medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20152140; this version posted July 27, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Complexity signatures in the COVID-19 epidemic: power-law behaviour in the saturation regime of fatality curves

Giovani L. Vasconcelos[1,*], Antônio M. S. Macêdo[2], Gerson C. Duarte-Filho[3], Arthur A. Brum[2], Raydonal Ospina[4], and Francisco A. G. Almeida[3]

[1]Departamento de Física, Universidade Federal do Paraná, 81531-990 Curitiba, Paraná, Brazil
[2]Departamento de Física, Universidade Federal de Pernambuco, 50670-901 Recife, Pernambuco, Brazil
[3]Departamento de Física, Universidade Federal de Sergipe, 49100-000, São Cristóvão, Sergipe, Brazil
[4]Departamento de Estatística, Universidade Federal de Pernambuco, 50740–540, Recife, Pernambuco, Brazil
[*]giovani.vasconcelos@ufpr.br

ABSTRACT

We apply a versatile growth model, whose growth rate is given by a generalised beta distribution, to describe the complex behaviour of the fatality curves of the COVID-19 disease for several countries in Europe and Asia. We show that the COVID-19 epidemic curves not only may present a subexponential early growth but can also exhibit a similar subexponential (power-law) behaviour in the saturation regime. We argue that the power-law exponent of the latter regime, which measures how quickly the curve approaches the plateau, is directly related to control measures, in the sense that the less strict the control, the smaller the exponent and hence the slower the diseases progresses to its end. The power-law saturation uncovered here is an important result, because it signals to policymakers and health authorities that it is important to keep control measures for as long as possible, so as to avoid a slow, power-law ending of the disease. The slower the approach to the plateau, the longer the virus lingers on in the population, and the greater not only the final death toll but also the risk of a resurgence of infections.

Introduction

The COVID-19 pandemic has been recognised as one of the gravest public health crises in recent history. As of this writing, more than 15 million infections have been confirmed worldwide and over 620,000 deaths have been attributed to the disease. Although many countries, notably in Far-East Asia and Europe, have been able to control somewhat the spread of the disease, the SARS-CoV-2 virus is still advancing in many parts of the world—an alarming situation that prompted the WHO Director-General to state on July 9, 2020, that "in most of the world the virus is not under control. It is getting worse."[1] Furthermore, countries that have gained some control over the epidemic are not entirely safe, as there persists a potential risk of resurgence of infections, since a large percentage of their population remains susceptible to the virus.

It is therefore important for governmental and health authorities not only to adopt measures to contain the early spread of the virus, but also to keep those measures in place, in one way or another, for as long as possible, even after the worst phase of the epidemic has been overcome, so as to bring the advance of the disease to a near halt as quickly as possible. In this context, where the epidemic remains a danger from beginning to end, it is important to have models that can give a full description of the epidemic dynamics: from its early rapid-growth phase to the late saturation regime, in which the cumulative number of cases

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/10.1101/2020.07.12.20152140; this version posted July 27, 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. All rights reserved. No reuse allowed without permission.

or deaths tends to a levelling plateau.

Epidemics, in general, and the COVID-19 pandemic, in particular, exhibit a complex dynamics that results from the intertwined interaction between the virus's modus operandi and the human responses: the virus propagates in a complex network of human contacts, which is affected by containment measures (or lack thereof), which in turn impacts back on the virus rate of spread, and so on. Power-law behaviour is a distinct signature of complexity, and in such a guise they are expected to be present in epidemic dynamics as well. Evidences of complexity in the COVID-19 epidemic dynamics have indeed been observed in a number of recent works, most noticeably in the form of a power-law behaviour in the early growth regime of case and death curves[2–7], but also in the long-time asymptotic of the probability to become infected[8–10]. It has also been pointed out that such complex behaviour lies outside the range of applicability of standard SIR-type models, and a number of alternative dynamics have been suggested, such as the recent proposal of a random walk on a hierarchic landscape of social clusters[11–13].

Phenomenological growth models offer a simplified but powerful tool[14] to describe the complexity of Covid-19 epidemic data. On the one hand, their simplicity stems from the fact they are formulated in terms of a single differential equation that often admits an exact analytical solutions, which renders such models quite amenable to both mathematical analysis and numerical application to empirical data. On the other hand, their power derives from the fact that the model parameters seek to capture (in an effective manner) the basic aspects of the underlying epidemic dynamics, but without having to describe epidemiological mechanisms that may be difficult to identify[15]. Nevertheless, variations in the mechanisms of the disease spread, say, owing to demographics or economic factors as well as to different response interventions, will be naturally reflected in changes of the model parameters from group to group. Thus, when properly applied and interpreted, growth models can provide useful insights into the spread of novel infectious diseases[16].

In this paper we apply a generalised logistic model[17], which we refer to as the *beta logistic model* (BLM), to study the fatality curves of COVID-19, as represented by the cumulative number of deaths as a function of time, for different countries. As will be argued later, the BLM is one of the simplest growth models that is capable of capturing the three main generic features of a cumulative epidemic curve: i) an initial subexponential growth; ii) an intermediate linear regime; and iii) a final subexponential saturation trend toward the plateau; see Fig. 1 for a schematic of a generic epidemic curve. (Exponential behaviour can be attained in both early and late regimes as limiting cases.) We show that the BLM describes very well the mortality curves of the six countries selected here to exemplify the model, namely: Germany, Italy, Japan, Netherlands, Sweden, and the United Kingdom. In particular we show that these countries exhibit a polynomially slow approach to the plateau, in the sense that the growth rate asymptotically decays in time as a power-law, rather than exponentially fast as predicted by mechanistic models of the SIR type.

In this context, it is important to bear in mind that there is an intrinsic connection between mechanistic epidemic models of the SIR type and growth models. For instance, it is sometimes possible to map the entire curve of some growth models onto the corresponding curve of a SIRD-type model[18]. We remark, however, that given the polynomial nature of the early and late time evolution that we find in the analysed data, the target SIRD-type model, to be put in correspondence with the BLM, would have to be modified, for example, by the inclusion of a power law in the incidence term, as proposed by[19], or by considering time-dependent parameters, as studied in[20]. This interesting possibility of connection between generalised SIRD models and the BLM is, however, beyond the scope of the present paper and is left for future studies. (For a brief preliminary discussion on this topic, see the Appendix.)

The results reported in the present paper are important for several reasons. First, we show that countries that have kept stricter measures for longer periods tend to experience a faster approach to the plateau; whereas countries with less strict measures throughout the epidemic or that relaxed those measures at

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20152140; this version posted July 27, 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. All rights reserved. No reuse allowed without permission.

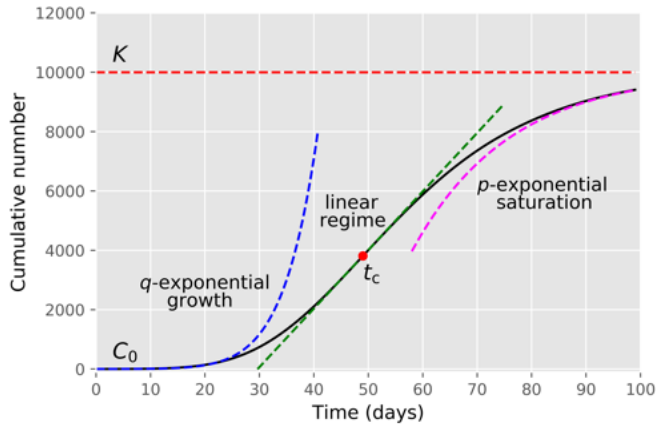

Figure 1. Schematic of a typical epidemic curve for the cumulative number of cases or deaths, where the three main regimes of growth are indicated: i) early q-exponential growth; ii) intermediate region of linear growth; and iii) p-exponential saturation towards the plateau. Here C_0 is the value at $t = 0$, K is the value at the end of the epidemic (plateau), and t_c denotes the inflection point of the curve.

an earlier stage usually display a slow saturation regime. This slow progress to the end of the epidemic represents an extra danger, for the longer the virus lingers on in a population, the greater the risk of a resurgence of cases. Second, our results are also relevant from a more fundamental viewpoint, since they reveal a seemingly new power-law dynamical regime in the final stages of the COVID-19 epidemics. Third, we hope that our findings will stimulate further research on the problem of better characterising the final phase of epidemic dynamics.

Results

Mathematical Results

We model the time evolution of the number of cases (infections, deaths, etc) in the epidemic by means of the beta logistic model (BLM), defined by the following differential equation[17]:

$$\frac{dC}{dt} = r\,[C(t)]^q \left[1 - \left(\frac{C(t)}{K}\right)^{\alpha}\right]^p, \tag{1}$$

where $C(t)$ is the cumulative number of cases at time t, r is the growth rate at the early stage, and K is the final epidemic size. The exponent q controls the initial growth regime and allows to interpolate from linear growth ($q = 0$) to subexponential growth ($q < 1$) to purely exponential growth ($q = 1$); whilst the exponent p controls the rate of approach to the plateau, with $p > 1$ implying a polynomial rate, whereas $p = 1$ yields an exponentially fast approach (see below). The exponent α is the so-called asymmetry parameter that controls the degree of asymmetry with respect to the symmetric S-shape of the logistic curve, which is recovered for $q = p = \alpha = 1$, so that the larger the exponent α, the quicker the curve bends towards the plateau.

The model (1) can be integrated exactly to yield the following analytical solution in implicit form, in terms of the curve $t(C)$:

$$t = \frac{C^{1-q}}{r(1-q)}\,{}_2F_1\left(p, \frac{1-q}{\alpha}; 1 + \frac{1-q}{\alpha}; \left(\frac{C}{K}\right)^{\alpha}\right) - t_i, \tag{2}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20152140; this version posted July 27, 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. All rights reserved. No reuse allowed without permission.

where ${}_2F_1(a,b;c;x)$ is the Gauss hypergeometric function[21] and

$$t_i = C_0{}^{1-q}\,{}_2F_1\left(p, \frac{1-q}{\alpha}; 1+\frac{1-q}{\alpha}; \left(\frac{C_0}{K}\right)^{\alpha}\right).$$

That the solution of the BLM comes in implicit form does not represent any numerical impediment, as the above solution can easily be applied, say, for curve-fitting purposes by viewing the empirical data in the same 'implicit' form $t(C)$. To the best of our knowledge, the exact solution for the BLM given in (2) has not appeared before in the literature. Exact solutions for certain particular cases have been obtained before[17,22], but it appears that an analytical solution for arbitrary values of q, p, and α was not known previously.

A detailed analysis of the BLM is left for the section Methods. For our purposes here it suffices to quote some important properties of the solution (17). First, the inflection point of the curve $C(t)$, which is defined as the point where the second derivative vanishes, i.e. $\ddot{C}(t_c) = 0$, is given by

$$t_c = \frac{K^{1-q}}{r(1-q)}\left(\frac{q}{q+\alpha p}\right)^{(1-q)/\alpha}\,{}_2F_1\left(p, \frac{1-q}{\alpha}; 1+\frac{1-q}{\alpha}; \frac{q}{q+\alpha p}\right) - t_i. \tag{3}$$

The small- and large-times asymptotic behaviour of $C(t)$ are found to be

$$C(t) \approx \begin{cases} At^{\mu}, & \text{for} \quad r^{-1} \ll t \ll t_c, \\ K - \dfrac{B}{t^{\nu}}, & \text{for} \quad t \gg t_c, \end{cases} \tag{4}$$

where $\mu = 1/(1-q)$, $A = [r(1-q)]^{1/(1-q)}$, $\nu = 1/(p-1)$, and $B = \left[K^{1-q}/(p-1)r\alpha^p\right]^{1/(p-1)}$. (For $q \to 1$ and $p \to 1$, one obtains exponential growth and exponential decay, respectively; see Methods.) The behaviour of $C(t)$ near the inflection point t_c is given by

$$C(t) = C(t_c) + \dot{C}(t_c)(t-t_c) + \frac{1}{6}\dddot{C}(t_c)(t-t_c)^3 + \mathscr{O}((t-t_c)^4), \tag{5}$$

where

$$C(t_c) = K\left(\frac{q}{q+\alpha p}\right)^{1/\alpha}, \tag{6}$$

$$\dot{C}(t_c) = \frac{rK^q q^{q/\alpha}(\alpha p)^p}{(q+\alpha p)^{p+q/\alpha}}, \tag{7}$$

$$\dddot{C}(t_c) = -\frac{r^3 K^{3q-2} q^{1+(3q-2)/\alpha}(\alpha p)^{3p}}{p(q+\alpha p)^{3p-1+(3q-2)/\alpha}}. \tag{8}$$

In particular note that $C(t_c)$ does not depend on the growth rate r, as expected, since r only sets the time scale of the problem. Furthermore, for large values of α and p, the growth 'velocity' $\dot{C}(t_c)$ at the inflection point becomes simply $\dot{C}(t_c) \approx rK^q$, which indicates that the slope of the curve $C(t)$ at the inflection point is mainly controlled by r, together with q and K, but largely independent of both α and p. Similarly,

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20152140; this version posted July 27, 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. All rights reserved. No reuse allowed without permission.

for large α, the jerk (or jolt) $\dddot{C}(t_c)$ does not depend on p: $\dddot{C}(t_c) \approx -r^3 K^{3q-2} q\alpha$, which confirms the fact anticipated above that (for fixed r, q, and K) the exponent α is the main responsible for controlling the 'bending' of the curve towards the plateau.

The above summarized analysis thus shows that each of the main features of an epidemic curve can be said to be mainly governed by a specific parameter of the BLM, as follows: i) K is, of course, the plateau height; ii) q determines the nature of the early subexponential growth; iii) r is responsible for determining (together with q and K) the slope of the linear region; iv) α controls the sharpness of the bending towards the plateau; and v) p characterises the polynomial saturation regime.

Application to Empirical Data

As discussed above, the BLM is capable of capturing both a subexponential growth and a subexponential approach to the plateau (with exponential behaviours contemplated in the limits $q, p \to 1$). This makes the BLM quite suitable for situations where the epidemic has reached the intermediate to final stage, such that the corresponding epidemic curve has a well defined inflection point and is clearly trending toward a possible plateau. With this in mind, we have chosen countries whose accumulated curves of death due to COVID-19 are deemed to be either near or at least not far from the plateau. As representative examples of these situations we have chosen the following countries for the present study: Germany, Italy, Japan, Netherlands, Sweden, and United Kingdom.

In Fig. 2 we show the epidemic curves (red circles) for the selected countries, together with the corresponding best fits (black solid curves) by the BLM. From these figures, one sees that the theoretical curves describe remarkably well the empirical data for all cases considered. The corresponding fitting parameters, together with their respective 95% Confidence Intervals (CI), are given in table 1, where we also show the respective inflection points t_c, as obtained from (3).

In each of the insets of the plots presented in Fig. 2, we show the end 'tail' of the quantity $\varepsilon(t) = K - C(t)$, where K is the plateau of the epidemic as estimated by the BLM. The red circles in the insets are the values of $\varepsilon(t)$ obtained from the empirical curves in the corresponding main plots and the black solid curves represent the respective theoretical predictions from the BLM. The insets are in log-log scale, so that a power law decay for $\varepsilon(t)$ should appear as a straight downward line. One sees from the insets of Fig. 2 that the slow, power-law approach to the plateau predicted by the BLM is indeed well observed in all selected countries.

We recall that the closer the exponent p is to 1, the closer the late dynamics is to an exponential approach to the plateau. Conversely, larger values of p indicate a slow polynomial saturation of the epidemic curve. If we use as a comparison measure the lower value of the 95% CI for the exponent p, which determines the exponent ν of the power-law decay of $\varepsilon(t)$, see (4), we can broadly divide the selected countries in three different groups: i) countries, such as Netherlands and Italy, with $1 < p < 1.1$ ($\nu > 10$), indicating a fast, near-exponential saturation approach to the plateau, in the sense that the power-law decay is so fast that it would be difficult to distinguish it from an exponential law within a limited data interval; ii) countries, such as Germany and the United Kingdom, with $1.1 < p < 2$ ($10 < \nu < 1$), indicating a power-law decay not as fast as the near-exponential of the previous case but faster than a hyperbolic decay; and iii) and countries, such as Sweden and Japan in the present study, for which $p > 2$ ($\nu < 1$), which implies a decay slower than hyperbolic. Of course, the 'boundaries' between these groups are admittedly somewhat arbitrary and are meant only to help to differentiate between a fast, intermediate, and slow approach to the plateau.

It is worth pointing out that although Sweden and Japan both have $p > 2$, the nature of the epidemic dynamics in these two countries are quite different, as can be seen from the shapes of the curves shown in Fig. 2e and 2f. In Sweden one sees a well formed power-law regime that starts approximately in the

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20152140; this version posted July 27, 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. All rights reserved. No reuse allowed without permission.

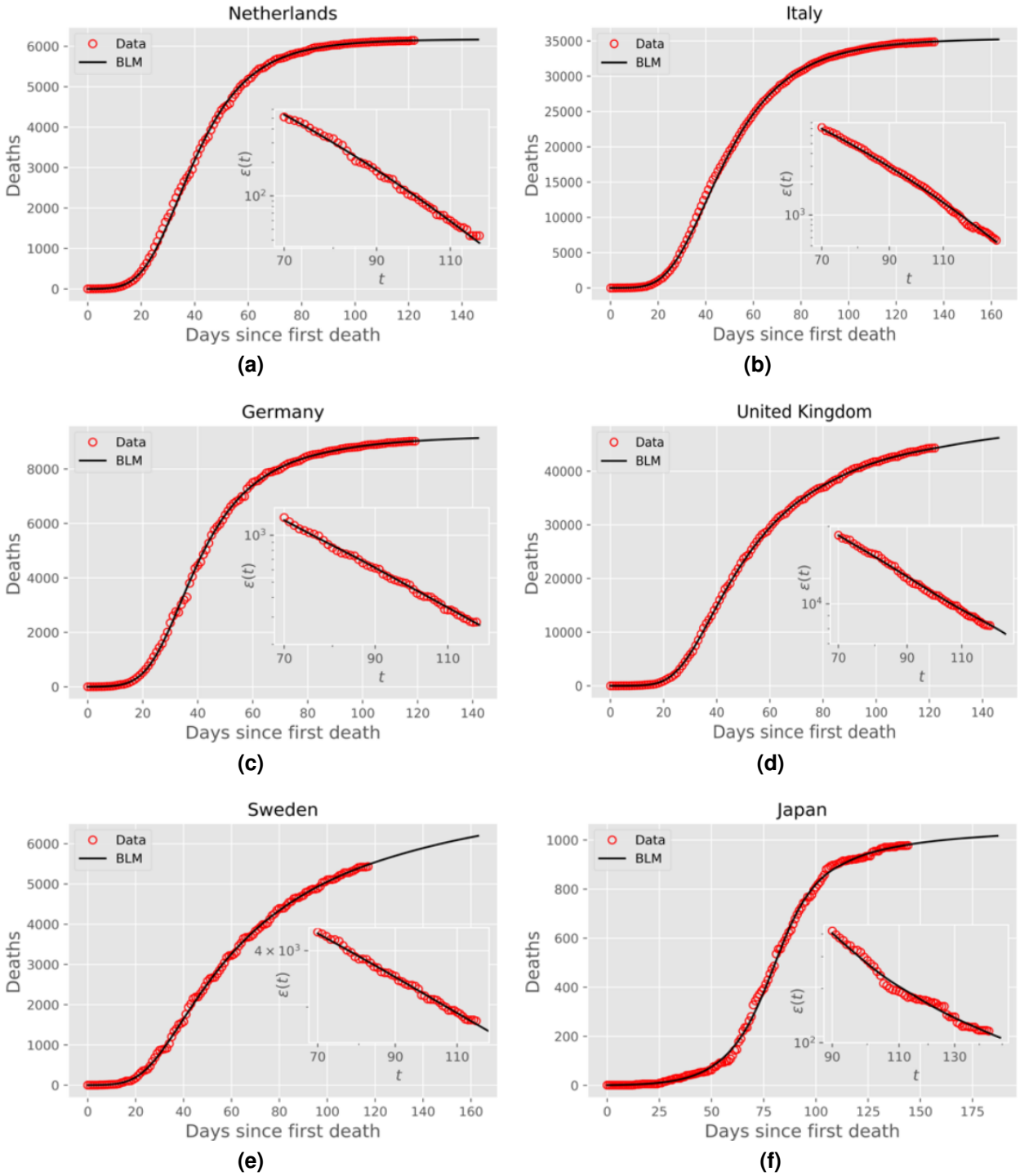

Figure 2. Cumulative number of deaths (red circles) attributed to COVID-19 up to July 7, 2020, for (a) Netherlands, (b) Italy, (c) Germany, (d) United Kingdom, (e) Sweden, and (f) Japan. The solid black curves are the best fits by the beta logistic model, with the corresponding parameters given in table 1. The insets show the respective plots, in log-log scale, of the tails of the quantity $\varepsilon(t) = K - C(t)$; in this scale, a power-law decay appears as a downward straight line, which seems to be the case for all cases shown.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20152140; this version posted July 27, 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. All rights reserved. No reuse allowed without permission.

middle of the curve; whereas Japan's empirical curve has a long linear region, as indicated by the large value of $t_c = 80$, followed by a somewhat sudden bend in the curve, only after which does the curve seem to enter a saturation regime. The fact that the a power-law regime is not yet quite well established in Japan is also reflected in the quite large upper value of the 95% CI for the p-exponent; see table 1. As more data is accumulated, a better estimate for p (in the sense of a smaller 95% CI) is expected to follow. We remark furthermore that the sharp knee seen in Japan's empirical curve around $t = 100$ is reflected in a high value of the exponent α, see table 1, which is the parameter that controls the quickness of the bending toward the plateau.

We have seen above that the BLM is a very flexible model which is capable of describing epidemic curves with quite distinct profiles: in all cases shown in Fig. 2 the model predicts very well the entire dynamical evolution of the fatality curves, from beginning to end. In cases where the epidemic curve is either in the early growth phase or has not clearly entered a saturation regime, the BLM is of course not appropriate, so that an attempt to fit the empirical data with formula (2) will either not converge (former case) or produce unreliable estimate for the tail exponent p (latter situation). In such cases, one can use other simpler growth models, such as the generalised growth model ($p = 0$) or the Richards model ($q = p = 1$), to gain some insight into the early to intermediate stages of the disease[7]. Once the epidemic reaches a later phase, the BLM should then be used for a more complete description of the full evolution. (An automated version of this approach is currently available via an online app[23].)

Discussion

We have seen above that the cumulative death curves of COVID-19 for many countries exhibit a slow, subexponential approach to the plateau, contrarily to the exponentially fast saturation predicted by most epidemiological models, such as mechanistic models of the SIR type. It is thus reasonable to argue that this slower than exponential saturation of the COVID-19 dynamics stems from the complex human response in the late phase of the epidemics. It is normally the case that countries impose stringent containment measures in the early and intermediate stages of the epidemics, which are then progressively relaxed once the worst phase of contagion has been left behind. It is therefore expected that countries that have kept stricter measures, of one sort or another, for longer periods should experience a faster approach to the end (plateau) of the epidemics; whereas countries with more relaxed sets of measures should display the opposite trend (slow saturation regime).

This assumption is largely supported by the data analysis reported above. For instance, Sweden, which has adopted relatively less strict measures throughout the pandemic, displays the slowest approach to the plateau, i.e., the highest exponent p, among the European countries shown in Fig. 2. Conversely, Italy, which enforced one of the longest and most stringent lockdown, has one of lowest value of p, indicating a fast, near-exponential saturation onto the plateau. The same can be said of the Netherlands that took quite severe measures to enforce social distancing (the so-called "1.5-meter rule", whereby failure to keep this safe distance could be punishable by fine[24]). Germany also shows a somewhat small p value, but significantly larger than Italy's—a difference that might be owing to the fact that Germany eased some of the more restrictive measures sooner than Italy. It is important to emphasise, however, that the exponent p characterises only the final part of the epidemic curve and hence it is not strongly related to how well countries have been able to control the epidemic in terms, say, of mortality rates. In other words, countries that have low mortality rates can nonetheless display comparatively higher values of p; and vice-versa[24].

The total number of fatalities that occur in the late phase of the epidemic is, of course, quite smaller than that in the early and intermediate stages. Nonetheless, it is important for authorities to aim at a high 'rate of approach' to the plateau (i.e., small p), because the longer the virus lingers on in the population,

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20152140; this version posted July 27, 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. All rights reserved. No reuse allowed without permission.

the greater the risk of a resurgence of infections, which may in turn prolong even more the epidemic. More to the point, once countries have manged to 'bend the curve', they should try to climb to the plateau as close to exponentially fast as possible. Failing to do that may significantly increase the total number of victims at the end of the pandemic. It is also worth noting that the subexponential growth recently documented in COVID-19 epidemic data has been usually attributed to the early adoption of mitigation measures[7,25,26]. On the flip side of the coin, the subexponentially slow approach to the plateau that we have reported here seems to be, in part, a consequence of relaxing these measures.

Table 1. Parameter estimates of the BLM for the selected countries, with the respective 95% Confidence Intervals given in parenthesis. The inflection point t_c is obtained after inserting the respective parameter estimates in formula (3).

Country	q	r	α	p	K	t_c
Netherlands	1	0.60 (0.48, 0.71)	0.15 (0.12, 0.29)	1.09 (1.07, 1.11)	6176 (6171, 6181)	34.1
Italy	1	0.959 (0.957, 1.29)	0.0839 (0.083, 0.173)	1.123 (1.098, 1.124)	35420 (35402, 35501)	39.9
Germany	1	0.35 (0.34, 0.37)	0.36 (0.33, 0.53)	1.36 (1.32, 1.40)	9300 (9275, 9334)	33.7
United Kingdom	1	0.43 (0.41, 0.45)	0.33 (0.30, 0.50)	1.79 (1.71, 2.02)	52548 (51528, 54397)	37.8
Sweden	1	0.36 (0.35, 0.40)	0.353 (0.346, 0.508)	2.25 (1.95, 2.41)	8234 (8213, 9187)	39.2
Japan	0.91 (0.88, 0.95)	0.106 (0.0976, 0.114)	2.9 (1.6, 4.7)	2.80 (1.40, 19.1)	1094 (1012, 1655)	80.2

Conclusion

The relevance of epidemiological models often lies in the dynamical regimes and scenarios they are able to predict and/or explain, rather than in the specific values of the parameters they produce. In this perspective, the main contribution of the present paper is to predict a novel behaviour for the COVID-19 epidemic, namely: a polynomially slow approach of the growth curve toward the plateau at the end of the epidemic. This implies in turn that the rate of growth (i.e., the daily number of deaths) asymptotically diminishes subexponentially in time, contrarily to the exponentially fast decay predicted by standard compartmental models.

Here we have modelled the fatality curves of COVID-19 by means of a generalised logistic model that is able to capture not only a possible subexponential early growth regime but also a subexponential saturation dynamics in the final phase of the epidemics. In particular, our model shows that the 'residual' number of deaths, defined as the difference between the expected final death toll and the current number of deaths, decays in time as a power-law. We have furthermore argued that the less strict the control measures in the late phase of the epidemic, the smaller the respective power-law exponent and hence the slower the approach to the plateau. This, in turn, implies a greater risk of a resurgence of infections, as the virus continues to propagate in the population (even if at a slower pace) for a longer period of time. Although here we have restricted ourselves to mortality data, it is expected that a similar power-law behaviour should also hold for epidemic curves of confirmed cases.

The results reported in the present paper are relevant both from for a basic viewpoint, in that they show that additional care is needed when modelling the terminal phase of epidemics, and from a practical perspective, for they remind policymakers and health authorities that it is important not only 'to bend the curve,' but also to make sure that it reaches its dynamical end as quickly as possible.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20152140; this version posted July 27, 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. All rights reserved. No reuse allowed without permission.

Methods

Mathematical Model

As indicated before, in this paper we model the fatality curves of the COVID-19 epidemic via the growth model defined in Eq. (1). This equation must be supplemented with an initial condition, which we take as follows:

$$C(0) = C_0, \tag{9}$$

for some given value of C_0. The model defined in (1) is one of the most general growth models in that it recovers, as particular cases, several other known models, namely: i) the logistic model ($q = p = \alpha = 1$); ii) the Richards model ($q = p = 1$); iii) the generalized Richards model ($p = 1$); iv) Blumberg's hyperlogistic equation ($\alpha = 1$); v) the generalised logistic model ($p = \alpha = 1$); and vi) the generalised growth model ($p = 0$).

Equation (1) was introduced by Tsoularis and Wallace[17] and was referred to by these authors as the generalised logistic function. However, the term generalised logistic model has been more commonly used in connection with the particular case $\alpha = p = 1$[27]. Here we propose instead the terminology *beta logistic model* (BLM), because the right-hand-side of (1) corresponds precisely to the generalised beta distribution of the first kind[28,29], which is defined by the following probability density function (pdf):

$$GB1(x;a,b,c,k) = \frac{ax^{ab-1}(1-(x/k)^a)^{c-1}}{k^{ab}B(b,c)}, \tag{10}$$

for $0 < x < k$ and $a,b,c,k > 0$, where $B(b,c)$ is the beta function defined as

$$B(b,c) = \frac{\Gamma(b)\Gamma(c)}{\Gamma(b+c)}, \tag{11}$$

with $\Gamma(\cdot)$ denoting the gamma function. In view of (10), Eq. (1) can be rewritten as

$$\frac{dC}{dt} = \frac{rK^{q+1}B\left(\frac{q+1}{\alpha}, p+1\right)}{\alpha} GB1\left(C;\alpha,\frac{q+1}{\alpha},p+1,K\right). \tag{12}$$

Viewing the right-hand side of (1) as a pdf is useful to gain some insights into the underlying growth dynamics. Seen in this perspective, the model (1) encodes the fact that the rate of growth of the disease (be it measured in number of cases or deaths) at a given time t is proportional to the product of two terms: i) the first term represents the current 'strength' of disease, thus being related to the number of cases in the population; and ii) the second term accounts for the available 'space' in the population for the disease to grow. Note, however, that in contrast to the standard logistic model where both such product terms are linear in $C(t)$, i.e., $q = p = \alpha = 1$, the BLM seeks to take into account, albeit in an effective and indirect manner, the fact that the virus propagates in a complex network of human contacts, which in turn is affected by the very measures adopted to curtail the spread of the disease. This complex dynamics is captured in the BLM via the introduction of the exponents q, p, and α, which results in the generalised beta distribution.

It turns out that the model (1) can be integrated exactly. To see this, we start by introducing the reduced number of cases

$$c(t) = \frac{C(t)}{K}, \tag{13}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20152140; this version posted July 27, 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. All rights reserved. No reuse allowed without permission.

so that (1) becomes

$$\frac{dc}{dt} = rK^{q-1}c^q\,(1-c^\alpha)^p\,, \tag{14}$$

After a trivial separation of variable, we write

$$\frac{dc}{c^q(1-c^\alpha)^p} = rK^{q-1}dt \tag{15}$$

which upon integration on both sides yields

$$c^{1-q}\,{}_2F_1\left(p,\frac{1-q}{\alpha};1+\frac{1-q}{\alpha};c^\alpha\right) = rK^{q-1}(1-q)t + \text{constant}, \tag{16}$$

where ${}_2F_1(a,b;c;x)$ is the Gauss hypergeometric function[21]. After fixing the constant of integration with (9), we obtain

$$c^{1-q}\,{}_2F_1\left(p,\frac{1-q}{\alpha};1+\frac{1-q}{\alpha};c^\alpha\right) = rK^{q-1}(1-q)(t+t_i), \tag{17}$$

where

$$t_i = \frac{C_0^{1-q}}{r(1-q)}\,{}_2F_1\left(p,\frac{1-q}{\alpha};1+\frac{1-q}{\alpha};\left(\frac{C_0}{K}\right)^\alpha\right). \tag{18}$$

Combining (17) and (18), and returning to the original variable $C(t)$, we obtain the solution given in (2).

The BLM as defined in (17) has two 'extensive' parameters, namely C_0 and K, which represent the initial and final number of cases, respectively, and four 'intensive' parameters, namely: r, q, α, and p, which are related to the intrinsic dynamics of the epidemic in each particular group of individuals. It is therefore instructive to seek a better understanding of the dynamical (and epidemiological) roles that these parameters play.

The growth rate r sets the time scale of the problem, and as such it reflects the epidemiological conditions in the early growth phase of the epidemic. In other words, r is expected to be directly related to the basic reproduction number, R_0, since the higher the R_0, the faster the virus spreads and the larger the growth rate r should be. (Broadly speaking, R_0 can be defined as the expected number of secondary infections caused by an primary infectious individual in an entirely susceptible population.)

The parameter q, as already anticipated, is related to the nature of the growth regime (if exponential or subexponential) in the beginning of the epidemic. To see this, first note that if $C \ll K$, the ODE (1) yields the so-called generalised growth model[27]:

$$\frac{dC}{dt} = r\,[C(t)]^q\,, \tag{19}$$

whose solution is simply

$$C(t;r,q,C_0) = \left[C_0^{1-q} + (1-q)rt\right]^{1/(1-q)} = C_0\exp_q\left(\frac{rt}{C_0^{1-q}}\right), \tag{20}$$

where the function $\exp_q(x) = [1+(1-q)x]^{1/(1-q)}$ is known in the physics literature as the q-exponential function[30,31]. The parameter q thus characterises the early growth dynamics, allowing the model to include

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20152140; this version posted July 27, 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. All rights reserved. No reuse allowed without permission.

from a linear growth ($q = 0$) to a subexponential regime ($0 < q < 1$) and to a fully exponential growth ($q = 1$). (Values of q greater than one in (20) are ruled out, because they would lead to a superexponential spread of the disease, which is neither epidemiological reasonable nor mathematically acceptable, as it would lead to a singularity in finite time.) In particular, note that according to (20) the early growth of the epidemic curve is approximately polynomial for $0 \leq q < 1$, as given in the top equation in (4). Polynomial early growth has been observed across a range of epidemics before[14]. It has also been recently identified in the COVID-19 data for several countries[7,25,27].

In the intermediate region of the epidemic curve, one observes an approximately linear regime around the inflection point t_c, with a small third-order deviation from a straight line. More specifically, from (1) one readily finds that the expansion of $C(t)$ about $t = t_c$ is given by

$$C(t) = K\left(\frac{q}{q+\alpha p}\right)^{1/\alpha} + \frac{rK^q q^{q/\alpha}(\alpha p)^p}{(q+\alpha p)^{p+q/\alpha}}(t-t_c) - \frac{r^3K^{3q-2}q^{1+(3q-2)/\alpha}(\alpha p)^{3p}}{6p(q+\alpha p)^{3p-1+(3q-2)/\alpha}}(t-t_c)^3 + \mathscr{O}((t-t_c)^4), \tag{21}$$

which corresponds to the expression given in (5)–(8).

Finally, in the long-time asymptotic limit, i.e., when $C(t) \to K$, we can investigate the behaviour of the BLM curve by introducing the small quantity $\varepsilon(t) = K - C(t)$. Inserting this definition in (1) and considering $\varepsilon \ll 1$, we obtain to leading order:

$$\frac{d\varepsilon}{dt} = -\gamma\varepsilon^p, \tag{22}$$

where $\gamma = r\alpha^p/K^{1-q}$, which upon integration yields

$$\varepsilon(t) = \frac{1}{[D+(p-1)\gamma t]^{1/(p-1)}} = D^{1/(1-p)}\exp_p\left(-\frac{\gamma t}{D}\right), \tag{23}$$

with D denoting an arbitrary constant of integration. This shows that in general the BLM curve $C(t)$ approaches the plateau K according to a p-exponential function (i.e., the q-exponential function defined before with $q = p$). [Eq. (22) also appears, e.g., in the problem of protein folding/unfolding, leading to q-exponential relaxation phenomena there as well.[32]]. In particular notice that, for sufficiently large times t, the curve $C(t)$ generically approaches the plateau polynomially slowly in time, in the sense that $\varepsilon(t)$ decays in time as a power law:

$$\varepsilon(t) \approx \frac{B}{t^\nu}, \qquad \text{for} \quad t \to \infty \quad \text{and} \quad p > 1, \tag{24}$$

$B = [(p-1)\gamma]^{1/(1-p)}$, which recovers the bottom equation in (4). In contrast, only for $p = 1$ does one get an exponentially fast saturation into the plateau: $\varepsilon(t) \propto \exp(-\gamma t)$. It follows from (24) that the growth rate asymptotically decays as a power-law for large times:

$$\frac{dC}{dt} \propto \frac{1}{t^{\nu+1}}, \qquad \text{for} \quad t \to \infty, \tag{25}$$

which shows that the corresponding daily curves also display power-law decay for large times. In this context, it is worth recalling that compartmental models of the SIRD type predict an exponential decay for the rates of both new infections and deaths toward the end of the epidemic.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20152140; this version posted July 27, 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. All rights reserved. No reuse allowed without permission.

Data and Statistical Fits

Here we first briefly note that because the number of confirmed cases is a poor metric for the advance of the COVID-19 epidemic, owing to the fact that a large proportion of infections go undetected, we have adopted the fatality curves as our primary objects of study, since it is widely recognised that the death count is a more reliable statistics in this case[7]. The data used in this study were obtained from the database made publicly available by the Johns Hopkins University[33]. We have used data for the total number of deaths up to July 07, 2020, for the following countries: Germany, Italy, Japan, Netherlands, Sweden, and United Kingdom.

To perform the statistical fits reported in this paper, we employed the Levenberg-Marquardt algorithm to solve the non-linear least square optimisation problem, as implemented in the *lmfit* package for Python[34], which not only provides parameter estimates but also determines their respective confidence levels. Technically, because the exact solution for the BLM is given in implicit form, we have to fit the analytical formula given in (2) to the empirical datasets seen as curves of the type $t(C)$. But this imposes no difficulty in the numerical routine whatsoever. To reduce the number of free parameters in the model we have set $C(0) = C_0$, where C_0 is the number of deaths recorded at the first day that a death was reported, so that we were left with five parameters, namely (r, q, α, p, K), to be determined numerically. Growth models with several parameters are known to be susceptible to the problem of overfitting and parameter redundancy[7,18,27,35]. To minimise this risk we have adopted some precautions, as discussed next.

First, we limited the exponent q to the mathematically and epidemiological acceptable range $0 \leq q \leq 1$. When the *lmfit* routine returns $q = 1$, it usually cannot estimate the confidence levels because this is the upper bound of q; in this case we assume that $q = 1$ is an 'exact value' for the country in question. As for the remaining parameters, no restriction was enforced *a priori* other than their natural ranges ($r, \alpha, K > 0$ and $p > 1$). Nevertheless, we checked the estimates returned by the curve-fitting routine against possible inconsistencies. For example, the growth rate r is expected to be typically within the interval (0,1)[7] and that has been the case for all countries considered here. Similarly, the asymmetry parameter α is typically smaller than 1[7], but countries that display a somewhat 'sharp' bending after the inflexion point (as Japan, see below) are expected to have higher values of α. In the same spirit, we checked the values of the exponent p, which determines the power-law decay of the $\varepsilon(t)$ given in (24), by comparing the tails of $\varepsilon(t)$ for both the empirical data and the theoretical prediction. Good agreements between the BLM and the data, over the entire, including the tails, were observed in all cases reported here, which validates the estimates of the model parameters.

As already indicated, the computer codes for the statistical fits were written in the *Python* language, and the plots were produced with the data visualisation library *Matplotlib*. All numerical codes and corresponding data sets used in this paper are available in our group's software repository at `fisica.ufpr.br/redecovid19/software` or can be requested from the authors. The fitting-curve routine for the BLM can be automatically tested for other countries (not included here) via an app available online[23].

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20152140; this version posted July 27, 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. All rights reserved. No reuse allowed without permission.

Appendix: A SIRD Model with Power-Law Behaviour

We start by recalling the standard Susceptible (S)-Infected (I)-Recovered (R)-Deceased (D) epidemiological model[36,37]

$$\frac{dS(t)}{dt} = -\frac{\beta S(t)I(t)}{N} \tag{26}$$

$$\frac{dI(t)}{dt} = \frac{\beta S(t)I(t)}{N} - (\gamma_1+\gamma_2)I(t) \tag{27}$$

$$\frac{dR(t)}{dt} = \gamma_1 I(t) \tag{28}$$

$$\frac{dD(t)}{dt} = \gamma_2 I(t), \tag{29}$$

where $S(t)$, $I(t)$, $R(t)$, and $D(t)$ are the number of individuals at time t in the classes of susceptible, infected, recovered, and dead respectively, while N is the constant total number of individuals in the population, so that $N = S(t)+I(t)+R(t)+D(t)$. The parameters β, γ_1 and γ_2 are the transmission, recovery and death rates respectively. The initial values can be chosen to be $S(0) = S_0$, $I(0) = I_0$, with $S_0 + I_0 = N$, and $R(0) = 0 = D(0)$.

We then consider a modified SIRD model, where in (26) and (27) we replace N with only the partial population in the S and I compartments[38]. Furthermore, we follow Refs.[19,39] and replace the term $I(t)$ on the right-hand side of all equations above by $[I(t)]^p$, to obtain

$$\frac{dS}{dt} = -\frac{\beta S(t)}{I(t)+S(t)}[I(t)]^p, \tag{30}$$

$$\frac{dI}{dt} = \frac{\beta S(t)}{I(t)+S(t)}[I(t)]^p - (\gamma_1+\gamma_2)[I(t)]^p, \tag{31}$$

$$\frac{dR}{dt} = \gamma_1[I(t)]^p, \tag{32}$$

$$\frac{dD}{dt} = \gamma_2[I(t)]^p. \tag{33}$$

Although this model is still not general enough to accommodate all the phenomenology of the intervention biased dynamics of the COVID-19 epidemics, it does nonetheless exhibit subexponential behaviour for both short and large time scales in all compartments. In order to show this, we define $y(t) = S(t)+I(t)$ and divide (31) by (30) to obtain

$$\frac{dy}{y} = \frac{1}{R_0}\frac{dS}{S}, \tag{34}$$

where $R_0 = \beta/(\gamma_1+\gamma_2)$. Integrating both sides of (34), and inserting the result into (30), yields an equation of the BLM type:

$$\frac{dS}{dt} = -r[S(t)]^q\left[1-\left(\frac{S(t)}{K}\right)^{\alpha}\right]^p, \tag{35}$$

where $r, K > 0$, $q = 1+(p-1)/R_0$ and $\alpha = 1-1/R_0$. It follows from (35), in comparison with model (1), that $S(t)$ exhibits power-law regimes (for both early and large times) akin to those described by the BLM. Furthermore, it is easy to see that all other compartments, $I(t)$, $R(t)$, and $D(t)$, inherit from $S(t)$ the

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20152140; this version posted July 27, 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. All rights reserved. No reuse allowed without permission.

power-law behaviour, even though their respective equations of motion are not of the BLM-type. Although the parameters q, p and α in (35) are not all independent of one another, the preceding qualitative argument shows nonetheless that the power-law dynamics of the sort predicted by the BLM can in principle be accommodated by compartmental models. We are currently conducting further research in this direction.

References

1. Ghebreyesus, T. A. WHO Director-General opening remarks at the Member State Briefing on the COVID-19 pandemic evaluation. https://www.who.int/dg/speeches/detail/who-director-general-opening-remarks-at-the-member-state-briefing-on-the-covid-19-pandemic/-evaluation---9-july-2020 (2020). Retrieval date: July 9, 2020.
2. Verma, M. K., Asad, A. & Chatterjee, S. COVID-19 pandemic: Power law spread and flattening of the curve. *Transactions Indian Natl. Acad. Eng.* 1–6 (2020).
3. Singer, H. M. The COVID-19 pandemic: growth patterns, power law scaling, and saturation. *Phys. Biol.* (2020).
4. Merrin, J. Differences in power-law growth over time and indicators of COVID-19 pandemic progression worldwide. *medRxiv medRxiv:2020.03.31.20048827* (2020).
5. Ziff, A. L. & Ziff, R. M. Fractal kinetics of COVID-19 pandemic. *medRxiv medRxiv:2020.02.16.20023820* (2020).
6. Chatterjee, S. *et al.* Evolution of COVID-19 pandemic: Power law growth and saturation. *medRxiv medRxiv:2020.05.05.20091389* (2020).
7. Vasconcelos, G. L. *et al.* Modelling fatality curves of COVID-19 and the effectiveness of intervention strategies. *PeerJ* **8**, e9421, DOI: 10.7717/peerj.9421 (2020).
8. Blasius, B. Power-law distribution in the number of confirmed COVID-19 cases. *arXiv preprint arXiv:2004.00940* (2020).
9. Fukui, M. & Furukawa, C. Power laws in superspreading events: Evidence from coronavirus outbreaks and implications for SIR models. *medRxiv medRxiv:2020.06.11.20128058* (2020).
10. Beare, B. K. & Toda, A. A. On the emergence of a power law in the distribution of COVID-19 cases. *arXiv preprint arXiv:2004.12772* (2020).
11. Khrennikov, A. Ultrametric model for COVID-19 dynamics: An attempt to explain slow approaching herd immunity in Sweden. *Preprints* DOI: 10.20944/preprints202007.0037.v1 (2020). 2020070037.
12. Volov, V. & Zubarev, A. Toward ultrametric modeling of the epidemic spread. *arXiv preprint arXiv:2005.08761* (2020).
13. Nguyen, N. *et al.* Modeling in the time of COVID-19: Statistical and rule-based mesoscale models. *arXiv preprint arXiv:2005.01804* (2020).
14. Chowell, G., Sattenspiel, L., Bansal, S. & Viboud, C. Mathematical models to characterize early epidemic growth: A review. *Phys. of Life Reviews* **18**, 66–97 (2016).
15. Bürger, R., Chowell, G. & Lara-Díaz, L. Y. Comparative analysis of phenomenological growth models applied to epidemic outbreaks. *Math. Biosci. Engin: MBE* **16**, 4250–4273 (2019).
16. Chowell, G. *et al.* Using phenomenological models to characterize transmissibility and forecast patterns and final burden of Zika epidemics. *PLOS Curr. Outbreaks* (2016).

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20152140; this version posted July 27, 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. All rights reserved. No reuse allowed without permission.

17. Tsoularis, A. & Wallace, J. Analysis of logistic growth models. *Math. Biosciences* **179**, 21–55 (2002).

18. Wang, X.-S., Wu, J. & Yang, Y. Richards model revisited: Validation by and application to infection dynamics. *J. Theor. Biol.* **313**, 12–19 (2012).

19. Chowell, G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: a primer for parameter uncertainty, identifiability, and forecasts. *Infect. Dis. Model.* **2**, 379–398 (2017).

20. Caccavo, D. Chinese and Italian COVID-19 outbreaks can be correctly described by a modified SIRD model. *medRxiv medRxiv:2020.03.19.20039388* (2020).

21. Aomoto, K., Kita, M., Kohno, T. & Iohara, K. *Theory of hypergeometric functions* (Springer, 2011).

22. Turner Jr, M. E., Bradley Jr, E. L., Kirk, K. A. & Pruitt, K. M. A theory of growth. *Math. Biosci.* **29**, 367–373 (1976).

23. Brum, A. A., Duarte-Filho, G. C. & Vasconcelos, G. L. Application Modinterv Covid-19. Available at: https://fisica.ufpr.br/modinterv/ (2020).

24. Netherlands-Government. Dutch measures against coronavirus, Government of the Netherlands. Available at: https://www.government.nl/topics/coronavirus-covid-19/ (2020).

25. Manchein, C., Brugnago, E. L., da Silva, R. M., Mendes, C. F. & Beims, M. W. Strong correlations between power-law growth of COVID-19 in four continents and the inefficiency of soft quarantine strategie. *Chaos: An Interdiscip. J. Nonlinear Sci.* **30**, 041102 (2020).

26. Maier, B. F. & Brockmann, D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in china. *Science* **368**, 742–746 (2020).

27. Wu, K., Darcet, D., Wang, Q. & Sornette, D. Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world. https://www.medrxiv.org/content/early/2020/03/16/2020.03.11.20034363.full.pdf, DOI: 10.1101/2020.03.11.20034363 (2020).

28. McDonald, J. B. Some generalized functions for the size distribution of income. *Econom. J. Econom. Soc.* 647–663 (1984).

29. Kotz, S. & Van Dorp, J. R. *Beyond beta: other continuous families of distributions with bounded support and applications* (World Scientific, 2004).

30. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **52**, 479–487 (1988).

31. Picoli Jr, S., Mendes, R., Malacarne, L. & Santos, R. q-distributions in complex systems: A brief review. *Braz. J. Phys.* **39**, 468–474 (2009).

32. Tsallis, C., Bemski, G. & Mendes, R. S. Is re-association in folded proteins a case of nonextensivity? *Phys. Lett. A* **257**, 93–98 (1999).

33. JHU. Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) . https://coronavirus.jhu.edu/map.html (2020). Accessed: 2020-03-29.

34. Newville, M., Stensitzki, T., Allen, D. & Ingargiola, A. Non-linear least-squares minimization and curve-fitting for Python. *Chicago, IL* (2015).

35. Hsieh, Y.-H. Richards model: a simple procedure for real-time prediction of outbreak severity. In *Modeling and dynamics of infectious diseases*, 216–236 (World Scientific, 2009).

36. Anderson, R. M. *The population dynamics of infectious diseases: theory and applications* (Springer, 2013).

medRxiv preprint doi: https://doi.org/10.1101/2020.07.12.20152140; this version posted July 27, 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. All rights reserved. No reuse allowed without permission.

37. Diekmann, O., Heesterbeek, H. & Britton, T. *Mathematical tools for understanding infectious disease dynamics*, vol. 7 (Princeton University Press, 2012).

38. Wang, X.-S., Wu, J. & Yang, Y. Richards model revisited: Validation by and application to infection dynamics. *J. Theor. Biol.* **313**, 12–19 (2012).

39. Tirnakli, U. & Tsallis, C. Epidemiological model with anomalous kinetics - the Covid-19 pandemics. *medRxiv medRxiv:2020.06.24.20139287* (2020).