cases in Orange County, the home of the authors, have been obtained from the daily updates provided by the website of the Orange County Health Care Agency (OCHCA).

3.2 Growth laws of the epidemic in different countries

Heterogeneity in growth laws across the different countries are observed. While some countries appear to show exponential growth of confirmed COVID19 cases, other countries appear to exhibit growth laws that deviate from exponential. Previous work [13] has suggested that a power law might be a better description of the cumulative COVID19 cases over time in China during the earlier stages of the epidemic (before it was controlled). Therefore, we hypothesized that for a subset of the countries that were analyzed, a power law is an appropriate description. To test this hypothesis, we fit both exponential and power law curves to the data for each country and determined the goodness of fit.

This data fitting was performed as follows: Only the data points were considered where the number of COVID19 cases had risen above a threshold, which we set at 1 case per million people (see Appendix for variations of this threshold). We fit both a power function and an exponential function to the data to determine which model fits the data better. For the power law function, a complication arose because fitting requires knowledge of the "zero" point, that is, the moment of time when the growth (according to the power law) began. The fits to the data change if the time scale is changed. Hence, we started by assuming the first data point to be the day when the infection frequency first exceeded 1 case per million, and fitted the power law, $a_1 x^{b_1}$, for some constants a_1 and

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



Country	Italy	US	UK	France	Germany	Spain	Orange County
Days behind Italy	0	15	16	10	9	4	20

Figure 2: The same data as in figure 1(bottom), presented by shifting individuals lines to match the Italy curve. The table shows the lag, that is, by how many days each country is behind Italy.

b_1. Then we shifted the time series incrementally by one day, and for each shift the power law was fitted. For each fitting frame, a different value of the power law exponent, b_1, was obtained. Subsequently, we fit an exponential function to the same data ($a_2 e^{b_2 x}$). The estimated exponent does not depend on the time shift, so fitting the exponential function was straightforward and yields a unique value b_2 for all the fitting frames. For both the exponential and the power law fits, we determined the sum squared error between observed and expected, and compared them. For more details of the fitting procedure, see Appendix A.

Figure 3 shows the fitting errors calculated for 75 countries; we included a country if the number of cases reached 20 per million, and excluded China and South Korea, since their epidemics clearly deviate from an exponential or a power law. The yellow horizontal lines represent the exponential fitting and the blue lines the power law fitting. We observe that there are several different configurations that are repeatedly encountered.

- For some countries (like the US, see also figure 4(a)), the power fitting error is always above the exponential fitting error. Such countries are clearly showing an exponential epidemic growth.
- There is another group of countries (such as Italy, see also figure 4(b)), where the power law fitting error is always below the exponential error; here we clearly have a power law growth.
- There are some other countries that we can classify as power law-like and exponential-

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

like. Suppose a power law error curve crosses the exponential error line (see Greece, figure 4(c)), at a given frame shift. In this case, we will classify the growth as power law-like if the value of the exponent b_1 that corresponds to this frame shift satisfies $b_1 < 5$. Otherwise, we will classify the growth law as exponential-like.

For the examples mentioned here, figure 4 shows the best fits obtained by this method. For (b) and (c) it is clear that the power law provides more satisfactory fits. More details are provided in Appendix.

Since in smaller countries (such as for example Luxembourg) the laws may be harder to determine and the data are subject to a higher degree of noise, for classification purposes we restricted the pool of countries to those with over a million inhabitants. Then the following country's epidemics were classified as exponential and as a power law, see also figure 5:

Exponential growth: Australia, Canada, Croatia, Israel, New Zealand, North Macedonia, Oman, United Arab Emirates, US, Austria, Dominican Republic, Ecuador, Ireland, Lithuania, Malaysia, Portugal, South Africa, United Kingdom.

Power law: Albania, Armenia, Belgium, Cyprus, Denmark, Georgia, Iran, Italy, Jordan, Mauritius, Moldova, Netherlands, Norway, Qatar, Slovakia, Slovenia, Sweden, Turkey, Uruguay, Bahrain, Bosnia and Herzegovina, Bulgaria, Chile, Costa Rica, Czechia, Estonia, Finland, France, Germany, Greece, Hungary, Kuwait, Latvia, Lebanon, Panama, Poland, Romania, Saudi Arabia, Serbia, Singapore, Spain, Switzerland, Trinidad and Tobago.

In the lists above, the countries that are classified as power law-like and exponential-like are printed in gray. There are 18 countries in the exponential law class, with 9 of them being truly exponential and the rest exponential-like. There are 43 countries in the power law class, with 20 of them being truly power law and the rest power law like.

3.3 Correlates of the growth law

It is interesting that the countries showing a power law (or power law-like behavior) are different with respect to some characteristics compared to the countries showing exponential or exponential-like behavior. Figure 6(a) shows a numerical probability distribution for the day when the infection in each country reached the level of 1 case per million. Blue represents the power law set and yellow the exponential set (grey means an overlap of the two colors). The average date of reaching 1 case per million (counting from Jan 22) is about 48 days for the power law and 52 days for the exponential set ($p = 0.035$ by T test). In other words, the countries that are demonstrating a power law infection spread have had the relative infection level slightly longer than the exponentially developing countries. This points us towards a hypothesis that perhaps it is typical to observe a transition between an early, exponential stage of growth, and a later, power-like stage of growth.

Figure 6(b) shows the difference between the two classes of countries in terms of

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

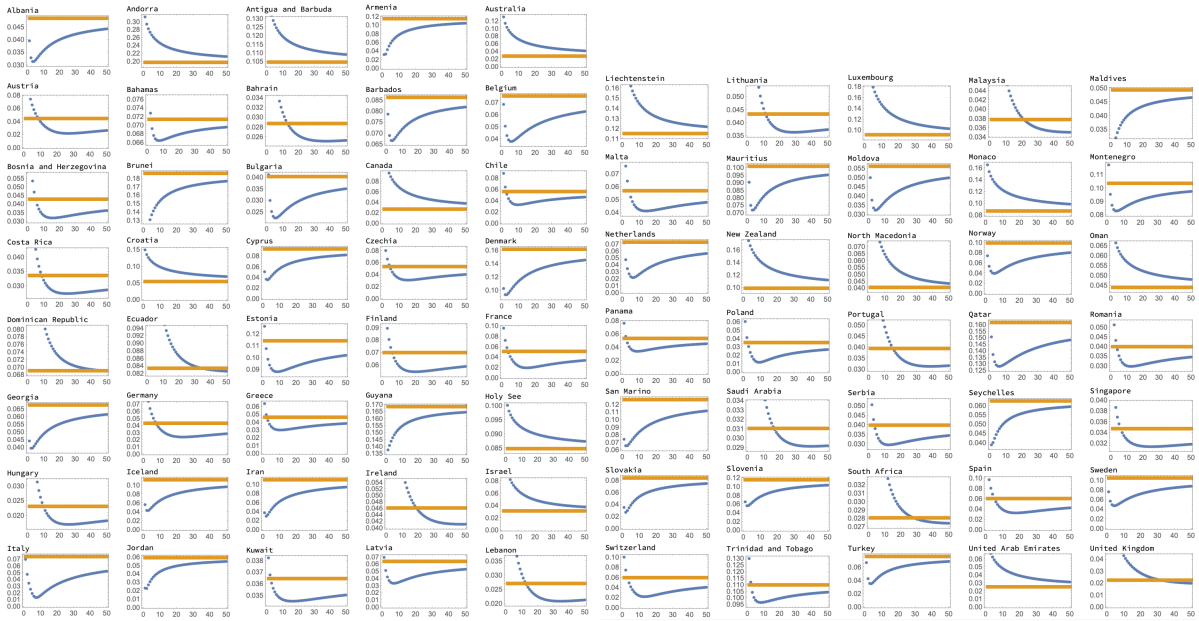


Figure 3: 75 countries' fitting results are presented as errors (blue for power law fits and yellow for exponential fits) as functions of the frame shift. Three distinct configurations can be observed: blue below yellow (a clear power law case), blue above yellow (a clear exponential case), and blue intersecting yellow. For such intermediate cases, we classified the growth as power-like if the power corresponding to the point of intersection corresponded to the power $b_1 < 5$. Otherwise it was classified as exponent-like.

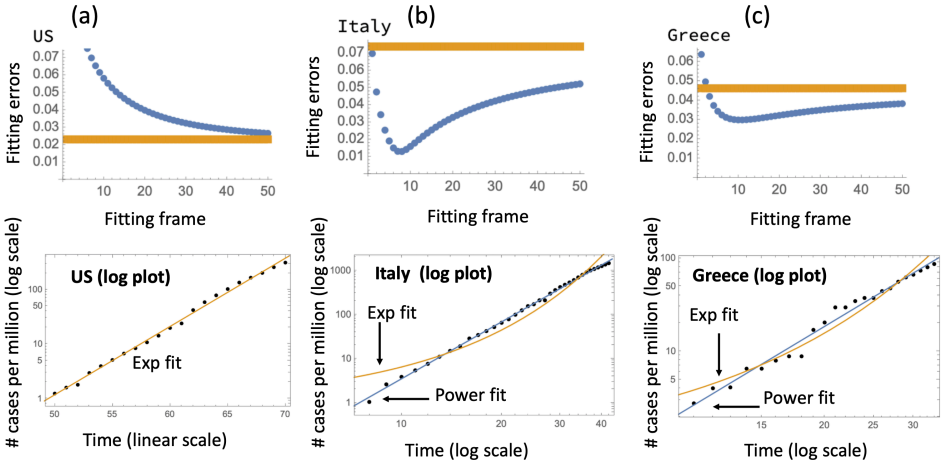


Figure 4: Examples of three error graph configurations. (a) US, exponential; a log plot of the data is presented with the exponential fit. (b) Italy, power-law; a log log plot of the data is presented with the best fitting power law and exponential fits. (c) Greece, exponential-like; as in (b), a log-log plot is presented.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

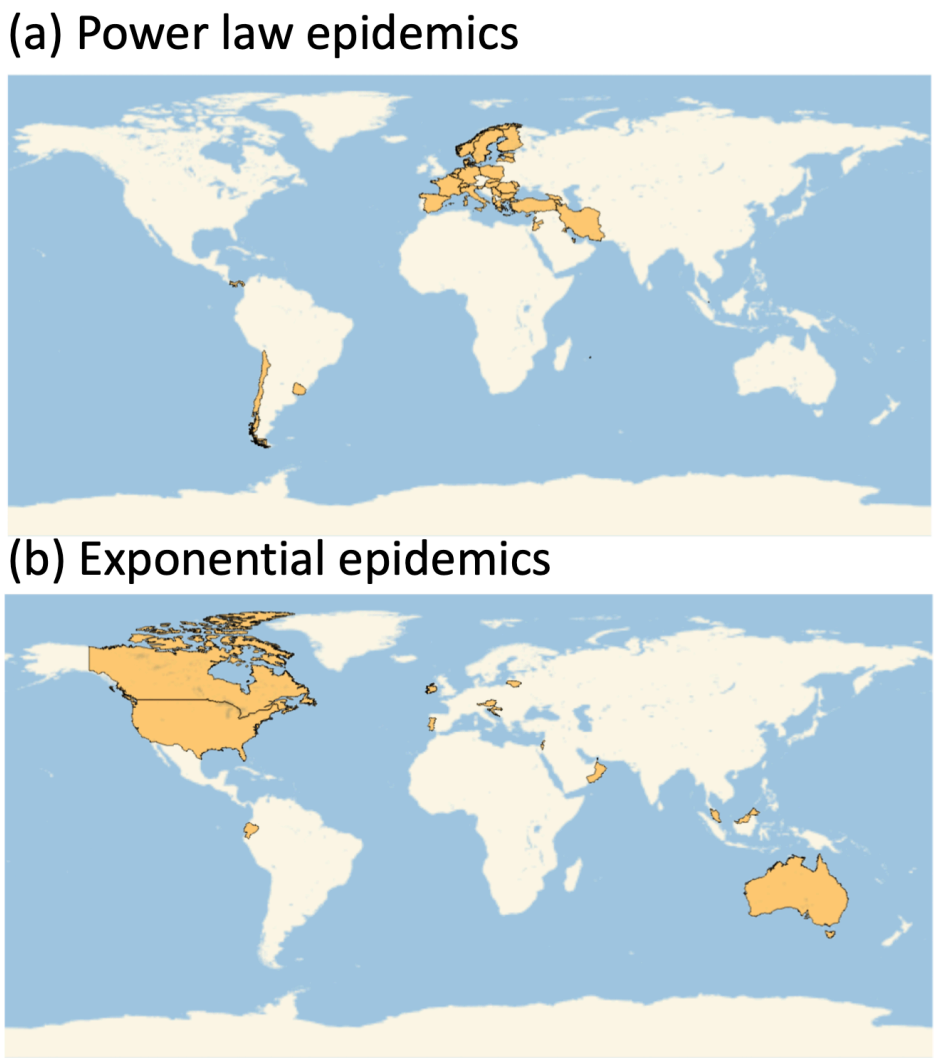


Figure 5: Geographic distributions of different epidemic growth laws. (a) Power law epidemics; (b) Exponential epidemics.

their area. We find that the exponentially growing infection class is associated with larger countries (mean area of about 1.7×10^6 km^2) compared to the power law class (mean area about 2.3×10^5 km^2, $p = 0.018$ by T test). Similarly, exponential epidemic spread tends to correlate with lower density countries (figure 6(c)). It is possible that it takes longer for a larger country of a lower density to transition to a power law growth. We provide a possible explanation of this correlation in the context of metapopulation modeling.

3.4 Possible explanations of the trends

It is important to understand the dynamics with which the cumulative case counts increase over time, such that we have a better ability to judge whether non-pharmaceutical

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

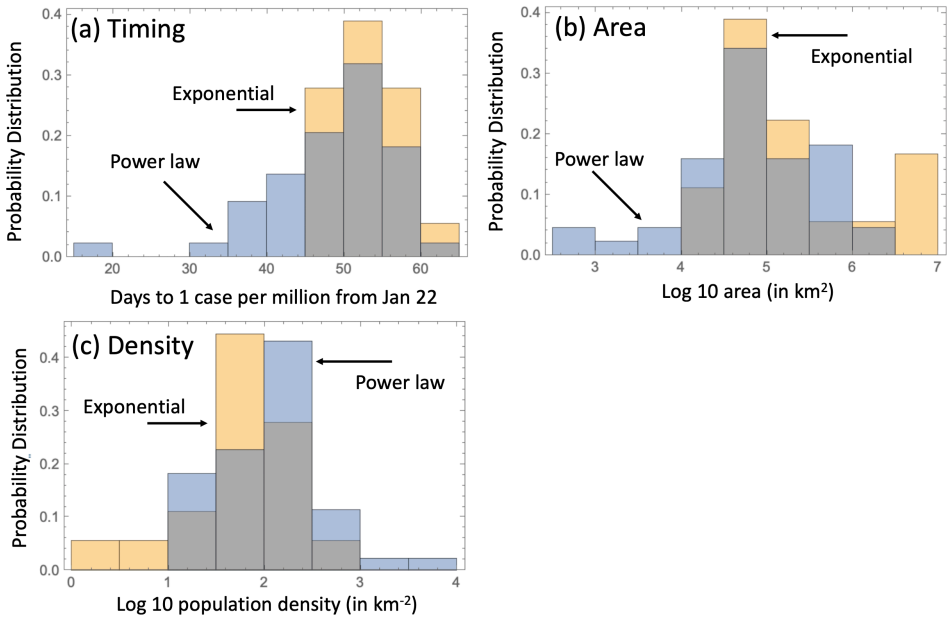


Figure 6: Comparison of the two classes of infection spread. (a) The timing of the infection: the distribution of time of reaching 1 case per million (counting from Jan 22), for the exponential and power law classes. The means are 48 days for the power law and 52 days for the exponential set ($p = 0.035$ by T Test). (b) Country size: the area of countries for the exponential and power law classes. The means are about 2.3×10^5 km^2 for the power law and 1.7×10^6 km^2 for the exponential, $p = 0.018$ by T test. (c) Country density: the means are about 374 people per km^2 for the power law and 98 people per km^2 for the exponential law, $p = 0.035$ by T test.

interventions (e.g. social distancing) have an impact on the course of the epidemic. The COVID19 epidemic is often thought to grow exponentially. If this is the case, a deviation from exponential growth following the introduction of non-pharmaceutical interventions can indicate success of those interventions. In contrast, if the infection grows according to a power law while growth is incorrectly assumed to be exponential, a slow down of the cumulative COVID19 cases over time on a log scale can result in the false conclusion that the non-pharmaceutical intervention methods are working. If the cumulative cases grow like a power law, successful intervention would result in the growth deviating from the power law, and not from exponential growth.

One interpretation of our analysis could be that in the absence of non-pharmaceutical interventions, the disease burden grows exponentially, but that over time, the non-pharmaceutical interventions slow down the spread. Hence, overall, the disease dynamics are described better by a power law than by an exponential function, because the power law is characterized by a slow-down of the growth rate over time. Italy, where one of the most pronounced power laws was observed, implemented some degrees of social distancing relatively early, although the more strict measures were implemented only recently.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

While this cannot be excluded, the comparative plots in figure 2 might argue against this hypothesis. These graphs shows the spread dynamics for a few countries, whose epidemics are relatively advanced, and they have all been shifted to be temporally synchronized with Italy. As the infection spreads to higher levels, the curves for the different countries converge towards a common trajectory of the per capita cases of COVID19. In other words, as the infection spreads, the per capita number of cases converge and follow the same power law. If this deviation from exponential growth was due to non-pharmaceutical interventions, then it is surprising that those countries follow the same trajectory because they implemented those interventions at different times and to different degrees.

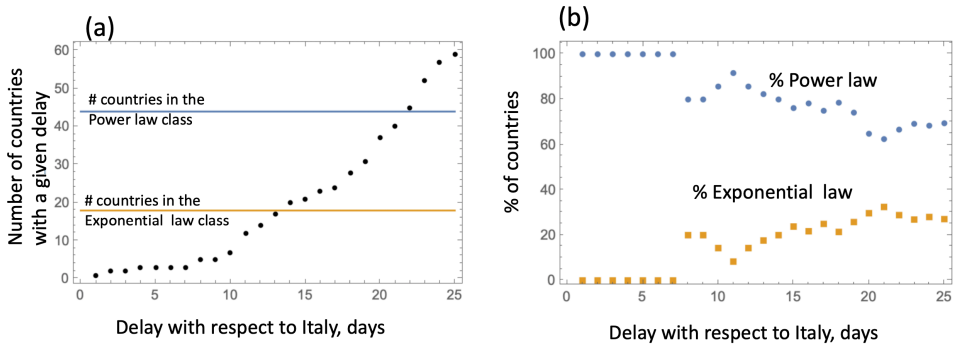


Figure 7: Temporal development of infection. The "age" of the epidemic is measured by the days of delay with respect to Italy. (a) The number of countries for each value of the time-lag with respect to Italy. The number of countries in the two growth law classes is shown for comparison. (b) The percentage of countries with a given delay that belong to the Power law group and to the Exponential law group. The trend that the percentage of exponential growth increases with the time lag (that is, decreases with the epidemic "age") is significant ($p < 10^{-4}$ by linear fitting).

Alternatively, it is possible that the local epidemics characterized by power laws demonstrate an intrinsic power law spread that is independent of interventions. Computational models suggest that infection spread across networks or in spatially structured populations can lead to dynamics that follow a power law rather than an exponential trajectory, see below.

If some countries truly show infection spread that is governed by a power law, the question arises why other countries show clear exponential spread, and why yet other countries are more difficult to classify. We hypothesize that different countries are at different stages of epidemic development, but they all roughly follow the same trajectory, where an initial exponential growth is gradually replaced by a more power like behavior. Figure 7 demonstrates evidence in favor of this theory. Panel (a) plots the number of countries classified by the number of days they are delayed with respect to Italy. As explained in section 3.1 (see also figure 2), we shifted the growth trajectories of all coun-

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

tries until, for each country, the best match with the Italian curve was obtained. As we can see in figure 7(a), there are only a few countries that are just behind Italy, and as the number of lag days increases, the number of countries grows. This corresponds to the world wide spread of the pandemic: more and more countries become affected as time goes by. Figure 7(b) calculates, for each lag time, the percentage of countries that were classified as following power law or exponential dynamics. We can see that for the countries that are just a few days behind Italy, 100% of them belong to the power law group. As the lag time increases, indicating an earlier stage of the epidemic, more and more countries exhibit exponential growth ($p < 10^{-4}$). The same trend is also corroborated by figure 6(a) that shows that the epidemics in the exponential group are "younger".

How can we explain the existence, to different extents, of power law like behavior in different countries? One possible reason could be non-pharmaceutical interventions that are only partially effective, see a schematic in figure 8. It is possible that in early stages of the infection, exponential virus spread occurs because people have not yet altered their behavior and continue to travel and socially mix with each other. Mass-action dynamics are expected to yield exponential growth. As a result of non-pharmaceutical interventions that are only partially effective, people might stop traveling and thus slow down large scale mixing. However, they would still be interacting locally within their social network, which would lead to a transition to power law dynamics. It is possible that a number of countries are difficult to classify because they exhibit early mass action dynamics (exponential), followed by network interactions that lead to a transition to a power law. This would indicate that stronger non-pharmaceutical intervention methods need to be implemented.

More generally, the results can be interpreted in the context of a minimally parameterized metapopulation model, see figure 8. Assume that within a local deme (such as a local community), people interact with each other, resulting in mass action dynamics. For the infection to spread further, however, people have to enter other demes, and seed the infection there. We have performed computer simulations of such a model to explore outcomes. The model is a two-dimensional metapopulation consisting of $N \times N$ patches. In each patch, i, the infection dynamics are given by a set of ordinary differential equations (ODEs) that take into account the population of susceptible (S_i), infected (I_i), recovered (R_i), and dead (D_i) individuals:

$$\frac{dS_i}{dt} = -\frac{\beta S_i I_i}{S_i + I_i + \epsilon} + f \sum_{j \in B_i^n} (S_j - S_i), \tag{1}$$

$$\frac{dI_i}{dt} = \frac{\beta S_i I_i}{S_i + I_i + \epsilon} - g I_i + f \sum_{j \in B_i^n} (I_j - I_i), \tag{2}$$

$$\frac{dR_i}{dt} = g(1-a) I_i + f \sum_{j \in B_i^n} (R_j - R_i), \tag{3}$$

$$\frac{dD_i}{dt} = g a I_i. \tag{4}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

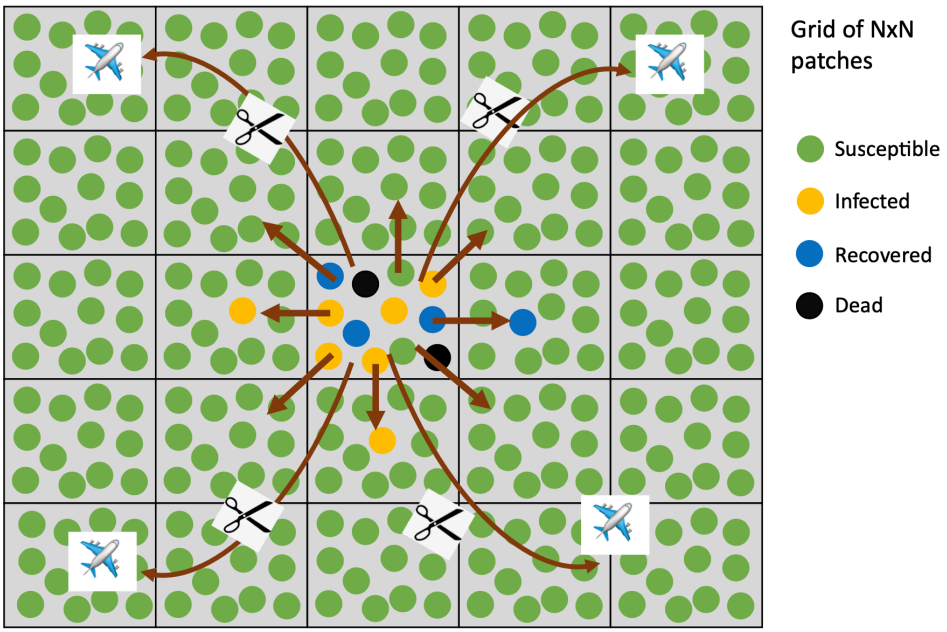


Figure 8: The concept of partial social distancing measures and the metapopulation model. There is a grid of $N \times N$ patches. Within each patch (which represents a local community), deterministic SIR dynamics are assumed (complete mixing). Infection can also spread by contact (mixing) with neighboring patches (demes). Global infection transfer is also possible, e.g. by air travel within the country and outside, but this is disrupted by partial social distancing measures. Equations (1-4) correspond to the situation where long-haul interactions are not prresent. This is what we implemented by simulations.

Here, infection is described by a frequency-dependent infection term [15], characterized but the rate constant β and a saturation constant ϵ. Infected individuals die with a rate ga and recover with a rate $g(1-a)$. The migration terms include the outward migration to n neighbors and an inward migration from all the n neighboring demes that belong to neighborhood B_i^n of deme i. The migration rate is denoted by f and we assume that each patch has eight direct neighbors, i.e. $n = 8$. The boundary demes are characterize by fewer inward/outward migrations (that is, they have smaller neighborhood sets).

Using this model, we track the predicted dynamics for $I + R + D$ over time, which represents the cumulative infection case counts. In a first scenario, we start the computer simulations with a small amount of infected individuals in a single patch, located in the center of the grid. All other patches contain only susceptible individuals. The resulting dynamics are shown in figure 9. We observe an initial exponential phase of infection spread, followed by a transition to a power-law spread. The spread is initially exponential, because within a single patch (the first patch), the dynamics are governed by well mixed populations. As the infection spreads to other patches by migration, the overall infection spread starts to be governed by spatial dynamics, which explains the

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

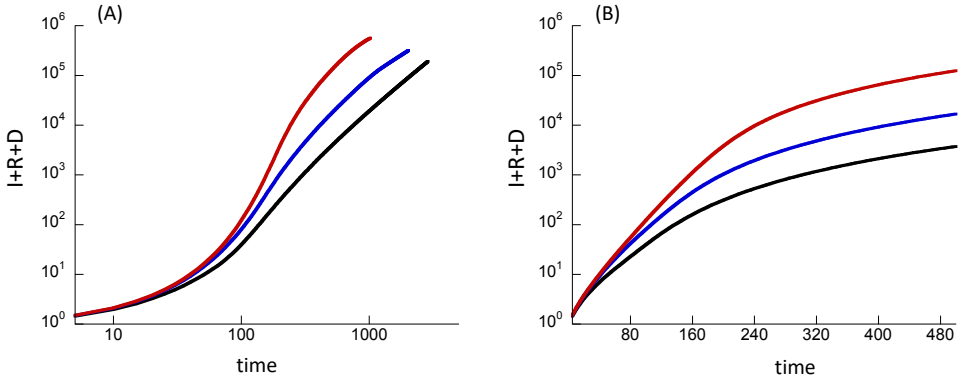


Figure 9: Results from implementing the metapopulation model, equations (1-4). The total number of cases (given by $I+R+D$) is plotted as a function of time, on a log log scale (a) and a log scale (b). The black line shows the dynamics where the simulation starts with 1/10 of individuals infected in a single patch in the middle. The blue line corresponds to the initial condition where 1/50 of the individuals are infected in 5 randomly chosen patches. The red line shows the consequence of 1/500 of the individuals infected in 50 randomly chosen patches. The rest of the parameters are $S = 10, R = 0, D = 0$ initially in all patches, $\beta = 0.1, g = 0.05, f = 0.001, a = 0.01, \epsilon = 1, N = 300$.

transition to the power law behavior. The key is the difference between the time scale of local spread and the time-scale of global mixing.

Next we assumed that instead of starting with infecteds only being present in a single patch, a small amount of infected individuals are initially present in more than one patch around the same time. This could correspond to larger countries, in which the infection is simultaneously seeded in multiple areas (e.g. due to travel from other places). Now, we observe overall growth dynamics that are are more exponential-like. The length of the predicted exponential phase becomes longer the more patches are initially seeded. The reason is that with more initial seeding events, the importance of spatial spread is de-emphasized. The metapopulation model can therefore predict an array of growth patterns where an exponential phase of varying length is followed by a transition to power law, depending on the initial conditions of the simulation.

4 Discussion and Conclusions

In this paper, we analyzed data that document the cumulative COVID19 case counts over time in a large number of countries around the world, and examined the laws according to which the infection spreads. This suggests that although the initial phase of the spread may be exponential, the longer term dynamics tend to be governed by a power law. The analysis indicates that the countries that display clear evidence for exponential growth are currently in a relatively early phase of the epidemic. The data further suggest that countries that are further along in the epidemic converge to a common power law

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

behavior, and cumulative per capita case counts appear to converge over time. These observations were interpreted by computer simulations of a metapopulation model that takes into account both local spread and spread across geographical space. This model predicts an initial exponential phase (due to local transmission events driving the dynamics), followed by a transition to a power law (once spatial dynamics significantly drive spread). The duration of the exponential phase is determined by the number of patches that are initially seeded with the infection. If the infection originates in a single location (patch), the exponential phase is likely not very pronounced, and most of the growth curve is predicted to follow a power law. If the infection is seeded simultaneously or nearly simultaneously in multiple locations, the duration of the exponential phase becomes longer. The larger the number of initially seeded locations, the more pronounced the exponential phase of the infection spread.

These predictions are further in agreement with the correlations that we found between the growth law of the infection and country size. Exponential growth was associated with countries that are characterized by a larger area. More pronounced power laws tend to occur in counties with a smaller area. If a country has a larger area, it is more likely that multiple locations are seeded with the infection around a similar time, for example due to travelers returning from a country with a larger disease incidence. For this scenario, the metapopulation model predicts more pronounced exponential growth. If the country has a smaller area, it is more likely that the infection is seeded within one geographical area and spreads outward from there. In this case, the metapopulation predicts an infection spread pattern that mostly follows a power law.

Beyond scientific interest, a better understanding of the laws according to which COVID19 spreads through populations is also of practical importance. Currently, emphasis is placed on non-pharmaceutical intervention measures to slow down the spread of the infection such that the ability of the health care system to cope with the number of incoming patients is preserved. The success of these intervention measures should be reflected in slowed infection spread. An understanding of the infection spread laws is crucial to interpret such data. If we assume that infection growth is exponential and if we plot the cumulative number of COVID19 cases on a log scale, a deviation and slow-down from exponential growth would lead to the conclusions that the non-pharmaceutical intervention measures are successful [16]. If, however, the true spread dynamics are characterized by a power law, we expect deviation and slow-down compared to exponential growth over time, even if the infection continues to spread at full force. In this case, a deviation from exponential spread cannot be interpreted to mean that non-pharmaceutical intervention measures are sufficient. To come to that conclusion, we would need to observe a deviation and slow-down of the infection spread compared to a power law null model.

As with any data and modeling studies, it is important to note that results can depend on assumptions and methodologies. These are clearly spelt our here. One of the larger challenges we faced in the data analysis is the lack of knowledge at what time the infection was initiated in the individual countries. This information is not available. The time frame in turn influences the fit of the power law to the data, which we have addressed

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

with our time shifting methodology. If further information becomes available about the time when infections are estimated to have originated in the individual countries, the methodology can be updated. Genetic studies could provide valuable data in this respect.

Another limitation of the data interpretation is the degree, to which different countries test for the coronavirus. If some countries test less than others, they will appear to be further behind Italy, while in reality the lag could be shorter. This type of uncertainty however does not change the central finding that the long term dynamics of COVID19 cases in different countries follow a power law, after an initial stage of exponential growth.

Acknowledgements. Support of grant NSF DMS 1662146/1662096 is gratefully acknowledged.

References

[1] Velavan TP, Meyer CG. The COVID-19 epidemic. Trop Med Int Health. 2020;25(3):278–280.

[2] Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. Journal of Autoimmunity. 2020;p. 102433.

[3] Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet. 2020;.

[4] Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al.. Fair allocation of scarce medical resources in the time of Covid-19. Mass Medical Soc; 2020.

[5] Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College, London DOI: https://doi org/1025561/77482. 2020;.

[6] Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? The Lancet. 2020;.

[7] Li Y, Liang M, Yin X, Liu X, Hao M, Hu Z, et al. COVID-19 Epidemic Outside China: 34 Founders and Exponential Growth. medRxiv. 2020;.

[8] Hilton J, Keeling MJ. Estimation of country-level basic reproductive ratios for novel Coronavirus (COVID-19) using synthetic contact matrices. medRxiv. 2020;.

[9] Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. Journal of travel medicine. 2020;.

[10] Choi SC, Ki M. Estimating the reproductive number and the outbreak size of Novel Coronavirus disease (COVID-19) using mathematical model in Republic of Korea. Epidemiology and Health. 2020;p. e2020011.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[11] Shao N, Cheng J, Chen W. The reproductive number R0 of COVID-19 based on estimate of a statistical time delay dynamical system. medRxiv. 2020;.

[12] Zhang S, Wang Z, Chang R, Wang H, Xu C, Yu X, et al. COVID-19 containment: China provides important lessons for global response. Frontiers of Medicine. 2020;p. 1–5.

[13] Li M, Chen J, Deng Y. Scaling features in the spreading of COVID-19. arXiv preprint arXiv:200209199. 2020;.

[14] Cookson C. Coronavirus may have infected half of UK population. Financial Times, March 24. 2020;.

[15] McCallum H, Barlow N, Hone J. How should pathogen transmission be modelled? Trends in ecology & evolution. 2001;16(6):295–300.

[16] Maier BF, Brockmann D. Effective containment explains sub-exponential growth in confirmed cases of recent COVID-19 outbreak in Mainland China. arXiv preprint arXiv:200207572. 2020;.

A Details of the fitting

Here we present the details of the fitting procedure used to determine the growth laws for different countries. It is illustrated with the example of Italy in figure 10. The full data for the number of cases per million in Italy are presented in figure 1(b), orange curve. In figure 10, the subset of same data starting from 1 case per million, is plotted on the log-log scale (panel (a)) and on a linear scale (panel (b)), with varying horizontal shift, which corresponds to changing the position of time zero. This is what we refer to as a "fitting frame". The fitting frame number one is when the first data point in the selected subset corresponds to day 1. The ith fitting frame shifts this point to day i. For each fitting frame, we obtained the best fit with function $a_1 x^{b_1}$, by using a built-in Mathematica routine "FindFit". Note that the natural logarithm of the data was fitted for the figures presented here.[1]

Clearly, some fits are better than others, see figure 10(a,b). The fitting error for each fitting frame, i, was calculated as the distance between the data and the fit, normalized by the number of points:

$$\mathrm{Err}_i = \frac{1}{N}\sqrt{\sum_{j=1}^{N}(\ln y_j^{(i)} - \ln x_j^{(i)})^2},$$

where $y_j^{(i)}$ is the jth component of the dataset in frame i, $x_j^{(i)}$ is the corresponding prediction of the fit and N is the number of points in the dataset. These errors are

[1] We also tried fitting the function without taking the logarithm; similar results were obtained (not shown).

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

shown for the different fitting frames in panel (c) of figure 10. The minimum error corresponds to the power $b_1 = 4.3$.

Exponential fits of the same data from Italy are shown in figure 11(a,b). In panel (a) we used a log scale, such that the exponential fits look like straight lines. It is clear that, first of all, these fits are all parallel lines and thus the error is exactly the same (thus the exponential fitting errors as functions of the fitting frame are horizontal lines, see figures 3 and 4). Second, we note that these fits are not very good for Italy, that's why the power fit errors is always below the exponential error, see figure 4(b). For comparison, panels (c) and (d) of figure 11 show the exponential fits for the US data. We can see that the quality of the fits is better, see also figure 4(b).

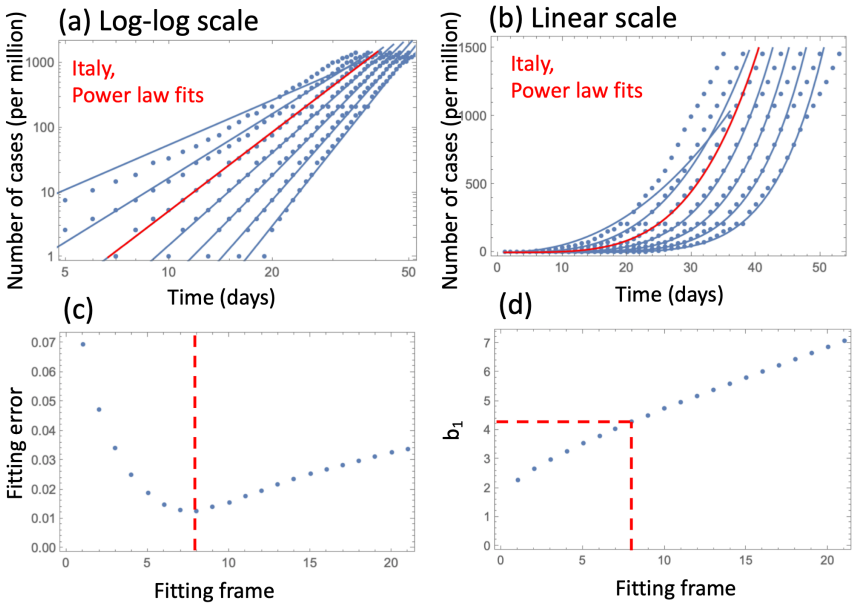


Figure 10: Details of the fitting procedure, using the example of Italy. Power law fits are presented for 7 choices of the fitting frame from 1 to 21, using (a) the log-log scale and (b) the linear scale. The data are plotted as dots and the fits as lines. The best fit is marked by the red line. (c) The power law fitting errors as a function of the frame. (d) The fitted value of the power exponent, b_1, as a function of the fit. The red dashed lines in (c) and (d) mark the best fit.

In the main text and in this Appendix so far, we describe a fitting procedure where the "confirmed cases" data for each country were used only if the numbers exceeded 1 case per million. Here we demonstrate how this changes if a different choice is made and a minimum of 5 cases per million is required for each data point to be included. In figure 12 we demonstrate the difference for 30 countries. The choice of countries for this graphics was somewhat arbitrary: we included the 30 countries out of the subset used in figure 3 that had the largest infection (cases per million).

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

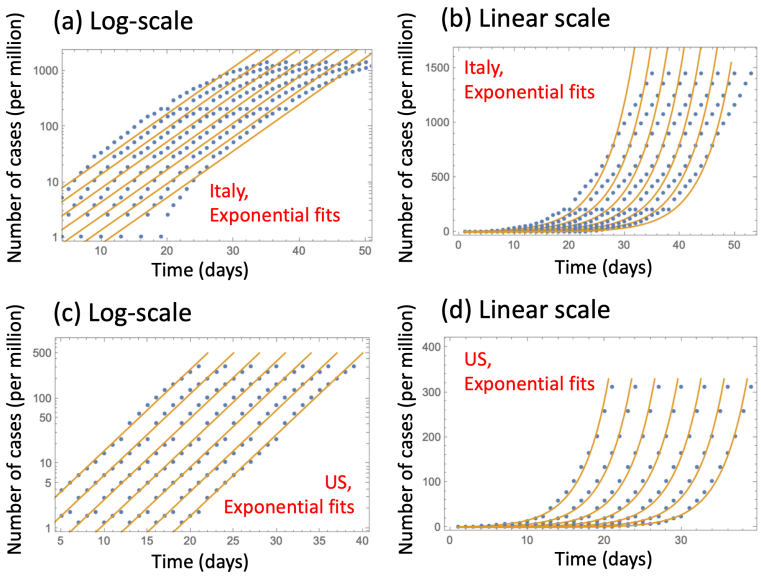


Figure 11: Examples of exponential fits for Italy (a,b) and the US (c,d), where the log scale is used in (a,c) and the linear scale in (b,d). The yellow lines show exponential fits for 7 different frames from 1 to 21, as in figure 10(a,b).

B Best fits for different countries

Here we present plots of the best fits for different classes of countries. Figure 13 shows the 19 countries that were classified as a power law countries. The plots are presented on a log-log scale, such that the power law fits are straight lines. We can see that the best power law fit (blue) is a visibly better match than the exponential fit (yellow). Note that for all of these countries the power law fitting error for any frame shift is smaller than that obtained by the exponential fitting. The rest of the power law countries (those that were classified as power law like) are shown in figure 14. This list contains 24 countries. For convenience, we present both a log log plot (such that the power law fits, blue, appear as straight lines) and a log plot (such that the exponential fits, yellow, appear as straight lines).

Figure 15 shows the 9 countries that are characterized by a straight exponential growth. For these data, we used a log scale, such that the exponential fits are straight lines. For all of these countries the power law fitting error for any frame shift is larger than that obtained by the exponential fitting. Note however that power fits that are almost as good as exponential fits can always be found, if we we shift the frame far enough. These fits correspond to very large values of the power coefficient b_1 in the power law, see for example figure 4(a) which presents the example of the US. As the fitting frame index increases, the power law fitting error (top, blue line) approaches the exponential fitting error (horizontal yellow line). This, however, is meaningless, and does not indicate the presence of a power law. Figure 16 presents the rest of the countries from the exponential class, that is, those that were classified as exponential-like.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

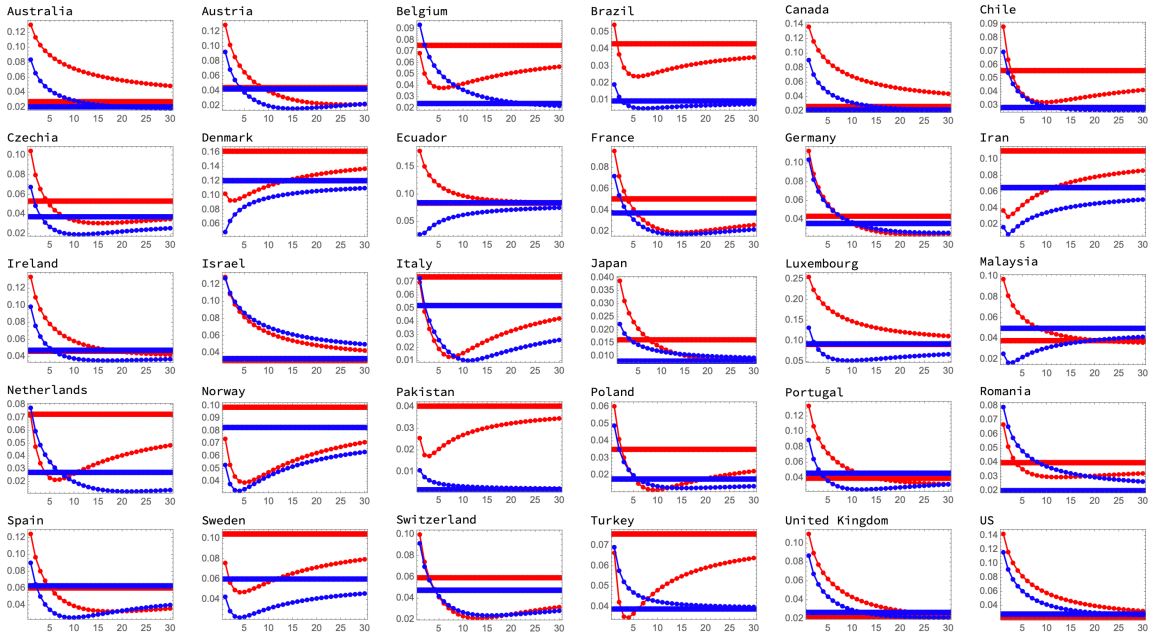


Figure 12: Comparison of two different choices of the fitting procedure, for 30 countries. Each panel represents a country, with the hrizontal axes being the fitting frame, and the vertical the fitting error. Red symbols correspond to the 1 case per million threshold, and blue circles to the 5 cases per million threshold. Circles represent the error of the power law fitting; they form non-constant functions. Squares represent the error of the exponential fitting and form horizontal lines, because these errors do not depend on the fitting frame.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

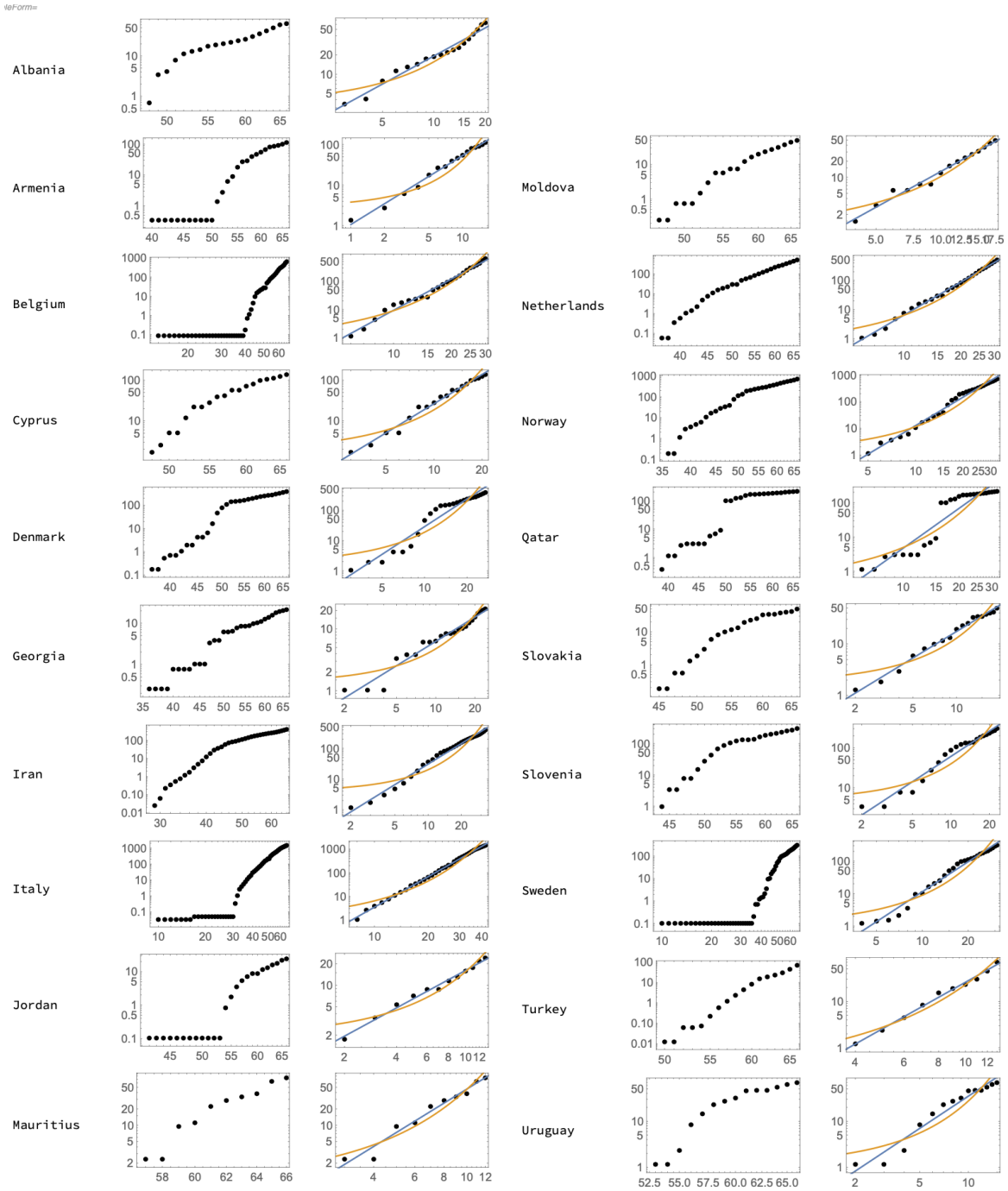


Figure 13: The 19 countries that were classified as those following a power law. For each country, two panels are presented. One is the full data (cases per million) plotted on a log log scale. The other is the subset of data (with 1 or more cases per mission) plotted on a log-log scale (black circles) together with the best power law (blue line) and exponential (yellow line) fits.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

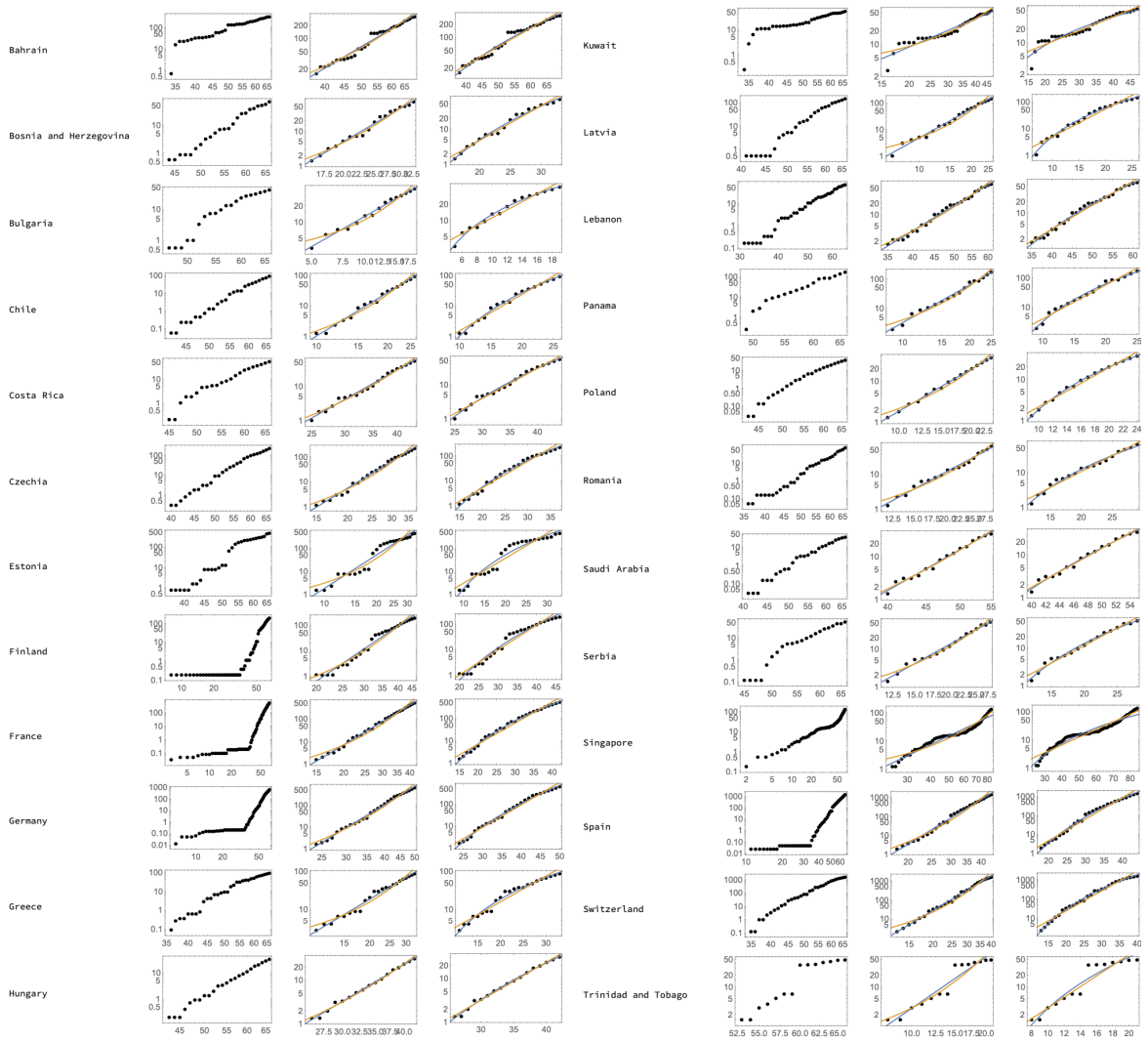


Figure 14: The 24 countries that were classified as power law like. For each country, three panels are presented: (1) is the full data (cases per million) plotted on a log log scale. (2) is the subset of data (with 1 or more cases per mission) plotted on a log-log scale (black circles) together with the best power law (blue line) and exponential (yellow line) fits. (3) is the same as (2) except on a log scale.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20047274; this version posted April 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

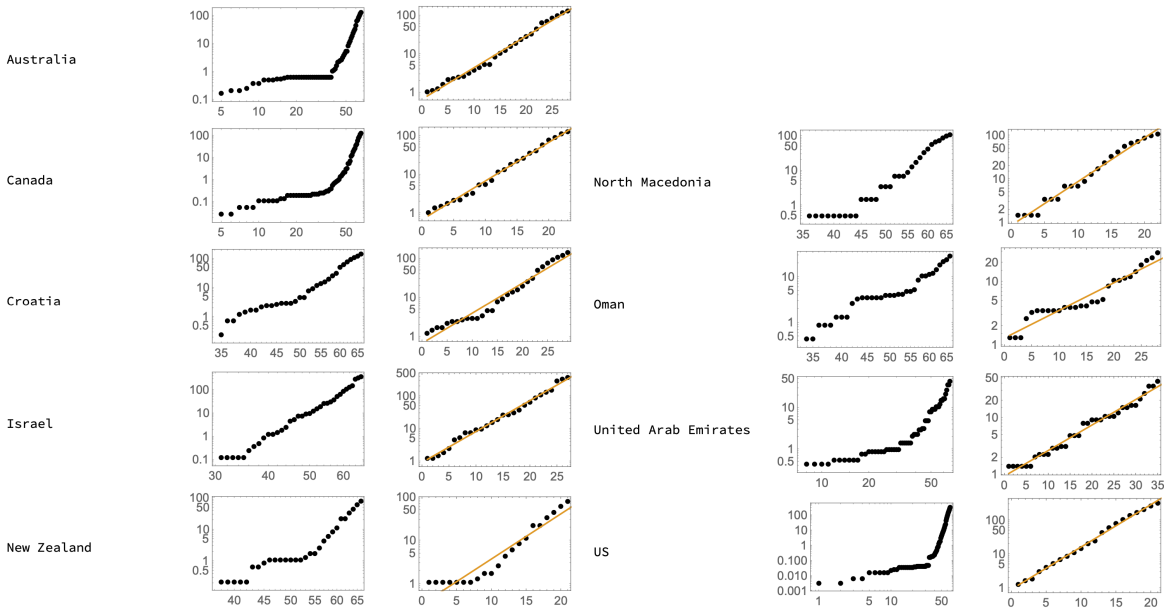


Figure 15: The 9 countries that were classified as those following an exponential law. For each country, two panels are presented. One is the full data (cases per million) plotted on a log log scale. The other is the subset of data (with 1 or more cases per mission) plotted on a log scale (black circles) together with the exponential fit (yellow line). Note that for these countries, the power law fits correspond to very high values of the exponent and are therefore not significantly different from the exponential fits.

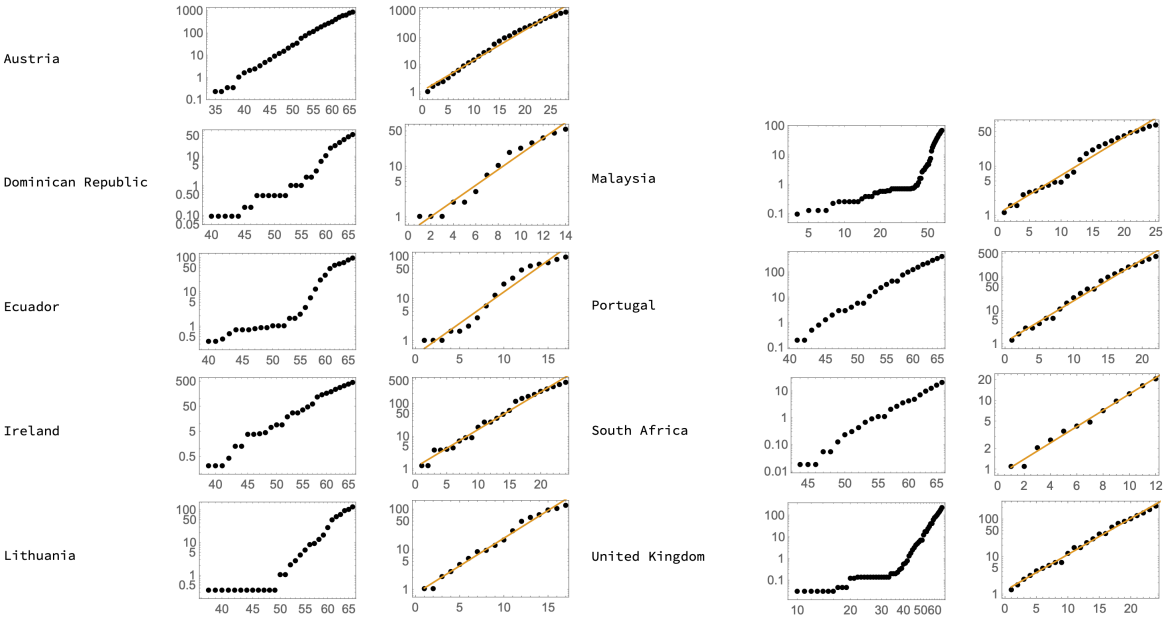


Figure 16: The 9 countries that were classified as exponential-like. Panels are as in figure 15.