Asymmetry in the peak in Covid-19 daily cases and the pandemic R-parameter

Sayali Bhatkar[1], Mingyang Ma[2], Mary Zsolway[3], Ayush Tarafder[3], Sebastian Doniach[4], Gyan Bhanot[2,5,*]

[1] Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 40005, India.
[2] Department of Physics and Astronomy, Rutgers University, Piscataway, NJ, 08854, USA
[3] School of Arts and Sciences, Rutgers University, Piscataway, NJ, 08854, USA
[4] Department of Applied Physics, Stanford University, Stanford, CA 94305, USA
[5] Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ, 08854, USA
[*] Corresponding Author: gyanbhanot@gmail.com, Tel: +1 848 391 7508, Address: 136 Frelinghuysen Rd, Busch Campus, Rutgers University, Piscataway, NJ 08854, USA.

Keywords: Pandemics; Universal Scaling Laws; SIR Model; Pandemic parameters; Covid-19 (SARS-CoV-2).

ABSTRACT

Within the context of the standard SIR model of pandemics, we show that the asymmetry in the peak in recorded daily cases during a pandemic can be used to infer the pandemic R-parameter. Using only daily data for symptomatic, confirmed cases, we derive a universal scaling curve that yields: (i) r_{eff} , the pandemic R-parameter; (ii) L_{eff}, the effective latency, the average number of days an infected individual is able to infect others and (iii) α, the probability of infection per contact between infected and susceptible individuals. We validate our method using an example and then apply it to estimate these parameters for the first phase of the SARS-Cov-2/Covid-19 pandemic for several countries where there was a well separated peak in identified infected daily cases. The extension of the SIR model developed in this paper differentiates itself from earlier studies in that it provides a simple method to make an a-posteriori estimate of several useful epidemiological parameters, using only data on confirmed, identified cases. Our results are general and can be applied to any pandemic.

INTRODUCTION

A pandemic occurs when a new pathogen enters a naïve population. The recent SARS-Cov-2 pandemic was caused by a Coronavirus, one of a family of large, enveloped, single-stranded RNA viruses that are widespread in animals and usually cause only mild respiratory illnesses in humans [1-5]. In 2003, a new coronavirus emerged, and was named SARS-CoV (Severe Acute Respiratory Syndrome – Corona Virus). This virus caused a life-threatening respiratory disease in humans, with a fatality rate of almost 10% [6,7]. In fact, after an initial burst of interest in development of treatment options, interest in this virus waned. The emergence of the novel coronavirus SARS-CoV-2, identified in December 2019 in Wuhan, China, 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, a long latency period and an asymptomatic infective phase [13].

The governments of several countries took significant measures to slow the infection rate of Covid-19, such as social distancing, quarantine, identification, tracking and isolation. However, there was no uniform policy, some governments reacted later than others, and some (e.g. Sweden) decided 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 to estimate pandemic characteristics, guide public policy and assess the efficacy of interventions [15]. In this paper, we use data from the WHO website: https://covid19.who.int/WHO-COVID-19-global-data.csv.

It is well known that in most pandemics, confirmed infected cases often 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-morbidities; i.e. COVID-19 increases susceptibility to other diseases and conditions [19]. Moreover, the virus can be transmitted by asymptomatic individuals, who can comprise a substantial portion of the infected population [20], militating against accurate estimates of total infection rates. In this context, as indicated in [21], analytical models can provide useful information.

Dynamical (mechanistic) models have been used for forecasting and for making projections. For example, projections and forecasting models of various types were used as early as February 2020 to determine a reproductive number for SARS-CoV-2 [13]. More generally, multiple research groups have models to estimate Case Fatality Ratios (CFRs) [22], to forecast and project the need for hospital beds [23] and to project and forecast mortality [24]. Among the many applications of models to COVID-19, four variable Susceptible-Exposed-Infective-Recovered (SEIR) models have been used to project the impact of social distancing on mortality [25], three variable Susceptible-Infective-Recovered (SIR) models have been used 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]. There is also some model based evidence that the transmission of the SARS-Cov-2 virus is regulated by temperature and humidity [29]. In this paper, we model the Covid-19 pandemic using an extension of the SIR model [30], which partitions the population into three compartments: Susceptibles (S),

Infectious (I) and Removed R. This and other models (using more variables) have been used in a variety of contexts to study the global spread of diseases (For some recent reviews, see [31-33]).

In this paper, within the context of an extension of the standard epidemiological SIR model [30], using only daily recorded case data of symptomatic individuals, we develop a method to estimate: (i) the pandemic parameter r_{eff} , which is the average number of individuals infected by a single infected individual, (ii) the effective latency L_{eff}, which is the average time an infected individual is able to infect others, and (iii) the probability α of infection from a single encounter between an infected and a susceptible individual.

METHOD

The spread of a virus depends on several factors, such as patterns of contacts among infectious and susceptible persons, the latency period between being infected and becoming infective and/or developing symptoms, the duration of infection/infectivity, immunity acquired by previous exposures, effects of vaccination, etc. In spite of these complications, some of the features of the spread are well described by a simple epidemiological model, the so-called SIR model [30]. In this model, at any given instant in time, the population is divided into three compartments : susceptible (S), infected (I) and recovered (R) with S+I+R = N. The parameter N is assumed to be constant in time and is the number of individuals exposed to the virus in a given region. It is assumed that individuals in the S compartment are equally susceptible to being infected by individuals in the I compartment and individuals in the R compartment are both immune to further infection and unable to infect others. The R compartment also includes those who are "removed" (dead, quarantined etc.).

At any given moment in time, the reported number of infected individuals is a subset of the actual number of infected individuals [16,17]. Let $I_0(t)$ and $I_1(t)$ be the number of individuals in the unidentifiable and identifiable compartments respectively, with $I_1(t) = \omega I(t)$ and $I_0(t) = (1-\omega)I(t)$. Similarly, let $R(t) = R_0(t) + R_1(t)$ where $R_0(t)$ are derived from $I_0(t)$, and $R_1(t)$ are derived from $I_1(t)$. We assume that individuals in the I_1 compartment, once identified, are no longer able to infect others because they would be isolated, confined, or quarantined. On the other hand, individuals in the I_0 compartment, being asymptomatic and unaware of their infected state, would continue to infect others until they become non-infective. Let L_0 be the average number of days an asymptomatic individual is infective and L_1 be the average number of days a symptomatic individual is infective. Let $\gamma_1 = \frac{1}{L_1}$ and $\gamma_0 = \frac{1}{L_0}$ be the rates at which these two types of infected individuals leave the infective pool. Under these assumptions, we propose the following simple extension of the SIR model for the pandemic dynamics:

$$\frac{dS}{dt} = -\left(\frac{\alpha}{N}\right)SI \tag{1}$$

$$\frac{dI}{dt} = \frac{dI_1}{dt} + \frac{dI_0}{dt} = \left(\frac{\alpha}{N}\right)SI - \gamma_{eff}I = \left(\frac{\alpha}{N}\right)SI - (\omega\gamma_1 + (1-\omega)\gamma_0)\, I \tag{2}$$

$$\frac{dR}{dt} = \frac{dR_1}{dt} + \frac{dR_0}{dt} = (\omega\gamma_1 + (1-\omega)\gamma_0)\, I = \gamma_{eff}I \tag{3}$$

α is the probability of infection in a single encounter between an infected and susceptible individual and

$$\gamma_{eff} = (\omega\gamma_1 + (1-\omega)\gamma_0) = (\omega/L_1 + (1-\omega)/L_0) \quad (4a)$$

is the rate at which an individual in the I compartment transitions to the R compartment. The reciprocal L_{eff} of γ_{eff} is the average effective latency, the average number of days that an infected individual (symptomatic or not) is infective. Thus,

$$L_{eff} = \frac{1}{\gamma_{eff}} = 1/(\omega/L_1 + (1-\omega)/L_0) \quad (4b)$$

Let X(t) be the observed rate at which daily cases are identified. From Eq. 3,

$$X(t) = \frac{dR_1}{dt} = \omega\gamma_1 I(t). \quad (5)$$

The key relationship we exploit in this paper is Eq. 5, the fact that the number of daily observed cases $X(t) = \frac{dR_1}{dt}$, is proportional to I(t). This proportionality implies that the width and location (in units of time t) of the peak in X(t) and I(t) are the same. The location of the peak in I(t) or X(t) is difficult to estimate because it depends on identifying the "start" of the pandemic, i.e., the day when the first person was infected. However, other characteristics of the peak in X(t) are easier to measure. As we show below, scaling laws relating the pandemic parameter R can be inferred using dimensionless quantities from the form of X(t) using Eq. 1-5. In this paper, we will use the left and right half-widths at the peak in X(t) to estimate parameters.

To do this, we rescale the time t to $\tau = \gamma_{eff}t = \frac{t}{L_{eff}}$ Eq. 1-3 can then be rewritten in terms of the fractions s = S/N, i = I/N, r= R/N, $r_1 = R_1/N$ and x = X/N as follows:

$$\frac{ds(\tau)}{d\tau} = -r_{eff}\, s(\tau)i(\tau) \quad (6)$$

$$\frac{di(\tau)}{d\tau} = r_{eff}\, s(\tau)i(\tau) - i(\tau) \quad (7)$$

$$\frac{dr(\tau)}{d\tau} = i(\tau) \quad (8)$$

$$x(\tau) = \frac{dr_1(\tau)}{d\tau} = \omega\gamma_1 i(\tau)/\gamma_{eff} \quad (9)$$

$$\text{with } r_{eff} = \alpha/\gamma_{eff} \quad (10)$$

Let W_L and W_R be the left and right half widths of the peak in X(t). From Eq. 5, we note that the dimensionless quantity W_R/W_L is the same for X(t), I(t). Furthermore, since $x(\tau)$ and X(t) are proportional and the relationship between τ and t is just a scale factor ($\tau = \gamma_{eff}t$),

W_R/W_L is the same for $x(\tau)$ and X(t). Finally, from Eq 6-9, it is obvious that W_R/W_L depends only on r_{eff}. The functional relationship between W_R/W_L and r_{eff}, can be obtained by numerically solving Eqs 6-9 and is shown in Fig. 1 (detailed values in Supplementary Table 1). Note that the result of Fig. 1 is a universal scaling law in the SIR model which applies to any pandemic. We will now demonstrate with an example how the parameters r_{eff}, L_{eff} and α can be obtained using only the results in Fig. 1 and Supplementary Table 1 and data for X(t). We note that the fraction $f_{inf} = 1 - s(\infty)$ of individuals infected at $t=\infty$ can also be computed from r_{eff} using the relationship $r_{eff} = -\frac{(s(\infty))}{1-s(\infty)}$ (Appendix A, Eq. A8)

Figs 2a, 2b, 2c show an example of the procedure we follow to find parameters from only X(t) data. Fig 2a shows X(t) from a numerical solution of Eqs. 1-4 for parameter values $N = 4.6\times10^5$, $L_0 = 10$ days, $L_1 = 5$ days, $\omega = 0.25$, $\alpha = 0.46875$. These parameter values correspond to $\gamma_{eff} = 0.125$ (Eq. 4a), $L_{eff} = 8$ (Eq. 4b) and $r_{eff} = 3.75$ (Eq. 10).

The procedure we follow to estimate the parameters from just the data for X(t) (Fig 2a) and the universal scaling results in Figure 1/Supplementary Table 1 is as follows:

- We first estimate the ratio W_R/W_L from the data for X(t) (Fig. 2a). We find $W_R = 10.65$ days, $W_L = 6.83$ days, and $W_R/W_L = 1.56$. We also obtain the maximum value of X(t) as $M_{X(t)} = 8768.8$ (shown as a blue dot in Fig. 2a).
- The value of W_R/W_L is used to infer r_{eff} by interpolating the data in Supplementary Table 1. This yields the value r_{eff}(estimated) = $r_{fit} = 3.74$.
- Using $r_{fit} = 3.74$, we solve Eqs. 6-9 to generate $i(\tau)$, shown in Figure 2b.
- The right and left widths w_R and w_L at half maximum are estimated from the data for $i(\tau)$. This gives $w_R = 1.33$, $w_L = 0.85$. Note that these quantities are dimensionless. Similarly, the maximum in $i(\tau)$ is measured to be $m_{i(\tau)} = 0.3859$. These quantities are measured very accurately because they result from solving a set of equations.
- An estimate L_{eff} is now obtained as the ratio $L_{eff} = (W_R+W_L)/(w_R+w_L) = 7.99$ days.
- To check whether the fit is good, we multiply $i(\tau)$ in Fig 2b by $M_{X(t)}/m_{i(\tau)}$, scale the τ axis by a factor of L_{fit} and shift the location of the maximum in X(t) to coincide with the peak in X(t). This fit is shown in Fig . 2c.
- In this example, we used a small time-step to generate accurate data for X(t). In reality, only daily data is available and there would be an error of approximately 0.5 days in estimates of W_R and W_L. Using this, we can estimate the error in the parameters. The final results for our example are: $W_R/W_L = 1.56 +/- 0.14$, $r_{eff} = 3.74 +/- 0.76$, $L_{eff} = 7.99 +/- 1.42$ days.

Note that in the procedure described above, we used only the observed daily cases X(t) and the universal scaling results in Figure 1 and Table 1.

When we apply this procedure to pandemic data, we have to first contend with one additional issue, which is that the data shows a three, five, or seven day cycle, depending on the country. This is presumably due to the logistics of data collection and reporting, and requires smoothing before our method can be applied. In this paper, we averaged the raw data for each

country over five, seven and nine days; inferred r_{eff} for each averaging scheme and used the average of the values obtained as the final result. Finally, as our method is sensitive to the position of the peak in daily cases, and since we are interested in the asymmetry near the peak, we fitted the data near the peak to a cubic to obtain the location (in time) and height of the peak and used these values and the smoothed data points to find W_R and W_L.

RESULTS:

Worldwide data for confirmed Covid-19 cases and deaths from January 3, 2020 was downloaded from the World Health Organization (WHO) website: https://covid19.who.int/WHO-COVID-19-global-data.csv (Supplementary Table 2). The data for daily cases obtained from this source directly measures the function X(t) in our analysis. Before performing any analysis, the data for daily cases was smoothed/averaged as described above.

Our model assumes a single circulating viral strain infecting a homogeneous set of individuals in a given region who were equally susceptible to infection (uniform immune response). The model also assumes that exposed individuals observed the same rules regarding the use of masks/isolation/quarantine, there was no significant variation in population density among them, little variation in their movements, equal vaccination status, and symptomatic cases once identified were quarantined. These requirements would most likely be valid, at least for some countries, for the first wave of the Covid-19 pandemic when the world population was naïve to the virus (no immunity) and everyone was equally susceptible. Furthermore, at least in some countries, many of the other assumptions of the model did apply, such as homogeneity of response, lack of vaccines resulting in no innate immunity, standard medical protocols, good testing practices, and a single circulating strain of the virus. Such countries would show an initial exponential rise in daily cases at small times followed by a clear peak with no overlap with subsequent peaks. We applied our method to analyze the first peak in daily cases for twenty-one countries that satisfied these conditions.

Supplementary Table 1 gives the scaling function data from solving Eq 6-9 for different values of r_{eff}. For each value of r_{eff} the columns show $w_L = W_L/L_{eff}$, $w_R = W_R/L_{eff}$, W_R/W_L, $m_{i(\tau)}$, $f_{inf} = 1-s(\infty)$, r_{eff} calculated from Eq. A8 and the ratio $W_R/W_L = w_R/w_L$. This data is universal and applies to all pandemics. Figs 3-5 show the fits for three countries from very different geographic regions Spain (Europe), Thailand (Asia) and Australia (Oceania). Results and figures for all twenty-one countries are in Table 1 and Supplementary Figures 1.

Using Eq A in Appendix A, we estimate the total fraction f_{tot} of the exposed subset of the population that was infected (Table 1).Note that this represents both the cases that were counted (symptomatic and/or tested) as well as those that were not identified. By our estimates, 50-100% of exposed individuals (fraction of N in the SIR model) in the countries we analyzed were infected.

Finally, we note that for the daily case data to be reliable, the total number of tests (including positives and negatives) should be sufficiently large. Using test data from (https://www.ecdc.europa.eu/en/publications-data/covid-19-testing) we checked that the number

of tests performed on each day were much higher than the number of cases for countries where test data was available for the time range of interest (Supplementary Figures 2).

DISCUSSION:

In this paper we show that the asymmetry of the peak in recorded daily cases in any pandemic contains important information. In particular, for the standard epidemiological SIR model [30], this asymmetry can be used to relate the pandemic R-parameter to the ratio of right and left half widths of the peak, which in turn provides an a-posteriori estimate of several important pandemic parameters, including the effects of both recorded (symptomatic/tested) cases as well as unmeasured/asymptomatic cases, namely: r_{eff}, the effective pandemic R-parameter, the effective latency L_{eff} (the average number of days an infected individual is able to infect others), and the infection probability α of transmission from an infected individual to a susceptible individual in a single encounter. These can be inferred using a universal scaling function (Figure 1 and Supplementary Table 1) that relates the ratio of the right and left half widths at full maximum of the peak in daily identified cases X(t) to the effective pandemic R-parameter (r_{eff}). Within the limits of the SIR model, our results are general and apply to any pandemic. We apply our method to worldwide country specific data to find r_{eff}, L_{eff} and α for the first phase (first peak in daily cases) for the SARS-COV-2 pandemic for twenty-one countries which had a clear, well separated peak in daily cases (Table 1, Supplementary Figures 1).

A novel feature of this analysis is that our result for f_{tot}, the fraction of the population that was infected, includes both those who were identified/counted as infected as well as those that were not identified/counted. Our results for f_{tot} suggest that in the 2019-2020 SARS-COV-2 pandemic, the original variant of the virus was highly effective in transmission. In some of the developed countries, we find that almost all exposed individuals were infected in the first phase of the pandemic (Table 1). The fact that f_{tot} is close to unity for many countries means there was a large fraction of infected in the population who were not identified.

Using data from https://www.worldometers.info/world-population/population-by-country and https://en.wikipedia.org/wiki/List_of_cities_by_average_temperature (Supplementary Table 3), we explored possible relationships among the values of the R parameter, the α coefficient and the latency period L_{eff} with the population density, mean temperature during the first peak and the median age of the population (Sup. Figs. 3a-i). Although the results are too noisy to make definitive statements about these relationships, overall, across all the countries analyzed, from the linear fits, the weighted average values of r_{eff}, α, and L_{eff} are 2.18, 0.23, and 15.35 days respectively.

Several countries, notably the United States, Canada, The Russian Federation, India, and Pakistan did not meet our criterion of a clear, well separated first peak in daily cases in 2020. This is most likely because they cannot be thought of as homogeneous in the sense of response from local authorities regarding the use of masks, quarantine etc. In the United States for example, the response was state and/or county specific. In principle, our method could be applied at the county or state level in the US or the province or sub-province level in Canada to determine parameters

from recorded case data, wherever these compartments had uniform rules for containment of the virus.

It would be interesting to apply our method to subsequent peaks in daily cases for the SARS-COV-2 pandemic, as the virus evolved into different variants across the globe. Comparing changes in the inferred parameters across countries would provide a country specific estimate of the effects of preventive measures such as the effectiveness/efficacy of vaccination, changes in behavior (mask use, testing/quarantine, work-from-home, social distancing, travel restrictions) etc. Furthermore, this method could be applied to other viral pandemics, such as the SARS-COV pandemic of 2003, and Influenza pandemics of the past, such as the H1N1 Spanish Flu pandemic of 1918-19 which recurred in 1950 and 1977, the H2N2 Asian Flu pandemic of 1957, the H3N2 Hongkong pandemic of 1968 and the more deadly H5N1 East Asian pandemic of 1997.

Summary: In this paper, we propose a simple method to estimate the R parameter of epidemics using the asymmetry around the peak in daily reported cases. We quantify this asymmetry using the ratio W_R/W_L of the right half width W_R to the left half width W_L and show how this dimensionless ratio can be used to estimate pandemic parameters. The function that makes this possible is shown in Fig. 1 and Supplementary Table 1. Within the SIR model, this relationship between r_{eff} and W_R/W_L is universal and can be used for any epidemic. The value of such estimates is of necessity limited by the accuracy and reliability of the available data. However, as we show using a simulated example (Fig. 2), given data of sufficient accuracy, our method is an extremely simple and accurate way to estimate the R parameter, α, and the effective latency.

<u>Figure Captions:</u>

Figure 1: Universal curve in the SIR model for reff as a function of the ratio W_R/W_L, where W_L and W_R are the left and right half widths of the peak in X(t). Note that this scaling function is universal in the SIR model and applies to any pandemic.

Figure 2 a-b: Results obtained by numerically solving the SIR equations (Eqs. 1-5) for parameter values $N = 4.6\times10^5$, $L_0 = 10$ days, $L_1 = 5$ days, $\omega = 0.25$, $\alpha = 0.46875$, which corresponds to $r_{eff} = 3.75$. The red dots in Fig 2a show daily case data (X(t)) from a numerical solution of Eqs. 1-5 using these parameter values. The blue line shows the location of the maximum and green dots are the locations of the right and left half width points. Using the values of W_R and W_L inferred from X(t), an estimate (r_{fit}) of the pandemic parameter (r_{eff}) was obtained by interpolation from Supplementary Table 1. Fig 2b shows $i(\tau)$ by solving Eq. 6-9 for the value $r_{fit} = 3.74$ obtained from Fig 2a. The solid line in Fig 2c shows the result of mapping this solution to the data for X(t) by matching the peak of X(t) to the peak of $i(\tau)$ and rescaling the τ axis to match the half width for

$i(\tau)$. The rescaling of the half width from τ to t provides the estimate $L_{eff} = (W_R+W_L)/(w_R+w_L) =$ 7.99 days.

Figure 3 a-c: Fit of our model to data for X(t) for Spain (data for X(t) from Supplementary Table 2) from the World Health Organization. Fig 3a is the raw X(t) data for Spain. Fig 3b shows the $i(\tau)$ curve for the r_{fit} value obtained from its average value from analysis of X(t) for five, seven and nine day averaging applied to the raw data in Fig 3a. Fig 3c shows the mapping of the $i(\tau)$ data to X(t) to match its peak location and height. The points plotted for X(t) are seven day averaged data. The scaling of the τ axis to the t axis allows an estimate of L_{eff} as in the example in Fig 2.

Figure 4 a-c: Analysis of X(t) data for Thailand as in Figure 3.

Figure 5 a-c: Analysis of X(t) data for Australia as in Figure 3.

Table Captions:

Table 1 : Results for W_R, W_L, r_{eff}, L_{eff}, f_{tot}, and α for twenty-one countries using the scaling law in Fig 1 and Supplementary Table 1. Here W_L and W_R are the left and right half widths of the peak in X(t), the number of recorded daily cases.

Supplementary Figure Captions:

Supplementary Figures 1: Raw data, $i(\tau)$ and fits of our model to data for X(t) for twenty-one countries. In addition to figures similar to Figs 3-5, we also include the fits to 5,7,9 day averaged data. These show that the r_{eff} values for these various averagings are in good agreement.

Supplementary Figures 2: Data for total number of tests carried out in 3 countries of interest for the period of interest. Data for tests is from https://www.ecdc.europa.eu/en/publications-data/covid-19-testing. The blue dots are the number of positive tests and the red dots are the total number of tests carried out (including both positive and negative cases). The positive fraction in all cases was sufficiently low for the testing to be considered adequate. This means that the X(t) data we used here should be reliable.

Supplementary Figures 3: Exploring possible relationships among r_{eff} , α and L_{eff} (in days) and a the population density, average temperature and median age of population. Demographic data from https://www.worldometers.info/world-population/population-by-country and temperature data from https://en.wikipedia.org/wiki/List_of_cities_by_average_temperature.

Supplementary Table Captions:

Supplementary Table 1 : Detailed results from numerical solutions of Eq 6-9 for different values of r_{eff}. These were used in Figure 1 and in inferring r_{eff} from X(t) data. Within the context of the SIR model, the relationship between r_{eff} and W_R/W_L in this Table is universal and applies to any pandemic.

Supplementary Table 2 : Data on recorded cases for the SARS-Cov2 pandemic from the World Health organization that was used in the analysis (from https://covid19.who.int/WHO-COVID-19-global-data.csv)

Supplementary Table 3: Expanded version of Table 1 including demographic and temperature data from https://www.worldometers.info/world-population/population-by-country and https://en.wikipedia.org/wiki/List_of_cities_by_average_temperature.

Author Contributions:
Data Analysis and coding: SB, MM, AT, MZ, GB
Ideas: SB, GB, SD
Paper writing: SB, GB

Author Approval: All authors have seen and approved the manuscript.

Competing Interests: The authors declare there are no competing interests.

Acknowledgements: GB was partly supported by grants from DoD (KC180159) and NIH (P01CA250957). GB thanks Professor Charles DeLisi for many discussions and collaboration on earlier (unpublished) work on the SARS-CoV-2 pandemic using the SIR model. GB also thanks the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611, the Geballe Lab for Advanced Materials at Stanford University, and Professor Gautam Mandal and the Tata Institute of Fundamental Research for their kind hospitality while this work was being done.

References:

1. Thiel V, *Coronaviruses: Molecular and Cellular Biology*, Caister Academic Press 2007, ISBN:978-1-904455-16-5.

2. Rabadan R, *Understanding Coronavirus*, University Printing House, Cambridge, UK (2020), ISBN: 978-1-108-920254.

3. Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, Liu W, Bi Y, Gao GF, *Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses,* Trends Microbiol. 2016 Jun; 24(6):490-502. doi: 10.1016/j.tim.2016.03.003. Epub 2016 Mar 21.

4. Lai MM, Cavanagh D, *The molecular biology of coronaviruses*, Adv Virus Res. 1997;48:1-100.

5. Masters PS, *The molecular biology of coronaviruses*, Adv Virus Res. 2006; 66:193-292; Artika IM, Dewantari AK, Wiyatnoc A, *Molecular biology of coronaviruses: current knowledge.* Heliyon. 2020 Aug; 6(8): e04743.

6. Cherry JD, Krogstad P, *SARS: the first pandemic of the 21st century*, Pediatric Research 2004 Jul; 56(1):1-5. Epub 2004 May 19.

7. Peiris J, Guan Y and Yuen K, *Severe acute respiratory syndrome*. Nat Med 10, S88–S97 (2004). https://doi.org/10.1038/nm1143.

8. J. A. Al-Tawfiq, *Asymptomatic coronavirus infection: Mers-cov and sars-cov-2 (covid-19)*, Travel Med Infect Dis. 2020 May-June; 35: 101608.

9. Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP, *The trinity of COVID-19: immunity, inflammation and intervention*, [published online ahead of print, 2020 Apr 28]. Nat Rev Immunol 20, 363–374 (2020). https://doi.org/10.1038/s41577-020-0311-8

10. Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, Liu S, Zhao P, Liu H, Zhu L, *et. al*., *Pathological findings of covid-19 associated with acute respiratory distress syndrome*, The Lancet Respiratory Medicine, 2020 Apr;8(4):420-422.

11. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li H, Wu X, Xu J, Tu S, Zhang Y, Chen H, Cao B, *Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study*. Lancet. 2020 Mar 28;395(10229):1054-1062. doi: 10.1016/S0140-6736(20)30566-3.

12. Sørensen MD, Sørensen B, Gonzalez-Dosal R, *et. al., Severe acute respiratory syndrome (SARS): development of diagnostics and antivirals*. Ann N Y Acad Sci. 2006;1067(1):500-505. doi:10.1196/annals.1354.072

13. Inglesby TV, *Public Health Measures and the Reproduction Number of SARS-CoV-2* [published online ahead of print, 2020 May 1], JAMA. 2020 Jun 2;323(21):2186-2187. doi: 10.1001/jama.2020.7878.; Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J, *The reproductive number of COVID-19 is higher compared to SARS coronavirus*, Journal of Travel Medicine, 2020 Mar 13;27(2):taaa021. doi: 10.1093/jtm/taaa021.

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

15. Flaxman, S, Mishra, S, Gandy, A et al. *Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe*. Nature 584, 257–261 (2020). https://doi.org/10.1038/s41586-020-2405-7; Davies NG, Kucharski AJ,Eggo RM, Gimma A, Edmunds WJ, *Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study*, The Lancet, July 01, 2020, Volume 5, Issue 7, e375-e385; Ferguson et al., Imperial College Covid-19 Response Team, Report 9(https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/).

16. Lu FS, Nguyen AT, Link NB, Davis JT, Chinazzi M, Xiong X , Vespignani A, Lipsitch M, Santillana M, *Estimating the Cumulative Incidence of COVID-19 in the United States Using Four Complementary Approaches*, medRxiv 2020.04.18.20070821; doi: https://doi.org/10.1101/2020.04.18.20070821.

17. COVID-19 Antibody Seroprevalence in Santa Clara County, California
Bendavid E, Mulaney B, Sood N, Shah S, Ling E, Bromley-Dulfano R, Lai C, Weissberg Z, Saavedra-Walker R, Tedrow J, Tversky D, Bogan A, Kupiec T, Eichner, Gupta R, Ioannidis J, Bhattacharya J, *COVID-19 Antibody Seroprevalence in Santa Clara County, California,*
Sood N, Simon P, Ebner P, et al. *Seroprevalence of SARS-CoV-2–Specific Antibodies Among Adults in Los Angeles County, California, on April 10-11*, JAMA. 2020;323(23):2425–2427. doi:10.1001/jama.2020.8279

18. https://www.washingtonpost.com/investigations/coronavirus-death-toll-americans-are-almost-certainly-dying-of-covid-19-but-being-left-out-of-the-official-count/2020/04/05/71d67982-747e-11ea-87da-77a8136c1a6d_story.html.

19. https://www.nytimes.com/interactive/2020/06/01/us/coronavirus-deaths-new-york-new-jersey.html

20. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J, *Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2)*, Science, 01 May 2020, Vol. 368, Issue 6490, pp. 489-493.

21. https://covid-19.bsvgateway.org

22. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et. al. *Estimates of the severity of coronavirus disease 2019: a model-based analysis*, The Lancet, June 01, 2020, Volume 20, Issue 6, p669-677.

23. https://penn-chime.phl.io/

24. https://www.kff.org/policy-watch/covid-19-models/

25. Branas CC, Rundle A, Pei S, Yang W, Carr BG, Sims S, Zebrowski A, Doorley R, Schluger N, Quinn JW, Shaman J, *Flattening the curve before it flattens us: hospital critical care capacity limits and mortality from novel coronavirus (SARS-CoV2) cases in US counties,* medRxiv 2020.04.01.20049759; doi:https://doi.org/10.1101/2020.04.01.20049759

26. Anastassopoulou C, Russo L, Tsakris A, Siettos C. *Data-based analysis, modelling and forecasting of the COVID-19 outbreak*. PLoS One. 2020, 15(3):e0230405. doi:10.1371/journal.pone.0230405.

27. Casella F. *Can the COVID-19 epidemic be controlled on the basis of daily test reports?* Preprint 2020: https://arxiv.org/abs/2003.06967.

28. Giordano, G., Blanchini, F., Bruno, R. et al. *Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy.* Nat Med 26, 855–860 (2020). https://doi.org/10.1038/s41591-020-0883-7

29. Raines K, Doniach S, Bhanot G, *The transmission of SARS-CoV-2 is likely comodulated by temperature and by relative humidity*, PLoS One. 2021 Jul 29;16(7):e0255212. doi: 10.1371/journal.pone.0255212. PMID: 34324570; PMCID: PMC8321224.

30. Kermack W, McKendric A (1991), *Contributions to the mathematical theory of epidemics – I. Bulletin of Mathematical Biology*. 53 (1–2): 33–55; ibid. *Contributions to the mathematical theory of epidemics – II. The problem of endemicity*", Bulletin of Mathematical Biology. 53 (1–2):

57–87. doi:10.1007/BF02464424. PMID 2059742; ibid. *Contributions to the mathematical theory of epidemics – III. Further studies of the problem of endemicity*. Bulletin of Mathematical Biology. 53 (1–2): 89–118. doi:10.1007/BF02464425. PMID 2059743.

31. Huppert A, Katriel G. *Mathematical modelling and prediction in infectious disease epidemiology*. Clin Microbiol Infect. 2013;19(11):999-1005. doi:10.1111/1469-0691.12308

32. Brauer F, *Mathematical epidemiology: Past, present, and future*. Infect Dis Model (2017) 2(2):113-127. Published 2017 Feb 4. doi:10.1016/j.idm.2017.02.001

33. Colizza V, Barrat A, Barthelemy M, Valleron AJ, Vespignani A, *Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions*. PLoS Med. 2007;4(1):e13. doi:10.1371/journal.pmed.0040013.

<u>Appendix A: Some useful results.</u>

The rescaled equations for the pandemic dynamics are:

$$\frac{ds(\tau)}{d\tau} = -r_{eff}\, s(\tau) i(\tau) \quad \text{(A1)}$$

$$\frac{di(\tau)}{d\tau} = r_{eff}\, s(\tau) i(\tau) - i(\tau) \quad \text{(A2)}$$

$$\frac{dr(\tau)}{d\tau} = i(\tau) \quad \text{(A3)}$$

The scaled quantities $s(\tau)$, $i(\tau)$ are related to S(t), I(t) of the SIR model by:

$$s(\tau) = S(t)/N \quad \text{(A4)}$$

$$i(\tau) = I(t)/N \quad \text{(A5)}$$

$$\text{with, } \tau = \gamma_{eff} t = \frac{t}{L_{eff}} \quad \text{(A6)}$$

Dividing (A2) by (A1) gives:

$$\frac{di(\tau)}{ds(\tau)} = \frac{1}{r_{eff} s} - 1 \quad \text{(A6)}$$

Using the large N boundary conditions s(0) = 1, i(0) = 0 generates the exact result:

$$i(\tau) = 1 - s(\tau) + \log(s(\tau))/r_{eff} \quad \text{(A7)}$$

At t=∞, $i(\tau) = 0$. Hence,

$$r_{eff} = -\frac{\log(s(\infty))}{1-s(\infty)} = -\frac{\log(1-f_{tot})}{f_{tot}} \quad \text{(A8)}$$

When $s(\infty) = 1$ (no pandemic), L'Hôpital's rule gives $r_{eff}(s(\infty) = 1) = 1$.

It is easy to see that for $0 \leq s(\infty) < 1, r_{eff} > 1$. Hence, a pandemic requires $r_{eff} > 1$.

From (A2), the maximum in $i(\tau)$ happens when $s(\tau) = 1/r_{eff}$. Hence:

Maximum value of $i(\tau) = 1 - \frac{[1+\log(r_{eff})]}{r_{eff}}, \; r_{eff} > 1$ (A9)

Note that because of (A5), this quantity is the same as the maximum of I(t)/N which is the quantity H_I/N in Eq. 12c in the main text. Hence,

$H_I/N = 1 - \frac{[1+\log(r_{eff})]}{r_{eff}}, \; r_{eff} > 1$ (A10)

For small τ, $s(\tau) \sim 1$. Hence, we can expand the right-hand side of (A7) in powers of $(1 - s(\tau))$. To lowest order,

$$\log(s(\tau)) = \log\left[1 - \left(1 - s(\tau)\right)\right] \cong -\left(1 - s(\tau)\right) \quad \text{(A11)}$$

Substituted into (A7) this gives,

$$i(\tau) = \frac{r_{eff}-1}{r_{eff}} \left(1 - s(\tau)\right) \quad \text{(A12)}$$

Substituting from (A12) into (A1) gives the Logistic Equation:

$$\frac{ds(\tau)}{d\tau} = -(r_{eff} - 1)s(\tau)(1 - s(\tau)) \quad \text{(A13)}$$

whose solution, with the boundary condition $s(0) = 1-\varepsilon$ is:

$$s(\tau) = \frac{1}{\left[1+\varepsilon\, e^{(r_{eff}-1)\tau}\right]} \quad \text{(A14)}$$

Hence, for $\tau \leq \frac{\log(\varepsilon)}{(1-r_{eff})}$,

$$s(\tau) = \left[1 - \varepsilon\, e^{(r_{eff}-1)\tau}\right], \quad \text{(A15)}$$

Combining (A12) and A(15) shows that for $\tau \leq \frac{\log(\varepsilon)}{(1-r_{eff})}$,

$$i(\tau) = \frac{r_{eff}-1}{r_{eff}}\, \varepsilon\, e^{(r_{eff}-1)\tau} \quad \text{(A16a)}$$

Hence, from Eq. 9,

$$x(\tau) = \omega\gamma_1 i(\tau)/\gamma_{eff} = \left(\frac{\omega\gamma_1}{\gamma_{eff}}\right) \frac{r_{eff}-1}{r_{eff}} \varepsilon\, e^{(r_{eff}-1)\tau} \qquad \text{(A16b)}$$

This shows that,

(i) $x(\tau)$ increases exponentially for small τ with exponent $(r_{eff} - 1)$ (A17a)
(ii) X(t) increases exponentially for small t with exponent $(r_{eff} - 1)\gamma_{eff}$ (A17b)

Finally, using similar arguments, it is easy to show that,

$$i(\tau) \rightarrow e^{-(1-r_{eff}s(\infty))\tau} \text{ as } \tau \rightarrow \infty \qquad \text{(A18)}$$

Hence,
(iii) $x(\tau)$ decreases exponentially for large τ with exponent $-(1 - r_{eff}s(\infty))$ (A19a)
(iv) X(t) decreases exponentially for large t with exponent $-(1 - r_{eff}s(\infty))\gamma_{eff}$ (A19b)

Figure 1

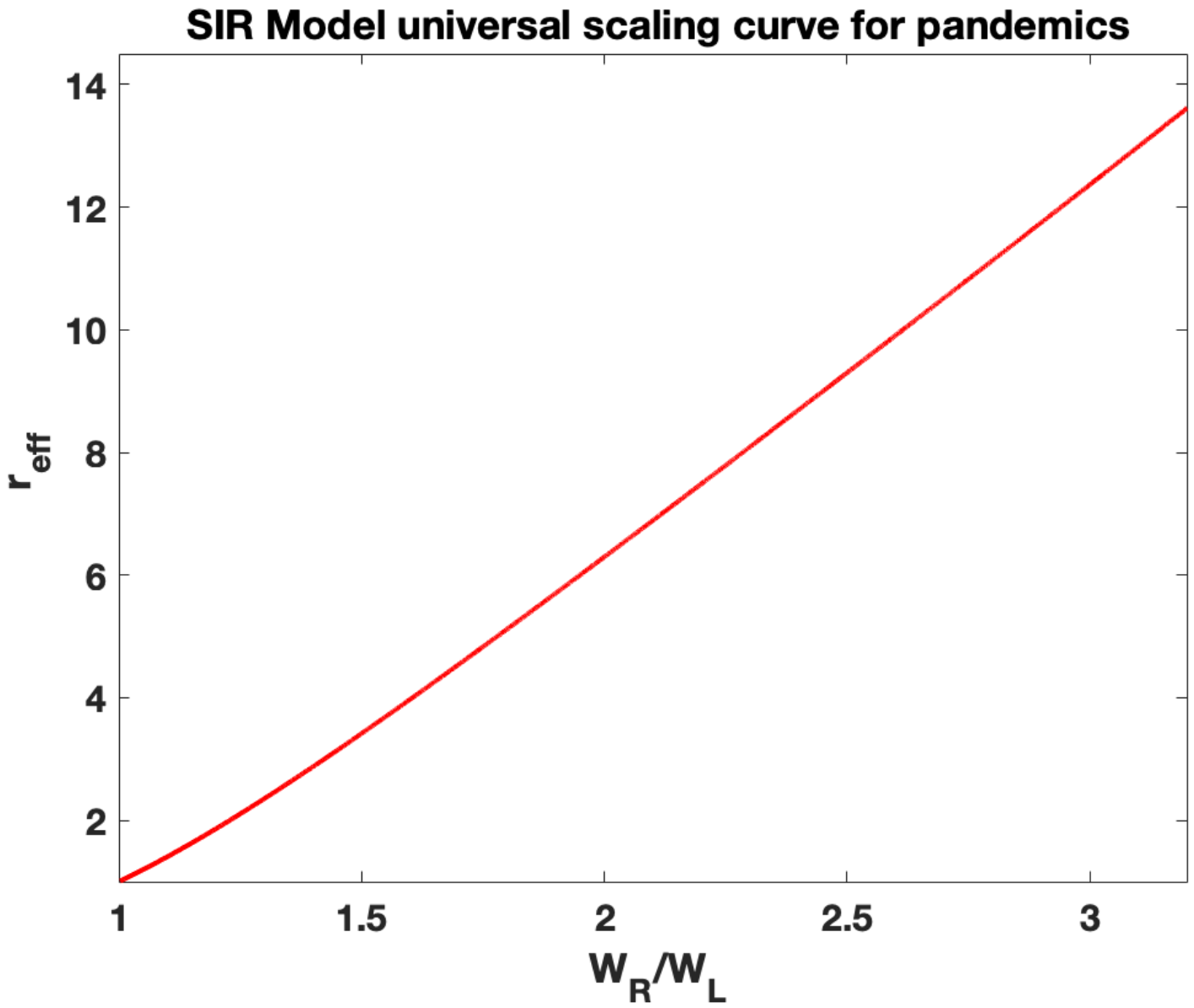
SIR Model universal scaling curve for pandemics
14
12
10
8
6
4
2
r_{eff}
1
1.5
2
2.5
3
W_R/W_L

Figure 2a

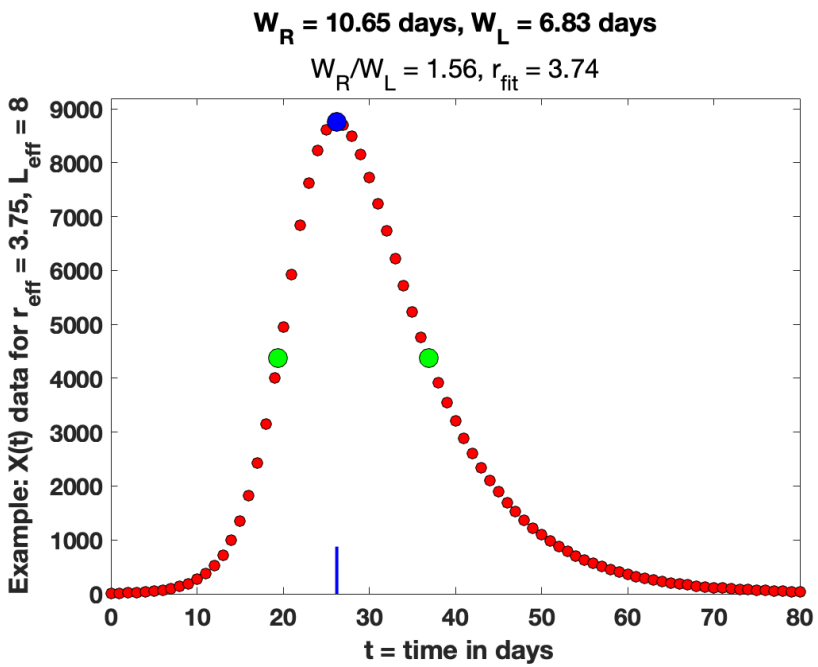

Figure 2b

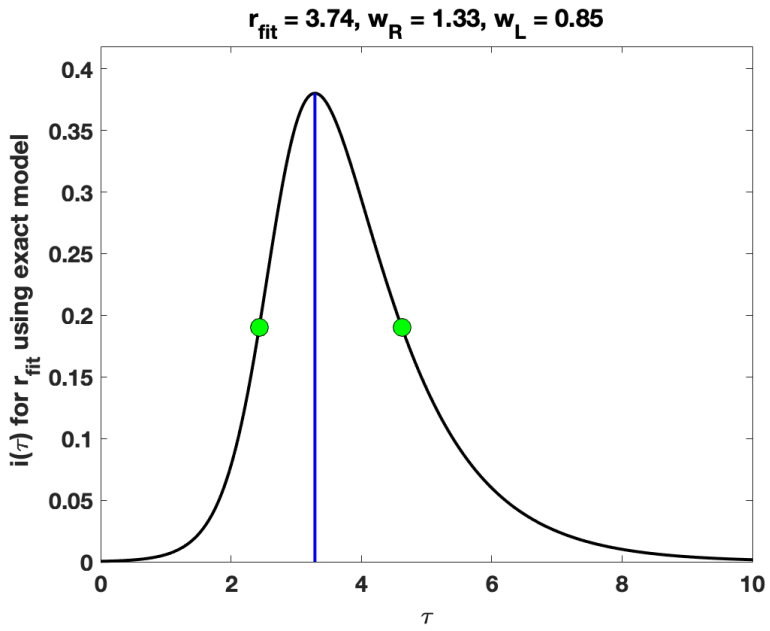

Figure 2c

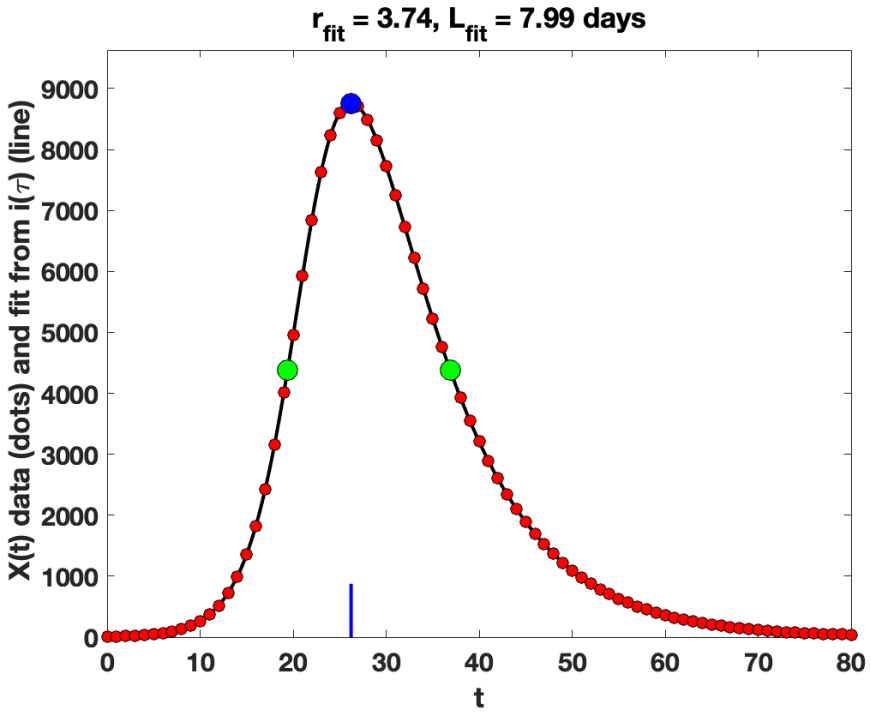

Figure 3a

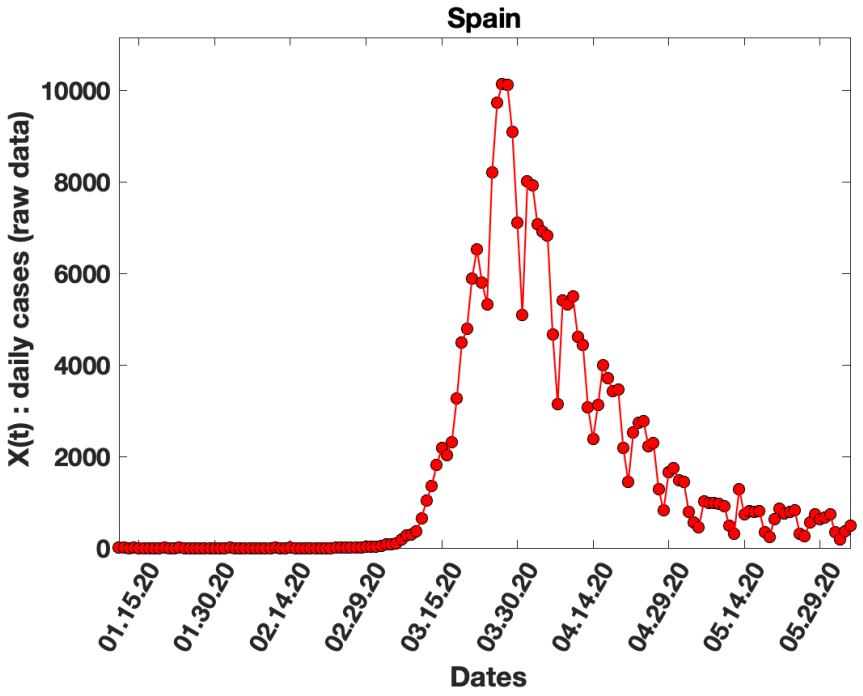

Figure 3b

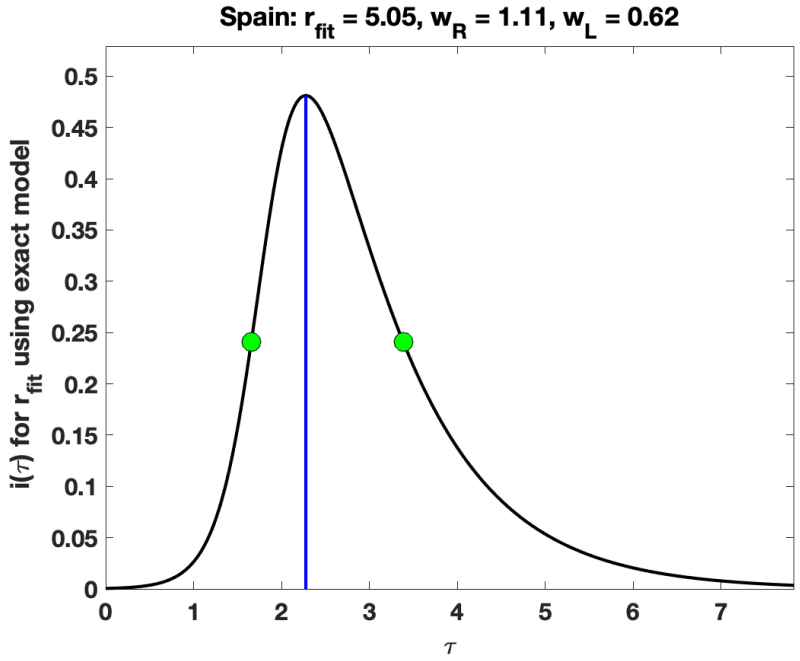

Figure 3c

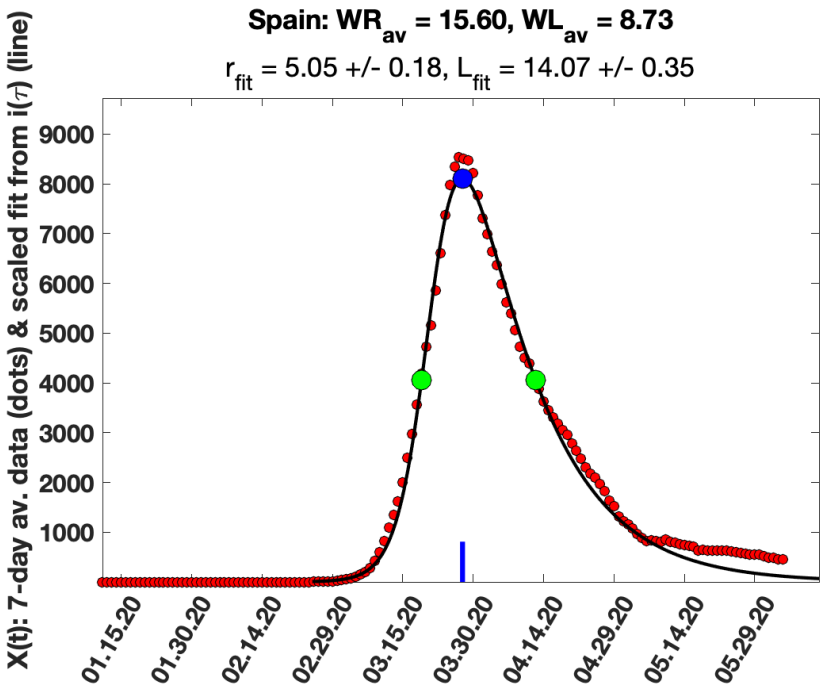

Figure 4a

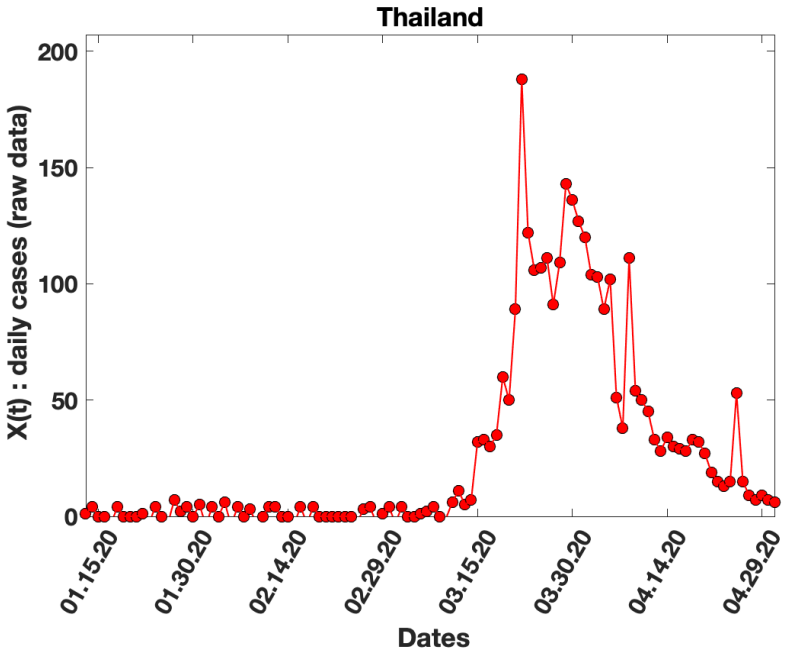

Figure 4b

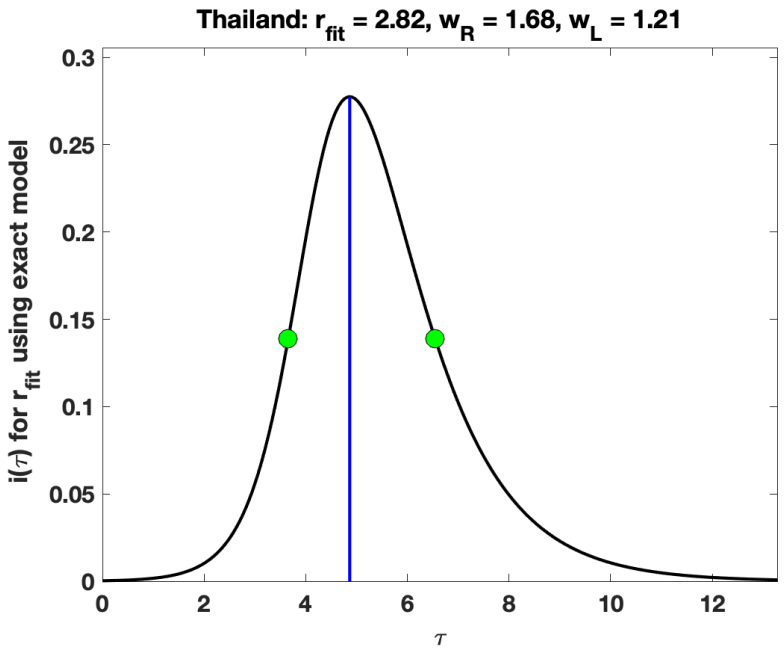

Figure 4c

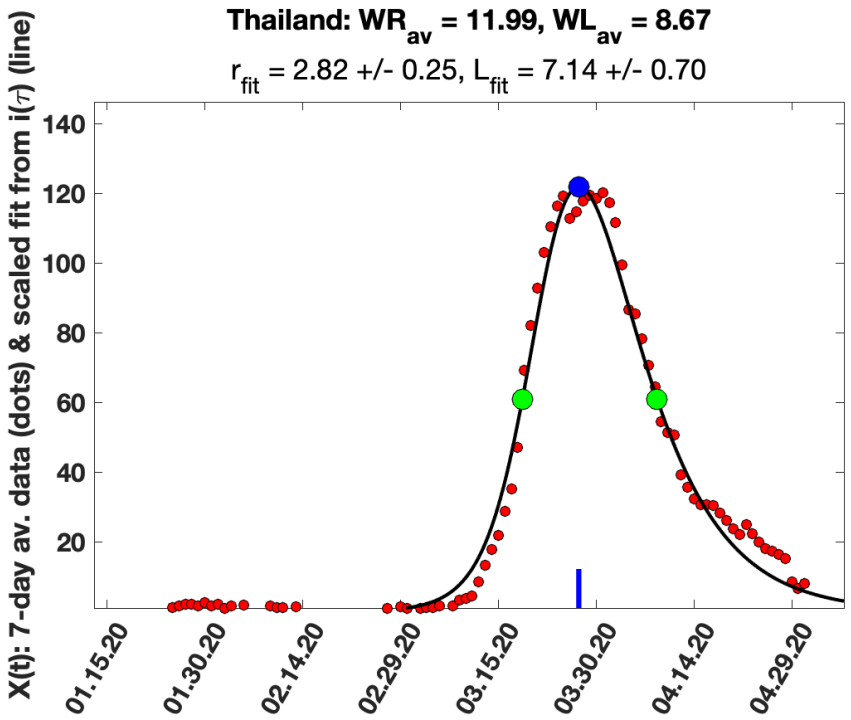

Figure 5a

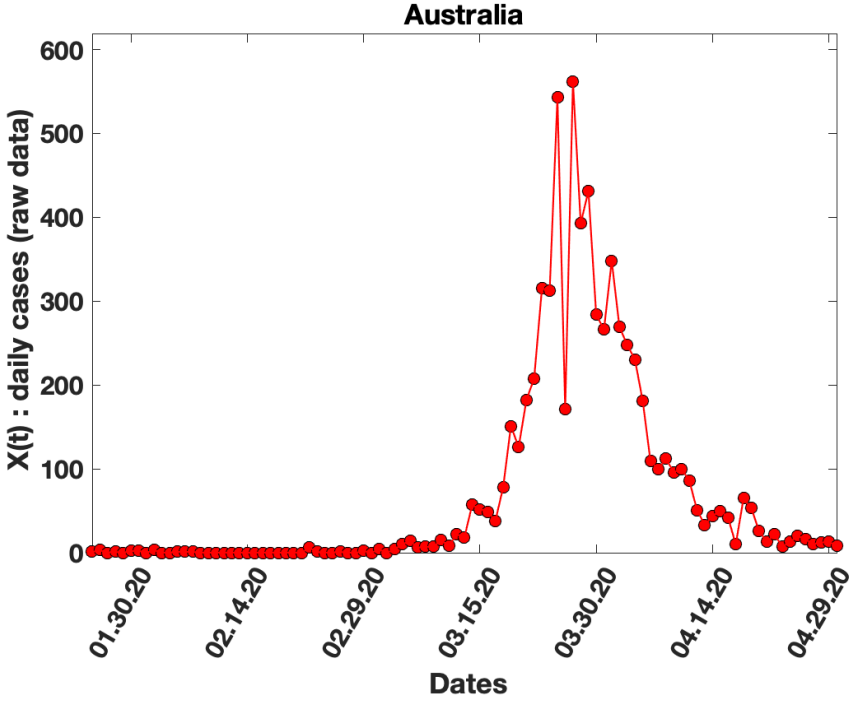

Figure 5b

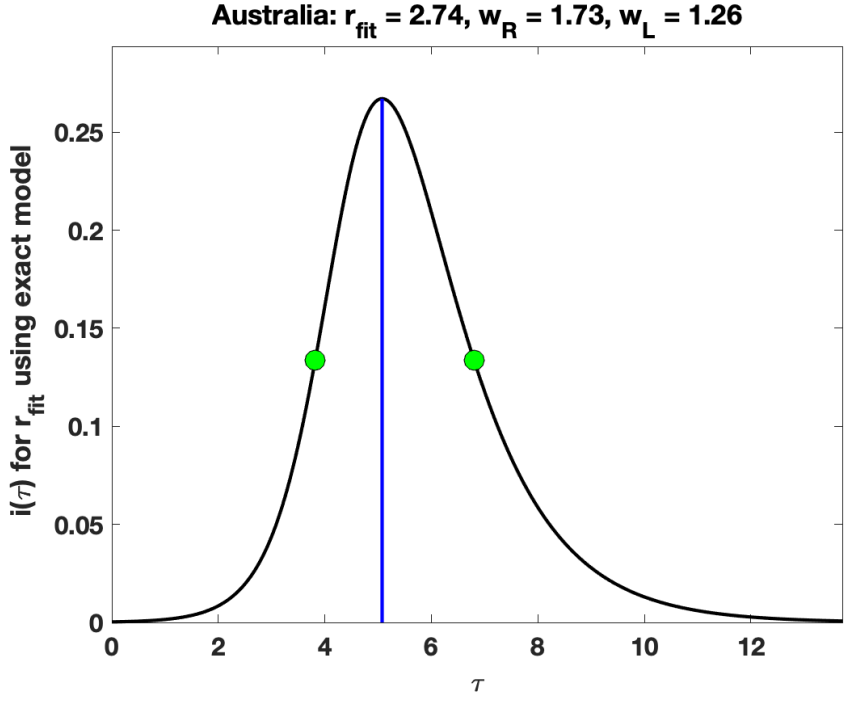

Figure 5c

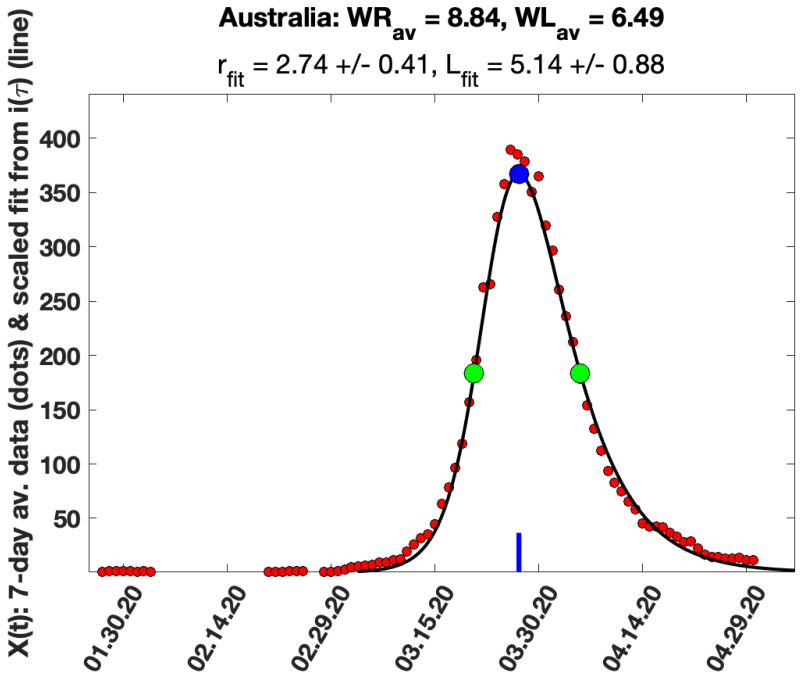

Table 1

Country	WR_av (days)	WL_av (days)	WR/WL	r_fit from WR/WL	L_fit (days)	alpha	f_tot (infected fraction)
Australia	8.84	6.49	1.362	2.74+/- 0.41	5.14 +/- 0.88	0.53 +/- 0.12	0.92
Austria	7.96	4.70	1.694	4.59 +/- 0.47	6.83 +/- 0.52	0.67 +/- 0.09	0.99
Belgium	17.37	15.63	1.111	1.48 +/- 0.23	3.91 +/- 1.67	0.38 +/- 0.17	0.57
Egypt	30.29	23.71	1.278	2.26 +/- 0.21	14.22 +/- 1.81	0.16 +/- 0.03	0.86
France	17.96	11.70	1.535	3.62 +/- 0.48	13.14 +/- 1.59	0.28 +/- 0.05	0.97
Germany	17.35	12.31	1.409	2.93 +/- 0.43	10.72 +/- 1.70	0.27 +/- 0.06	0.94
Ireland	14.68	11.32	1.297	2.35 +/- 0.15	7.23 +/- 0.60	0.33 +/- 0.03	0.87
Israel	19.05	8.95	2.128	7.05 +/- 0.09	19.99 +/- 0.15	0.35 +/- 0.01	1.00
Italy	28.95	12.38	2.338	8.32 +/- 0.22	32.19 +/- 0.42	0.26 +/- 0.01	1.00
Japan	15.21	9.79	1.554	3.78 +/- 0.76	11.51 +/- 2.03	0.33 +/- 0.09	0.98
Netherlands	18.53	17.81	1.040	1.16 +/- 0.09	1.58 +/- 0.83	0.73 +/- 0.39	0.27
New Zealand	9.25	6.42	1.441	3.14 +/- 0.30	6.07 +/- 0.58	0.52 +/- 0.07	0.95
Portugal	25.54	13.80	1.851	5.42 +/- 0.38	23.89+/- 1.13	0.23 +/- 0.02	1.00
Qatar	31.52	22.15	1.423	3.00 +/- 0.05	19.84 +/- 0.37	0.15 +/- 0.00	0.94
Republic of Korea	7.41	5.59	1.326	2.54 +/- 0.40	3.99 +/- 0.77	0.64 +/- 0.16	0.90
South Africa	21.64	21.69	0.998	1.05 +/- 0.03	1.00 +/- 0.33	1.00 +/- 0.35	0.09
Spain	15.60	8.73	1.787	5.05 +/- 0.18	14.07 +/- 0.35	0.36 +/- 0.02	0.99
Switzerland	15.01	10.66	1.408	2.97 +/- 0.46	9.39 +/- 1.54	0.32 +/- 0.07	0.94
Thailand	11.99	8.67	1.383	2.82 +/- 0.25	7.14 +/- 0.70	0.39 +/- 0.05	0.93
Turkey	18.31	13.02	1.406	2.91 +/- 0.07	11.22 +/- 0.31	0.26 +/- 0.01	0.93
UK	33.43	20.90	1.600	3.98 +/- 0.27	26.17 +/- 1.51	0.15 +/- 0.01	0.98