Article

A Parsimonious Description and Cross-Country Analysis of COVID-19 Epidemic Curves

Kristoffer Rypdal * and Martin Rypdal [ID]

Department of Mathematics and Statistics, UiT—The Arctic University of Norway, 9019 Tromso, Norway;
martin.rypdal@uit.no
* Correspondence: kristoffer.rypdal@uit.no

check for updates

Abstract: In a given country, the cumulative death toll of the first wave of the COVID-19 epidemic follows a sigmoid curve as a function of time. In most cases, the curve is well described by the Gompertz function, which is characterized by two essential parameters, the initial growth rate and the decay rate as the first epidemic wave subsides. These parameters are determined by socioeconomic factors and the countermeasures to halt the epidemic. The Gompertz model implies that the total death toll depends exponentially, and hence very sensitively, on the ratio between these rates. The remarkably different epidemic curves for the first epidemic wave in Sweden and Norway and many other countries are classified and discussed in this framework, and their usefulness for the planning of mitigation strategies is discussed.

Keywords: COVID-19; epidemic curve; logistic curve; Gompertz model; death toll

1. Introduction

At the time of writing, the COVID-19 pandemic is expanding exponentially in many countries, particularly in the tropics and the southern hemisphere. On the other hand, in Southeast Asia, Europe, and North America, the first wave of infection is mostly over. In some countries, e.g., the United States, Israel, Iran, and Spain, a second and stronger wave develops. For the first wave, the general shape of the epidemic curves, the daily number of confirmed cases or COVID-19 related deaths, is one of rapid growth followed by a slower decay. However, even though this general characteristic is ubiquitous, the total death toll per million inhabitants in comparable countries varies by more than an order of magnitude. For instance, by 5 August 2020, Sweden had registered 569 deaths per million, while neighboring Norway only 47.

The debate about the causes of this pronounced variability between countries has involved a plethora of explanations [1] based on correlation techniques such as multiple regression [2], or on models that involve statistical modelling of the dynamics of transmission [3]. Suggested predictors (explaining variables) in such analyses have encompassed demographic, socioeconomic, and health-related characteristics, and strength and timing of border closures and social distancing measures. Some success in assessing the effect of non-pharmaceutical interventions in 11 European countries was obtained in [3], and the effect of social distancing measures in Sweden, Denmark and Norway was studied by a considerably simpler approach in [4]. The three countries are similar in terms of health care, language, culture, climate, economy, and institutional framework. Moreover, due to the simultaneous import of infected cases via ski tourists returning from Austria and Italy in early March 2020, it is plausible that community spread of the disease started at approximately the same time in these countries. On the other hand, while Norway and Denmark followed the general lockdown of most of the European Union countries on 12 March 2020, Sweden followed a different path, largely using voluntary measures [5].

Int. J. Environ. Res. Public Health **2020**, *17*, 6487

2 of 23

On a world scale, cross-country comparative analysis is extremely challenging for at least two reasons. One is that the relevant predictors vary across the world and may attain an unmanageable number. Another is that the same problem could pertain to the predictands (explained variable). How do we best describe the impacts of the disease in a way that allows comparison between countries in different parts of the world? Some studies employ regression analysis to a sample containing the majority of the world's countries, but consider the accumulated COVID-19 cases and/or deaths up to a certain date as predictand [6]. This variable may be very misleading, however, since the virus was not introduced simultaneously in all countries. In principle, we need to predict and compare the full epidemic curves across countries, but in order to extract information from from such comparison, we need a simplified characterization of these curves, i.e., we need to reduce the information in the curves to a few numbers. This paper is about how we can obtain and make use of such a characterization.

Our main finding is that the epidemic curves for COVID-19 related deaths for most countries with a reliable reporting system are surprisingly well described by the so-called Gompertz growth model [7]. This model contains only two essential free parameters to be determined by the data. These parameters determine the initial relative growth rate of the number of daily deaths and the decay rate of this number as the first epidemic wave comes to an end. The mathematical properties of the Gompertz model imply that the total death toll in the first wave is given by the exponential of the ratio between these two rates. Countries with rapid initial growth and slow later decay suffer the higher death toll, but there are variations among countries with a similar number of deaths. For instance, while Spain experienced very fast initial growth followed by a rapid decay due to a very forceful lockdown, Swedish death numbers rose and decayed considerably more slowly. Other countries, like Norway, experienced slower initial growth than Sweden, and faster decay, and thereby obtained death numbers one order of magnitude lower. Since the initial growth rate in Norway was so much lower than in Spain, Norway could obtain these results with slower decay, i.e., with a considerably milder lockdown, than Spain.

In this paper, we make a detailed comparison of epidemic curves for Sweden and Norway, including computation and plotting the respective time evolution of the relative growth rates and time-dependent reproduction numbers. We further fit the Gompertz model to the epidemic curves (deaths) for a large number of countries that have completed the first wave, and plot the model parameters in a diagram highlighting their differences. The two rates are predictands that correlate in different ways to various predictors of the epidemic curve, and via these predictands they determine the total death toll in the first wave. Regressions that determine the coefficients of these predictors will be useful in the design of optimal mitigation strategies for upcoming waves of the epidemic.

2. Materials and Methods

2.1. Data

Data employed in this paper are downloaded from Our World in Data [8]. Their data for COVID-19 related deaths are retrieved from the European Centre for Disease Prevention and Control. As explained by the COVID-19 Health System Response Monitor [9], figures may vary among countries and may complicate cross-country analysis. This problem is most serious for the headline figures presenting the most recent day-to-day data. We use data published at the end of July, but only data up to the first week of July. In Section 4 we will discuss the implications of such uncertainties for the conclusions of the analysis.

The official figures from China suffer from 40% discontinuous increase in death numbers on 17 April. The official explanation is that home deaths before that date were included. In our analysis, we account for this by adjusting the numbers before 17 April by a factor of 1.4, thus removing the discontinuity.

Int. J. Environ. Res. Public Health **2020**, *17*, 6487

3 of 23

2.2. The Gompertz Model

The Gompertz model is a special case of the class of logistic growth models [10]. This general class are solutions $J(t)$ to the nonlinear, separable differential equation

$$\frac{dJ}{dt} = \gamma(J)\, J, \tag{1}$$

where t is the time coordinate and J is a quantity undergoing monotonic growth that saturates at the carrying capacity J_∞ in the limit $t \to \infty$. In the Gompertz model the instantaneous relative growth rate $\gamma(J) = d_t J/J = d_t(\ln J)$ is (see Appendix A),

$$\gamma(J) = \gamma_\infty \ln\left(\frac{J_\infty}{J}\right), \tag{2}$$

and Equation (1) can be integrated to yield

$$\ln J(t) = \ln J_\infty - \ln\left(\frac{J_\infty}{J_0}\right) \exp\left(-\gamma_\infty t\right), \tag{3}$$

or equivalently,

$$J(t) = J_\infty \left(\frac{J_0}{J_\infty}\right)^{\exp(-\gamma_\infty t)}. \tag{4}$$

Note that in this limit $\gamma(J)$ diverges as $J \to 0$, so Equation (1) should be understood as an initial value problem with $J(0) = J_0$ and an initial growth rate

$$\gamma_0 = \gamma_\infty \ln\left(\frac{J_\infty}{J_0}\right). \tag{5}$$

The logarithmic form of Equation (3) shows that the Gompertz model describes a growth where the logarithm of J converges exponentially towards the limit $\ln J_\infty$, i.e., the difference $J_\infty - J$ decays with the decay rate γ_∞. The growth curve is completely determined by the initial value J_0, the carrying capacity J_∞, and the asymptotic decay rate γ_∞. These features make it natural to plot $J(t)$ in a logarithmic plot.

From Equation (3) we have $\ln(J_\infty/J) = \exp\left(-\gamma_\infty t\right) \ln(J_\infty/J_0)$, and Equation (2) then yields γ as a function of time,

$$\gamma(t) = \gamma_0 \exp\left(-\gamma_\infty t\right). \tag{6}$$

This is an alternative way of expressing Equations (1) and (2), which shows that the relative growth rate as a function of time decays exponentially with rate γ_∞ for all $t > 0$ and is a remarkable result since it implies that $\gamma(t)$ depends only on γ_0 and γ_∞. Equation (5) shows that γ_0 is determined by the three model parameters J_0, J_∞, and γ_∞, and hence can replace any of these as a model parameter. The parameters J_∞, and γ_∞ are more fundamental than the others, however, since the latter depend on the choice of the time origin. Equation (4) implies that a translation of the time variable, $t \to t - t_0$, corresponding to a shift of the time origin, leads to a shift of the initial value; $J_0 \to J_\infty (J_0/J_\infty)^{\exp(-\gamma_\infty t_0)}$, and Equation (6) leads to a shift of the initial growth rate; $\gamma_0 \to \gamma_0 \exp(\gamma_\infty t_0)$. A natural choice is to choose the time origin to be the first day the observed value of J exceeds one death per million nhabitants.

In the fitting procedure we do not fix $J_0 = 1$, thus allowing the modeled value J_0 to be slightly different from one and from the observed value at $t = 0$. Equation (5) can now be rewritten to give us the total death toll in terms of the rates γ_0 and γ_∞,

$$J_\infty = J_0 \exp\left(\frac{\gamma_0}{\gamma_\infty}\right), \tag{7}$$

where J_0 is close to unity. The significance of Equation (7) is that the accumulated death toll over the first epidemic wave depends exponentially on the ratio of the growth rate γ_0 at the time when the number of deaths exceeds one per million, to the asymptotic decay rate γ_∞ as the epidemic burns out. A more asymmetric epidemic curve, rapid rise and slow decay, intuitively leads to a larger number of deaths, but the exponential dependence on the ratio of rates suggests that the sensitivity of the death toll with respect to countermeasures is surprisingly high.

The monotonically increasing sigmoid shape of $J(t)$ suggests that it models the evolution of accumulated quantities like the cumulative number of infected individuals or cumulative deaths. In the literature one also frequently deals with the daily numbers, i.e. with the derivative $d_t J$. Figure 1a shows an example of $J(t)$ given by the Gompertz function, and Figure 1b its derivative, both in logarithmic plots. In Figure 1b

$$\Gamma(t) = d_t \ln(d_t J) = \gamma_\infty \left(e^{-\gamma_\infty t} \ln\left(\frac{J_\infty}{J_0}\right) - 1 \right) = \gamma_\infty \left[\ln\left(\frac{J_\infty}{J(t)}\right) - 1 \right]. \tag{8}$$

is the relative growth rate of the daily number $d_t J$ which is a skew, bell-shaped function. $\Gamma(t)$ is a monotonically decreasing function starting out at the positive value $\Gamma_0 = \gamma_\infty (\ln(J_\infty/J_0) - 1)$, crossing zero at the peak of the $d_t J$-curve, and converging towards the negative value $-\gamma_\infty$ as $t \to \infty$. From Equation (8) we find an alternative to Equation (7)

$$J_\infty = J_0 \exp\left(1 + \frac{\Gamma_0}{\gamma_\infty}\right) = (J_0 e) \exp\left(\frac{\Gamma_0}{\gamma_\infty}\right), \tag{9}$$

and it follows that
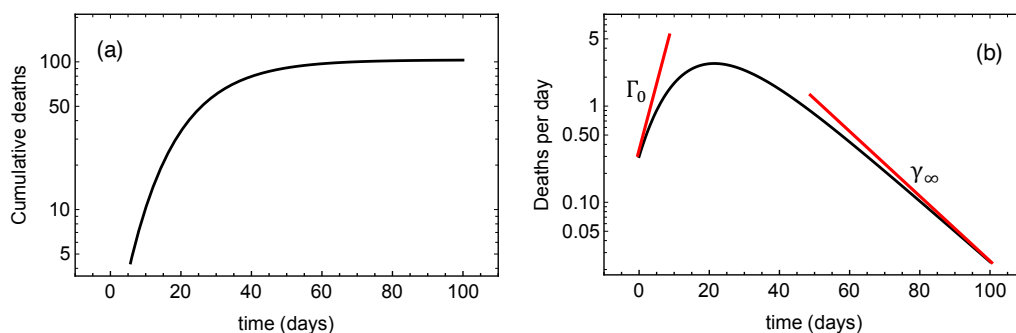
$$\Gamma_0 = \gamma_0 - \gamma_\infty. \tag{10}$$

Figure 1. (**a**): Evolution of cumulative deaths per million $J(t)$ in a log-plot expressed as a Gompertz function. (**b**): The evolution of the daily death number $d_t J(t)$ in a log-plot with the tangents at $t = 0$ and $t = \infty$ marked. The slopes of these tangents are Γ_0 and γ_∞, respectively.

Thus, we observe that the total death toll depends exponentially on the ratio Γ_0/γ_∞, which is the absolute value of the ratio of the initial and the final slopes of the curve in Figure 1b.

The parameters J_∞ and γ_∞ are intrinsic in the sense that they do not depend on the time origin, while γ_0 and Γ_0 do exhibit such a dependence. This is reflected by the inconvenient fact that the estimated J_0 is different for each country and is determined by the choice we have made for the time origin (the first day the number of deaths per million exceeds one). For comparison of countries, it would be more correct to use the growth rate γ_1 at the time t_1 when $J(t_1) = 1$. This growth rate is found by putting $J = 1$ in Equations (2) and (8), yielding

$$\gamma_1 = \gamma_\infty \ln J_\infty, \quad \Gamma_1 = \gamma_1 - \gamma_\infty, \tag{11}$$

such that Equations (7) and (9) reduce to,

$$J_\infty = \exp\left(\frac{\gamma_1}{\gamma_\infty}\right) = \exp\left(1 + \frac{\Gamma_1}{\gamma_\infty}\right) = e\exp\left(\frac{\Gamma_1}{\gamma_\infty}\right), \tag{12}$$

or, by taking the logarithm,

$$\ln J_\infty = \frac{\gamma_1}{\gamma_\infty} = 1 + \frac{\Gamma_1}{\gamma_\infty} \tag{13}$$

2.3. The Evolution of the Reproduction Number

In Appendix A the Gompertz model is derived heuristically from the SIR dynamical model with some auxiliary assumptions. In the SIR model, the variable $J(t)$ is the cumulative number of infections in a country's population. This quantity cannot be observed directly, because the majority of infected individuals are never tested for the virus. The number of cases confirmed by tests is not a reliable proxy, because testing regimes change over the course of the epidemic, and vary among countries as well. The number of COVID-19 related deaths is also unreliable as a proxy, since infection death rates vary from one country to another, but as a measure of the time development of the cumulative number of infections it is probably the best statistic that is generally available. In this paper, we are interested in the overall shape of the epidemic curve, and not the time for epidemic onset or details depending on the distribution of the time between infection and death. For most countries the shape of the curves for infections and deaths are very similar in the sense that the former can be obtained by a trivial time translation and re-scaling of the latter

$$J(t) \rightarrow s_d J(t + t_d), \tag{14}$$

where the scaling factor s_d in Scandinavia is of the order 150 and the time lag between infection and death t_d is about 20 days. After performing this re-scaling of the curve of observed deaths, we interpret the result as the curve of cumulative infection cases that appears in the SIR model (see Appendix A). In that model we encounter the time-dependent reproduction number $\mathcal{R}(t)$, which is more commonly used in epidemiology than the instantaneous relative growth rate $\gamma(t)$. Consider now the the linearized SIR Equations (A7) and (A8) for the cumulative number of infected, $J(t)$, and the number of active transmitters of the infection, $I(t)$, as described in Appendix A. Introducing the reproduction number $\mathcal{R}(t) = \beta(t)/\alpha$, these equations take the form

$$\frac{dJ}{dt} = \alpha \mathcal{R} I, \tag{15}$$

$$\frac{dI}{dt} = \alpha(\mathcal{R} - 1)I. \tag{16}$$

Here, the reproduction number is typically reduced with time due to countermeasures and the gradual removal of superspreaders from the susceptible population, while the recovery rate α remains constant. By solving Equation (16) and inserting into Equation (15), the latter can be written,

$$\frac{dJ}{dt} = \alpha I_0 \mathcal{R} \exp\left(\int_0^t \alpha(\mathcal{R} - 1)dt'\right). \tag{17}$$

Taking the logarithm of this equation and differentiating with respect to t, yields the differential equation

$$\frac{d\mathcal{R}}{dt} + \alpha \mathcal{R}^2 - \alpha \mathcal{R} - \Gamma(t)\mathcal{R} = 0, \tag{18}$$

where $\Gamma(t)$ is given by Equation (18). The equation can be solved numerically with the proper initial condition $\mathcal{R}(0) = \mathcal{R}_0$. In the asymptotic limit $t \to \infty$, $\Gamma(t) \to -\gamma_\infty$ and $\mathcal{R}(t) \to \mathcal{R}_\infty$, where

$$\mathcal{R}_\infty = 1 - \frac{\gamma_\infty}{\alpha}. \tag{19}$$

Let us assume that nobody infected has recovered at the time $t = 0$, implying that $J(0) = I(0)$. From Equation (15) we then find that,

$$\gamma_0 = \alpha \mathcal{R}_0. \tag{20}$$

Thus, by using Equations (19) and (20), Equation (7) becomes

$$J_\infty = \exp\left(\frac{\mathcal{R}_0}{1 - \mathcal{R}_\infty}\right). \tag{21}$$

2.4. Fitting Procedure

The essential methodological idea in this work is to fit a mathematical model for the epidemic curve to an observed time series; in this case the cumulative number of COVID-19 related deaths. This curve has a sigmoid shape with near-exponential growth in the initial growth phase. In such situations a usual least-square fitting procedure will give a poor representation of the initial growth because large relative errors will give small contributions to the absolute least-square error. Much better representation is obtained by fitting the logarithm of the model to the logarithm of the data. This is what we do here, by fitting the function for $\ln J(t)$ given by Equation (3) to the logarithm of the cumulative death per million time series starting at the first day this number exceeds one. We use the built-in fitting routine FindFit in *Mathematica* 12.0.0.0, which employs a least-square optimation to estimate the parameters J_∞, γ_∞, and J_0.

We have typically used about 100 days of the time series for this fitting, but have adjusted this manually for each country to make sure that we only include the first wave of the epidemic. We have only included countries which have clearly completed a first wave, so that we can reliably estimate the decay rate γ_∞. For this reason, we have not analyzed many countries in the tropics or the southern hemisphere. The data for the 73 countries we have analyzed, and the fits to them, are shown in Figures A2–A7 in Appendix B. The fits are surprisingly good, and demonstrate the usefulness of the Gompertz model to give an analytical representation of these data.

3. Results

In this section we first present a detailed comparison of Sweden and Norway in order to highlight some of the useful information that can be drawn from the analytical representation of the epidemic curves through the fitted Gompertz functions. Next we estimate the Gompertz model parameters from 73 countries and classify them in terms of total death toll and ratio between rates of initial growth and later decay.

3.1. Epidemic Curves of Sweden and Norway

Figure 2a shows the development of cumulative deaths per million inhabitants in Norway and Sweden in a logarithmic plot, and the corresponding fitted Gompertz function. Sweden data converge to a total death toll which is more than ten times higher than that of Norway. The relative growth rates $\gamma(t) = d_t \ln J(t)$ given as the slope of each curve are plotted in Figure 2b. Equation (6) shows that these growth rates decay exponentially from an initial value γ_0 towards zero at a rate γ_∞. Observe that the gap between Swedish and Norwegian death numbers continues to grow throughout April and May, consistent with the higher Swedish growth rate throughout this period.
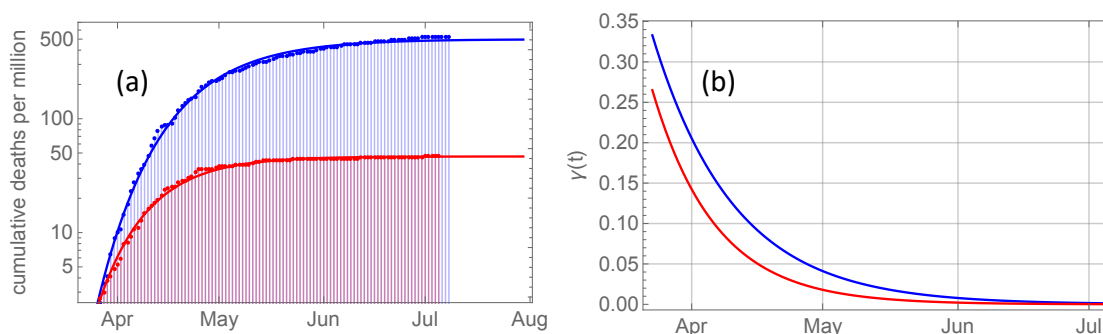
Figure 2. (**a**): Red bullets represent cumulative COVID-19 related deaths per million inhabitants in Norway from March 23 and onwards, and the full curve is the Gompertz curve $J(t)$ fitted to these data. The blue bullets and curve are the corresponding for Sweden (shifted 3 days forwards). Note that the plot is logarithmic, and that cumulated death toll per million in Sweden in early July is 12 times that in Norway. (**b**): The relative growth rate $\gamma(t) = d_t(\ln J)$ as given by Equation (6) for Norway (red) and Sweden (blue). These growth rates are the slopes of the curves in (**a**).

Figure 3a shows the derivative $d_t J(t)$ for Sweden an Norway in logarithmic plot. The growth rate $\Gamma(t)$ of this derivative is the slope of these curves, and is plotted in Figure 3b. It is given by Equation (8) and decays exponentially from $\Gamma_0 = \gamma_0 - \gamma_\infty$ to $-\gamma_\infty$. As shown by Equation (12) the total death toll J_∞ is determined by the ratio of the initial and final slope, i.e., the ratio Γ_1/γ_∞. The steeper initial growth in Sweden signified through the higher Γ_1 obviously contributes to the total death toll, but so does also also the lower rate on the decaying slope in June and July.
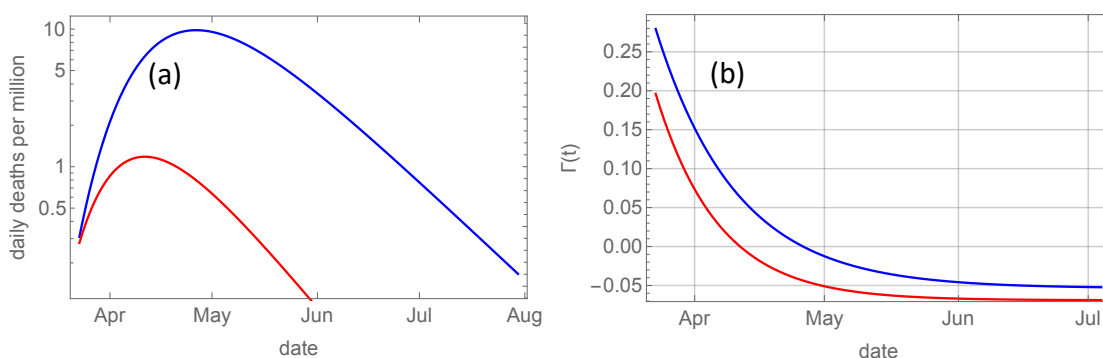

Figure 3. (**a**): Evolution of daily deaths per million, dJ/dt, computed numerically from Equation (12) for Norway (red) and for Sweden (blue). (**b**): The relative growth rate $\Gamma(t) = d_t(\ln d_t J)$ for Norway (red) and for Sweden (blue).

Before we try to quantify these contributions, let us look at the reconstructed evolution of the reproduction number $\mathcal{R}(t)$ for the two countries by solving Equation (18) with $\Gamma(t)$ shifted 17 days backwards in time to account for the average delay between infection and death, and $\alpha^{-1} = 5$ days representing the average infectious period. The initial value $\mathcal{R}_0$ is given by Equation (20). The 17-days delay is uncertain, so the shape of $\mathcal{R}_0$-curves shown in Figure 4a may be more accurate than the absolute dating. The important observation is that the relatively small, but consistent, difference between the reproduction number in Sweden and Norway throughout the entire epidemic wave has been sufficient to produce a tenfold higher number of deaths in Sweden.

Figure 4b shows that the ratio of the cumulative deaths in Sweden and Norway increases strongly throughout the entire epidemic wave, which indicates that the higher $\mathcal{R}$ in Sweden during the decaying phase has had a strong effect on the total death toll, i.e., people have continued dying in Sweden in May and June, long after the death rate was close to zero in Norway.
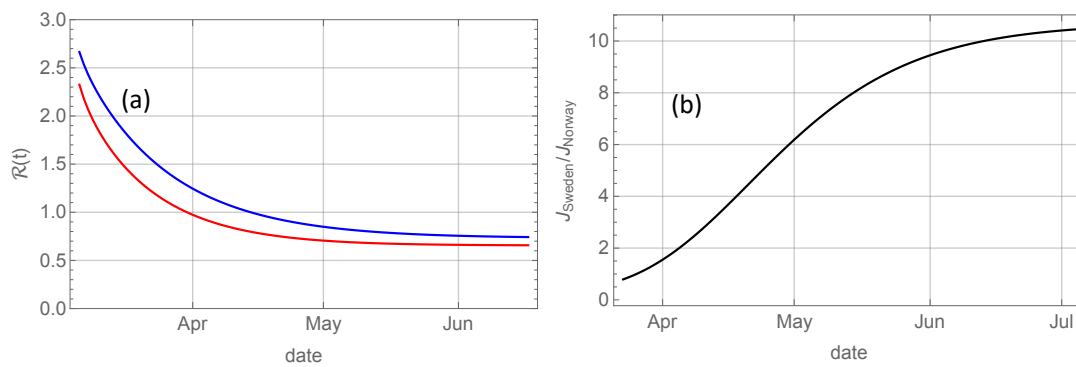
Figure 4. (**a**): Evolution of $\mathcal{R}(t)$ computed numerically from Equation (12) for Norway (red) and for Sweden (blue). (**b**): Evolution of ratio between accumulated deaths in Sweden and Norway.

Further insight into the differences of the epidemic curves between countries can be obtained by the graphics demonstrated in Figure 5. Here we have marked the point $(\Gamma_1, \gamma_\infty)$ for Sweden and Norway and represented the function $\ln J_\infty = 1 + \Gamma_1/\gamma_\infty$ in a density plot. The diagonal lines represent isolines for constant J_∞. We observe that Sweden is located on the iso-line where $J_\infty = 520$ and Norway on the line where $J_\infty = 49$. In addition, it is seen, as already noticed, that Sweden exhibits a considerably higher initial growth rate Γ_1 than Norway, and also a lower decay rate γ_∞. The main value of the diagram, however, is that it allows us to explore the effect on the death toll of hypothetical action in the two countries that did not materialize.
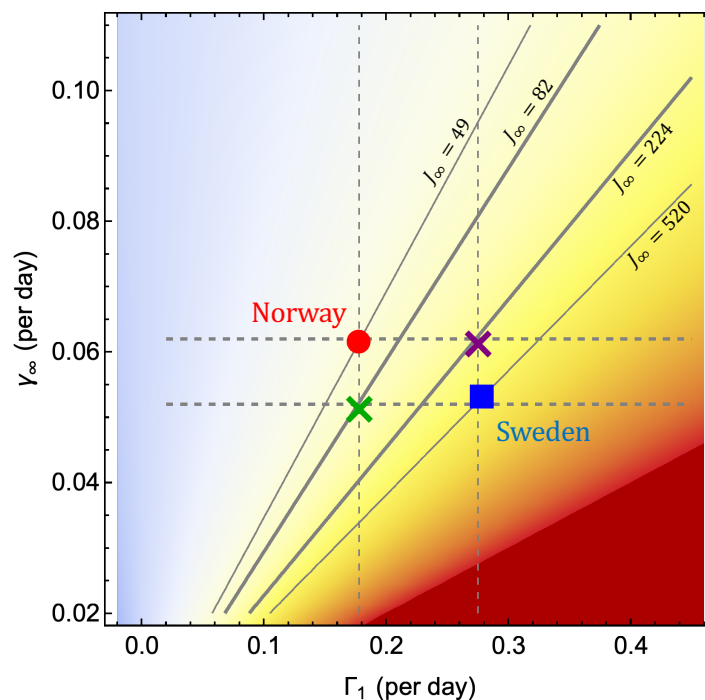

Figure 5. Density plot of $J_\infty(\Gamma_1, \gamma_\infty)$ with some isolines marked, along with the points for Sweden and Norway. The crosses mark the end states for paths where the inital and final phase of the two countries are mixed.

Judging from the final death tolls, the Swedish path does not seem as a very attractive option. However, what if the countries had chosen to follow the Swedish path initially, with strong growth in the early phase and then followed the Norwegian path with fast decay in the late phase? The end state would then be the purple cross which is located on the isoline with totally 224 deaths per million.

Another option would be to follow the Norwegian path with slow initial growth followed by the Swedish path with slow decay. This would lead to the blue cross which lies on the 82 deaths per million line. This suggests that reduction of the initial growth is the most effective way to bring the total death toll in the first wave down. More precisely, the lower initial growth in Norway reduced the number of deaths by a factor of six compared to Sweden, and the faster decay in the late phase by nearly another factor of two.

It should be mentioned that the hypothetical paths discussed above may not be realizable in practice since there may be dynamical constraints that do not allow Γ_1 and γ_∞ to be varied independently of each other. However, as described in Section 3.2 by studying a large number of countries we find a wide range of possible states in the $(\Gamma_1, \gamma_\infty)$-plane, and no sign of "forbidden" areas within this range.

3.2. Analysis of Data from 73 Countries

Figure 6 shows the same as Figure 5, where the data for Sweden and Norway has been supplemented by 71 other countries. The epidemic curves and their Gompertz fits are shown in Figures A2–A7. Figure 6 shows that the death toll J_∞ in the first wave covers a range of more than two orders of magnitude, from just above 3 in China to more than 800 in Belgium. Ordered by increasing death toll (see Figure 7c for the full ordered list), we have identified a few countries in the legend. China, New Zealand, Slovakia, South Korea and Japan belong to the group of countries with less than 10 deaths per million. In the group with $\sim 10^2$ deaths, we have Iceland, Norway, Finland, Austria, Denmark, and Germany. Furthermore, in group with $\sim 10^3$ deaths, United States, Sweden, Spain, Italy, United Kingdom, and Belgium.
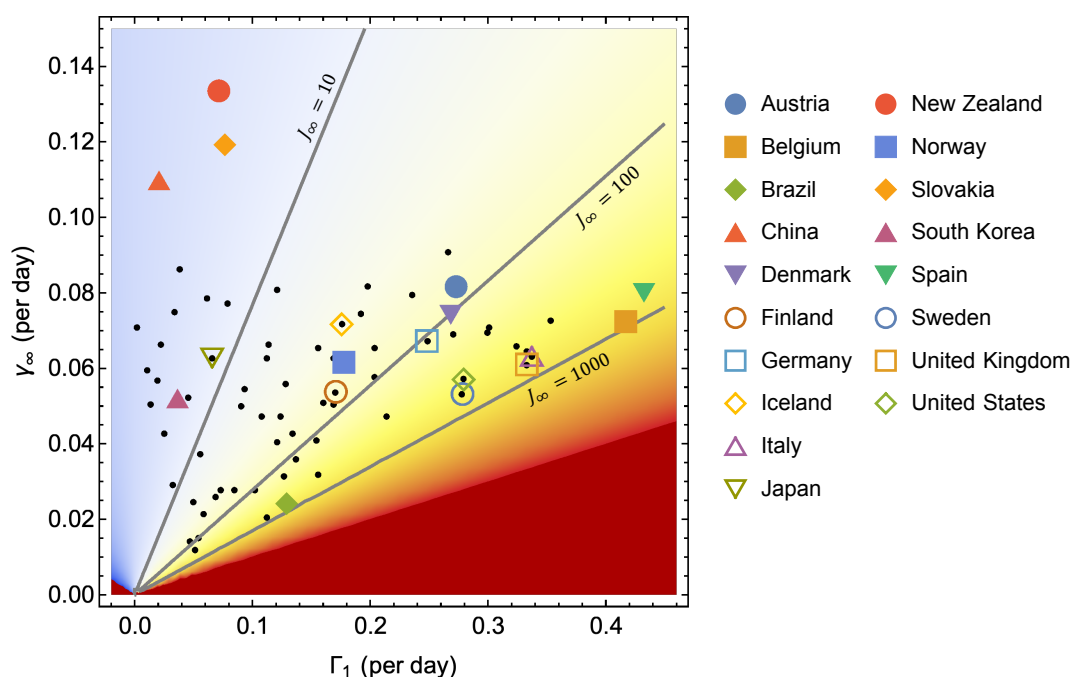

Figure 6. Density plot of $J_\infty(\Gamma_1, \gamma_\infty)$ with some isolines marked, along with the points for 73 countries. The legend shows the positions for some selected countries.

From Figure 7c one could be tempted to conclude that the epidemic curves are very similar for countries with similar death toll, but Figures 6 and 7a,b show that this is not so. The magnitudes of the growth rates vary considerably among countries with similar $J_\infty = \exp\left(\Gamma_1/\gamma_\infty\right)$. Some countries within one group have a rapid initial rise Γ_1 which is compensated by a rapid fall (high rate γ_∞) in the late phase, while others have a slow initial growth and a slower decay, resulting in approximately the

Int. J. Environ. Res. Public Health **2020**, *17*, 6487

10 of 23

same death toll. As mentioned in the introduction, Sweden experienced considerably slower initial growth than Spain, but also achieved much slower decay than Spain due to weak measures to contain the epidemic compared to Spain's radical lock-down. The result is that the two countries are located very close to each other in Figure 7c, but are strongly separated along the same J_∞ isoline in Figure 6.
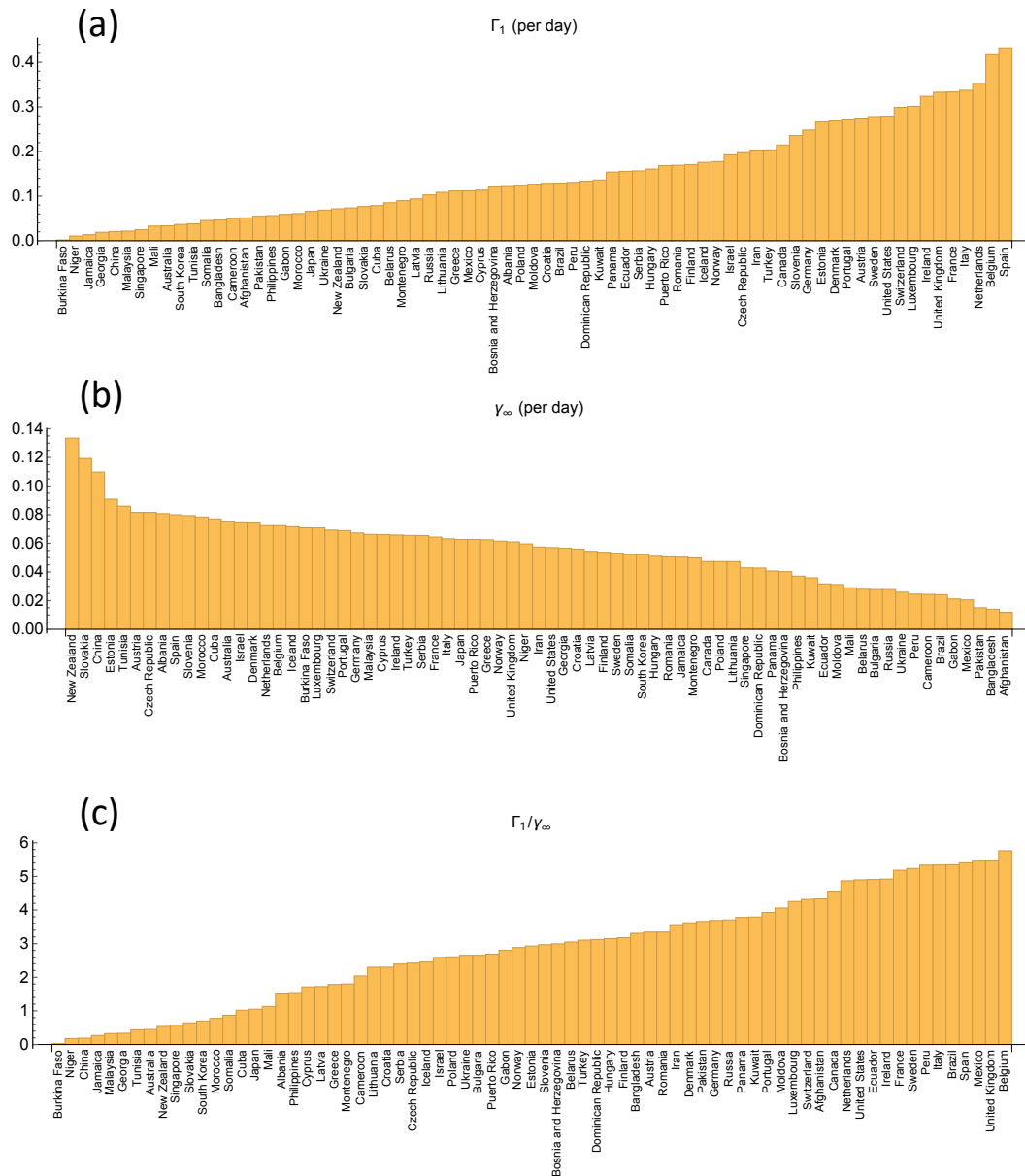

Figure 7. (**a**): Estimated Γ_1 for the 73 countries ranked from lower to higher values. (**b**): Estimated γ_∞ ranked from high to low values. (**c**): Estimated $\Gamma_1/\Gamma_\infty = \ln J_\infty$ ranked from lower to higher values.

Mathematically, Figure 6 locates each country in a 3D plot which specifies the three Gompertz parameters $(\Gamma_1, \Gamma_\infty, \ln J_\infty)$. Hence, it contains all the information contained in the Gompertz description of the epidemic curve. Figure 7 can be conceived as the projections of the points in Figure 6 onto the three axes.

4. Discussion

The main purpose of this paper has been to extract objective information from the epidemic curves for COVID-19 related deaths on the country level by utilizing the observational fact that the

Int. J. Environ. Res. Public Health **2020**, *17*, 6487

11 of 23

curves for the first wave are quite accurately represented by the Gompertz function for many countries of interest. The results may be subject to different interpretations and raise many questions that we cannot answer in this paper. In this section, we will limit ourselves to present some subjective views on why the Gompertz model seems to fit these data so well and some viewpoints on the information presented in Figures 6 and 7.

4.1. How to Interpret the Gompertz Model

By taking the time derivative of Equation (3), the Gompertz model can be written in the form

$$\gamma(t) = \frac{d_t J}{J} = d_t \ln J(t) = \gamma_\infty \ln \left(\frac{J_\infty}{J_0} \right) \exp(-\gamma_\infty t) = \gamma_0 \exp(-\gamma_\infty t), \tag{22}$$

which is identical to Equation (6). The relative growth rate γ represents how much one unit of J grows per unit time. It is closely related to the reproduction number $\mathcal{R}$, which can be interpreted as the average number of new infections caused by one infected individual. If the disease transmission mechanisms remain constant over time, including the density of susceptible individuals, γ will remain constant, and the growth of $J(t)$ will be exponential. If a large fraction of the population becomes infected, and eventually immune, herd immunity will appear as a nonlinearity in the SIR-equations (see Appendix A). This effect has not yet appeared on the country level in the COVID-19 epidemic, so changes in the transmission mechanism must cause a reduction of $\gamma(t)$ over time. One effect that is seen in all epidemics even without societal action is that individuals who are particularly active in spreading the disease (so-called super-spreaders) catch the disease early and are removed from the susceptible population. As time goes, this reduces the effective reproduction number and γ. In addition, society acts in complex ways to resist the disease. If the total effect of this resistance on the rate of change of γ is proportional γ itself (which is a common property of a dissipative force), the equation for γ would take the form

$$\frac{d\gamma}{dt} = -\eta\gamma, \tag{23}$$

which is what we will find by taking the time derivative of Equation (22) with the resistance $\eta = \gamma_\infty$. Thus, we have arrived at a straightforward interpretation of the Gompertz model as a mathematical description of an unstable system where a quantity $J(t)$ naturally grows exponentially in the absence of dissipation. The dissipative force in this system does not act on $J(t)$ itself, but rather on its growth rate. Society does not take action in response to the death toll itself but to its growth.

4.2. How to Interpret Figures 6 and 7

The interpretation of Figures 6 and 7 is sensitive to how the sample of countries has been selected. A complete picture will emerge as more countries complete the first wave. We have excluded countries that are not well beyond the first peak in daily deaths, or countries that have entered a second wave much before the first was completed. Countries that have not reached the first peak are Argentina, India, South Africa, and Colombia, and countries that have entered an early second wave are Iran, Saudi Arabia, Iraq, and Chile.

We have also excluded countries that we strongly suspect have unreliable or irregular reporting or epidemic curves that, by visual inspection, are not well fitted by a sigmoid function. However, we emphasize that no country has been excluded from the sample based on the results of the analysis.

The relative effect of the two parameters on the total death toll is seen by traversing the sample in Figure 6 in the vertical and horizontal direction. We observe that the distribution in the γ_∞-direction is much wider for small Γ_1 than for large Γ_1. In the vertical direction, J_∞ varies two orders of magnitude among countries with $\Gamma_1 < 0.15$, i.e., there is a high variability of the decay rate γ_∞ among countries with low initial growth. Some countries, like China, New Zealand, Slovakia, South Korea, and Japan, have taken strong action throughout the first wave, despite low initial growth of the death numbers. On the other hand, other countries like Brazil, have used relatively low initial growth as an excuse

for non-action, resulting in a very high death toll. For countries with $\Gamma_1 > 0.15$ the range of γ_∞ becomes narrower with increasing Γ_1. In this group, we find most Western-European countries and the United States. Finland, Norway, Iceland, Germany, Denmark, and Austria have Γ_1 in the lower end of this range and $J_\infty \sim 10^2$ or less, while for Sweden, United States, United Kingdom, Italy, Belgium, and Spain, Γ_1 is higher and $J_\infty \sim 10^3$. The overall impression is that the variability of the death toll in Western Europe and the United States is largely due to variations of the initial growth rate, and to a lesser extent due to variations of the later decay. There are exceptions, however. For instance, Sweden and Austria have almost the same Γ_1, but γ_∞ on opposite tails of the distribution, yielding about ten times higher death toll in Sweden.

In the rest of the world, the picture is more mixed. Some low-income countries, with low initial growth rates, still end up with a high death toll because the decay rate is also low. On the other hand, high-income countries in South-East Asia suffer few deaths due to a combination of low initial growth combined with moderate or high decay rate.

The exceptionally low Γ_1 for China is due to the strong confinement of the epidemic to the Hubei province, which constitutes only 4.3 percent of the Chinese population. If we had treated Hubei as a country, the death toll per million would have been 23 times higher, and Γ_1 would have increased by a factor of $\ln 23 \approx 3$ and become more like that of Japan. Remember that Γ_1 is defined as the growth rate at the time the death toll exceeds one per million, and this time comes earlier if the population is considered to be only that of Hubei. This observation underscores that geographic isolation of the epidemic to limited regions within a country is crucial in reducing the initial growth rate and the total death toll. The success of the Chinese strategy in limiting the first wave is the effective isolation of Hubei from the rest of China and the very strict lockdown within the province.

A caveat of this entire discussion is that the death rates reported from the various countries may be inaccurate. Systematic under-reporting will influence the estimate of Γ_1, but not that of γ_∞. At present, we have not been able to make systematic corrections to these figures for all countries. Corrections based on figures for excess mortality is a possibility for some countries, but such figures do not exist for many of countries for which the official COVID-19 death rates are least reliable. Excess mortality may also give rise to under-estimation of COVID-19 deaths in countries with effective interventions, because these interventions reduce mortality from other diseases.

5. Conclusions

The huge variability of the death toll among countries in the first wave of the COVID-19 epidemic is puzzling. Many causal explanations have been suggested in the media, but on a world scale, little systematic work has been published in the peer-reviewed literature. The present work does not attempt a causal explanation, but seeks a quantitative characterization of the epidemic curve in terms of as few parameters as possible. We demonstrate that this is possible in terms of the Gompertz function, which describes a three-parameter sigmoid curve. One of these parameters is redundant because it only determines the time origin that determines the point in time when the death toll per million inhabitants exceeds one. The two parameters that determine the total death toll is the initial relative growth rate of the number of daily deaths and the decay rate of this number as the first wave subsides. The salient property of the Gompertz function is that the total death toll of the first wave depends exponentially on the ratio of those rates. It hence describes the exceptional sensitivity of the death toll to the values of these rates.

We illustrate these features by applying the analysis to compare the effect of initial growth rates and later decay rates of the neighbouring countries Norway and Sweden, where more than ten times higher death toll in Sweden is attributed to a higher early death rate and, to a lesser extent, a lower decay rate in the subsiding phase. It is also shown to be related to a somewhat higher reproduction number in Sweden throughout the entire wave.

The Gompertz function is shown to be well fitted to the death data for 73 countries around the world, and plots of the Gompertz parameters for these countries show that they are broadly distributed

in this parameter space, suggesting that there are many routes to a given death toll. A task for future work is to find out more about commonalities between countries that are close to each other in the parameter space, and about features that distinguish countries that appear similar but end up at different locations in this space. The most straightforward approach would be some sort of multiple regression analysis by which the growth rate Γ_1 and the decay rate γ_∞ are linked to socioeconomic variables and intervention measures.

At the time of writing, many countries are entering the second epidemic wave. It will be interesting to investigate if this wave can be analyzed by the same method and to study if parameters change from the first to the second wave. We already observe marked changes from the first to the second wave in Europe, in particular that death tolls in the second wave so far are much lower compared to the number of confirmed infections. Some of this effect may be related to more frequent testing, but the most important difference seems to be a much higher epidemic activity among young adults, who experience a very low death rate. Social interventions are also very different in the second wave, with more emphasis on testing, tracking, isolation, and quarantining of close contacts. These factors will probably make it necessary to analyze confirmed cases data, in addition to death data, in order to fully grasp the most important characteristics of the second wave. A more complete data set for the first wave, involving more countries, will also become available, facilitating a more complete classification of the epidemic curves of the COVID-19 pandemic.

Appendix A. From the SIR Model to the Gompertz Model

Let N be the total population, S the number of susceptible individuals, and I the number of infected. We then have,

$$\frac{dS}{dt} = -\beta \frac{IS}{N}, \tag{A1}$$

$$\frac{dI}{dt} = \beta \frac{IS}{N} - \alpha I, \tag{A2}$$

where β is the rate by which the infection is being transmitted, and α is the rate by which the infected are isolated from the susceptible population. Decoupled from these, we have the equation for the number R of individuals that are quarantined, deceased, and recovered;

$$\frac{dR}{dt} = \alpha I. \tag{A3}$$

These three equations constitute the SIR model.

Appendix A.1. Nonlinear Evolution of an "Old" Disease

During the development of an epidemic, health authorities typically report the accumulated number of registered infected $J = N - S$. The equations for J and I then take the form

$$\frac{dJ}{dt} = \beta \left(1 - \frac{J}{N} \right) I, \tag{A4}$$

$$\frac{dI}{dt} = \left(\beta - \alpha - \beta \frac{J}{N} \right) I. \tag{A5}$$

When one studies the flare-up of an old pathogen, there may be a considerable herd immunity, i.e., J may be of the same order of magnitude as N. Then, the peak growth rate of I is attained when

$$J = J_s = N \left(1 - \frac{\alpha}{\beta} \right), \tag{A6}$$

and the saturation and decay of the epidemic is a nonlinear effect.

Appendix A.2. Evolution of an Epidemic due to a "New" Pathogen

If the pathogen is new, there is no immunity in the population from the start, and if we assume that only a small fraction of the population is affected by the epidemic, we have $J \ll N$ for all times and we can linearize;

$$\frac{dJ}{dt} = \beta I, \tag{A7}$$

$$\frac{dI}{dt} = (\beta - \alpha) I. \tag{A8}$$

Initial growth, and later decay, of I now requires that α and β evolve in time, and that $\beta - \alpha$ changes sign from positive to negative at a certain time t_s. Since $J(t)$ is monotonically increasing with t, there is a bijective mapping between t and J, and we can think of α, β, and I as functions of J rather than t.

Appendix A.3. The Reproductive Number

If we define $\mathcal{R}(J) = \beta/\alpha$ as the reproductive number, Equations (A1) and (A2) reduce to

$$\frac{dI}{dJ} = 1 - \frac{1}{\mathcal{R}(J)}. \tag{A9}$$

It is not unreasonable to assume that $\mathcal{R}(J)$ is a monotonically decreasing function, since measures to reduce the transmission rate will lower β as J increases and measures to diagnose and quarantine infected individuals will increase α. $I(J)$ has its maximum at $J = J_s = J(t_s)$ for which $\mathcal{R}(J_s) = 1$. This is also the inflection point for $J(t)$;

$$\left(\frac{d^2 J}{dt^2} \right)_{t_s} = 0. \tag{A10}$$

Appendix A.4. Reduction to a Generalized Logistic Growth Model

Since $I = I(J)$, the equation

$$\frac{dJ}{dt} = -\beta I, \tag{A11}$$

can be written as

$$\frac{dJ}{dt} = \gamma(J)\, J, \tag{A12}$$

where

$$\gamma(J) = \beta \frac{I}{J}, \tag{A13}$$

is the relative growth rate for J. Equation (A11) has the general form of a nonlinear growth model with a growth rate depending on the growing variable itself.

Int. J. Environ. Res. Public Health **2020**, *17*, 6487

15 of 23

Appendix A.5. Richards' Growth Model

$J(t)$ takes some sort of sigmoid form, which means that $\gamma(J)$ is monotonically decreasing from $\gamma(J = 0) = \gamma^{(0)}$ to $\gamma(J_\infty) = 0$, where $J_\infty = J(t \to \infty)$. A common form is to express $\gamma(J)$ by a function depending on the three parameters $\gamma^{(0)}$, J_∞, and ν;

$$\gamma(J) = \gamma^{(0)} \left[1 - \left(\frac{J}{J_\infty} \right)^\nu \right], \tag{A14}$$

and hence Equation (A11,) reduces to Richards' equation,

$$\frac{dJ}{dt} = \gamma^{(0)} \left[1 - \left(\frac{J}{J_\infty} \right)^\nu \right] J. \tag{A15}$$

Note that $\gamma^{(0)} = \gamma(J = 0)$ is the growth rate for $J = 0$, not the growth rate at time $t = 0$. The latter is

$$\gamma_0 = \gamma(t = 0) = \gamma^{(0)} \left(1 - \left(\frac{J_0}{J_\infty} \right)^\nu \right), \tag{A16}$$

The normalized growth rate $\gamma(J/J_\infty)/\gamma^{(0)}$ is plotted for varying ν in Figure A1a, and the corresponding Richards curves in Figure A1b. The solution satisfying the initial condition $J(t = 0) = J_0$ is

$$J(t) = \frac{J_\infty}{\left(1 + ((J_\infty/J_0)^\nu - 1)e^{-\gamma^{(0)}\nu t} \right)^{1/\nu}}, \tag{A17}$$

leaving us with a model with four free parameters to be determined by fitting the model to the time series for the death toll; $\gamma^{(0)}$, J_∞, ν, and J_0. In the asymptotic limit $t \to \infty$, $J(t)$ converges monotonically towards the limit J_∞, and the difference from this limit takes the form

$$J_\infty - J(t) \to \frac{1}{\nu} \left[\left(\frac{J_\infty}{J_0} \right)^\nu - 1 \right] e^{-\gamma^{(0)}\nu t}, \tag{A18}$$

hence it decays exponentially, and so does the derivative $J'(t)$ (the daily number of deaths), at a rate
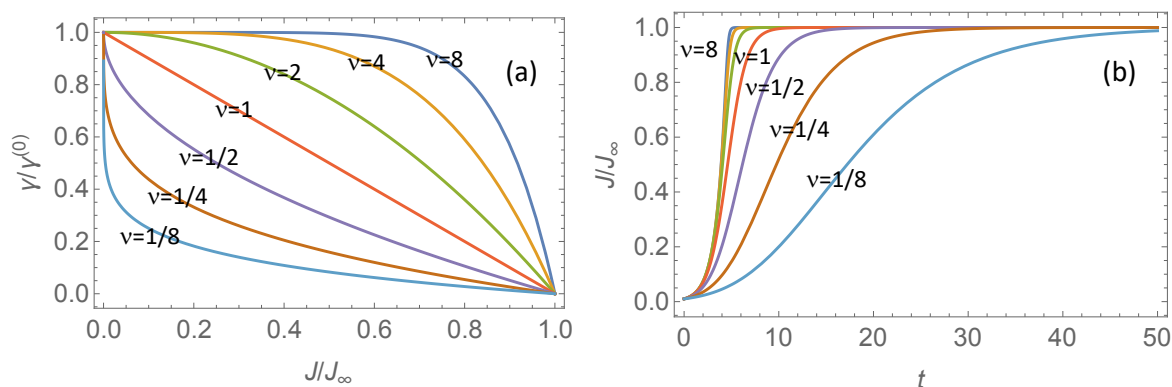
$$\gamma_\infty = \gamma^{(0)}\nu. \tag{A19}$$

Figure A1. (**a**): Normalized growth rate $\gamma/\gamma^{(0)}$ as a function of J/J_∞ for varying ν. Light blue; $\nu = 1/8$. Brown; $\nu = 1/4$. Magenta; $\nu = 1/2$. Red; $\nu = 1$. Green; $\nu = 2$. Orange; $n = 4$. Dark blue; $\nu = 8$. For $\nu = 1$ we have the symmetric logistic growth curve [11]. Increasing $\nu > 1$ shifts the inflection point J_s closer to the limiting value J_∞. Decreasing $\nu < 1$ shifts it towards J_0. (**b**): Normalized Richards growth curve J/J_∞ as a function of time t for varying ν with same color coding as in (a).

Appendix A.6. The Gompertz Limit

It is observed from Figure A1 that the limit $\nu \to 0$ corresponds to a rapid initial decay of the relative growth rate $\gamma(t)$, followed by a slower decay towards zero as $J \to J_\infty$. By fitting the Richards model to time series for countries that have gone through the first wave of the epidemic, we typically find small values for the exponent ν. This suggests that for countries with a reasonably rapid response to the increasing death tolls it may be reasonable to look at this limit, and hence reduce the number of free parameters by one. The resulting model is the one suggested by Gompertz [7]. Using the relation $\lim_{\nu \to 0} \nu^{-1}[1 - (J/J_\infty)^\nu] = \ln(J_\infty/J)$, it is easily shown that in this limit Equation (A14) reduces to

$$\gamma(J) = \gamma_\infty \ln\left(\frac{J_\infty}{J}\right), \tag{A20}$$

and Equation (A17) to

$$J(t) = J_\infty \left(\frac{J_0}{J_\infty}\right)^{\exp(-\gamma_\infty t)}, \tag{A21}$$

Note that in the Gompertz limit, $\gamma(J)$ diverges as $J \to 0$, while for finite ν (Richards' model) it converges towards $\gamma^{(0)}$. The important feature, however, is not the growth during the zoonotic stage, or the stage when the epidemic is dominated by imported cases, but rather the stages when the pathogen is transmitted in the population. By defining the time origin $t = 0$ as the time when J exceeds a certain threshold J_0, and considering only $t \geq 0$, we are concerned with the initial growth rate

$$\gamma_0 = \gamma(J_0) = \gamma_\infty \ln\left(\frac{J_\infty}{J_0}\right), \tag{A22}$$

which is completely determined by the model parameters; J_0, J_∞, and γ_∞.

Appendix B. Gompertz fits to 73 countries

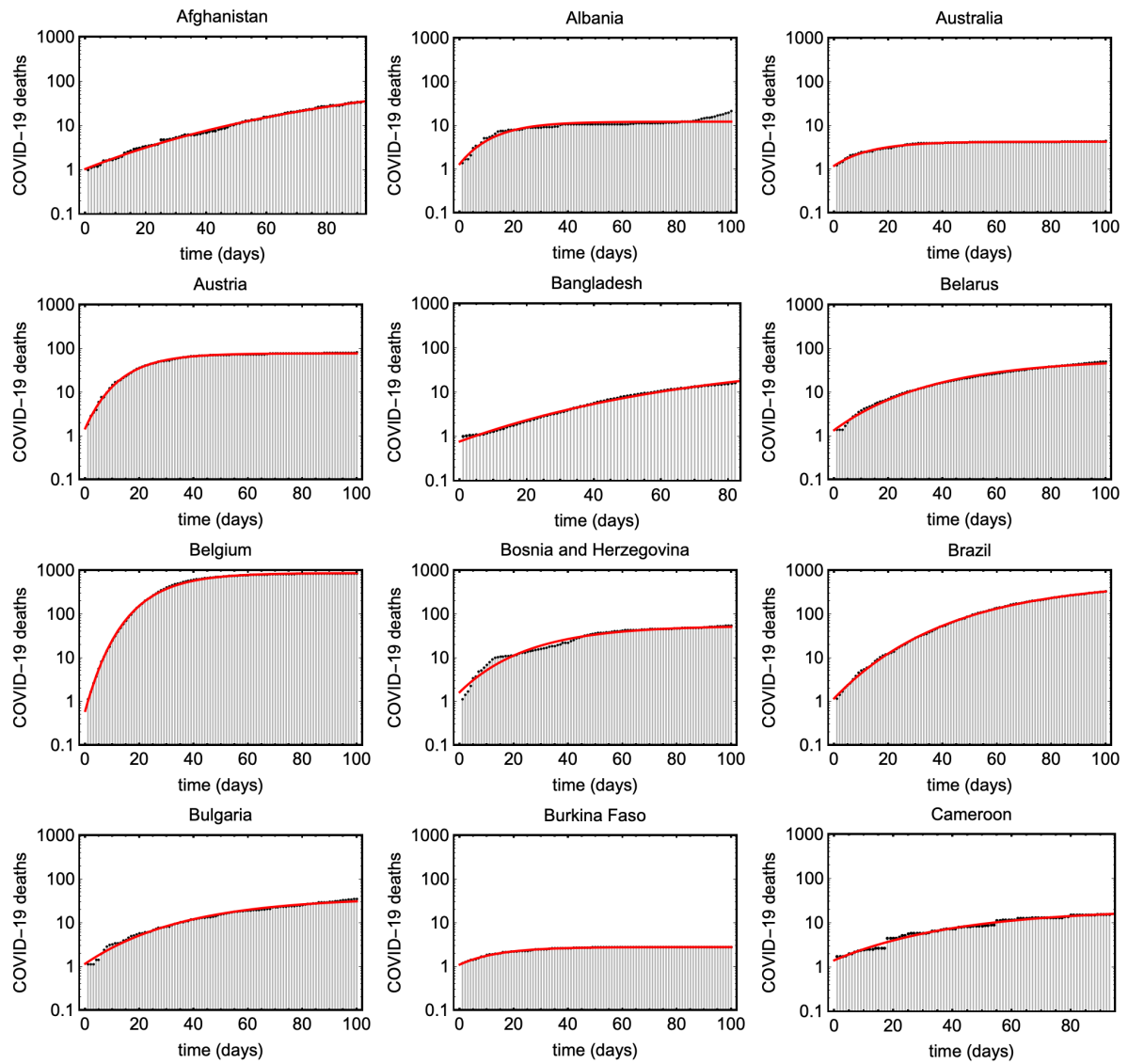Figures A2–A7 presents the Gompertz fits to the 73 countries with parameters plotted in Figures 6 and 7.

Int. J. Environ. Res. Public Health **2020**, *17*, 6487
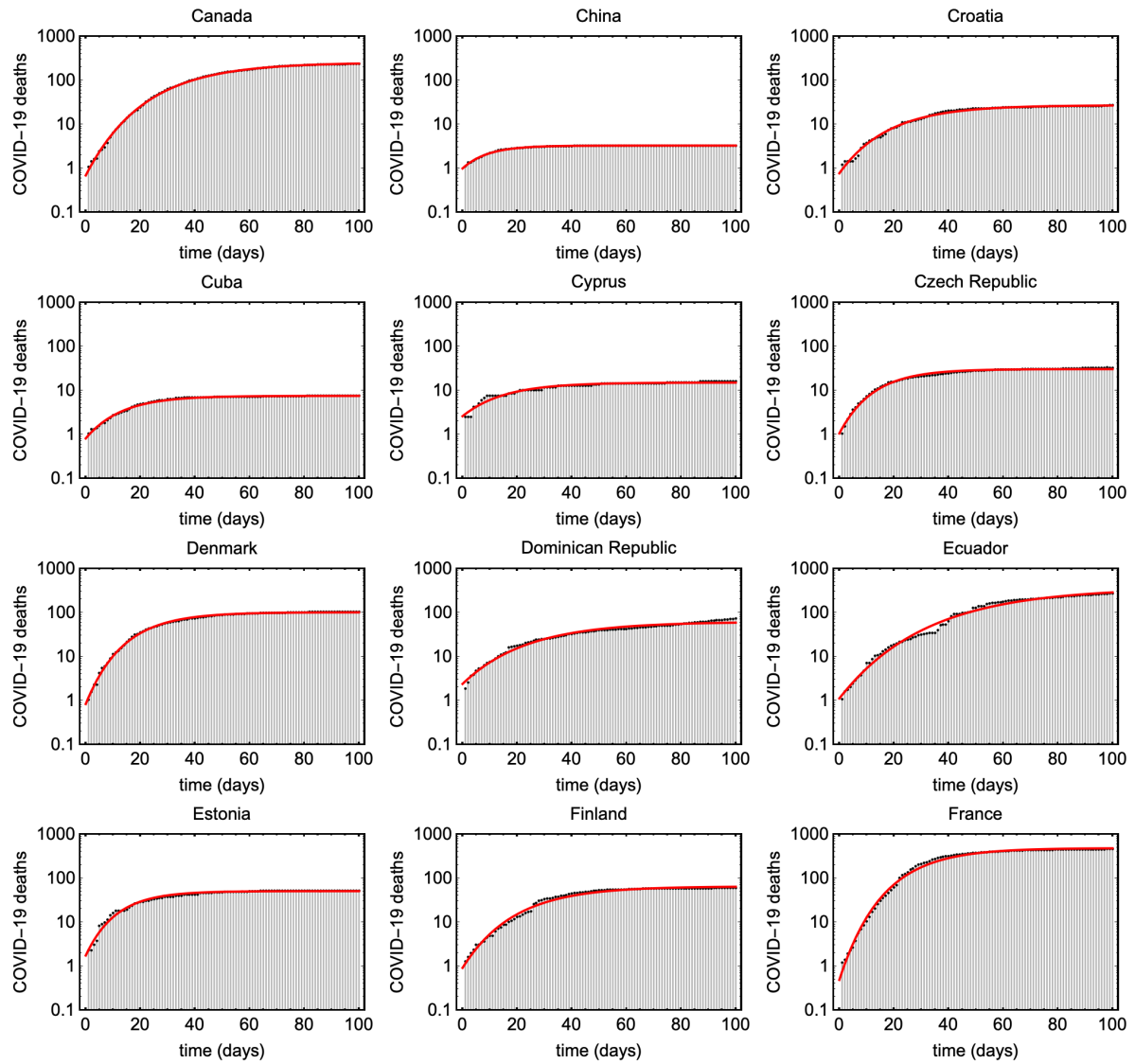
17 of 23

Figure A2. Shows cumulative COVID19 deaths (black) and the estimated Gompertz curves for different countries (red).

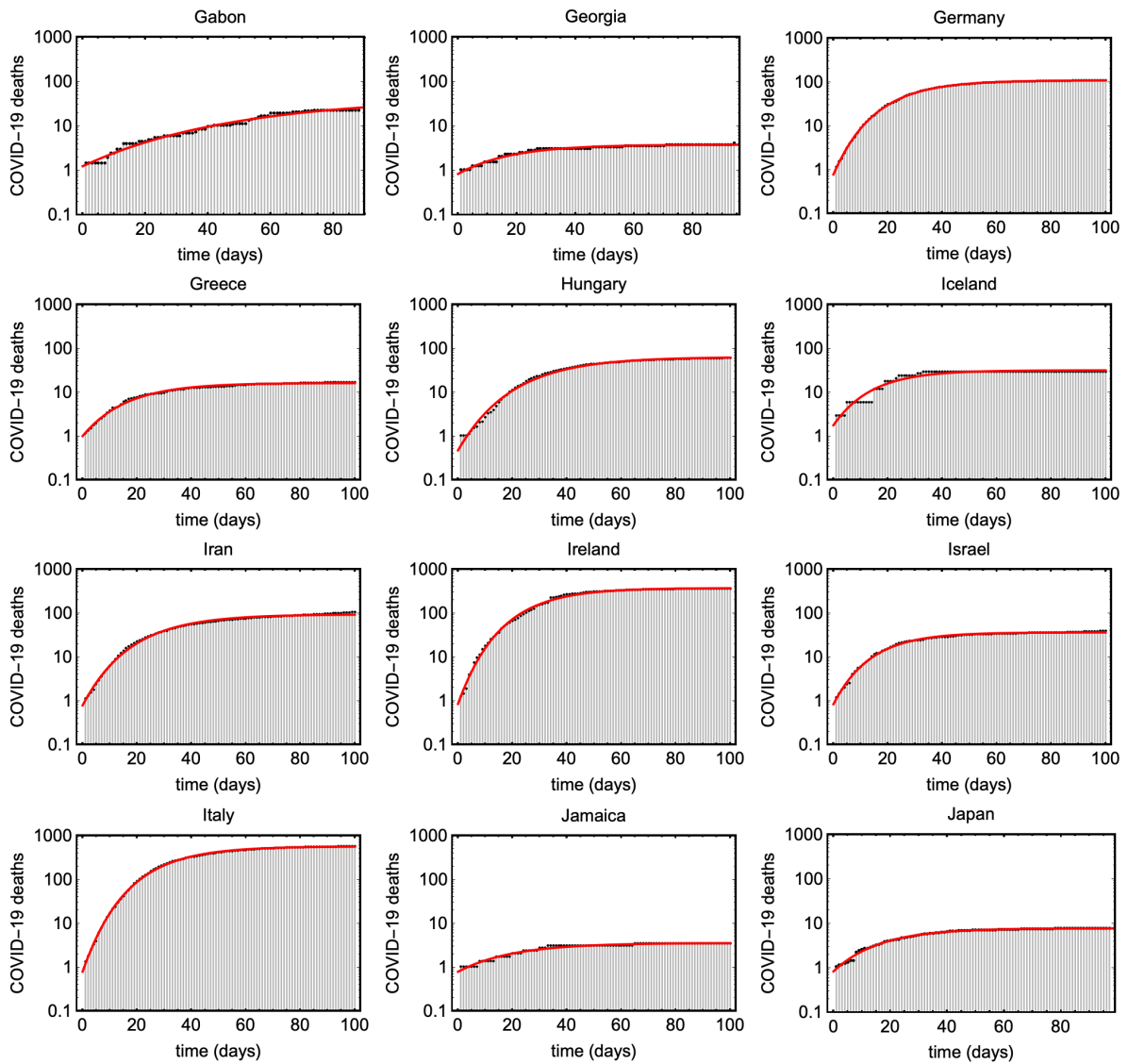**Figure A3.** Shows cumulative COVID19 deaths (black) and the estimated Gompertz curves for different countries (red).
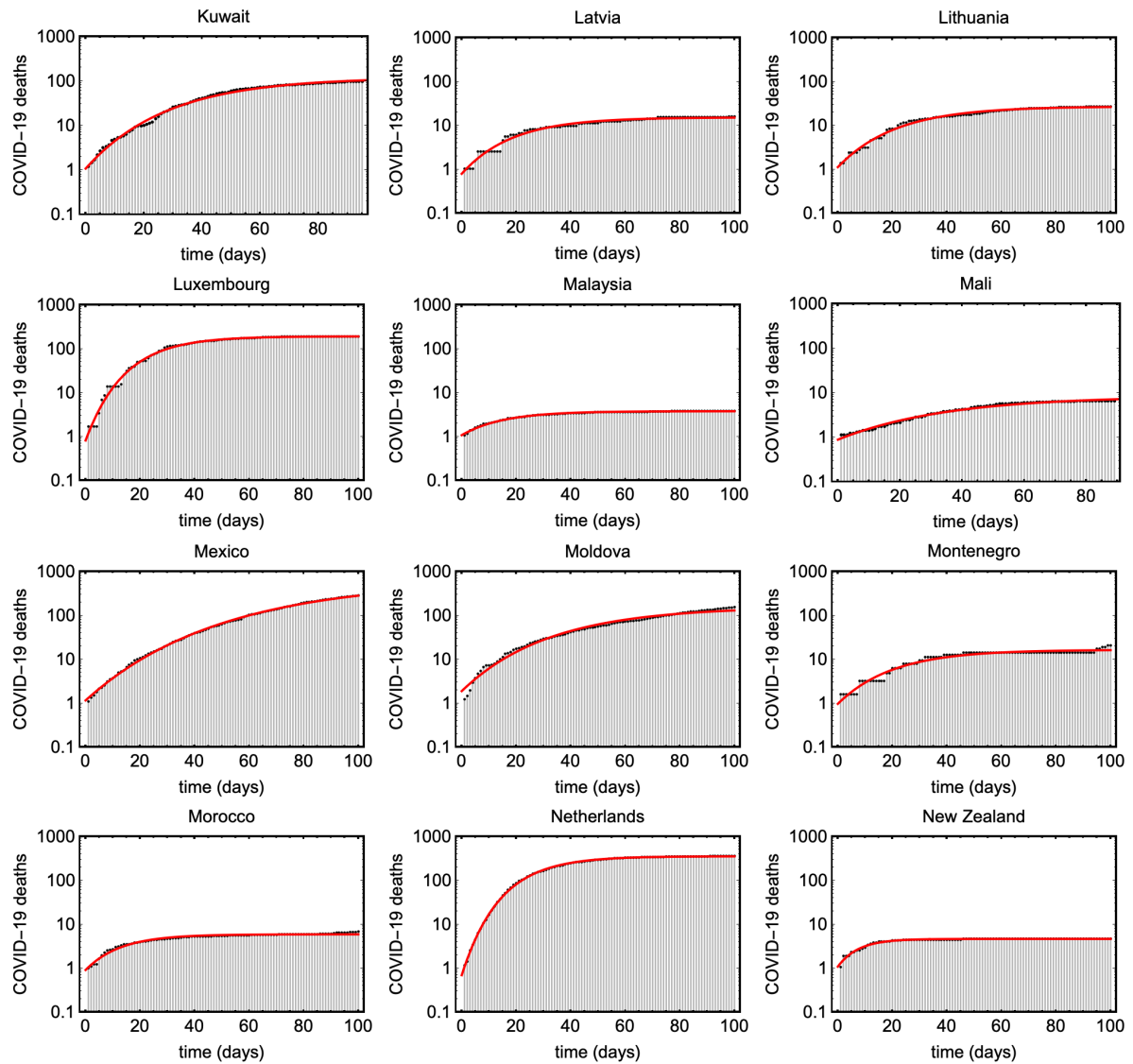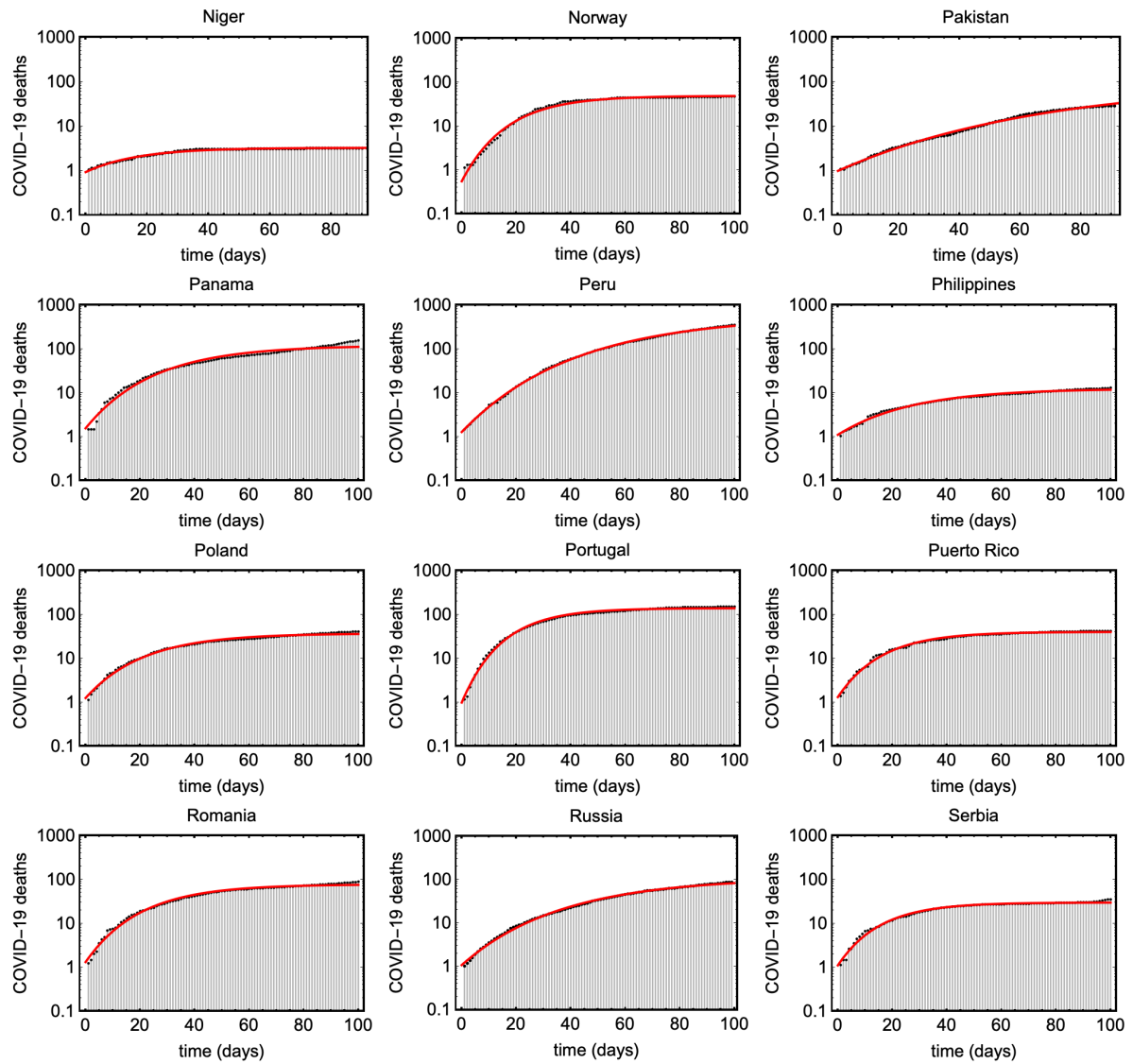
Int. J. Environ. Res. Public Health **2020**, *17*, 6487

19 of 23

Figure A4. Shows cumulative COVID19 deaths (black) and the estimated Gompertz curves for different countries (red).

Int. J. Environ. Res. Public Health **2020**, *17*, 6487

20 of 23

Figure A5. Shows cumulative COVID19 deaths (black) and the estimated Gompertz curves for different countries (red).

Figure A6. Shows cumulative COVID19 deaths (black) and the estimated Gompertz curves for different countries (red).
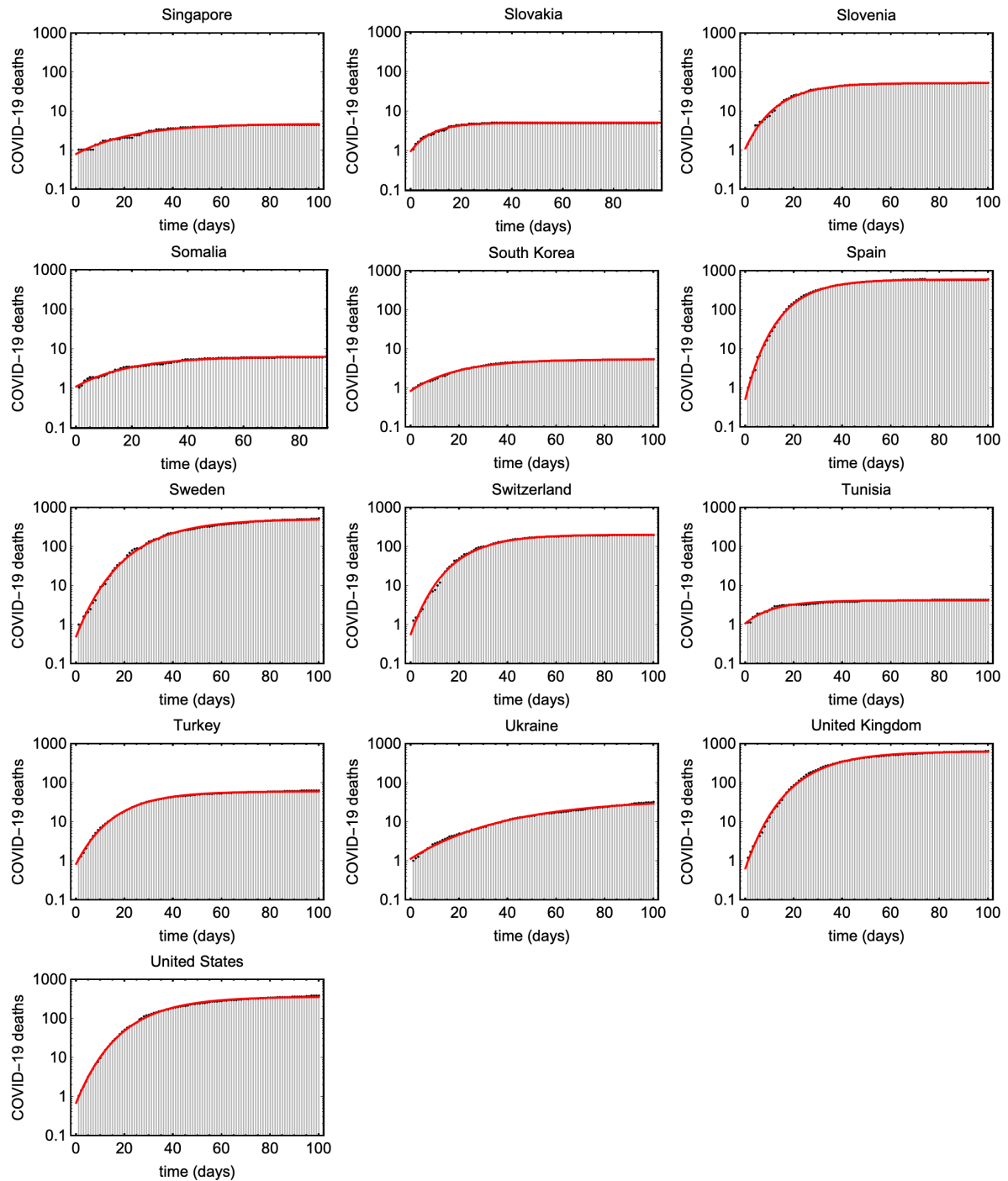
Figure A7. Shows cumulative COVID19 deaths (black) and the estimated Gompertz curves for different countries (red).

References

1. Why Do Countries' COVID-19 Death Rates Vary So Much? Multiple Factors at Play: Testing Capacity, Case Definitions, Age Distribution, Preparedness. MEDPAGE TODAY 15 May 2020. Available online: https://www.medpagetoday.com/infectiousdisease/covid19/86527 (accessed on 4 September 2020).
2. Chaudhry, R.; Dranitsaris, G.; Mubashir, T.; Bartoszko, J.; Riazi, S. A country level analysis measuring he impact of governmental actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes. *EclinicalMedicine* **2020**, 100464. [CrossRef] [PubMed]

Int. J. Environ. Res. Public Health **2020**, *17*, 6487

23 of 23

3. Flaxman, S.; Mishra, S.; Gandy, A.; Unwin, H.J.T.; Mellan, T.A.; Coupland, H.; Whittaker, C.; Zhu, H.; Berah, T.; Eaton, J.W.; et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **2020**, *584*, 257–261. [CrossRef] [PubMed]

4. Juranek, S.; Zoutman, F.T. *The Effect of Social Distancing Measures on Intensive Care Occupancy: Evidence on COVID-19 in Scandinavia. Discussion Papers 2020/2*; Norwegian School of Economics, Department of Business and Management Science: Bergen, Norway, 2020.

5. Paterlini, M. Sweden's coronavirus strategist. *Nature* **2020**, *580*, 574. [CrossRef] [PubMed]

6. Globalization and Outbreak of COVID-19: An Empirical Analysis. Available online: https://www.cesifo.org/en/publikationen/2020/working-paper/globalization-and-outbreak-covid-19-empirical-analysis (accessed on 4 September 2020).

7. Gompertz, B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philos. Trans. R. Soc. Lond.* **1825**, *115*, 513–585.

8. Our World in Data. Available online: https://ourworldindata.org/coronavirus-data (accessed on 4 September 2020).

9. Karanikolos, M.; McKee, M. How Comparable is COVID-19 Mortality Across Countries? COVID-19 Health System Monitor. 4 June 2020. Available online: https://analysis.covid19healthsystem.org/index.php/2020/06/04/how-comparable-is-covid-19-mortality-across-countries/ (accessed on 4 September 2020).

10. Tsoularis, A. Analysis of logistic growth models. *Res. Lett. Inf. Math. Sci.* **2001**, *2*, 23–46. [CrossRef]

11. Verhulst, P.-F. Notice sur la loi que la population poursuit dans son accroissement. *Corresp. Math. Phys.* **1838**, *10*, 113.