Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Caveats on COVID-19 Herd Immunity Threshold: the Spain case

David Garcia-Garcia
University of Alicante: Universitat d'Alacant
Enrique Morales
University of Alicante: Universitat d'Alacant
Eva S. Fonfría
University of Alicante: Universitat d'Alacant
Isabel Vigo
University of Alicante: Universitat d'Alacant
Cesar Bordehore (cesar.bordehore@ua.es)
University of Alicante: Universitat d'Alacant https://orcid.org/0000-0002-2816-0538

Research Article

Keywords: COVID-19, Herd immunity, Basic reproductive number, R0, dynamic model, Infectious disease modelling

Posted Date: June 7th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-596691/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at Scientific Reports on January 12th, 2022.
See the published version at https://doi.org/10.1038/s41598-021-04440-z.

Title: Caveats on COVID-19 Herd Immunity Threshold: the Spain case

Authors:

David García-García[1] https://orcid.org/0000-0002-7273-9037

Enrique Morales[1] https://orcid.org/0000-0002-7531-4974

Eva S. Fonfría[2] https://orcid.org/0000-0001-5593-8838

Isabel Vigo[1] https://orcid.org/0000-0002-6102-946X

Cesar Bordehore[2,3,*] https://orcid.org/0000-0002-2816-0538

Affiliations:

1. Department of Applied Mathematics. University of Alicante, Alicante, Spain.

2. Multidisciplinary Institute for Environmental Studies "Ramon Margalef", University of Alicante, Spain.

3. Department of Ecology, University of Alicante, Spain.

***Corresponding author:** Cesar Bordehore

Corresponding author email: cesar.bordehore@ua.es

Abstract

After a year of living with the COVID-19 pandemic and its associated consequences, hope looms on the horizon thanks to vaccines. The question is what percentage of the population needs to be immune to reach herd immunity, that is to avoid future outbreaks. The answer depends on the basic reproductive number, R_0, a key epidemiological parameter measuring the transmission capacity of a disease. Besides the virus itself, R_0 depends on the characteristics of the population and their environment. Additionally, the estimate of R_0 depends on the methodology used, the accuracy of data, and the generation time distribution. The aim of this study is to provide a herd immunity threshold for Spain, for which we considered the different combinations of these elements to obtain the R_0 for the Spanish population. Estimates of R_0 range from 1.39 to 3.10, with the largest differences produced by the choice of the methodology to estimate R_0. With these values, the herd immunity threshold ranges from 28.1% to 67.7%, which makes 70% a realistic upper bound for Spain.

Keywords: COVID-19, Herd immunity, Basic reproductive number, R_0, dynamic model, Infectious disease modelling.

1. Introduction

On 11 March 2020, the World Health Organization declared the COVID-19 pandemic, and by 11 March 2021, 2.63 million people had died because of it (https://ourworldindata.org/coronavirus-data-explorer). However, although these are the declared numbers, there were probably many more deaths due to the virus that were not recorded as such due to lack of tests. After a year of struggling, restrictions to lessen the spread of the virus, a downturn in the economy and the cost of human lives, most people are wondering when the pandemic will end. The year 2020 ended with the hopeful approval of some vaccines (https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/coronavirus-disease-covid-19/treatments-vaccines/covid-19-vaccines), but how many people must be vaccinated to return to pre-pandemic life? Under the assumptions that recovered and vaccinated people get permanent immunisation against the different mutations of the SARS-CoV-2 virus, there is a general opinion in Spain that the *herd immunity threshold* (HIT) will be reached when 70% of the population becomes immune. Note that there is no single definition of HIT [1] and this can lead to misunderstandings. In this study the HIT will refer to the minimum proportion of the immune population that will produce a monotonic decrease of new infections, even if restrictions are lifted and society returns to a pre-pandemic level of social contact. The question is whether 70% is a realistic HIT for Spain.

The HIT is usually defined in terms of the *effective reproduction number*, $R_e(t)$, which is the average number of secondary infections produced by an infected individual at time t. Any outbreak starts with $R_e>1$, stabilizes with $R_e=1$, and declines with $R_e<1$. Therefore, the HIT will be reached when $R_e=1$ and $R_e<1$ afterwards. Given the number of susceptible individuals, that is, those that can get infected, $R_e(t)$ can be estimated in an unmitigated epidemic as [2, 3]

$$R_e(t) = R_0 \cdot \frac{S(t)}{N}, \quad (1)$$

where $S(t)$ is the number of susceptible individuals at time t; N is the total number of the population; and R_0 is the *basic reproductive number*, that is, the expected number of secondary infections produced by an infected individual in a population where all individuals are susceptible and there are no measures to reduce transmission [2, 4]. The proportion of susceptible, $S(t)/N$, can be written as $1–q$, where q is the proportion of immune population. Then, if $R_e(t)=1$ (and $R_e(t)<1$ afterwards), HIT equals q by definition. Replacing these equalities in Eq. 1 and operating, we get [5,6])

$$HIT = 1 - \frac{1}{R_0}. \quad (2)$$

Note the direct relationship: the larger the R_0, the larger the HIT. R_0 is used to quantify the transmissibility of the virus, which depends on the virus itself and the characteristics of the population that is being infected. Regarding other infectious diseases, typical values of R_0 are 0.9-2.1 for seasonal flu and 1.4-2.8 for 1918 flu [7], ~3 for SARS-CoV-1 [8] and <0.8 for MERS [9]. For COVID-19, a systematic review of 21 studies [10] found R_0 ranging from 1.9 to 6.5, which means HIT values between 47% and 84%. In 62% of these studies, R_0 was between 2 and 3 (HIT between 50% and 67%). Therefore, 70% is an upper bound of HIT in most of the cases, but not in all.

This study encompasses a detailed analysis of the HIT from different approaches and quantifies the influence of three key factors: (i) source/quality of data; (ii) infectiousness time evolution; and (iii) methodology to estimate R_0.

2. Data

Three COVID-19 daily infection datasets for Spain were used, from 1 January to 29 November 2020: (i) Official infections published by the Instituto de Salud Carlos III (https://cnecovid.isciii.es/covid19/#documentaci%C3%B3n-y-datos); and Infections estimated with the REMEDID algorithm [11] from (ii) official COVID-19 deaths and (iii) excess of all-causes deaths (ED) from European Mortality Monitoring surveillance

system (MoMo, www.euromomo.eu). The REMEDID-derived infection data are more realistic than official infection data since they assimilate seroprevalence studies data [12] and known dynamics of COVID-19 (see [10], for further discussion).

3. Intrinsic growth rate

At the beginning of an outbreak the infections, *I(t)*, increase exponentially [2, 6] and can be fitted to the model

$$I(t) = ae^{rt} + \varepsilon(t), \tag{3}$$

where $\varepsilon(t)$ accounts for errors in the fitting; t is time; a is a positive number determining the point where the function crosses the ordinate axis, and then depends on where the origin of time has been set; and r is a positive number called *intrinsic growth rate* or *Malthusian number*, that defines the increasing rate of the exponential. r is usually the first property that epidemiologists estimate in an outbreak. The higher the r, the higher the speed in the increase of cases. When comparing diseases, r is an indicator of contagiousness, as is R_0. In fact, with enough information about the latent and infectious periods, r (t^{-1} units) can be used to estimate R_0 (dimensionless), although the relationship is not simple [13]. In the *latent period* (*exposed* in a Susceptible-Exposed-Infected-Recovered (SEIR) model), an infected individual cannot produce a secondary infection, unlike in the *infectious period*, where secondary infections may be produced.

When estimating r, it must be kept in mind that $I(n)$, where n denotes time discretized in days, increases exponentially during a short period of time. Consequently, the first problem is to figure out the latest day, n_0, before $I(n)$ will abandon the strictly exponential growth because of the diminishing the number of susceptible individuals. To estimate n_0, we use the property that during the exponential growth $I(n)$ is not only rising, but is accelerating with an increasing acceleration. Then, n_0 is the day where the first maximum of $I''(n)$, the second (discrete) derivative of $I(n)$, is reached. For REMEDID $I(n)$, from both official and MoMo data, n_0 is 23 February 2020 (Figure 1c). Figure 2 shows the least-squares best fit of Equation 3 to REMEDID $I(n)$ truncated at n_0, whose parameters are:

(1) a=11.86 (CI=[11.01, 12.70]) and r=0.1592 (CI=[0.1576, 0.1609]), when MoMo ED are used;

(2) a=10.11 (CI=[9.25, 10.96]) and r=0.1591 (CI=[0.1571, 0.1610]), when official deaths are used.

In both cases, there are similar growth rates. Hereafter, REMEDID $I(n)$ will be estimated from MoMo ED. If the same analysis were carried out with official $I(n)$, which were not reliable at the beginning of the pandemic, we would get r = 0.2336 (CI=[0.2157, 0.2515]) and the end of the exponential growth on 5 March 2020. Despite the larger value of r, the fitted exponential is smaller than those estimated from REMEDID $I(n)$ (Figure 2) because of the horizontal shift due to differences in the a parameter. The end of the exponential growth has been estimated from 7-days running averaged versions of $I(n)$, $I'(n)$, and $I''(n)$. It has to be said that at the beginning of the outbreak, the official data underestimated the number of infections due to the low sampling capability.

[Figures 1 and 2]

4. Estimates of R_0

4.1 Generation time

During the infectious period, an infected individual may produce a secondary infection. However, the individual's infectiousness is not constant during the infectious period, but it can be approximated by the probability distribution of the *generation time* (GT), which accounts for the time between the infection of a primary case and the infection of a secondary case. Unfortunately, such distribution is not as easy to estimate as that of the *serial interval*, which accounts for the time between the onset of symptoms in a primary case to the onset of symptoms of a secondary case. This is because the time of infection is more difficult to detect than the time of symptoms onset. Ganyani et al. [14] developed a methodology to estimate the distribution of the GT from the distributions of the incubation period and the serial interval. Assuming an incubation period following a gamma distribution with a mean of 5.2 days and a standard deviation (SD) of 2.8 days, they estimated the serial interval from 91 and 135 pairs of documented infector-infectee in Singapore and Tianjin (China). Then, they found that the GT followed a gamma distribution with mean=5.20 (CI=[3.78, 6.78]) days and SD=1.72 (CI=[0.91, 3.93]) for Singapore (hereafter GT_1), and with mean = 3.95 (CI=[3.01, 4.91]) days and SD=1.51 (CI=[0.74, 2.97]) for Tianjin (hereafter GT_2). Ng et al. [15] applied the same methodology to 209 pairs of infector-infectee in Singapore and determined a gamma distribution with mean=3.44 (CI=[2.79, 4.11]) days and SD 2.39 (CI=[1.27, 3.45]; hereafter GT_3). Figure 3 shows the probability density functions (PDF) of such distributions, f_{GT}. The differences between them are remarkable. For example, the 54.5%, 81.0%, and 80.7% of the contagions are produced in a pre-symptomatic stage (in the first 5.20 days after primary infection) assuming GT_1, GT_2, and GT_3, respectively.

Theoretically, assuming that the incubation periods of two individuals are independent and identically distributed, which is quite plausible, the expected/mean values of the GT and the serial interval should be equal [16; 17]. The mean of the serial interval is easier to estimate than that of the GT. For that reason, we assume a mean serial interval as estimated from a meta-analysis of 7 studies, which is 4.97 days [18], is more reliable than the aforementioned means of the GT. This value is within the error estimates of the

means of GT_1 and GT_2, but not for GT_3. Then, we construct a theoretical distribution for the GT that follows a gamma distribution (hereafter GT_{th}) with mean=4.97 days and SD=1.88 days. This theoretical distribution can be seen in Figure 3 and approximates the average PDF of three gamma distributions with mean=4.97 and the SD of GT_1, GT_2, and GT_3. GT_{th} shows 63.1% of pre-symptomatic contagions.

[Figure 3]

4.2 R_0 from r

In theory, the basic reproduction number R_0 can be estimated as far as the intrinsic growth rate r, and the distributions of both the latent and infectious periods are known [13, 19-21]. The latent period accounts for the period during which an infected individual cannot infect other individuals. It is observed in diseases for which the infectious period starts around the end of the incubation period, as happens with influenza [22] and SARS [23]. However, from Figure 3 it is inferred that COVID-19 is transmissible from the moment of infection, and we will assume a null latent period. Then, if the GT follows a gamma distribution, R_0 can be estimated from the formulation of Anderson and Watson [19], which was adapted to null latent periods by Yan [13] as

$$R_0 = \frac{mean_{GT}}{1-\left(1+mean_{GT}\cdot r\cdot \frac{1}{shape_{GT}}\right)^{-shape_{GT}}}\cdot r, \quad (4)$$

Where $mean_{GT}$ is the mean GT and $shape_{GT}$ is one of the two parameters defining the gamma distribution, which can be estimated as

$$shape_{GT} = \frac{(mean_{GT})^2}{(SD_{GT})^2}. \quad (5)$$

For GT_{th}, we get R_0 = 1.50 (CI=[1.48, 1.54]) for REMEDID $I(n)$ and R_0 = 1.76 (CI=[1.71, 1.82]) for official $I(n)$. For the other three GT distributions, R_0 ranges from

1.39 (CI=[1.27, 1.58]) to 1.51 (CI=[1.34, 1.80]) for REMEDID *I(n)* and from 1.59 (CI=[1.40, 1.88]) to 1.78 (CI=[1.51, 2.22]) for official *I(n)* (Table 1). In all cases, R_0 from GT_{th} are within those from the other three GT distributions. The lower (upper) bound of the CI is estimated as the minimum (maximum) R_0 obtained from all the possible combinations of all the CI bounds of *r*, $mean_{GT}$ and SD_{GT}, except for *GTth* where $mean_{GT}$ is fixed. In general, all estimates are lower than those summarised by Park et al. [10].

Alternatively, R_0 can be estimated by applying the Euler-Lotka equation [16, 20],

$$R_0 = \frac{1}{\int_0^{+\infty} e^{-rt}.f_{GT}(t)dt.} \quad (6)$$

In this case, we get values closer to previous estimates [10]. In particular, for GT_{th}, we get R_0=2.12 (CI=[2.05, 2.16]) for REMEDID *I(n)* and R_0=2.90 (CI=[2.71, 3.07]) for official *I(n)*. For the other three GT distributions, R_0 ranges from 1.63 (CI=[1.43, 1.90]) to 2.21 (CI=[1.59, 2.95]) for REMEDID *I(n)* and from 1.96 (CI=[1.59, 2.54]) to 3.10 (CI=[1.84, 4.89]) for official *I(n)* (Table 1).

4.3 R_0 from a dynamical model

We designed a dynamic model with Susceptible-Infected-Recovered (SIR) as stocks that accounts for the infectiousness of the infectors. Such a model is a generalisation of the Susceptible-Exposed-Infected-Recovered (SEIR) model [24]. Births, deaths, immigration and emigration are ignored, which seems reasonable since the timescale of the outbreak is too short to produce significant demographic changes. For the sake of simplicity, the recovered stock includes recoveries and fatalities, and it is denoted as *R(t)*. A random mixing population is assumed, that is a population where contacts between any two people are equally probable. Time is discretized in days, so the real time variable *t* is replaced by the integer variable *n*. As a consequence, the derivatives in the differential equations defining the dynamic model explained below are discrete derivatives.

The size of the population is fixed at N=100,000, and then, for any day n we get

$$\tilde{S}(n) + (\sum_{k=0}^{20} \tilde{I}(k)) + \tilde{R}(n) = N, \quad (7)$$

where $\tilde{S}(n)$, $\tilde{I}(n)$, and $\tilde{R}(n)$ are the discretized versions of $S(t)$, $I(t)$, and $R(t)$. The summation is a consequence of the infectiousness, which is approximated according to the GT, whose PDF is discretized as

$$\widetilde{f_{GT}}(n) = \int_{n-1}^{n} f_{GT}(t)\, dt, \quad (8)$$

from n=1 to 20. Figure 3 shows $\widetilde{f_{GT}}(n)$ for GT_{th}. Truncating at n=20 accounts for 99.99% of the area below the PDF of all the GT. Then, an infected individual at day n_0 is expected to produce on average

$$\widetilde{R_e}(n_0 + n) \cdot \widetilde{f_{GT}}(n) \quad (9)$$

infections n days later, where $\widetilde{R_e}(n)$ is the discretized version of $R_e(t)$. From this expression, it is obvious that values of $\widetilde{R_e}(n) < 1$ will produce a decline of infections. Conversely, infections at day n_0 are produced by all individuals infected during the previous 20 days as

$$\tilde{I}(n_0) = \tilde{R}_e(n_0) \cdot \left(\sum_{n=1}^{20} \tilde{I}(n_0 - n) \cdot \widetilde{f_{GT}}(n)\right), \quad (10)$$

whose continuous version has been reported in previous studies [16, 25]. The expression in brackets is called total infectiousness of infected individuals at day n_0 [26]. According to Eq. 1, Eq. 10 can be expressed in terms of R_0 as

$$\tilde{I}(n_0) = R_0 \cdot \frac{\tilde{S}(n_0)}{N} \cdot \left(\sum_{n=1}^{20} \tilde{I}(n_0 - n) \cdot \widetilde{f_{GT}}(n)\right). \quad (11)$$

As we want a dynamic model capable of providing $\tilde{I}(n_0)$ from the stocks at time step $n_0 - 1$, we replaced $\tilde{S}(n_0)$ by $\tilde{S}(n_0 - 1)$ in Eq. 11. This assumption makes sense in a discrete domain since the infections at time n_0 take place in the susceptible population at

time $n_0 - 1$. Then, assuming that all stocks are set to zero for negative integers, our dynamic model can be expressed in terms of Eq. 7 and the following differential equations:

$$d\tilde{I}(n_0) = R_0 \cdot \frac{\tilde{S}(n_0-1)}{N} \cdot \left(\sum_{n=1}^{20} \tilde{I}(n_0 - n) \cdot \widetilde{f_{GT}}(n)\right) - \tilde{I}(n_0 - 1), \quad (12)$$

$$d\tilde{S}(n_0) = -\tilde{I}(n_0), \quad (13)$$

$$d\tilde{R}(n_0) = \tilde{I}(n_0 - 21), \quad (14)$$

where $d\tilde{I}$, $d\tilde{S}$, and $d\tilde{R}$ are the (discrete) derivatives of $\tilde{I}$, $\tilde{S}$, and $\tilde{R}$, respectively. Applying the initial conditions $\tilde{S}(0) = N - 1$, $\tilde{I}(0) = 1$, and $\tilde{R}(0) = 0$, it is assumed that the outbreak was produced by only one infector. The latter is not true in Spain, since several independent introductions of SARS-CoV-2 were detected [27]. However, for modelling purposes it is equivalent to introducing a single infection at day 0 or M infections produced by the single infection n days later. Then, the date of the initial time n=0 is accounted as a parameter $date_0$, which is optimised, as well as R_0, to minimise the root-mean square of the residual between the model simulated $\tilde{I}(n)$ and the REMEDID and official estimated $I(n)$ for the period from $date_0$ to n_0.

The model was implemented in Stella Architect software (www.iseesystems.com) and exported to R software with the help of *deSolve* and *stats* packages, and the Brent optimisation algorithm was implemented. For REMEDID $I(n)$ and GT_{th}, we obtained $date_0$=13 December 2019 and R_0=2.71 (CI=[2.67, 2.76]). Optimal solutions combine lower/higher R_0 and earlier/later $date_0$ (Figure 4), which highlights the importance of providing an accurate first infection date to estimate R_0. When the other three GT distributions were considered, we obtained similar $date_0$, ranging from 12 to 17 December 2019, and R_0 values ranging from 2.08 (CI=[1.86, 2.42]) to 2.85 (CI=[2.05, 3.25]; see Table 1). For official infections, $date_0$ was set to 1 January 2020 for all cases, and R_0 ranged from 1.81 (CI=[1.64, 2.07]) to 2.41 (CI=[1.80, 2.91]).

[Figure 4]

5. Herd immunity threshold and discussion

HIT was estimated from R_0 via Eq. 2 and values are shown in Table 1, which range between 28.1% (CI=[21.3, 36.7]) and 64.9% (CI=[51.2, 69.2]) for REMEDID *I(n)* (hereafter HIT_R), and between 37.0% (CI=[28.4, 46.7]) and 67.7% (CI=[45.5, 79.6]) for official *I(n)* (Hereafter HIT_O). The differences between the estimations are determined by three key factors: (i) source/quality of data; (ii) GT distribution; and (iii) methodology to estimate R_0.

In general, official infection data are of poor quality, but if death records and seroprevalence studies were available, the REMEDID algorithm would provide more reliable infections time series [11]. The maximum difference between HIT_R and HIT_O is 14 percentage points. Moreover, official data vary depending on the date of publication. For example, the maximum HIT_O is 67.7%. from data available in February 2021, and 80.1% from data available a year before, in March 2020. The latter is similar to the 80.7% published by Kwok et al. [28] in March 2020, which was obviously based on data available at that time. The February 2021 version of the data is more realistic than the March 2020 one, and the REMEDID-derived infections are more realistic than both of them [11]. In consequence, results based on REMEDID data should be more reliable.

The most influential factor for estimating the HIT is the methodology to estimate R_0, which may produce differences of ~30 percentage points for the same dataset and GT distribution. For each GT, the lowest HIT values were obtained from Eq. 4, but the largest HIT_R and HIT_O are obtained from the dynamic model and Eq. 6, respectively. The CI from Eq. 6 and the dynamic model are longer than those from Eq. 4, meaning that the former are more sensitive to errors in the involved parameters. Moreover, errors from Eq. 6 significantly increase for official data, meaning that it is the methodology most sensitive to data quality. In general, results from Eq. 6 are reconcilable with the other two within the error estimates, but Eq. 4 and the dynamic model are only reconcilable for official data (Table 1).

The selection of a GT produces HIT differences up to 6 percentage points when R_0 is estimated from Eq. 4; 18.7 from Eq.6; and 13.7 from the dynamic model. It is more

difficult to estimate the GT than the serial interval. For that reason, many studies approximate the GT by a serial interval (e.g. [26, 28]). However, though GT and serial interval have the same mean, serial interval presents a larger variance [17], which will underestimate R_0 when using Eq. 6 [16]. Results from Eq. 4 are reconcilable for all GT. On the contrary, results from Eq. 6 and the dynamic model are only reconcilable for some GT. The smallest errors are from GT_{th}, which makes sense since its mean is fixed during the error estimates.

In summary, accurately estimating HIT is quite complicated. In any case, assuming that REMEDID-derived infection data are more accurate than official data, 70% seems to be a good upper bound of HIT. However, the upper bound increases to 80% (accounting for the CI) if we trust in official data. These results are valid for a randomly mixing population with a spread dynamic similar to Spain as a whole. However, even Spanish regions show different dynamics between themselves [11]. Possibly, reaching HIT will be complicated due to new SARS-CoV-2 variants, vaccine hesitancy, and the delayed vaccination for children [29]. But even if it is reached, it will not be the panacea. First, if HIT is reached in most places in a country but there are some specific regions or population subgroups in a region with a percentage of immune individuals below HIT, local outbreaks will be possible for those regions or subgroups. Second, the final size of an epidemic in a randomly mixing population with HIT=70% is reached at 94.7% of infections [5, 24]. This means that the decreasing rate of infections after the HIT may still produce a non-negligible 27% of infections, that is 12.7 million infections in Spain. Third, interpretation of HIT values must be done carefully and overoptimistic messages should be avoided as has been learnt from Manaus in the Brazilian state of Amazonas. In October 2020, it was thought that Manaus had reached the HIT with 76% of infected population [30], which led to a relaxation of the control measures. However, either because the percentage of infected population was not accurately estimated or because the new SARS-CoV-2 P.1 variant was capable of re-infecting, Manaus had a second wave in January 2021 with a higher mortality rate than in the first one [31]. Therefore, health authorities should strictly ensure an adaptive and proactive management of the new situation after theoretical herd immunity is reached.

Conflict of interest: None declared.

Funding statement: This work was supported by the University of Alicante [COVID-19 2020-41.30.6P.0016 to CB] and the Montgó-Dénia Research Station (Agreement Ajuntament de Dénia-O.A. Parques Nacionales, Ministry of the Environment, Spain) [2020-41.30.6O.00.01 to CB].

References

1. Fine P, Eames K, Heymann DL. "Herd Immunity": A Rough Guide. Clinical Infectious Diseases. 2011; 52 (7), 911–916. https://doi.org/10.1093/cid/cir007

2. Anderson RM, May RM. Infectious diseases of humans: Dynamics and control. Oxford: Oxford University Press. 1991; 757 p.

3. Hannon B, Ruth M. Dynamic Modeling of Diseases and Pests. Springer Publishing Company, Incorporated; 2009.

4. Heesterbeek JAP. A brief history of R_0 and a recipe for its calculation. Acta Biotheoretica. 2002; 50, 189-204.

5. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc R Soc Lond A. 1927; 115: 700–721.

6. Keeling M, Rohani P. Modeling Infectious Diseases in Humans and Animals. Princeton; Oxford: Princeton University Press. 2008. doi:10.2307/j.ctvcm4gk0

7. Coburn BJ, Wagner BG, Blower S. Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1). BMC Med. 2009;7:30. doi:10.1186/1741-7015-7-30

8. Bauch CT, Lloyd-Smith JO, Coffee MP, Galvani AP. Dynamically modeling SARS and other newly emerging respiratory illnesses: past, present, and future. Epidemiology 2005;16:791–801, doi:http://dx.doi.org/10.1097/01. ede.0000181633.80269.4c.

9. Breban R, Riou J, Fontanet A. Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. Lancet 2013; published online July 5. http://dx.doi.org/10.1016/S0140- 6736(13)61492-0.

10. Park M, Cook AR, Lim JT, Sun Y, Dickens BL. A Systematic Review of COVID-19 Epidemiology Based on Current Evidence. J. Clin. Med. 2020, 9, 967. https://doi.org/10.3390/jcm9040967

11. García-García, D., María Isabel Vigo, Eva S. Fonfría, Zaida Herrador, Miriam Navarro, Cesar Bordehore. Retrospective Methodology to Estimate Daily Infections from Deaths (REMEDID) in COVID-19: the Spain case study. Scientific Reports. Forthcoming 2021. Preprint available from https://www.medrxiv.org/content/10.1101/2020.06.22.20136960v3

12. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Hernán MA, Pérez-Olmeda M, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. The Lancet. 2020; 396 (10250), 535-544. https://doi.org/10.1016/ S0140-6736(20)31483-5

13. Yan, P. Separate roles of the latent and infectious periods in shaping the relation between the basic reproduction number and the intrinsic growth rate of infectious disease outbreaks. Journal of Theoretical Biology. 2008; 251, 238–252. doi:10.1016/j.jtbi.2007.11.027

14. Ganyani T, Kremer C, Chen D, Torneri A, Faes C, Wallinga J, et al. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data. Euro Surveill. 2020;25:2000257. https://doi.org/10.2807/1560-7917.ES.2020.25.17.2000257

15. Ng S, Kaur P, Kremer C, Tan W, Tan A, Hens N, et al. Estimating Transmission Parameters for COVID-19 Clusters by Using Symptom Onset Data, Singapore, January–April 2020. Emerg Infect Dis. 2021;27(2):582-585. https://dx.doi.org/10.3201/eid2702.203018

16. Britton T, Tomba GS. Estimation in emerging epidemics: biases and remedies. J. R. Soc. Interface 2019; 16, 20180670. http://doi.org/10.1098/rsif.2018.0670

17. Lehtinen S, Ashcroft P, Bonhoeffer S. On the relationship between serial interval, infectiousness profile and generation time. J. R. Soc. Interface. 2021; 18: 20200756. https://doi.org/10.1098/rsif.2020.0756

18. Fonfría ES, Vigo MI, García-García D, Herrador Z, Navarro M, Bordehore C. COVID-19 epidemiological parameters for clinical and mathematical modeling: mini-review and meta-analysis from Asian studies during early phase of pandemic. Frontiers in Medicine. Forthcoming 2021. Preprint available from https://www.medrxiv.org/content/10.1101/2020.06.17.20133587v1

19. Anderson D, Watson R. 1980. On the spread of a disease with gamma distributed latent and infectious periods. Biometrika 1980; 67 (1), 191–198. https://doi.org/10.1093/biomet/67.1.191

20. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive number. Proc. R. Soc. B. 2007; 274, 599–604.

21. Roberts MG, Heesterbeek JAP. Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. J. Math. Biol. 2007; 55, 803–816.

22. Lau LL, Cowling BJ, Fang VJ, Chan KH, Lau EHY, Lipsitch M, et al. Viral shedding and clinical illness in naturally acquired influenza virus infections. J Infect Dis. 2010;201(10):1509–1516. DOI: 10.1086/652241

23. Peiris JS, Chu CM, Cheng VC, Chan KS, Hung IFN, Poon LLM, et al. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. Lancet. 2003; 361(9371):1767–1772.

24. Ma J, Earn DJD. Generality of the Final Size Formula for an Epidemic of a Newly Invading Infectious Disease. Bull. Math. Biol. 2006; 68, 679–702. https://doi.org/10.1007/s11538-005-9047-7

25. Park SW, Sun K, Champredon D, Li M, Bolker, BM, Earn DJD, Weitz, et al. Forward-looking serial intervals correctly link epidemic growth to reproduction numbers. PNAS. 2021; 118 (2) e2011548118; https://doi.org/10.1073/pnas.2011548118

26. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol. 2013 Nov 1;178(9):1505-12. doi: 10.1093/aje/kwt133. Epub 2013 Sep 15. PMID: 24043437; PMCID: PMC3816335

27. Gómez-Carballa A, Bello X, Pardo-Seco J, Pérez del Molino ML, Martinón-Torres F, Salas A. Phylogeography of SARS-CoV-2 pandemic in Spain: a story of multiple introductions, micro-geographic stratification, founder effects, and super-spreaders. Zoological Research. 2020; 41(6): 605-620. doi: 10.24272/j.issn.2095-8137.2020.217

28. Kwok KO, Lai F, Wei WI, Wong SYS, Tang JWT. Herd immunity – estimating the level required to halt the COVID-19 epidemics in affected countries. Journal of infection. 2020; 80 (6), e32-e33. https://doi.org/10.1016/j.jinf.2020.03.027

29. Aschwanden C. Five reasons why COVID herd immunity is probably impossible. Nature 2021; 591, 520-522. doi: https://doi.org/10.1038/d41586-021-00728-2

30. Buss LF, Prete Jr CA, Abrahim CMM, Mendrone Jr A, Salomon T, de Almeida-Neto C, et al. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. Science 2021; 371, 288–292. DOI: 10.1126/science.abe9728

31. Taylor L. Covid-19: Is Manaus the final nail in the coffin for natural herd immunity? BMJ. 2021; 372 :n394 doi:10.1136/bmj.n394

Table 1. R_0 and HIT values estimated from GT_1, GT_2, GT_3, and GT_{th}, and REMEDID and official infections. For $date_0$, Dec. means December 2019, and Jan. means January 2020. Lower (higher) bound of any R_0 confidence interval (CI) is estimated conservatively as the minimum (maximum) of the R_0 estimated from all the combinations of the CI bounds of the involved parameters

Generation time	PDF gamma distribution		R_0 (conservative CI) for Eq.4 and Eq.6; $R_0/date_0$ (conservative CI) for Dyn. model			HIT from Eq. 2 in percentage (conservative CI)	
	Mean	SD		REMEDID [r=0.1592]	Official data [r=0.2314]	REMEDID	Official data
GT_1: Ganyani et al. (2020), Singapore	5.20	1.72	Eq. 4	1.51 (1.34, 1.80)	1.78 (1.51, 2.22)	33.9 (25.6, 44.5)	43.7 (33.7, 55.0)
			Eq. 6	2.21 (1.59, 2.95)	3.1 (1.84, 4.89)	54.7 (37.1, 66.1)	67.7 (45.5, 79.6)
			Dyn. model	2.85/13 Dec (2.05/16 Dec, 3.25/13 Dec)	2.41/1 Jan (1.80/1 Jan, 2.91/1 Jan)	64.9 (51.2, 69,2)	58.5 (44.4, 65.5)
GT_2: Ganyani et al, (2020), Tianjin	3.95	1.51	Eq. 4	1.39 (1.27, 1.58)	1.59 (1.40, 1.88)	28.1 (21.3, 36.7)	37 (28.4, 46.7)
			Eq. 6	1.82 (1.48, 2.19)	2.36 (1.69, 3.16)	45.2 (32.4, 54.3)	57.6 (40.7, 68.4)
			Dyn. Model	2.34/14 Dec (1.90/16 Dec, 2.76/12 Dec)	2.01/1 Jan (1.68/1 Jan, 2.33/31 Dec)	57.3 (47.4, 63.8)	50.2 (40.5, 57.1)
GT_3: Ng et al, (2021), Singapore	3.44	2.39	Eq. 4	1.42 (1.28, 1.58)	1.62 (1.41, 1.86)	29.7 (21.9, 36.7)	38.4 (29.0, 46.3)
			Eq. 6	1.63 (1.43, 1.90)	1.96 (1.59, 2.54)	38.5 (30.0, 47.3)	49 (37.0, 60.7)
			Dyn. Model	2.08/15 Dec (1.86/17 Dec, 2.42/14 Dec)	1.81/1 Jan (1.64/1 Jan, 2,07/1 Jan)	51.9 (46.2, 58.7)	44.8 (39.0, 51.7%)
GT_{th}: theoretical	4.81	1.88	Eq. 4	1.50 (1.48, 1.54)	1.76 (1.71, 1.82)	33.3 (32.3, 35.0)	43.0 (41.5, 45.1)
			Eq. 6	2.12 (2.05, 2.16)	2.90 (2.71, 3.07)	52.7 (51.3, 53.8)	65.5 (63.1, 67.4)
			Dyn. model	2.71/13 Dec (2.67/14 Dec, 2.76/13 Dec)	2.32/1 Jan (2.27/1 Jan, 2.34/1 Jan)	63.1 (62.5, 63.8)	56.8 (55.9, 57.3)

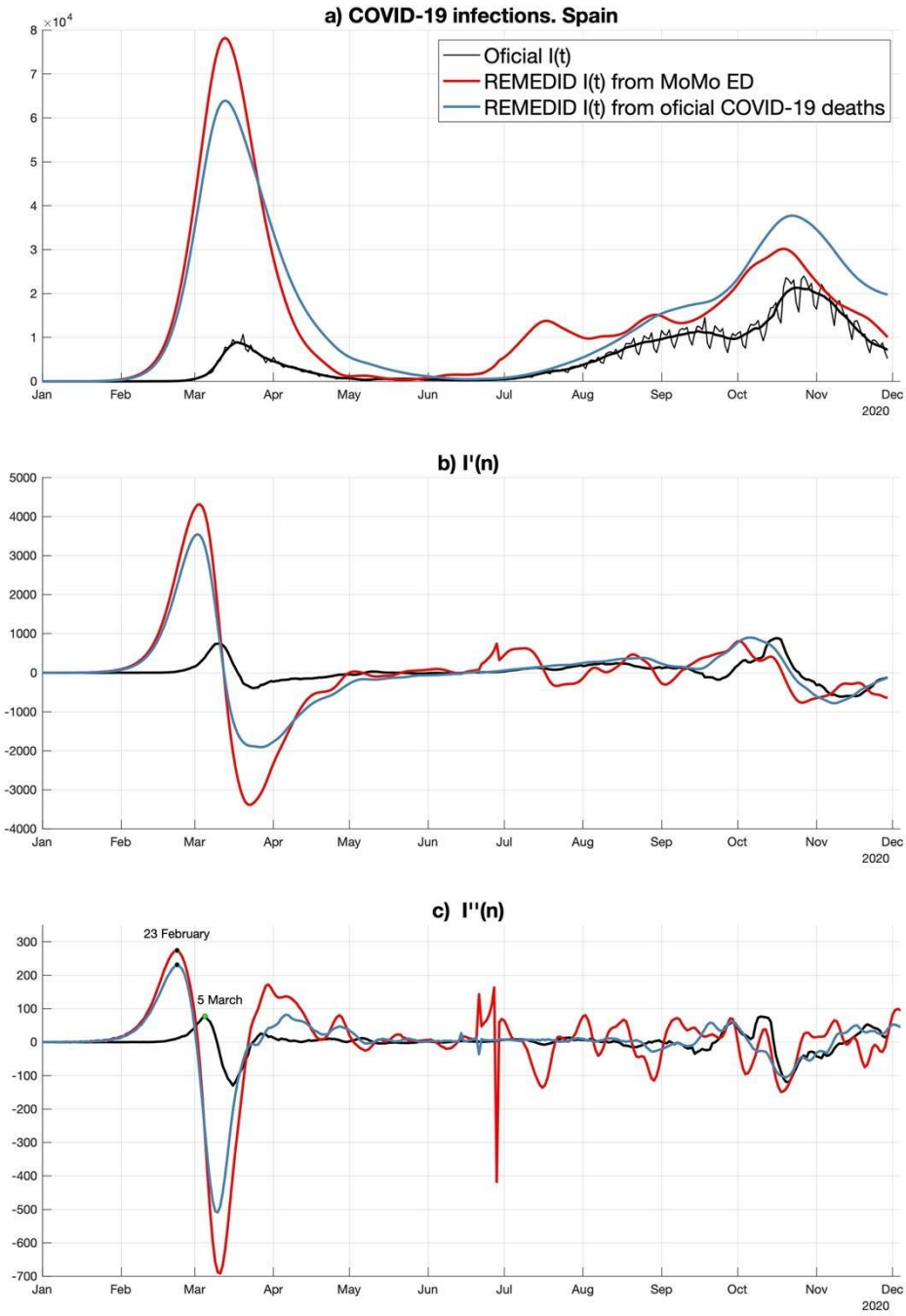

Figure 1. a) Daily new infections: black thin line reflects official data, and thick black line is its 7-days moving average; red and blue lines are infections inferred from REMEDID methodology applied to MoMo excess of dead and to official COVID-19 deaths, respectively. b) and c) are the first and second discrete derivative of time series shown in a). Official *I'(n)* (*I''(n)*) is estimated from the 7-days running mean of official *I(n)* (*I'(n)*). Panels b) and c) show the smoothed versions of *I'(n)* and *I''(n)*, respectively.

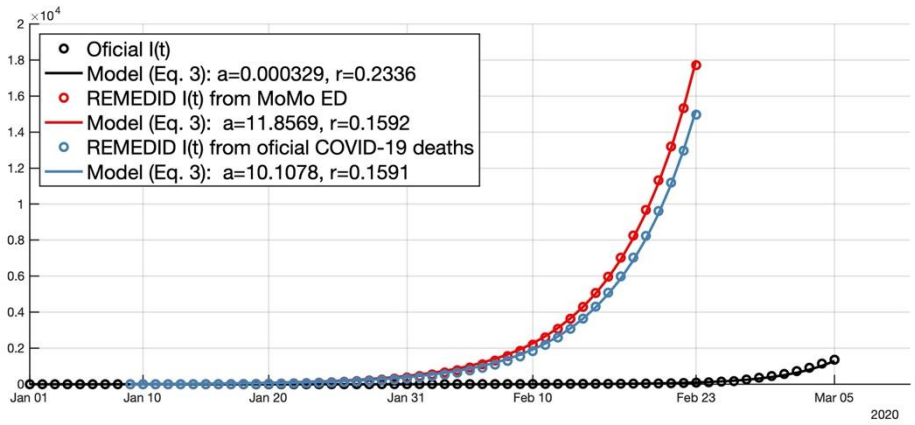

Figure 2. Daily new infections from official data (black dots) for a period of time of 65 days (1 January to 5 March 2020) and inferred from REMEDID applied to MoMo excess of dead (red dots) and official COVID-19 deaths (blue dots) for a period of time of 46 days (January 9 to February 23). Solid lines are the exponential fitting (Eq. 1) to them.

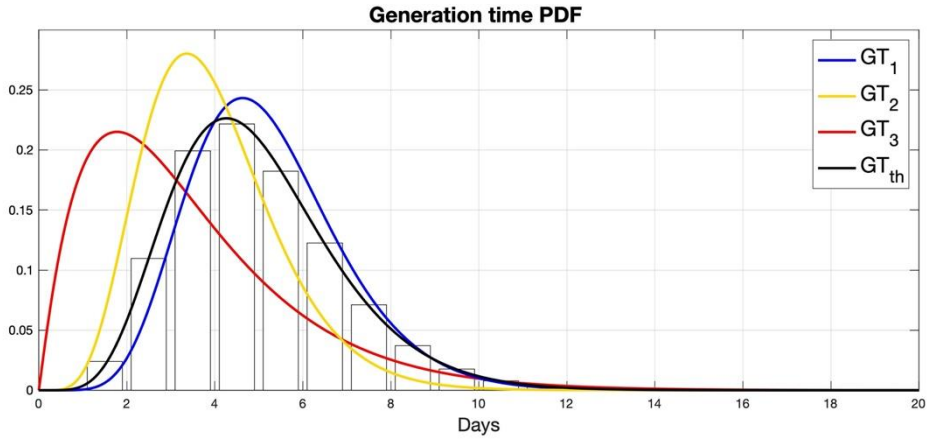

Figure 3: Probability density function of the generation time distribution, $f_{GT}(t)$, of GT_1 (blue line; Ganyani et al. (2020), Singapore), GT_2 (yellow line; Ganyani et al. (2020), Tianjin), GT_3 (red line; Ng et al. (2021), Singapore), and GT_{th} (black line; theoretical distribution). Bars are the discretized version, $\widetilde{f_{GT}}(n)$, of the PDF of GT_{th}.

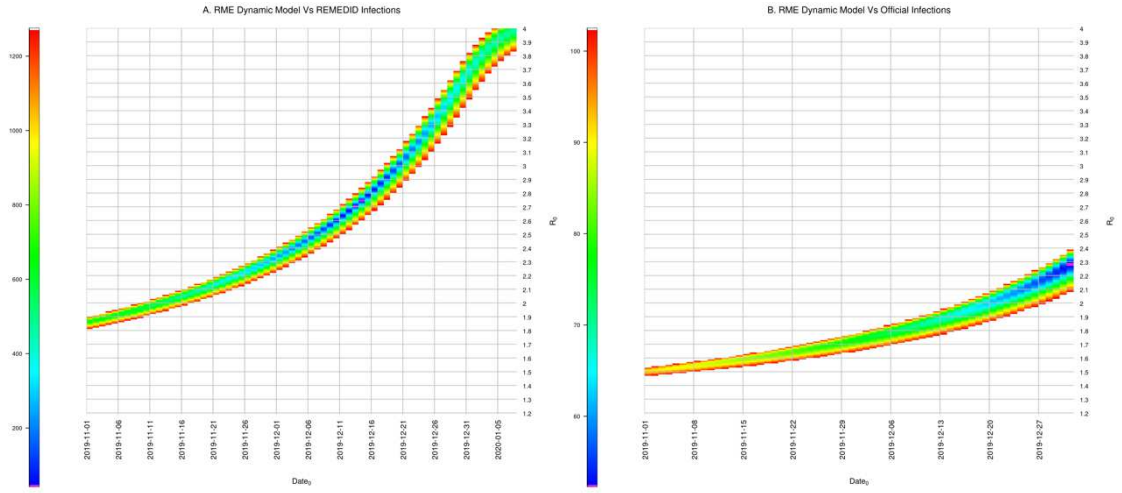

Figure 4. Root-mean square (RMS) of the residuals between infections from the model, which depends on $date_0$ (x-axis) and R_0 (y-axis), and REMEDID (from MoMo ED) and official infections. Parameters optimizing the model are highlighted in purple. RMS larger than 1275 (left panel) and 103 (right panel) are saturated in white.