Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Effective epidemic model for COVID-19 using accumulated deaths

G. Nakamura\textsuperscript{a, b, *}, B. Grammaticos\textsuperscript{a, b}, C. Deroulers\textsuperscript{a, b}, M. Badoual\textsuperscript{a, b}

\textsuperscript{a} Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405 Orsay, France
\textsuperscript{b} Université de Paris, IJCLab, 91405 Orsay, France

\begin{abstract}
The severe acute respiratory syndrome COVID-19 has been in the center of the ongoing global health crisis in 2020. The high prevalence of mild cases facilitates sub-notification outside hospital environments and the number of those who are or have been infected remains largely unknown, leading to poor estimates of the crude mortality rate of the disease. Here we use a simple model to describe the number of accumulated deaths caused by COVID-19. The close connection between the proposed model and an approximate solution of the SIR model provides estimates of epidemiological parameters. We find values for the crude mortality between 10^{-4} and 10^{-2} which are lower than estimated numbers obtained from laboratory-confirmed patients. We also calculate quantities of practical interest such as the basic reproduction number and subsequent increment after relaxation of lockdown and other control measures.
\end{abstract}

1. Introduction

Outbreaks of infectious diseases have been a common occurrence throughout history, often linked to or followed by disruptions in societies and human activities [1]. There are several ways to measure the impact of outbreaks but death tolls are the most relevant ones whenever the disease can threaten lives. For instance, 32 million persons have died between 1981 and 2019 in the ongoing HIV epidemics, 700000 in 2018 alone [2]. Aside from this large-scale epidemic, the world has experienced several other recent outbreaks with varying degrees of severity and scale: Zika fever, whose symptoms are mild but can produce long-lasting effects in newborns (microcephaly) [3,4]; Ebola virus disease, with a high mortality rate estimated between 20 and 75% [5,6]; Swine flu/H1N1, which became a pandemic in 2009-2010 although with a lower mortality rate than regular flu [7]. In 2019-2020, the severe acute respiratory syndrome COVID-19 has emerged as the most recent pandemic, caused by the virus named SARS-CoV-2 [8,9]. Due to its novelty and lack of previous exposition, humans have no immunity against this threat, leading to an increased number of infections. At the time of this writing, the specifics of the pathogen transmission are still being investigated, as well as the complete infection process once the virus enters the host. However, it has been shown that the main human-to-human transmission mode occurs by the spreading of contaminated droplets, similar to other flu-like diseases [8]. In sharp contrast with H1N1, however, the mortality rate of COVID-19 was first estimated in the range 1-4%, with higher ill-outcomes among persons of old age [10,11]. This situation creates a unique scenario where healthcare facilities and workers can be overwhelmed in a short period, ultimately leading to untreated patients of COVID-19 as well as other diseases [12].

To make matters even worse, asymptomatic patients can spread the pathogen for an extended period of time, showing none or mild symptoms during the course of the infection. As a result, laboratory tests to detect the viral load are necessary to identify the correct number of cases outside hospital environments. Similar to the H1N1 pandemic, the required number of tests far exceeds the current amount available in most countries. Without timely tracking of new cases, contact tracing becomes a challenging task, hindering estimates of new cases per infection summarized by the basic reproduction number, $R_0$. The significance of this parameter lies in the fact that it provides a way to glimpse the values of the transmission rates. Those can then be used in compartmental models – mathematical models that describe the evolution of epidemics assuming nearly homogeneous populations [13,14]. Earlier estimates for $R_0$ using epidemiological data from Wuhan, China, set $R_0$ between 1.5 to 3.7 without additional measures to restrict the spreading [15–17]. With measures in place – such as lockdown, self-isolation, and social distancing – $R_0$ was estimated to be around 1.05 [15]. More importantly, the insufficient number of tests, in addition to the long waiting time for lab results, affects the quality of epidemic models.

In the absence of mass testing, the death toll can be used as an alternative metric to probe the extension of the epidemic. The medical staff can assess the cause of death from clinical re-
ports, which may contain test results or not, using the best of their knowledge. Additional tests may also be appended to reports to further specify the cause of death.

Both numbers of cases and deaths are publicly available as part of a global effort to tackle the pandemic. Here, we study the evolution of COVID-19 deaths in order to reduce the issues caused by the limited number of laboratory-confirmed tests. We show that the accumulated deaths can be effectively described by simple functions, namely, sigmoids whose parameters are explained in terms of the SIR epidemic model. The SIR model is a compartmental model with the following health states: susceptible, infective, and removed [18]. The removed state represents those who have passed away or have developed immunity, either by recovering from the disease or any other method such as vaccination. The SIR model was chosen in preference to epidemic models with additional health states or reinfections because it is the simplest model that addresses immunity, requiring fewer parameters and whose solutions can be approximated by simpler functions. Among our results, we show that crude mortality rates can be computed from the parameters of sigmoids and that the rates can change up to one order of magnitude, depending on the severity of the outbreak in a given region. The paper is organized as follows. Sec. 2 contains the description of the data and variables used along the text. Sec. 3 explains how the SIR model is reduced from a system of differential equations to a single non-linear differential equation, with emphasis on the expansion around equilibrium. Data is modeled in Sec. 4 via sigmoidal functions, whose parameters are explained in terms of the epidemiological parameters of the SIR model. Time windows are addressed in Sec. 5, with special emphasis on $R_0$, crude mortality rate, and quantitative effects of the confinement. In Sec. 6, we investigate the evolution of $R_0$ once control measures are relaxed using the results from previous sections as input. Final comments and conclusions are listed in Sec 7.

2. Data

The European Center for Disease Prevention and Control (ECDC) provides COVID-19 data updated in a daily schedule [19]. The daily reports portray the distribution of new confirmed cases and new deaths presented as time series. The dataset also displays the population size $N$ according to 2019 World Bank census for each geographical region.

To better grasp the nature of the data, consider the number of accumulated deaths in France, as shown in Fig. 1. Similar to other European countries, France was heavily affected by the pandemic, above the mark of 30 000 deaths with a sharp increase in deaths of infected patients around April-May. These deaths are linked to infections that took place between 3 to 6 weeks prior. On March 16, the French government implemented measures to mitigate the propagation of the disease. The measures included confinement of non-essential workers, temporary closures of schools and universities as well as commercial stores and services. The impact of said measures did not show up immediately on the data (see Fig. 1b) but instead they appeared after some time had passed, around 4 weeks, reducing the number of daily fatalities. The death toll experienced a rapid growth between March and April (see Fig. 1). This regime is the hallmark of epidemics and denotes the exponential phase. In general, the growth rate in compartmental epidemic models is summarized by $R_0$ that is the ratio of the transmission rate $\beta$ to the removal rate $\lambda$. As the name implies, the transmission rate dictates the average number of persons a given infective typically infects during a fixed time interval. The inverse of the removal rate is the characteristic time in which a person remains contagious. Using Fig. 1 as reference, the number of accumulated deaths describes a monotonic function with early exponential growth followed by a rapid deceleration. This pattern suggests sigmoids as ideal candidates to model the data.

In addition to interventions designated to curb local transmissions, travel restrictions were put in place to reduce the circulation of the virus between cities and to/from different countries. Although a clear quantification of the impact of these restrictions still requires the complete picture of the epidemic, they serve as a mechanism to address and limit the influence of spatial effects for a fixed period of time. After that, we must assume that spatial effects are part of the spreading dynamics and become once again relevant. Finally, we used France as a concrete example but in reality the vast majority of countries adopted similar measures in the same time frame, with few notable exceptions like Sweden.

3. SIR model

We seek general features that can be used to model outbreaks with the available data. These features concern the spreading regime of the disease among the population but exclude demographics and spatial effects. Demographics entail the various age brackets, their distribution, their average number of contacts with elements of the population, and effects of the disease on each age bracket. Spatial effects arise due to heterogeneous distribu-
tions of population densities, as well as patterns and routes of transmissions, or imported cases. Although the inclusion of data portraying the characteristics of each age bracket and the spatial distribution of the population provide a far more realistic description of the problem, it also implies additional parameters in the model and methods to extract them.

Instead, simpler models involving statistically equivalent individuals are useful to explain general features of the outbreak. That is the case of the classical compartmental models. In the case of COVID-19, the health-disease process can be crudely described by a progression through a sequence of different health stages. Starting from a healthy individual, exposition to the pathogen may trigger an infection with either mild or harsh symptoms. In both cases, virus shedding can occur before symptoms become apparent resulting in asymptomatic transmissions. If the infection persists, severe clinical symptoms may follow, including pneumonia and shortness of breath which may require hospitalization and, eventually, lead to death.

For the sake of simplicity, let us assume the spreading dynamics is approximated by the standard SIR model in a homogeneous population of size $N$ [20,21]. The model comprises a population whose subjects can be classified in three distinct health states, namely, susceptible, infective, and removed. The removed state includes individuals that have either died or recovered from the disease. The latter are assumed to not be infective anymore, nor susceptible to become sick again because of some kind of immunity. The fraction of individuals in each compartment is, respectively, $S(t)$, $I(t)$, and $R(t)$, at time instant $t$. The dynamic goes as follows. Infective subjects in the population transmit the pathogen to susceptible ones, under adequate conditions. The transmission occurs with rate $\beta$, and we assume the homogeneous mixture of the population, that is, each person in the population is statistically equivalent to each other [14]. Once infected, the person remains infective for an average period $1/\lambda$, where $\lambda$ is the removal rate. The dynamical equations that describe the model are:

$$\frac{dS}{dt} = -\beta S(t)I(t),$$

$$\frac{dI}{dt} = +\beta S(t)I(t) - \lambda I(t),$$

$$\frac{dR}{dt} = +\lambda I(t),$$

with the constraint $S + I + R = 1$, i.e., conservation of the population size. The model simplifies or neglects aspects of the COVID-19 pandemic such as differentiation between asymptomatic and symptomatic transmission or age-dependent rates [22,23]. In fact, research on the biological characteristics of the pandemic are still ongoing [24], but evidence indicates re-infections should be minimal in recovered patients [25]. Therefore, we make a case for an approximate description of the problem via the SIR model over the inclusion of extra complexities and uncertainties, aiming to capture dominant aspects.

The system of differential equations (1a–1c) can be further reduced to a single first-order differential equation as follows. From (1c) and (1a), one finds $S(t) = S_0 e^{-\beta \lambda R(t)}$. The constant $S_0 = S(0) e^{-(\beta \lambda /R(0))}$ depends on the initial conditions $S(0)$ and $R(0) = 1-S(0)-I(0)$. Usually, we are more interested in scenarios in which $R(0) = 0$ and thus $S_0 = 1-I(0)$, similar to the onset of an emerging disease. The complete expression for $S_0$ must be used for different initial conditions, which becomes a problem whenever the ratio $\beta /\lambda$ is unknown. Ignoring constant solutions, it can be shown [26] that the equation can be further reduced to

$$\frac{dR}{dt} = -\lambda S_0 e^{-\beta \lambda R(t)} - \lambda R + \lambda,$$

whose general solution can be obtained by quadrature

$$t - t_0 = \frac{1}{\lambda} \int_{R(0)}^{R(t)} \frac{dr}{1 - r - S_0 e^{-\beta \lambda r}}. $$

(3)

The stationary condition in (2) gives the value $R_\infty$ as a solution of the transcendental equation $R_\infty = 1 - S_0 e^{-\beta \lambda R_\infty}$. The solution $R(t)$ can be obtained by inverting (3), for which a general method remains unknown. To circumvent the issue, one may expand (2) around points of interest, for example, around $R = 0$ or $R = R_\infty$. Each expansion has advantages and issues. The former was first introduced by Kermack and McKendrick [18] to study small outbreaks and replaces $e^{-\beta \lambda R}$ by its expansion around $R = 0$ up to second order in $(2)$. The linear term in the expansion dictates the initial exponential growth of $R(t)$ with an effective rate $\beta S_0 - \lambda$ while the quadratic term produces the asymptotic value $R_\infty \approx (2/\beta S_0)(\beta S_0 - \lambda)(\lambda/\beta)^2$. The remaining constant term $\lambda(1-S_0)$ accounts for a temporal shift but, in general, $R = 0$ approximates the onset of the epidemic so that $S_0 \approx 1$. Crucially, the condition $R(t) \ll 1$ must hold throughout the full duration of the outbreak which is incompatible with large scale epidemics. Examples are quite common, in fact. One may consider the common cold or seasonal influenza whose typical reproduction number $R_0 \approx 1.3$ [27] implies around 40% of the population become part of the removed compartment by the end of the outbreak as illustrated in Fig. 2. Unfortunately, for epidemics, the competition between the early exponential growth and relaxation towards the asymptotic solution is often difficult to assess near the onset and requires further powers of $R(t)$.

Alternatively, we can get a better picture of the problem by expanding $R(t)$ around $R = R_\infty$. There are two immediate advantages in this approach. First, it already carries information about the asymptotic solution $R_\infty$, so the constraint $R(t) \ll 1$ is no longer necessary. In practice, this means the expansion can be used to model large scale epidemics. Second, as we shall see further below, the convergence to the asymptotic solution does not depend on the initial conditions, which is the case of the growth rate $\beta S_0 - \lambda$ in the previous expansion. One could argue that is just a matter of a shift in time. However, communicable diseases with varying degree of symptoms complicate the problem and introduce an new source of uncertainty to the data. More specifically, patients with mild symptoms are unlikely to search for immediate care and be included in surveillance databases during the early stages of the outbreak. The missing data leads to the under-notification of cases and outcomes which, coupled with large ratios ($\beta /\lambda$), creates large deviations in $S_0$ and predictions of the model, except in the case in which the onset of the outbreak is known a priori.

As in the previous case, we shall limit the expansion around $R = R_\infty$ up to quadratic terms in order to produce analytical results and look for possible applications. Define $\delta R(t) = R_\infty - R(t) \geq 0$. The expansion of (2) near $\delta R \ll 1$. together with the transcendental equation for $R_\infty$, gives

$$1 + \frac{d}{dt} \delta R = -\left[1 - (1 - R_\infty) \frac{\beta}{\lambda}\right] \delta R + \frac{1 - R_\infty}{2} \left(\frac{\beta}{\lambda}\right)^2 \delta R^2 + o(\delta R^3).$$

(4)

Keeping terms up to $o(\delta R^2)$, one converts (4) into a Bernoulli equation with relaxation time $\tau = [\lambda - \beta (1 - R_\infty)]^{-1}$. Solving for $\delta R$ and transforming back to $R$, we find the approximate solution (see Fig. 2)

$$R(t) = R_\infty - A e^{-t/\tau} \frac{1 + B e^{-t/\tau}}{1 + B e^{-t/\tau}},$$

(5)

with $A = R_\infty - (1 + B) R(0)$, $B = c_0 \tau /2\lambda$, $c_0 = (1 - R(0))^{-1} - c_0$, and $c_0 = (\lambda/2)(1 - R_\infty)(\beta/\lambda)^2$. 

G. Nakamura, B. Grammaticos, C. Deroulers et al.
Chaos, Solitons and Fractals 144 (2021) 110667
3
4. Effective model

The sigmoidal expression in (5) satisfies the requirements listed in Section 2 and is an excellent candidate to model accumulated deaths. However, the removed compartment corresponding to \( R(t) \) holds both recovered and deceased fractions of the population, i.e., all the infected who are unable to spread the disease. We can simplify this issue by assuming the existence of a simple relation between \( R(t) \) and the accumulated number of deaths divided by the population size, \( g(t) \), at the time instant \( t \). To keep the model as simple as possible, we neglect temporal delays and impose

\[
g(t) = fR(t),
\]

where the crude mortality rate \( f \) is the ratio between deceased and infected. As such, it also can be used as an estimator for the likelihood to die after contracting the disease.

The equilibrium value \( g_\infty \) varies from country to country and can be used to characterize the outbreak, more specifically, to assess the impact of the outbreak in the afflicted population. The monotonic nature of cumulative quantities together with the upper bound \( g_\infty \) restrict the possible functional forms for \( g(t) \). Sigmoidals are natural candidates to describe \( g(t) \) since they are bounded and monotonic. Here we consider the following expression in tandem with (5):

\[
g_{\text{eff}}(t) = g_\infty - a e^{-t/\tau} \cdot \frac{1}{1 + b e^{-t/\tau}}.
\]

The timescale \( \tau \) dictates the relaxation of \( g(t) \) towards \( g_\infty \), with infection at \( t_e = \tau \ln b \). Equation (7) has 3 or 4 parameters depending on whether \( g_{\text{eff}}(t) \) must pass by the initial entry \( g_{\text{eff}}(0) = g(0) = g_0 \) or not. The former implies the constraint \( a = g_\infty - g_0(1 + b) \), whereas the later \( g_{\text{eff}}(0) = (g_\infty - a)/(1 + b) \) with \( a < g_\infty \). Since the data are noisy, we do not require the fitting curve to pass by \( g_0 \).

The relationship between \( g(t) \) and \( R(t) \) is consistent if the following equations hold:

\[
\frac{1}{\tau} = \lambda - \beta \left( 1 - \frac{g_\infty}{f} \right),
\]

\[
g_\infty = f - fS_0 e^{-\beta/\lambda} g_\infty,
\]

\[
b = \frac{(f - g_\infty)(g_\infty - g_{\text{eff}}(0))\beta^2}{2\omega \lambda^2 f^2 - (f - g_\infty)(g_\infty - g_{\text{eff}}(0))\beta^2}.
\]

The system of equations (8a-8c) connects the epidemiological parameters \( (\beta, \lambda, f, S_0) \) with the parameters \( (\tau, g_\infty, b) \) of the sigmoid curve (5), which can be estimated by a fitting procedure, for instance, least-square. However, there are more variables than equations so at least one epidemiological variable must be fixed. There are two reasonable choices, namely, either \( S_0 \) or \( \lambda \). The first choice should be selected if the data includes the onset of the outbreak because \( S_0 \approx 1 \), since the majority of the population should be in the susceptible state at the early stage of the epidemic. However, if the beginning of the outbreak is unknown, the assumption \( S_0 \approx 1 \) is no longer valid. Alternatively, one may consider an estimate for the removal rate \( \lambda \), or its probability distribution. In this case, the input parameter to solve (8a-8c) depends solely on the characteristics of the disease and local demographics. It does not require details concerning the disease spreading nor the moment in which the outbreak begins. Therefore, evidence-based values for \( \lambda \) from patient data are far more suitable for the purposes of this study, and shall be used hereafter.

5. Time windows

Fig. 3 shows the French death toll from March 16 to May 25. We use the opportunity to check whether the sigmoid (7) or the SIR model (1) can describe the data. It becomes clear that the agreement between data and (7) (performed with a standard nonlinear least-square fitting procedure) is poor. In particular, the fitted curve becomes negative in early March. Parameter optimization of the complete SIR model (1) produce satisfactory results, except for the first 20 days. Furthermore, the optimized parameter \( \lambda_{\text{opt}} = 0.103 \text{ days}^{-1} \) does not match the signature value for COVID-19 \( \lambda = 0.2 \text{ day}^{-1} \). If the value \( \lambda = 0.2 \text{ days}^{-1} \) is fixed during the optimization, then the optimal set of parameters become highly sensitive to the initial guess in the optimization algorithm.

The experiment suggests that the SIR model (1) approximates the data but naive optimizations can return unrealistic epidemiological parameters. On the other hand, the sigmoid (7) only describes a portion of the data. This was not the case with the synthetic SIR data shown in Fig. 2. The inability to reproduce the entire real-world dataset can be explained by assuming that epidemiological parameters have changed at some point during the course of the data collection. In a sense, the simplicity of (7) lacks the flexibility to balance two sets of epidemiological parameters, which can be exploited as a tool to detect changes in
epidemiological regimes. Indeed, the assumption finds support in the data of daily number of deaths (see Fig. 1b and Fig. 4). An asymmetry is observed around the peak of daily deaths, which is in sharp contrast with the corresponding curve in the theoretical SIR model (see Fig. 2 b). The asymmetry may be explained by a variation in the transmission rate. Unlike other recent epidemics, several countries have adopted control measures such as lockdowns of non-essential workers, flight and other travel restrictions. The efforts effectively reduce the transmission rate of COVID-19 once they are in place. Thus, the data must be divided in non-overlapping time windows, each with its own set of epidemiological parameters.

As a first example, consider the Danish death toll between April 4 and May 31 as depicted in Fig. 5. The time window is within the second epidemiological regime observed in Fig. 4 which simplifies the analysis. The data are fitted via (7) with parameters \( g_\infty, a, b \) and \( \tau \), using trust region reflective algorithm (Python/Scipy). It is convenient to fit the data using \( \tau^{-1} \) instead of \( \tau \) and a finite interval \( 0 \leq \tau^{-1} \leq 1 \), so the infection lasts at least one day in the mathematical model. In the fitting procedure, we use bounds for \( b \) which depends on the time interval \( t_c \) between the infection point and the initial entry, more specifically, \( t_c = r \ln b \). For \( t_c > 0 \) we restrict the search to \( t_c/\tau < 10 \) so that \( b \sim o(10^3) \); for \( t_c < 0 \), i.e., starting the counting after the inflection point, the parameter space of \( b \) is limited between 0 and 1. The remaining parameters are restricted to \( 0 \leq g_\infty, a \leq 1 \).

The fitting (7) is in excellent agreement with data for \( g_\infty = (1.03 \pm 0.01) \times 10^{-4} \) (approximately 600 deaths) and \( \tau = 15.84 \pm 0.65 \) days. Next, we solve the system (8a-8c) for \( \lambda \) distributed according to some probability distribution, centered around \( \lambda = 0.2 \) days \(^{-1} \). In practical terms, the width of the probability distribution forms the ground to compute uncertainties of the transmission and crude mortality rates. However, the finer details of the distribution of \( \lambda \) remain unknown so we resort to a uniform distribution for virus shedding, whose interval lies between 3 and 7 days [22,28]. By doing so, we overestimate the uncertainties of the remaining epidemiological parameters of the SIR model and find \( R_0 = 0.81 \pm 0.07 \). Our estimate for the crude mortality rate \( f = (6.8 \pm 1.5) \times 10^{-4} \) is compatible with the estimate \( f_{DKN} = 8.2 \times 10^{-4} \) (confidence interval: [5.9 – 15.4] \times 10^{-4} \) obtained by screening antibodies of 20 640 blood donors below the age of 70, in Denmark [29]. Also, the asymptotic value \( R_0 = g_\infty/f = 0.15 \pm 0.03 \) places the fraction of infected well below the threshold (60%) for herd immunity.

We can now move to more complicated cases, in which the data themselves contains artifacts and the fitting procedure can be tricky. This is the case of France, for instance. The lockdown was issued on March 16, and unaccounted deaths in nursing homes were added to the official statistics on April 3, and 4, producing a large fluctuation in the number of daily deaths. In this case, we separate the time windows according to the number of deaths. The first time window lies between 100 and 5000 deaths (March 16 - April 3), whereas the remaining days until May 31 comprises the second time window, as shown in Fig. 6.

The curve fitting in the second time window is far more stable and insensitive to initial guesses, or bounds, with equilibrium death toll nearing 29440 ± 158 and time scale \( \tau = 15.12 \pm 0.72 \) days, which approaches the typical recovery time for mild COVID-19 infections. The solution of the system (8a-8c) returns \( R_{0,k} = 0.74 \pm 0.08 \leq 1 \) in agreement with the decline of new infections, where the notation \( R_{0,k} \) with \( k = 1 \) and 2 refers to the first and second time windows, respectively. In addition, the crude mortality rate in the second time window reads \( f_2 = (4.67 \pm 1.16) \times 10^{-3} \), one order of magnitude higher than Denmark or Germany (see Table. 1).
The curve fitting in the first time window, which includes the phase with exponential growth, can be challenging though. Even more so if the fitting data contains a large number of entries around the inflection point induced by the introduction of control measures. The inflection point is not the natural inflection point which occurs in an uncontrolled epidemic with constant parameters. Due to the significance of inflection points in the curve fitting, the induced inflection point artificially lowers $g_\infty$ in the first time window. Furthermore, multiple solutions can be found for (8a–8c), several of which are not realistic. Instead of using brute force, it is far more convenient to approximate the solution and set $S_0 = 0.99$. In such case, (8c) is discarded and the remaining equations produce $R_{0,1} = 3.25 \pm 1.71$. The crude mortality rate in the first time window, $f_1 = (4.23 \pm 0.58) \times 10^{-3}$, shares the same order of magnitude as $f_2$, indicating that overall the French health care system remained responsive throughout the epidemic. If control measures had not been implemented, the expected number of deaths would have soared and reached 240 000 and $R_\infty = 0.96$, assuming the mortality rate would have remained roughly the same.

6. What happens next?

In the previous section, the second window was delimited from April 4 to May 31 which included lockdowns, prohibition of large gatherings, and other non-pharmaceutical interventions to control the virus circulation. The window also coincides with the period in which spatial effects have less influence on the spreading dynamics. In general, spatial effects are important in outbreaks following changes in the chain of transmissions when compared to the random mixing hypothesis. In essence, they introduce correlations between persons which breaks a central assumption in compartmental models. The second window can be seen as the regime in which compartmental models are closest to describe real data. Therefore, estimates of epidemiological parameters produced during the second window should approximate their correct values and remain roughly unchanged throughout the full interval.

What happens after the restrictions are lifted or relaxed? With reduced power to limit virus circulation it is natural to assume that the transmission rate should increase. According to our findings so far (7) should be unable to handle two sets of epidemiological parameters, deviating from the data. Indeed, this outcome is seen in Fig. 7 when one attempts to fit (7) via the least-square procedure from April 4 to September 1, i.e., the second window is extended by three months. By late July, the fitted curve and the data grow with different rates, with the latter being approximately linear.

The next question follows on what are the new parameters on the third window which starts around June 30 and lasts until September 15, the last day in our data. Here we exploit the early growth observed during the last days in Fig. 7 and match that behavior with the expansion of the SIR model around the beginning of the third window. In what follows, primed (non-primed) variables or parameters belong to the third (second) window.

We start by expanding (2) around the beginning of the third window, $t = t_0^\prime$. As briefly explained in Sec. 3, the expansion produces, among other corrections, a linear term whose coefficient is $C'' = S_0'' \beta - \lambda'$. The linear fit of the data produces $C = d'/f'$ in which $d'$ is the linear coefficient and $f'$ is the crude mortality rate for the window. The constant $S_0'' = S(t_0^\prime) e^{d' / f'} R(t_0^\prime)$ depends on $R_0'' = \beta' / \lambda'$ and also on $S(t_0^\prime)$ and $R(t_0^\prime)$, which again are not known directly from the data. With hospitals and other healthcare services operating below their maximum capacity, it is reasonable to assume both removal and crude mortality rates remain unchanged, $\lambda' = \lambda$ and $f' = f$, so that $C'' = d'/f$ and $R(t_0^\prime) = g(t_0^\prime)/f$.  

| Table 1 | Parameters of the SIR model and sigmoid from April 4 to May 31 for selected countries, using a uniform distribution for $\lambda$. |
|---------|------------------------------------------------------|
| $\beta$  | $R_0$      | $f$ ($10^{-3}$) | $S_0$ | $g_\infty$ ($10^{-4}$) |
| France    | 0.17 ± 0.07 | 0.74 ± 0.08 | 4.67 ± 1.16 | 0.97 ± 0.02 | 4.40 ± 0.02 |
| Italy     | 0.19 ± 0.07 | 0.84 ± 0.07 | 4.17 ± 0.78 | 0.96 ± 0.02 | 5.75 ± 0.01 |
| Spain     | 0.19 ± 0.07 | 0.86 ± 0.06 | 1.93 ± 0.54 | 0.87 ± 0.08 | 6.02 ± 0.02 |
| UK        | 0.18 ± 0.07 | 0.82 ± 0.07 | 5.86 ± 1.08 | 0.97 ± 0.01 | 6.26 ± 0.05 |
| Germany   | 0.18 ± 0.07 | 0.83 ± 0.07 | 0.39 ± 0.11 | 0.90 ± 0.06 | 1.05 ± 0.01 |
| Belgium   | 0.17 ± 0.07 | 0.76 ± 0.08 | 4.75 ± 1.35 | 0.93 ± 0.04 | 8.27 ± 0.02 |

Fig. 6. Evolution of COVID-19 in France between March 16 and May 31. Reduced number of data points for clarity. Data (full circles) and sigmoidal fit (solid line) between March 16 and April 3, with $R_{0,1} = 3.25$ and crude mortality rate $f_1 = 4.23 \times 10^{-3}$. The effects of confinement emerge shortly after April 3, with a significant reduction in $R_{0,1} = 0.74$, as indicated by the sigmoidal fit (dashed line) of data in the second time window (empty squares). The mortality rate remains nearly unchanged $f_2 = 4.67 \times 10^{-3}$.

Fig. 7. Deaths in France between March 16 and September 1. The shaded region indicates the second time window. Reduced density of data points (circles) for clarity. The least-square fit of (7) deviates from the data around late June. (inset) Linear fit approximates the growth observed between June 30 and September 1.
Analogously, we evaluate $S(t_0') = S_0 e^{-(\beta'/\lambda)t_0'}$ using the results from the previous time window as input. After plugging the expressions we find

$$\frac{\beta'}{\lambda} e^{(\beta'-\beta)/\lambda} g(t_0') = \lambda \cdot f + a' \lambda / S_0. \quad (9)$$

It is convenient to redefine $x = (\beta'/\lambda)g(t_0')$ and $y = [g(t_0')/S_0](1 + a'/f) e^{(\beta'/\lambda)g(t_0')}$ so that (9) becomes $xe^\lambda = y$, which is the equation satisfied by Lambert W functions. The solution is obtained numerically or graphically and gives the updated value $R_0'$. Table 2 displays $R_0'$ obtained for selected European countries. The analysis shows that the relaxation of control measures lead to an increase in the transmission rate for all countries investigated.

7. Conclusion

The tragic developments in the COVID-19 pandemic have exposed flawed aspects in protocols used to assess large scale epidemics. Despite the various improvements in the global capacity to produce laboratory tests, usually based on reverse-transcription polymerase chain reaction (rRT-PCR), the majority of afflicted countries were unable to enforce early mass testing policies. This shortcoming has also been experienced in 2002 with the SARS epidemic but in a smaller scale. The lack of mass testing contributed in keeping the number of cases unknown, affecting the accuracy of disease spreading models. In this paper, we resort to the death toll as our primary data source as death certificates are mandatory and contain other medical assessments linking the cause of death with COVID-19 outbreaks. We emphasize that this methodology is not immune itself from sub-notification, nor unforeseen delays on death certificates.

We model the death toll via the sigmoid curve in (7), which requires the parameters $g_\infty$ (capacity) and $\tau$ (time scale). The capacity limits the growth of the outbreak, which is expected to vary with geographic region, healthcare quality, and efforts to control the spreading of the virus. Together with the crude mortality rate $f$, which also depends on healthcare and population demographics, they give access to a far more credible estimate for the infected fraction of the population $R_\infty = g_\infty / f$. To keep the model as simple as possible, we also neglect temporal delays between $g(t)$ and $R(t)$. The inclusion of said delay may introduce additional effects, but those are not expected to be dominant in this case since spatial effects are not being investigated.

The advantage of our approach relies on the curve fitting of a monotonic curve (death toll) to a sigmoid in sharp contrast to complex optimization schemes for models with multiple health states [22,23]. Epidemiological parameters are extracted by solving the algebraic system of equations (8a-8c) for transmission rate, crude mortality rate and initial condition, respectively, $\beta$, $f$, and $S_0$. Alternatively, the epidemiological parameters calculated from the curve fitting can be used as educated guesses in optimization algorithms reducing the likelihood of obtaining unrealistic optimal solutions. We stress that if $f$ is known, then the fraction of reported cases can be easily computed $R(t) = g(t)/f$. We find that $f$ ranges from $o(10^{-4})$ to $o(10^{-3})$, depending on geographic location. Such values lie well below recent estimates [30] using only rRT-PCR tests from hospitalized cases (7.2%), whereas it is more in line with the values obtained from antibody screening with large sample size in Denmark [29]. Interestingly, countries lightly afflicted by the epidemic exhibit lower $f$ even though the transmission of the virus is similar to neighbouring countries. Taking Germany or Denmark as examples, we see that both have $f \sim o(10^{-4})$ while maintaining $R_0$ compatible with other European countries. The result is likely attributed to healthcare resources and services either being more accessible or efficient, including early surveillance.

The sigmoidal fitting becomes far more reliable once the outbreak has been active for some time as the data start to move away from the exponential phase. The distance from the inflection point is another important factor that affects the quality of the fit, especially $g_\infty$. The introduction of lockdowns and other social distancing measures reduce the transmission rate and effectively create a new inflection point, which is different from the one expected without control measures. Thus, the fitting curve may converge to an incorrect equilibrium value if the fitting data include points around the induced inflection point. This can only be solved by introducing disjoint time windows for different regimes of $\beta$ as Fig. 6 shows. The difference between equivalent parameters in subsequent windows returns the effectiveness of interventions. Taking France as a concrete example, confinement reduced the basic reproduction number by $\delta R_{21} = -2.51 \pm 1.63$ while lifting and relaxation of restrictions resulted in $\delta R_{12} = 0.27 \pm 0.04$. We stress that parameters calculated within the second time window do not require additional data other than the removal rate. That is not the general case, especially in regimes with exponential growth and an unknown number of susceptible and infected individuals. Strict control measures mitigate uncertainties concerning $g_\infty$ and $\tau$ that often arise in curve fitting procedures. As a result, parameter extraction within time windows with strict control measures can also be used as input in subsequent windows which greatly simplify the analysis if they exhibit exponential growth.

Concerning deviations in parameters estimated through (8a-8c), they can be tracked to the uncertainties in the removal rate. In general, the inverse of the removal rate describes the average time required for an infective person to change to the removed state. For COVID-19, virus shedding occurs more prominently between 3 and 7 days, with a peak at day 5 since the onset of symptoms [28]. The exact probability distribution for $\lambda$ remains an open issue so our analyses assume a uniform distribution.

Finally, the SIR model rests on the random mixing hypothesis but deviations are expected with stronger effects in populations with varying demographics or populations at risk. In particular, communicable respiratory diseases become major issues in correctional facilities, given the lack of adequate environmental and sanitary conditions. The combination of higher transmission rate of pathogens and reduced healthcare can increase the crude mortality rate for incarcerated individuals. By a similar argument, outbreaks in nursing homes may affect estimates of epidemiological parameters because the disease becomes disproportionately more lethal for older patients. In this study, spatial effects are neglected to understand the disease spreading of the average population with the minimal number of parameters possible.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
References

[1] Morabia A. A history of epidemiologic methods and concepts. Basel Boston: Birkauser Verlag; 2004.

[2] WHO. World Health Statistics 2018 : monitoring health for the SDGs : sustainable development goals. Geneva, Switzerland: World Health Organization; 2018.

[3] Mlakar J, et al. Zika virus associated with microcephaly. N Engl J Med 2016;374:951–8. doi:10.1056/NEJMoa1600653.

[4] de Oliveira WK, et al. Infection-related microcephaly after the 2015 and 2016 zika virus outbreaks in brazil: a surveillance-based analysis. Lancet 2017;390:861–70. http://www.sciencedirect.com/science/article/pii/S0140673617313685.

[5] Team WER. West african ebola epidemic after one year - slowing but not yet under control. N Engl J Med 2015;372:584–7. doi:10.1056/NEJM41414992.

[6] Van Kerkhove MD, Bento AI, Milhs HL, Ferguson NM, Donnelly CA. A review of epidemiological parameters from ebola outbreaks to inform early public health decision-making. Sci Data 2015;2:150019 EP –

[7] Simonsen L, et al. Global mortality estimates for the 2009 influenza pandemic from the glamor project: A modeling study. PLOS Med 2013;10:1–17. doi:10.1371/journal.pmed.1001558.

[8] Chan JFW. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 2020;395:514–23. doi:10.1016/S0140-6736(20)30154-9.

[9] Chen N. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study. Lancet 2020;395:507–13. doi:10.1016/S0140-6736(20)30211-7.

[10] WHO. Report of the WHO-china joint mission on coronavirus disease 2019 (covid-19), 2020. Access April 20, 2020.

[11] Baud D. Real estimates of mortality following COVID-19 infection. Lancet Infect Dis 2020. doi:10.1016/S1473-3099(20)30195-X.

[12] Walker P, et al Report 12: The global impact of COVID-19 and strategies for mitigation and suppression. Tech. Rep., Imperial College London 10.25561/77735. (2020).

[13] Keeling M, Eames K. Networks and epidemic models. J R Soc Interface 2005;2:295–307.

[14] Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneus and network models in epidemiology. J R Soc Interface 2007;4:879–91.

[15] Kucharski AJ. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 2020;20:553–8. doi:10.1016/ S1473-3099(20)30144-4.

[16] Li Q. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. N Engl J Med 2020;382:1199–207. doi:10.1056/NEJMoa2001316.

[17] Sanchez S. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. Emerg Infect Dis 2020;26. doi:10.3201/eid2607.200282.

[18] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc R Soc A 1927;115:700–21. http://rspa.royalsocietypublishing. org/content/115/772/700.

[19] European-Center of Disease Prevention and C. Geographic distribution of COVID-19 cases worldwide. 2020. Dataset retrieved from https://opendata. ec.europa.eu/ covid-19/casedistribution/csv.

[20] Cooper I, Mondal A, Antonopoulos CG. A SIR model assumption for the spread of COVID-19 in different communities. Chaos Solitons Fractals 2020;139. doi:10.1016/j.chaos.2020.105657.

[21] Saad-Roy CM. Immune life history, vaccination, and the dynamics of SARS-CoV-2 over the next 5 years. Science 2020. doi:10.1126/science.aba7343.

[22] Prem K. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in wuhan, china: a modelling study. Lancet Public Health 2020. doi:10.1016/S2468-2567(20)30073-6.

[23] Giordano G. Modelling the COVID-19 epidemic and implementation of population-wide interventions in italy. Nat Med 2020. doi:10.1038/s41591-020-0883-7.

[24] Gandhi RT, Lynch JR, del Rio C. Mild or moderate covid-19. N Engl J Med 2020. doi:10.1056/NEJMc2009249.

[25] Ota M. Will we see protection or reinfection in COVID-19? Nat Rev Immunol 2020. doi:10.1038/s41577-020-0316-3.

[26] Harko T, Lobo FS, Mak M. Exact analytical solutions of the susceptible-infected-recovered (sir) epidemic model and of the sir model with equal death and birth rates. Appl Math Comput 2014;236:184–94. http://www.sciencedirect. com/science/article/pii/S009630031400383X.

[27] Biggerstaff M. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. BMC Infect Dis 2014;14:465–9. doi:10.1186/1471-2334-14–480.

[28] Wölfel R. Virological assessment of hospitalized patients with COVID-19. Nature 2020;581:465–9. doi:10.1038/s41586-020-2196-x.

[29] Erikstrup C. Estimation of SARS-cov-2 infection fatality rate by real-time antibody screening of blood donors. Clin Infect Dis 2020. doi:10.1093/cid/ciaa849.

[30] Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in italy. JAMA 2020. doi:10.1001/jama.2020.4683.