Analysis of Covid-19 Data for Eight European Countries and the United Kingdom Using a Simplified SIR Model

Gyan Bhanot1,2*, Charles DeLisi3

1 Department of Physics and Astronomy, Rutgers University, Piscataway, NJ, 08854, USA; Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ, 08854, USA; Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, 08903, USA
2 Moores Cancer Center, University of California San Diego, La Jolla, CA 92103, USA.
3 Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA

* Gyan Bhanot, PhD (corresponding author)
136 Frelinghuysen Road, Busch Campus, Rutgers University, Piscataway, NJ 08854, USA
Phone: (848) 391-7508
Fax: (732) 235-5331
Email: gyanbhanot@gmail.com

Charles DeLisi, PhD
24 Cummington Mall
Boston University, Boston, Ma 02215
Phone 617 875 5138
Email: DeLisi@bu.edu

Abstract
Understanding the characteristics of the SARS-Cov-2/Covid-19 pandemic is central to developing control strategies. Here we show how a simple Susceptible-Infective-Recovered (SIR) model applied to data for eight European countries and the United Kingdom (UK) can be used to forecast the descending limb (post-peak) of confirmed cases and deaths as a function of time, and predict the duration of the pandemic once it has peaked, by estimating and fixing parameters using only characteristics of the ascending limb and the magnitude of the first peak. As with all epidemiological analyses, unanticipated behavioral changes will result in deviations between projection and observation. This is abundantly clear for the current pandemic. Nonetheless, accurate short-term projections are possible, and the methodology we present is a useful addition to the epidemiologist's armamentarium. Since our predictions assume that control measures such as lockdown, social distancing, use of masks etc. remain the same post-peak as before peak, deviations from our predictions are a measure of the extent to which loosening of control measures have impacted case-loads and deaths since the first peak and initial decline in daily cases and deaths. The predicted and actual case fatality ratio, or number of deaths per million population from the start of the pandemic to when daily deaths number less than five for the first time, was lowest in Norway (pred: 44 ± 5 deaths/million; actual: 36 deaths/million) and highest for the United Kingdom (pred: 578 +/- 65 deaths/million; actual 621 deaths/million). The inferred pandemic characteristics separated into two distinct groups: those that are largely invariant across countries, and those that are highly variable. Among the former is the infective period, $T_L (\overline{T_L} = 16.3 \pm 2.7 \text{ days})$; the average time between contacts, $T_R (\overline{T_R} = 3.8 \pm 0.5 \text{ days})$. 

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
and the average number of contacts while infective, $R$ ($R = 4.4 \pm 0.5$). In contrast, there is a highly variable time lag $T_D$ between the peak in the daily number of confirmed cases and the peak in the daily number of deaths, ranging from a low of $T_D = 2.4$ days for Denmark and Italy respectively, to highs of $T_D = 12, 15$ for Germany and Norway respectively. The mortality fraction, or ratio of deaths to confirmed cases, was also highly variable, ranging from low values of 3%, 5% and 5% for Norway, Denmark and Germany respectively, to high values of 18%, 20% and 21% for Sweden, France, and the UK respectively. The probability of mortality rather than recovery was a significant correlate of the duration of the pandemic, defined as the time from 12/31/2019 to when the number of daily deaths fell below 5. Finally, we observed a small but detectable effect of average temperature on the probability $\alpha$ of infection per contact, with higher temperatures associated with lower infectivity. Policy implications of our findings are also briefly discussed.

**Keywords:** Epidemic modelling, SARS-Cov-2/Covid-19 pandemic, SIR model, predictions.
I. Introduction

Coronaviruses are large, enveloped, single-stranded RNA viruses which are widespread in animals and usually cause only mild respiratory illnesses in humans [1-5]. However, in 2003, a new coronavirus SARS-CoV emerged, which caused a life-threatening respiratory disease, with a fatality rate of almost 10% [6,7]. Unfortunately, after an initial burst of interest in development of treatment options, interest in this virus waned. The emergence of a novel coronavirus SARS-CoV-2, identified in January 2020 as the likely causative agent of a cluster of pneumonia cases which first appeared in Wuhan, China in December 2019, has since caused a worldwide pandemic [8-13]. SARS-CoV-2 is the seventh known coronavirus to cause pathology in humans [1]. The associated respiratory illness, called COVID-19, ranges in severity from a symptomless infection [8], to common-cold like symptoms, to viral pneumonia, organ failure, neurological complications and death [9-11]. While the mortality in SARS-CoV-2 infections is lower than in SARS-CoV [9-12], it has more favorable transmission characteristics, a higher reproduction number [13], and as we will show, a long latency period and asymptomatic infective phase.

The governments of several countries have taken significant measures to slow the infection rate of Covid-19, such as social distancing, quarantine, identification, tracking and isolation. However, there is no uniform policy, some governments have reacted later than others and some (e.g. Sweden) made a deliberate decision to keep the country open, leaving counter-measures up to individuals.

A large amount of consistent public data is now available on the number of tests performed, the number of confirmed infected cases, and the number of deaths in different contexts, such as locations and health conditions [14]. These provide important sources of information for the development and testing of models that can identify pandemic characteristics affecting viral dynamics, and guide public policy by predicting the impact of various interventions [15].

All data, of course, have limitations, and it’s ultimately the completeness and quality of data that limit the success of models. It is well known that confirmed infected cases seriously underestimate the actual number of infections [16,17]: not everyone who is infected is symptomatic, and not everyone who dies from the disease has been tested [18]. Even the number of reported deaths may be underestimated because of co-mortalities; i.e. COVID-19 increases susceptibility to other diseases and conditions [19]. Moreover, the virus can be transmitted by asymptomatic individuals – who comprise a substantial portion of the infected population [20]—mitigating against accurate estimates of transmission probability. Nonetheless as indicated in [21] and by our own verified forecasts, models can provide useful information.

Dynamical (mechanistic) models, such as the one presented here, have been used for forecasting (meaning that once initial conditions are set, there are no changes in the model) and for projecting (the outcome is changed by intervention strategies). For example, projections and forecasting models of various types have been used as early as February, 2020 to determine a reproductive number [13]. More generally multiple research groups have used them to estimate Case Fatality Ratios (CFRs) [22], to forecast and project the need for hospital beds [23] and to project and forecast mortality [24]. More specifically, among the many applications to COVID-19, four variable Susceptible-Exposed-Infective-Recovered models have been used to project the
impact of social distancing on mortality [25], three variable Susceptible-Infective-Recovered models to estimate case fatality and recovery ratios early in the pandemic [26], and a time delayed SIR has been used to evaluate the effectiveness of suppression strategies [27]. One of the most ambitious dynamical models, which includes 8 state variables and 16 parameters was fruitfully applied to evaluate intervention strategies in Italy, in spite of the fact that parameter identifiability could not be assured [28].

Our simplified version of the original SIR model [29] differentiates itself from the studies done so far by a unique methodological approach to rigorous identification of parameters, and by making a number of useful predictions very easily, such as the duration of the pandemic in different countries and case fatality ratios, using only characteristics of the increasing portion of verified cases vs time, and the amplitude of the peak in the daily cases. Although it is sometimes said that using daily mortality data is more reliable than daily verified cases, we find that at least in the countries we analyzed, the patterns for these two quantities are virtually identical, differing only in amplitude and a rigid time translation.

It is important to note that extrapolating our model predictions for daily cases and deaths past the peak assumes that control measures such as lockdown, social distancing, use of masks etc and care given to Covid-19 patients remained the same after the peak in daily cases as before the peak. Observed deviations from our predictions are a measure of the extent to which loosening of control measures or changes in care of infected individuals impacted case-loads and deaths past the peaks in the number of cases and deaths.

II. Methods:

We model the Covid-19 pandemic using a simplified version of the SIR model [29], which partitions the population into three compartments, Susceptibles (S), Infectious (I) and Removed R: Recovered or Dead after being infected. This and other models to study the global spread of diseases have been used in a variety of contexts (For some recent reviews, see [30-32]).

So far, the Covid-19 pandemic, at least in the developed countries in Europe where we will apply this model, seems to have the following dynamics: After being infected, an individual remains able to infect others for an average of $T_L$ days. After a time $T_L$, the infected individual becomes sick, gets tested, is identified as infected and is removed from the pool by quarantine or hospitalization. Thus, in our context, the SIR model dynamics can be defined as follows: At t=0, there is a pool of $N$ interacting individuals, almost all of whom are in the S compartment, except for the few infected cases in the I compartment. The R compartment is empty at t=0. Over time, individuals move from S to I and from I to R. In R, they either recover or die. Since the Recovered pool is populated only from the Infected pool, on average, the number removed each day must equal the number infected sometime in the past; i.e. the two are related by a fixed time displacement and a “mortality probability” factor. We assume that the number of deaths and the number of individuals recovered each day are proportional to the number Removed each day by fixed probabilities, that remain invariant over the course of the epidemic; i.e. that the number dead or recovered each day are proportional to the number infected on some previous day, with different time delays and probabilities.
The Model:

Let,

\[ X_1(t) = \text{number of Susceptible individuals at time } t, \]  
\[ X_2(t) = \text{number of Infected individuals at time } t, \]  
\[ X_3(t) = \text{number of individuals that are Removed at time } t \]  
\[ X_1(t) + X_2(t) + X_3(t) = N \]  

A fraction \( \delta \) of the infected individuals will die after being identified as infected. On average, there will be a time delay \( T_D \) between when a person is identified to be infected (tests positive) and when he/she dies of the disease. \( T_D \) will depend on a variety of factors, such as quality of care, age, severity of disease, co-morbidities, immune status etc.

Under these assumptions, the number of deaths \( X_4(t) \) at time \( t \) is related to the number of Infected cases \( X_2 \) by:

\[ X_4(t) = \delta X_2(t - T_D) = \text{number of individuals that died on day } t \]  

(5a)

Similarly, the number of Recovered at time \( t \) will be:

\[ X_5(t) = \beta X_2(t - T_{Re}) = \text{number of individuals that recovered on day } t \]  

(5b)

Let,

\( \alpha \) be the transmission rate, the probability of infection per day per contact.  
and let \( \gamma \) be the rate at which individuals leave the infected population = \( \frac{1}{T_L} \)  
with \( T_L \) being interpreted as the latency, or the average time interval during which an infected individual can infect a naive individual.

The equations governing the dynamics are:

\[ \frac{dx_1}{dt} = -\alpha X_1 X_2 \]  

(8)

\[ \frac{dx_2}{dt} = \alpha X_1 X_2 - \gamma X_2 \]  

(9)

The initial conditions at \( t = 0 \) are:

\[ X_1(0) = (N - a) \sim N \text{ and } X_2(0) = a; \text{with } a \ll N \]  

(10)

An equation relating the state variables \( X_1 \) and \( X_2 \) can be obtained by dividing (8) by (9) and integrating. This gives,

\[ X_2(t) = N + \frac{N}{R} \log \left( \frac{X_1(t)}{N} \right) - X_1(t) \]  

(11)
where \( R \equiv \frac{\alpha N}{\gamma} \) \hspace{1cm} (12)

Hence,

\[
dt \frac{dx_1}{dt} = -\alpha x_1 \left[ N + \frac{N}{R} \log \left( \frac{X_1(t)}{N} \right) - X_1(t) \right]
\] \hspace{1cm} (13)

From (9) the maximum in \( x_2 \) is at \( x_1 = \frac{\gamma}{\alpha} = \frac{N}{R} \) \hspace{1cm} (14)

Substituting this into (13) gives:

Maximum value of \( x_2 \equiv P = N - \frac{N}{R} \left[ 1 + \log(R) \right] \] \hspace{1cm} (15)

At \( t = \infty, x_2 = 0 \). Hence, from (13), we get:

\[ R = \frac{\log(S_1(\infty))}{1 - S_1(\infty)}, \text{ where } S_1(\infty) = \frac{x_1(\infty)}{N} \] \hspace{1cm} (16)

as the fraction of susceptible individuals at \( t = \infty \).

When \( S_1(\infty) \to 1, \log(S_1(\infty)) \sim -(1 - S_1(\infty)) \) and \( R \to 1 \)

This says that when nobody is infected, \( S_1(\infty) = 1 \), i.e. there is no pandemic. This corresponds to \( R = 1 \). It is easy to show that if \( R < 1 \), there are no solutions to (16) that satisfy \( 0 \leq S_1(\infty) \leq 1 \).

**Fitting the Model to Data**

We focus on four North European countries, Netherlands, Denmark, Sweden and Norway, denoted by EN, four South European countries, France, Italy, Spain and Germany, denoted by ES, and the United Kingdom (UK). Data for the number of confirmed cases and deaths were obtained from https://ourworldindata.org/coronavirus-source-data, from an EU agency established in 2005 and based in Stockholm with the aim to strengthen Europe’s defense against infectious diseases. The data for the number of tests was obtained from https://ourworldindata.org/coronavirus-testing [33].

We note that the data identify the daily number of Confirmed cases, whereas the SIR model we described above requires the number of Infected cases. However, the number of confirmed cases at time \( t \) are derived from the infected cases at some previous time \( t' \), where the time lag between \( t \) and \( t' \) is fixed but unknown. Since Equations (1-5) are invariant under a time translation, we can use the data for daily confirmed cases to represent \( x_2 \) with the understanding that there is an implicit shift in time between a person becoming infected and being identified as such in the data. This shift in time would depend on how quickly infected individuals are identified and included in the case count, and hence may vary from country to country. However, we assume that within a
given country, this time interval is fixed on average. With this caveat, in the results below, we will use the confirmed daily cases in our analysis to represent $X_2(t)$, and use either the term “infected” or “confirmed” when discussing $X_2$.

The first date for which data was available for these countries was 12/31/2019, which we denote as day number 0 in our analysis and in the plots to follow. Among the countries considered here, the earliest cases were identified in France on 1/25/2020, which corresponds to day 25. In the other countries, the earliest cases were identified on the following days, counting from 12/31/2019: Netherlands: day 59, Denmark: day 58, Sweden: day 32, Norway: day 58, UK: day 31, Spain: day 32, Germany: day 28, Italy: day 31.

The available data includes
(a) The cumulative number $I(t)$ and daily number $X_2(t)$ of confirmed cases
(b) The cumulative number $D(t)$ and daily number $X_4(t)$ of deaths

These are related by:
\[ X_2(t) = I(t) - I(t-1) \]  
and,
\[ X_4(t) = D(t) - D(t-1). \]

It is worth noting that the dynamics of the number removed each day and the total number of removed do not enter our analysis explicitly. Some connection to the full SIR model [29] can be made by noting that the total number of removed individuals increases asymptotically at large times to $N(1 - S_1(\infty))$.

We determine and analyze the following parameters country by country: $N$, as defined in (4); $\alpha$, the transmission rate; i.e. the number of infections per day per contact; $\gamma$, the average rate at which individuals leave the infected pool; $R$, the average number of transmissions per individual; $T_R$, the time between transmissions; $\delta$, the fraction of individuals in the infected pool who will, on average, die $T_D$ days later; $\beta = 1 - \delta$, the fraction of individuals in the infected pool who will, on average, recover $T_{Re}$ days later; and $T_L$, the infective period. We also determine the duration of the infection, defined as the number of days from 12/31/19 until the number of deaths drops below 5 per day.

The parameters $N, \alpha, \gamma, R$ were obtained using (5), (8) and (9) to do numerical fits. Because of the definition of R (see (12)), only three of the parameters, $N, \alpha, \gamma, R$, need to be determined from the data. Consequently, we proceed as follows:
1. Using (12), we define $\alpha$ in terms of $N, \gamma, R$. This eliminates $\alpha$.
2. Estimating $P = \text{maximum value of } X_2(t)$ from the data, we determine $N$ in terms of $R$ using (15). This eliminates $N$.
3. $\gamma(R - 1)$ is determined as the coefficient of $t$ in the exponential rise of $X_2(t)$ for small $t$ (see Appendix A). This eliminates $\gamma$.
4. Using a numerical solver, we vary $R$ to fit the observed data for $X_2(t)$.
We emphasize that in determining the parameters, we are only using the ascending limb and the peak in the data for \(X_2(t)\) and use these to predict how \(X_2(t)\) will evolve in time past the peak. Using the data for \(X_4(t)\) and the fitted model, we can also determine \(T_D\) and \(\delta\) using (5a) by a simple translation and scaling of the fitted model for \(X_2(t)\).

Furthermore, once \(N, \alpha, \gamma, R\) are determined, (16) determines \(S_1(\infty) = \frac{X_1(\infty)}{N}\), the fraction of the pool of interacting individuals who are naive (uninfected) at the end of the pandemic.

Some other useful parameters we obtain are:

The infective period: \(T_L = \frac{1}{\gamma}\),

the average time between transmissions: \(T_R = \frac{1}{N\alpha}\),

and the average number of transmissions while infective = \(N_C = \frac{T_L}{T_R} = \frac{N\alpha}{\gamma} = R\) \(\text{(23)}\)

### Data and Fitting Methodology

Although data on deaths is unambiguous, the data for the number of cases is trustworthy only when a sufficient number of tests is performed. It is therefore important to determine whether adequate testing was done to ensure the reliability of case data. Figure 1a and 1b show the cumulative number of tests performed in the countries analyzed, starting from 12/31/2019. We see that in the EN countries, the ratio of cumulative tests to cumulative cases always exceeded five, whereas in the ES countries and the UK, it always exceeded three. Consequently, we expect that the reported number of cases is reliable. It is also important to note that while replicate testing provides some assurance, it leaves unaddressed the problem of quality (reliability), breadth (the percentage of the population reached) and speed (the elapsed time between infection and identification of disease) of testing.

One interesting observation from the testing data (Figure 1a, 1b) was that initially, **public concern spread more rapidly than the virus**. As the disease spread and testing ramped up, the initial value of \(T_C\) (number of tests/case) was quite large (over 650 in the UK), suggesting that the exponential growth in infections was slower than the exponential public awareness of the disease, resulting in the testing of a large number of people with no disease. Later, as appropriate protocols to identify diseased individuals were established, \(T_C\) decreased then flattened out when the testing capacity adequately accounted for new cases, and finally rose again as the pandemic waned.

The values of \(X_2(t)\) and \(X_4(t)\) were extracted from the data for the cumulative number of cases and the cumulative number of deaths. To reduce fluctuations in these quantities, the data was averaged over 7 days. Once \(B = \gamma(R - 1)\) and \(P\) were determined from the exponentially increasing region and the peak in \(X_2(t)\) respectively (see discussion in previous section), the only undetermined parameter was \(R\). Several values of \(R\) in the range \(R = 1.5 - 6.0\) were then tested using the following procedure: For each parameter set, we fitting these data for \(X_2(t)\) from small \(t\) to the peak and beyond, by finding numerical solutions of (8) and (9) using the Matlab Solver myode2 to determine \([X_1(t), X_2(t)]\) as a function of time, with the initial conditions, \([X_1(t_0) = N - a, X_2(t_0) = a]\), starting from a value \(t_0\) of \(t\) such that \(X_2(t_0) = a \sim 10\).
To determine an error on the fitted parameters, the parameters were varied until a range of values was found that fit the data for $X_2(t)$ including fluctuations. Finally, $\delta, T_D$ were empirically determined (see eq. (5a)) by scaling and shifting the numerical fits for $X_2(t)$ to fit the data for $X_4(t)$. Using the fitted parameters, the numerical solutions for $[X_1(t), X_2(t)]$ were extended to estimate the number of days to the end of the pandemic, which we define as when the number of daily deaths are less than 5. The fitted solutions for $[X_2(t), X_4(t)]$ were also used to compute the predicted total cases per million and the predicted total deaths per million population and compared to their actual values in each country. Finally, the values of $T_L$ and $T_R$ were determined from the fitted data using (21) and (22) respectively.

### III. Results and Discussion

It is important to note that the parameter values we determine are derived solely from: (i) the ascending (small $t$) limb of $X_2(t)$ and (ii) the first peak value of $X_2(t)$. Using the fitted parameters to derive results past the peak allows us to: (i) predict how the pandemic would have continued beyond the peak, if control measures such as social distancing, use of PPEs and masks etc. had remained the same as before the peak; (ii) determine deviations of the actual data from our model predictions to estimate the effect of changes in control measures post peak to pandemic dynamics.

The results obtained for $N, \alpha, \gamma, R, \delta, T_L, T_R, T_D$ are given in Tables A. Figures 2 and 3 and Supplementary Figure S1a-i shows the data and fits of our model for $X_2$ and $X_4$ respectively. We note that for Sweden and the UK, there is no clear first peak in $X_2(t)$ but rather a plateau and a subsequent increase in the number of daily cases. For these countries, we estimated the value of $P$ (Eq. 15) from the plateau in Fig. 2c and Fig 2e respectively. For all the other countries, there was a clear first peak in $X_2(t)$ whose amplitude was used to estimate $P$. Unfortunately, for all countries, even those that had a well defined first peak in followed by a decline over time, $X_2(t)$ showed significant increase soon after the peak. In some cases, subsequent increases in daily deaths surpassed these first peaks by factors of 2-6 (Figure 1, S1). These suggest that in spite of significant early efforts to control infection rates using social distancing, quarantine, use of masks, testing and other containment measures, there was poor success in limiting the spread of the disease past the first peak in $X_2$ (Figure 2 a-i), most likely because of a combination of premature lifting of quarantine, inadequate use of facemasks and difficulties in limiting physical contacts, exacerbated by the advent of summer weather and holiday travel.

In contrast, for the number of daily deaths $X_4$ (Figure 3 a-i), there were significantly smaller increases past the peak for all countries, suggesting that measures to save lives worked better than measures to contain the number of infections. This is likely due to the efficiency of the health care systems in these countries and the competence and professionalism of health care workers.

**Results for $T_D$, $\delta$, $T_L$, the time delay between case identification and death and the time interval $T_R$ between transmissions while infective:**

The fraction $\delta$ of identified cases who died after a time interval $T_D$ (Figure 4a, Table A) also shows significant variation by country. Norway, Germany and Denmark have the smallest values: $\delta =$
0.03, 0.050 and 0.050 respectively, and the UK, France and Sweden have the highest: \( \delta = 0.21, 0.20 \) and 0.18 respectively. Assuming that most of the deaths occurred in hospitals, the average time \( T_D \) from case identification to death for those who died was highest (12, 15 days) for Germany and Norway respectively and lowest (2, 4 days) for Denmark and Italy respectively (Figure 4b, Table A).

It is significant that in each country, the relation between the daily number of cases and the daily deaths is described by just two parameters, \( \delta \) and \( T_D \) (Eq. 5a), for the entire duration of the pandemic. This is not required a priori. For example, pressure on resources during a peak period of infectivity might have caused a transient increase in the number of deaths per day relative to the number of recovered per day. In such a situation, a single probability \( \delta \) and a single value of \( T_D \) need not have sufficed for the entire epidemic. The fact that there is little evidence of variations in these parameters over the pandemic, to within the quality of the data, suggests that although the effectiveness of life saving measures may differ between countries, these measures seem to be relatively insensitive to changes in the case burden.

The average infective period \( T_L = 1/\gamma \) was approximately constant for all countries, with an average value: \( T_L = 16.3 \pm 2.7 \) days. The average time between transmissions \( T_R = \frac{1}{N_\alpha} \) was also remarkably uniform across countries, averaging: \( T_R = 3.8 \pm 0.5 \) days. Finally, the average number of transmissions while infective or \( N_C = \frac{T_L}{T_R} \), which is also the average value of \( R \), varied only in a narrow range for all countries, averaging: \( N_C = R = 4.4 \pm 0.5 \).

The duration of the pandemic, the case fatality ratio (deaths/million) and cases/million.

We define the end of the pandemic as the day when daily deaths becomes less than 5 for the first time after first peaking. The duration of the pandemic in each country is shown in Figure 4c as the number of days from 12/31/2019 till the day that daily deaths became < 5. With this definition, the predicted and actual dates are (Table A): Netherlands: (6/18 +/- 10 days, 6/17), Denmark: (5/11 +/- 4 days, 5/10), Sweden: (6/28 +/- 10 days, 7/26), Norway: (4/26 +/- 3 days, 4/28), UK: (8/5 +/- 13 days, 8/21), Spain: (6/29 +/- 10 days, 6/29), Germany: (6/24 +/- 9 days, 7/14), France: (7/12 +/- 12 days, 7/19), Italy: (7/28 +/- 15 days 8/2). Note that the model predictions for pandemic duration are in quite good agreement with observation.

The measured case fatality ratio or the total number of deaths per million population (Table A) ranged from low values of 36 and 89 for Norway and Denmark, to high values of 621, 604 and 582, 562 for the UK, Spain, Italy and Sweden respectively. The model predicted values for this quantity are in good agreement with the actual values (Figure 4d).

The measured total number of cases per million population (Table A) ranged from low values of 1367 and 1774 for Norway and Denmark to high values of 7375, 5293, 4806 and 4092 for Sweden, Spain, Sweden, the UK and Italy respectively. In UK and Spain, the model predictions for this quantity were in sharp disagreement with the actual numbers (Figure 4e), likely the result of a significant relaxation of containment measures after the peak in the daily cases was reached.
SARS-Cov-2 may transmit less effectively at higher temperatures.

As was also noticed in [34], an interesting finding was a weak “Temperature Effect” on the value of the infectivity parameter $\alpha$. Figure 4f shows the average temperature in February 2020 for the principal cities in each country versus the fitted value of $\alpha$, and suggests that SARS-Cov-2 may transmit less efficiently at higher temperatures. However, this result is far from final and needs further testing and validation because of many confounding factors, such as humidity, use of air conditioning and exhaust fans, crowding, population density etc. that may also have significant effects on $\alpha$.

In summary, the simple extension of the original SIR model [29] that we propose in this paper is able to identify the full set of pandemic parameters and fully characterize all three compartments S, I and R. It would be useful to try to understand what policy measures would result in improved values for some of these parameters. However, it is also important to recognize that some of the parameters cannot be controlled by policy changes, and some that can be will not all respond to the same strategy. For example, $T_L$ can be shortened by increasing the frequency and number of tests, since both these strategies increase the number of people removed from the infecting pool. On the other hand, the probability that an infected individual will die as opposed to recover might be affected by resources such as ventilators and masks. These considerations lead us to ask whether the degree of correlation between the parameters of the model, and features of the pandemic such as duration and deaths per million population, can account for the wide intercountry variation in pandemic features. A low degree of correlation for a particular parameter would indicate that it had little influence on variation of the feature, whereas a high degree of correlation would be informative.

To estimate some dependencies that may have useful policy implications, we computed the Spearman Correlation between the derived parameters $T_L$, $N\alpha$, and $\delta$ and some metrics of value in determining policy – namely “PD = pandemic duration” and “CFR = case fatality ratio or deaths per million.” We found that $N\alpha$ was not correlated with these metrics (Correlation with PD: -0.20, p-value 0.60; Correlation with CFR: -0.02, p-value 0.98). However, $T_L$ was somewhat correlated (Correlation with PD: 0.64, p-value 0.07; Correlation with CFR: 0.533, p-value 0.15) and $\delta$ was highly correlated (Correlation with PD: 0.86, p-value 0.005; Correlation with CFR: 0.76, p-value 0.02). These results suggest that the optimum method to decrease both the duration of the pandemic and the case fatality ratio would be to decrease the infective period $T_L$ and $\delta$, which are achievable by early case identification, contact tracing and quarantine (which would reduce $T_L$) and improving quality of care for identified cases (which would reduce $\delta$). Whereas these relationships may seem obvious, and indeed have been used as mechanisms to limit the impact of the pandemic in many countries, the fact that they are also readily identifiable in the data is, in our opinion, quite interesting.

The value of $N\alpha$, the average number of new infections per day, has clear implications for the time it will take a randomized trial to establish therapeutic efficacy. While a statistical analysis is outside the scope of this manuscript, we note that the wide inter-country dispersion in this parameter indicates that the choice of population could have a pronounced effect on the time to establish efficacy, and it might therefore be best to select a test population deliberately from a country or
countries with large and similar values of this parameter, rather than mix populations from
different countries with no consideration of the average rate of infectivity.

Declarations

Funding and Acknowledgments
GB was partly supported by grants from M2GEN/ORIEN, DoD/ KRCP (KC180159) and NIH/NCI (1R01CA243547-01A1). He thanks Professors Pablo Tamayo and Jill Mesirov for their kind hospitality at UC San Diego during his sabbatical year 2019-2020 when this work was done. GB thanks Kevin Raines, Sebastian Doniach, Alain Billoire and Carl Dietrich Foerster for many helpful discussions and suggestions.

Conflicts of interest/Competing interests
The authors declare no conflict of interests.

Ethics approval
Not applicable

Consent to participate
Not applicable

Consent for publication
Not applicable

Availability of data and material
The data used in this paper were all derived from public sources. Links to these data are included in the paper.

Code availability
The Matlab codes used to analyze the data along with all data files will be provided on request. Please send an email to gyanbhanot@gmail.com if you want these codes.

Author Contributions
GB: Idea development, analysis, manuscript.
CD: Idea development, analysis, manuscript.
**Figure Legends**

**Figure 1:** Ratio of the cumulative number of tests to number of identified Covid-19 positive cases as a function of time: (a) In the EN countries and (b) in the UK and ES countries. The dashed lines in (a) and (b) represent 5 and 3 tests/case respectively.

**Figure 2:** Parameter fits for the number of daily cases: Blue circles are observed data for $X_2(t)$, the number of cases per day. Solid lines are fits obtained by solving (8) and (9) using the ODE solver ode45 in Matlab. The values of the parameters are shown in the insets and represent the fit shown as the solid red line. The method used for the fits was to find $\gamma(R-1)$ from the exponential rise in $X_2$ at early times (Appendix A), estimate the peak value $P$ of $X_2$ (which gives the value of $N$ using (15)) and find an $R$ value that best fits the data (red solid line). The black lines represent solver results for $R$ varying by 0.5 from the best fit value.

**Figure 3:** Number of deaths per day and fits for $\delta$ and $T_D$: Red circles are observed data for $X_4(t)$, the number of deaths per day. The solid lines were obtained by shifting the fits for $X_2(t)$ in Figure 2 forward in time by an amount $T_D$ and scaling the results by $\delta$ (see eq. 5a).

**Figure 4:** Predictions from fits. (a) Best fits for $T_D$ (see eq. 5a and Figure 3), the average number of days from case identification to death, if death occurs. Values range from a high of 15 days for Norway and a low of 4 days for Italy and Denmark. (b) Best fits for $\delta$, (see eq. 5a and Figure 3), the fraction of identified cases that died an average of $T_D$ days after case was recorded. Values range from low values 0.032, 0.05 and 0.05 for Norway, Denmark and Germany to high values 0.21, 0.20 and 0.20 for UK, Sweden and France. (c) Predicted and actual number of days from 12/31/2019 to when number of daily deaths < 5. Note that for all countries, our predictions are very close to the actual results. This is also evident from the extrapolation of the fits past the peak in Figure 3. This suggests that measures to limit fatalities were just as effective after the peak as before. (d) Predicted and actual case fatality ratios (number of deaths per million) from start of the pandemic to when daily deaths < 5. Actual values for the case fatality ratios range from lows of 18 and 76 deaths/million population for Norway and Denmark and highs of 617, 604 and 579 deaths/million population for UK, Spain and Italy. Note that the predicted values from the fits are in good agreement with the actual values for all countries, suggesting consistent medical care throughout the pandemic. (e) Identification of a possible temperature effect on the infectivity parameter $\alpha$. The x-axis shows the temperature for February 2020 for the principal cities of the countries studied on a linear scale. The results suggest that the infectivity of SARS-CoV-2 decreases with increasing temperature (see also [34]).

**Supplementary Figure Legend**

**Supplementary Figure S1:** Parameter fits for the number of daily cases plotted on a linear scale: These plots are the same as those in Figure 2 except that the y axis is shown in a linear scale, rather than a log scale. This is to show the accuracy of the fits which are shown as solid lines. Blue circles are observed data for $X_2(t)$, the number of cases per day. Solid lines are fits obtained by
solving (8) and (9) using the ODE solver *ode45* in Matlab. The values of the parameters are shown in the insets and represent the fit shown as the solid red line. The method used for the fits was to find \( \gamma(R - 1) \) from the exponential rise in \( X_2 \) at early times (Appendix A), estimate the peak value \( P \) of \( X_2 \) (which gives the value of \( N \) using (15)) and find an \( R \) value that best fits the data (red solid line). The black lines represent solver results for \( R \) varying by 0.5 from the best fit value.

**Table A Legend:**
Table showing a summary of all results for all 9 countries.

**References**


14. For example, see: https://Ourworldindata.org/coronavirus from The European Center for Disease Control and Prevention.


34. Raines K, Doniach S, Bhanot G, *The transmission of SARS-CoV-2 is likely comodulated by temperature and by relative humidity*, medRxiv 2020.05.23.20111278; doi: https://doi.org/10.1101/2020.05.23.20111278.
All Figures
Black dashes correspond to 5 tests per case

Red = Netherlands
Blue = Denmark
Green = Sweden
Magenta = Norway

Log(Tc = Cumulative Tests/Cumulative cases) vs. time in days from 12/31/2020
Black dashes correspond to 3 tests per case

Red = UK, Blue = Spain
Green = Germany
Magenta = France
Black = Italy
Fig. 2a

Netherlands

![Graph showing data for Netherlands with labels R, P, γ, α, and N.]

- \( R = 3.9+/−0.5 \)
- \( P = 1200+/−120 \)
- \( γ = 0.0655+/−0.0116 \)
- \( α = 8.40e−05+/−1.70e−06 \)
- \( N = 3041+/−203 \)

Data points are plotted against time starting on 12/31/2019.
Norway

$R = 3.4+/−0.5$
$P = 260+/−26$
$\gamma = 0.0833+/−0.0181$
$\alpha = 3.77e-04+/−1.64e-05$
$N = 752+/−83$

Fig. 2d
Fig. 2e

United Kingdom

\[ X_2 \]

- \( R = 4.9 \pm 0.5 \)
- \( P = 4700 \pm 470 \)
- \( \gamma = 0.0538 \pm 0.0070 \)
- \( \alpha = 2.65 \times 10^{-5} \pm 1.47 \times 10^{-7} \)
- \( N = 9967 \pm 211 \)

Time starting on 12/31/2019
Spain

Fig. 2f

\[ X_2 \]

\begin{align*}
R &= 5 +/- 0.5 \\
P &= 7900 +/- 790 \\
\gamma &= 0.0675 +/- 0.0086 \\
\alpha &= 2.04e-05 +/- 1.48e-07 \\
N &= 16524 +/- 300
\end{align*}

time starting on 12/31/2019
Fig. 2g

Germany

\[ X^2 \]

\[ R = 4.5 +/- 0.5 \]
\[ P = 5600 +/- 560 \]
\[ \gamma = 0.0686 +/- 0.0100 \]
\[ \alpha = 2.44e-05 +/- 6.36e-08 \]
\[ N = 12626 +/- 445 \]

Time starting on 12/31/2019
Fig. 2h

France

$X_2$

$10^0$ $10^1$ $10^2$ $10^3$ $10^4$

$R = 4.3+/-.5$
$P = 4400+/-.440$
$\gamma = 0.0667+/-.0103$
$\alpha = 2.79e-05+/-.212e-07$
$N = 10275+/-.453$

time starting on 12/31/2019
Fig. 2i  

Italy  

$R = 5+/-0.5$  
$P = 5600+/-560$  
$\gamma = 0.0488+/-0.0062$  
$\alpha = 2.08e-05+/-1.51e-07$  
$N = 11713+/-213$  

(time starting on 12/31/2019)
Fig. 3b

Denmark

$X_4$

$\delta = 0.05$

$T_D = 2$

Time starting on 12/31/2019
Norway

\[ X_4 \]

**Fig. 3d**

\[ \delta = 0.03 \]
\[ T_D = 15 \]
Fig. 3f

Spain

$X_4$

$\delta = 0.11$

$T_D = 5$

time starting on 12/31/2019
Fig. 3h
Fig. 4a

- Norway
- Germany
- France
- United Kingdom
- Netherlands
- Sweden
- Spain
- Italy
- Denmark

$T_D$ : Case identification to death (days)
Fig. 4d

cyan bars = actual, red dots = predicted

case fatality ratio (deaths/million)

Norway, Denmark, Germany, Netherlands, France, Sweden, Italy, Spain, United Kingdom
green bars = actual, red dots = predicted

Fig. 4e
Temperature affects Covid-19 Infectivity

Fig. 4f
<table>
<thead>
<tr>
<th></th>
<th>NETHERLANDS</th>
<th>DENMARK</th>
<th>SWEDEN</th>
<th>NORWAY</th>
<th>UK</th>
<th>SPAIN</th>
<th>GERMANY</th>
<th>FRANCE</th>
<th>ITALY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Latitude</td>
<td>52.133</td>
<td>56.26</td>
<td>60.13</td>
<td>60.17</td>
<td>52.36</td>
<td>40.46</td>
<td>51.17</td>
<td>46.23</td>
<td>41.87</td>
</tr>
<tr>
<td>Population Millions</td>
<td>17.3</td>
<td>5.81</td>
<td>10.2</td>
<td>5.4</td>
<td>66.65</td>
<td>46.9</td>
<td>83</td>
<td>67</td>
<td>60.4</td>
</tr>
<tr>
<td>Major City</td>
<td>Amsterdam</td>
<td>Copenhagen</td>
<td>Stockholm</td>
<td>Oslo</td>
<td>London</td>
<td>Madrid</td>
<td>Berlin</td>
<td>Paris</td>
<td>Rome</td>
</tr>
<tr>
<td>Average Temp Feb 2020 (deg F)</td>
<td>38</td>
<td>34</td>
<td>30.5</td>
<td>28</td>
<td>44.5</td>
<td>46</td>
<td>34.5</td>
<td>42.5</td>
<td>47.5</td>
</tr>
<tr>
<td>ρ</td>
<td>0.19</td>
<td>0.19</td>
<td>0.16</td>
<td>0.2</td>
<td>0.21</td>
<td>0.27</td>
<td>0.24</td>
<td>0.22</td>
<td>0.195</td>
</tr>
<tr>
<td>P</td>
<td>1,200</td>
<td>310</td>
<td>500</td>
<td>260</td>
<td>4,700</td>
<td>7,900</td>
<td>5,600</td>
<td>4,400</td>
<td>5,600</td>
</tr>
<tr>
<td>N</td>
<td>3,041 +/-203</td>
<td>738 +/-37</td>
<td>1,214+/-66</td>
<td>752 +/-83</td>
<td>9,967+/-211</td>
<td>16,524 +/-300</td>
<td>12,626 +/-445</td>
<td>10,275+/-453</td>
<td>11,713 +/-213</td>
</tr>
<tr>
<td>R</td>
<td>3.9 +/-0.5</td>
<td>4.2 +/-0.5</td>
<td>4.1 +/-0.5</td>
<td>3.4 +/-0.5</td>
<td>4.9 +/-0.5</td>
<td>5.0 +/-0.5</td>
<td>4.5 +/-0.5</td>
<td>4.3 +/-0.5</td>
<td>5.0 +/-0.5</td>
</tr>
<tr>
<td>α 100,000</td>
<td>8.40 +/-0.17</td>
<td>33.80 +/-0.37</td>
<td>17.44 +/-0.23</td>
<td>37.7 +/-1.64</td>
<td>2.65 +/-0.02</td>
<td>2.04 +/-0.02</td>
<td>2.44 +/-0.06</td>
<td>2.79 +/-0.02</td>
<td>2.08 +/-0.02</td>
</tr>
<tr>
<td>υ</td>
<td>0.066 +/-0.012</td>
<td>0.059 +/-0.010</td>
<td>0.052 +/-0.009</td>
<td>0.083 +/-0.018</td>
<td>0.054 +/-0.007</td>
<td>0.068 +/-0.009</td>
<td>0.069 +/-0.01</td>
<td>0.067 +/-0.01</td>
<td>0.049 +/-0.006</td>
</tr>
<tr>
<td>δ</td>
<td>0.130</td>
<td>0.050</td>
<td>0.180</td>
<td>0.030</td>
<td>0.210</td>
<td>0.110</td>
<td>0.050</td>
<td>0.200</td>
<td>0.140</td>
</tr>
<tr>
<td>T₀</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>15</td>
<td>5</td>
<td>6</td>
<td>12</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>T₀ = 1/υ</td>
<td>15.3</td>
<td>16.8</td>
<td>19.4</td>
<td>12.0</td>
<td>18.6</td>
<td>14.8</td>
<td>14.6</td>
<td>15.0</td>
<td>20.5</td>
</tr>
<tr>
<td>Tₐ = 1/(σN₀)</td>
<td>3.9</td>
<td>4.0</td>
<td>4.7</td>
<td>3.5</td>
<td>3.8</td>
<td>3.0</td>
<td>3.2</td>
<td>3.5</td>
<td>4.1</td>
</tr>
<tr>
<td>Actual Max daily cases at first peak</td>
<td>1,120</td>
<td>328</td>
<td>532</td>
<td>282</td>
<td>4,751</td>
<td>7,902</td>
<td>5,596</td>
<td>4,482</td>
<td>5,644</td>
</tr>
<tr>
<td>Model Fitted Max daily cases at first peak</td>
<td>1,219 +/-58</td>
<td>313 +/-16</td>
<td>536 +/-21</td>
<td>265 +/-13</td>
<td>4,716 +/-234</td>
<td>7,928 +/-393</td>
<td>5,601 +/-277</td>
<td>4,410 +/-218</td>
<td>5659 +/-274</td>
</tr>
<tr>
<td>Nₐₓ₋₁(P) = Predicted Number of days from 12/31/2019 to pandemic end defined as day Daily Deaths &lt; 5 after first peak</td>
<td>169 +/- 10 days</td>
<td>132 +/- 4 days</td>
<td>179 +/- 10 days</td>
<td>117 +/- 3 days</td>
<td>217 +/- 13 days</td>
<td>180 +/- 10 days</td>
<td>175 +/- 9 days</td>
<td>193 +/- 12 days</td>
<td>209 +/- 15 days</td>
</tr>
<tr>
<td>Tₚ₋₁(P) = Predicted date pandemic ends = defined as day Daily Deaths &lt; 5 after first peak</td>
<td>6/18/2020 +/- 10 days</td>
<td>5/11/2020 +/- 4 days</td>
<td>6/28/2020 +/- 10 days</td>
<td>4/26/2020 +/- 3 days</td>
<td>8/5/2020 +/- 13 days</td>
<td>6/29/2020 +/- 10 days</td>
<td>6/24/2020 +/- 9 days</td>
<td>7/12/2020 +/- 12 days</td>
<td>7/28/2020 +/- 15 days</td>
</tr>
<tr>
<td>Nₐₓ₋₁(A) = Actual number of days after first peak when Daily Deaths &lt; 5</td>
<td>168</td>
<td>131</td>
<td>207</td>
<td>119</td>
<td>233</td>
<td>180</td>
<td>195</td>
<td>207</td>
<td>214</td>
</tr>
<tr>
<td>Tₚ₋₁(A) = Actual date after first peak when Daily Deaths &lt; 5</td>
<td>6/17/20</td>
<td>5/10/20</td>
<td>7/26/20</td>
<td>4/28/20</td>
<td>8/21/20</td>
<td>6/29/20</td>
<td>7/14/20</td>
<td>7/19/20</td>
<td>8/2/20</td>
</tr>
<tr>
<td>Predicted Cases upto Tₚ₋₁(P)</td>
<td>45,015 +/- 5,413</td>
<td>10,645 +/- 1,199</td>
<td>22,799 +/- 2,657</td>
<td>7,844 +/- 881</td>
<td>183,393 +/- 20,530</td>
<td>242,705 +/- 26,975</td>
<td>181,375 +/- 20,844</td>
<td>151,738 +/- 17,726</td>
<td>238,163 +/- 26,426</td>
</tr>
<tr>
<td>Actual Cases upto Tₚ₋₁(A)</td>
<td>48,921</td>
<td>10,306</td>
<td>75,228</td>
<td>7,382</td>
<td>320,301</td>
<td>248,240</td>
<td>198,770</td>
<td>179,059</td>
<td>247,164</td>
</tr>
<tr>
<td>Predicted Deaths upto Tₚ₋₁(P)</td>
<td>5,852 +/- 704</td>
<td>533 +/- 60</td>
<td>4,104 +/- 479</td>
<td>236 +/- 27</td>
<td>38,513 +/- 4,312</td>
<td>26,698 +/- 2,968</td>
<td>9,069 +/- 1,043</td>
<td>30,348 +/- 3,546</td>
<td>33,343 +/- 3,705</td>
</tr>
<tr>
<td>Actual Deaths upto Tₚ₋₁(A)</td>
<td>6,066</td>
<td>515</td>
<td>5,732</td>
<td>192</td>
<td>4,384</td>
<td>28,340</td>
<td>9,062</td>
<td>30,180</td>
<td>35,134</td>
</tr>
<tr>
<td>Predicted Cases/Million population upto Tₚ₋₁(P)</td>
<td>2602 +/- 313</td>
<td>1832 +/- 206</td>
<td>2235 +/- 260</td>
<td>1453 +/- 163</td>
<td>2752 +/- 308</td>
<td>5175 +/- 575</td>
<td>2185 +/- 251</td>
<td>2265 +/- 265</td>
<td>3943 +/- 438</td>
</tr>
<tr>
<td>Actual Cases/Million population upto Tₚ₋₁(A)</td>
<td>2,828</td>
<td>1,774</td>
<td>7,375</td>
<td>1,367</td>
<td>4,806</td>
<td>5,293</td>
<td>2,395</td>
<td>2,673</td>
<td>4,092</td>
</tr>
<tr>
<td>Predicted CFR (deaths/million) upto Tₚ₋₁(P)</td>
<td>338 +/- 41</td>
<td>92 +/- 10</td>
<td>402 +/- 47</td>
<td>44 +/- 5</td>
<td>578 +/- 65</td>
<td>569 +/- 63</td>
<td>109 +/- 13</td>
<td>453 +/- 53</td>
<td>552 +/- 61</td>
</tr>
<tr>
<td>Actual CFR (deaths/million) upto Tₚ₋₁(A)</td>
<td>351</td>
<td>89</td>
<td>562</td>
<td>36</td>
<td>621</td>
<td>604</td>
<td>109</td>
<td>450</td>
<td>582</td>
</tr>
</tbody>
</table>
Supplementary Figures
Netherlands

\[
\begin{align*}
R &= 3.9 \pm 0.5 \\
\gamma &= 0.066 \pm 0.012 \\
\alpha &= 8.40 \times 10^{-5} \pm 1.70 \times 10^{-6} \\
N &= 3041 \pm 203
\end{align*}
\]
Denmark

\[ R = 4.2+/0.5 \]
\[ P = 310 +/- 31 \]
\[ \gamma = 0.059 +/- 0.010 \]
\[ \alpha = 3.38e-04 +/- 3.52e-06 \]
\[ N = 738 +/- 37 \]
Fig. S1c

Sweden

- $R = 4.1 \pm 0.5$
- $P = 500 \pm 50$
- $\gamma = 0.052 \pm 0.009$
- $\alpha = 1.74e-04 \pm 2.34e-06$
- $N = 1214 \pm 66$

$X_2$ vs. time starting on 12/31/2019
Norway

Fig. S1d

\[ R = 3.4 \pm 0.5 \]
\[ P = 260 \pm 26 \]
\[ \gamma = 0.083 \pm 0.018 \]
\[ \alpha = 3.77 \times 10^{-4} \pm 1.64 \times 10^{-5} \]
\[ N = 752 \pm 83 \]
Fig. S1e

United Kingdom

$R = 4.9 \pm 0.5$

$P = 4700 \pm 470$

$\gamma = 0.054 \pm 0.007$

$\alpha = 2.65 \times 10^{-5} \pm 1.47 \times 10^{-7}$

$N = 9967 \pm 211$
Germany

Fig. S1g

\[ R = 4.5 +/- 0.5 \]
\[ P = 5600 +/- 560 \]
\[ \gamma = 0.069 +/- 0.010 \]
\[ \alpha = 2.44e-05 +/- 6.36e-08 \]
\[ N = 12626 +/- 445 \]
France

R = 4.3 +/- 0.5
P = 4400 +/- 440
γ = 0.067 +/- 0.010
α = 2.79e-05 +/- 2.12e-07
N = 10275 +/- 453
Fig. S1i

Italy

R = 5 +/- 0.5
P = 5600 +/- 560
\( \gamma = 0.049 +/- 0.006 \)
\( \alpha = 2.08e-05 +/- 1.51e-07 \)
N = 11713 +/- 213