

The SIR model estimates incorrectly the basic reproduction number for the covid-19 epidemic

Hyun Mo Yang^{a*}, Luis Pedro Lombardi Junior^b,
Ariana Campos Yang^c

^{a,b}Department of Applied Mathematics, State University of Campinas
Praça Sérgio Buarque de Holanda, 651; CEP: 13083-859,
Campinas, SP, Brazil

^cDivision of Allergy and Immunology,
General Hospital of the Medicine School of University of São Paulo
Av. Dr. Eneas Carvalho de Aguiar, 255; CEP: 05403-000,
São Paulo, SP, Brazil

emails: ^ahyunyang@ime.unicamp.br, ^bluispedro_jr@hotmail.com,
^carianacy@gmail.com

Abstract

The transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) becomes pandemic, but presents different patterns in the world. To characterize the epidemic of coronavirus disease 2019 (covid-19) in each countries and regions, mathematical models were formulated aiming the estimation of the basic reproduction number R_0 . Simple mathematical model, the SIR model, provided lower estimation for R_0 , ranging from 1.5 to 3.0. However, more elaborate model presented here estimated higher value for

*Corresponding author: tel. + 55 19 3521-6031

R_0 , 9.24 and 8.0 respectively, for São Paulo State (Brazil) and Spain. Additionally, SIR model estimated R_0 using the severe covid-19 cases, which are not participating in the SARS-CoV-2 transmission chain.

Keywords: mathematical model; estimation; basic reproduction number, São Paulo State (Brazil); Spain

Introduction

At the beginning of the epidemic, and also during the first months, we have two sets of covid-19 data: Severe covid-19 cases (those in hospitals, where they were tested and confirmed), and deaths due to covid-19. Mathematical models were formulated and used to fit the epidemic curve to estimate the basic reproduction number R_0 .

In the literature, the usually assumed basic reproduction number R_0 is around 2.5, see for instance [1] and [2]. The reliable estimation of R_0 is important because this number determines the magnitude of effort to eradicate infection. For instance, the efforts of vaccination to eradicate an infection must be vaccinating a fraction equal or greater than $1 - 1/R_0$ of susceptibles [3]. In [4], analyzing vaccination as a control mechanism, if R_0 is reduced by the vaccine to value lower than one, the number of cases decreased following exponential-type decay.

The estimation of the basic reproduction number R_0 , however, depends on the mathematical model. We consider two examples. The first model is the simplest susceptible - infectious - recovered (SIR) model, and the second is a model taking into account more aspects of the natural history of covid-19 infection encompassing the lethality rate depending on the age, called the elaborated model. These models are fitted taking into account the severe covid-19 data collections from São Paulo State (Brazil) [5] and Spain [6].

São Paulo State, Brazil, has 44.6 million inhabitants (demographic density, $177/km^2$) with 15.3% of the population comprised of elder individuals. The first case was registered on February 26, and partial quarantine was implemented on March 24. Considering that there is a delay around 9 days since the infection and the onset of covid-19, we estimate the basic reproduction number considering collected data from February 26 to April 2.

Spain has 47.4 million inhabitants (demographic density, $92.3/km^2$) with 25.8% of the population comprised of elder individuals. The first case was registered on January 31, and lockdown was implemented on March 16. Considering a delay of around 9 days, we estimate the basic

reproduction number considering collected data from January 31 to March 25.

Yang *et al.* [7], using the elaborated model, estimated R_0 using the first period of covid-19 data representing the natural (without any interventions) epidemic, obtaining 9.24 (São Paulo State) and 8.0 (Spain). The estimation of R_0 based on the first period is in concordance with the definition of the basic reproduction number: Completely susceptible population without constraints (interventions) [3].

1 The SIR model

In the SIR model, the dynamical system is (the varying total population N is $N = S + I + R$)

$$\begin{cases} \frac{d}{dt}S &= \phi N - \beta \frac{I}{N}S - \mu S \\ \frac{d}{dt}I &= \beta \frac{I}{N}S - (\gamma + \mu + \alpha) I \\ \frac{d}{dt}R &= \gamma I - \mu R, \end{cases} \quad (1)$$

where the total population N obeys

$$\frac{d}{dt}N = \phi N - \mu N - \alpha I, \quad (2)$$

and ϕ and μ are, respectively, the birth and natural mortality rates, β , γ , and α are, respectively, the transmission rate, the infectious rate, and the additional mortality (lethality) rate. If $\phi = \mu + \alpha I/N$, then N is constant.

In Appendix A, the steady-state corresponding to the system of equations (1) was analyzed to obtain the basic reproduction number R_0 . For simplicity, let us consider $\phi = \mu$. The system of equations (1), using $R = N - I - S$, can be rewritten as

$$\begin{cases} \frac{d}{dt}S &= \mu N - \beta \frac{I}{N}S - \mu S \\ \frac{d}{dt}I &= (\gamma + \mu + \alpha) (R_{ef} - 1) I \end{cases} \quad (3)$$

where the effective reproduction number R_{ef} is defined by

$$R_{ef} = R_0 \frac{S}{N} = R_0 s, \quad (4)$$

with the basic reproduction number R_0 being given by equation (A.2) changing ϕ by μ , that is, $R_0 = \beta / (\gamma + \mu + \alpha)$.

Let us analyze the system of equations (3) at two boundaries. Let us assume that the first

case of covid-19 is introduced at $t = 0$, that is, the initial conditions supplied to equation (3) are $S(0) = N - 1$ and $I(0) = 1$. For a large population, we can approximate $S \sim N$ and the system of equations can be approximated by

$$\begin{cases} \frac{d}{dt}S & \sim -\beta I \\ \frac{d}{dt}I & \sim (\gamma + \mu + \alpha)(R_0 - 1)I, \end{cases}$$

with $R = N - S - I$. Hence, at the beginning of the epidemic, if we estimate the transmission rate β , we can calculate R_0 using the expression obtained from the steady-state analysis. The system of equations (3) does not approach to a steady-state, but attains it when $\alpha = 0$. In this case, asymptotically ($t \rightarrow \infty$), we have $dI/dt = 0$ if $R_{ef} = R_0 S/N = 1$, that is, $S \rightarrow S^* = s^* N$ and $I \rightarrow I^* = i^* N$, where i^* and s^* are given by equations (A.6) and (A.7), respectively. Therefore, when $\alpha = 0$, at $t = 0$, $R_{ef} = R_0$, and when $t \rightarrow \infty$ (steady-state), $R_{ef} = 1$ (see equation (A.7)), from which we retrieve the well known relationship $s^* = 1/R_0$ [3].

Therefore, based on R_{ef} given by equation (4) when $\phi = \mu$ and $\alpha = 0$, the basic reproduction number R_0 obtained from mathematical modelings provides two useful information: At the beginning of the epidemic ($t = 0$), R_0 gives the magnitude of the initial takeoff of the epidemic, and when epidemic reaches the steady-state (after many waves of the epidemic, that is, $t \rightarrow \infty$), R_0 measures its severity providing the fraction of susceptible individuals, that is, $s^* = 1/R_0$. Between these two extremes, the effective reproduction number R_{ef} dictates the course of an epidemic, which follows decaying oscillations around $R_{ef} = 1$ [4]. It is worth stressing the fact that R_{ef} given by equation (4) is valid only when $\phi = \mu$ and $\alpha = 0$, and when one of these conditions is not valid, R_{ef} given by equation (4) can be used as an approximated value.

The number of accumulated severe covid-19 cases Ω is given by the exits from S , and entering into classes I , that is,

$$\frac{d}{dt}\Omega = \beta \frac{I}{N} S, \quad \text{with } \Omega(0) = 0. \quad (5)$$

To evaluate the parameter β , we calculate

$$Sum = \sum_{i=1}^n [\Omega(t_i) - \Omega^{ob}(t_i)]^2, \quad (6)$$

where Ω^{ob} is the accumulated severe covid-19 registered cases. The value of β that minimizes Sum is the fitted value. We stress the fact that we are estimating the model parameters against the severe covid-19 cases.

We estimate the basic reproduction number using equation (1) with different infective persons at $t = 0$. The model parameters are $\gamma = 1/10$, and, for São Paulo State, $\phi = \mu = 1/(78.4 \times 365)$ and $\alpha = 0.002$, and for Spain, $\phi = \mu = 1/(83.4 \times 365)$ and $\alpha = 0.002$ (all in $days^{-1}$). The transmission rate β is estimated and the basic reproduction number R_0 is calculated using equation (A.2).

For the data collected from São Paulo State, we obtained for $I(0) = 1$, $R_0 = 3.14$ with $Sum = 8.02 \times 10^5$, for $I(0) = 10$, $R_0 = 2.4$ with $Sum = 8.49 \times 10^5$, for $I(0) = 25$, $R_0 = 2.11$ with $Sum = 1.86 \times 10^6$, and for $I(0) = 100$, $R_0 = 1.62$ with $Sum = 5.87 \times 10^6$. Other initial conditions are $S(0) = 44.6$ million and $R(0) = 0$. The lowest Sum occurs when $R_0 = 3.14$.

For the data collected from Spain, we obtained for $I(0) = 1$, $R_0 = 2.97$ with $Sum = 4.16 \times 10^8$, for $I(0) = 10$, $R_0 = 2.5$ with $Sum = 6.32 \times 10^8$, for $I(0) = 25$, $R_0 = 2.3$ with $Sum = 1.13 \times 10^9$, and for $I(0) = 100$, $R_0 = 2.06$ with $Sum = 1.19 \times 10^9$. Other initial conditions are $S(0) = 47.4$ million and $R(0) = 0$. The lowest Sum occurs when $R_0 = 2.97$.

We observed that the larger the value of $I(0)$, the smaller is the estimated R_0 . By the stringent definition of R_0 , we must consider $I(0) = 1$. However, the initial condition $I(0) > 1$ mimics the first case of covid-19 occurring earlier than the time $t = 0$. The Singapore University of Technology and Design [8] estimated R_0 using $I(0) = 100$ for different countries.

2 The elaborated model

One of the main aspects of covid-19 is increased lethality in elder individuals. For this reason, a population is divided into two groups, composed of young (60 years old or less, denoted by subscript y), and elder (60 years old or more, denoted by subscript o) individuals. The vital dynamic of this community is described by the per-capita rates of birth (ϕ) and mortality (μ), and φ is the aging rate, that is, the flow from young subpopulation y to elder subpopulation o .

Since we are dealing with the initial phase of the epidemic, the model is that presented in [7], but the compartments related to quarantine and mass testing are removed. Hence, for each subpopulation j ($j = y, o$), individuals are divided into six classes: susceptible S_j , exposed and incubating E_j , asymptomatic A_j , symptomatic persons in the initial phase of covid-19 (or pre-diseased) D_{1j} , and symptomatic persons with severe covid-19 D_{2j} , mild covid-19 Q_{2j} . However, all young and elder individuals in classes A_j , Q_{2j} , and D_{2j} enter into the same immune class I (this is the 7th class, but common to both subpopulations).

The natural history of new coronavirus infection is the same for young ($j = y$) and elder ($j = o$) subpopulations. We assume that persons in the asymptomatic (A_j), pre-diseased (D_{1j}),

and a fraction z_j of mild covid-19 (Q_{2j}) classes are transmitting the virus, and other infected classes $((1 - z_j) Q_{2j}$ and D_{2j}) are under voluntary or forced isolation. Susceptible persons are infected according to $\lambda_j S_j$ (known as the mass action law [3]) and enter into class E_j , where λ_j is the per-capita incidence rate (or force of infection) defined by $\lambda_j = \lambda (\delta_{jy} + \psi \delta_{jo})$, with λ being

$$\lambda = \frac{1}{N} (\beta_{1y} A_y + \beta_{2y} D_{1y} + \beta_{3y} z_y Q_{2y} + \beta_{1o} A_o + \beta_{2o} D_{1o} + \beta_{3o} z_o Q_{2o}), \quad (7)$$

where δ_{ij} is Kronecker delta, with $\delta_{ij} = 1$ if $i = j$, and 0, if $i \neq j$; and β_{1j} , β_{2j} and β_{3j} are the transmission rates, that is, the rates at which a virus encounters a susceptible people and infects him/her. After an average period $1/\sigma_j$ in class E_j , where σ_j is the incubation rate, exposed individuals enter into the asymptomatic class A_j (with probability p_j) or pre-diseased class D_{1j} (with probability $1 - p_j$). After an average period $1/\gamma_j$ in class A_j , where γ_j is the recovery rate of asymptomatic individuals, asymptomatic individuals acquire immunity (recovered) and enter into immune class I . Possibly asymptomatic individuals can manifest symptoms at the end of this period, and a fraction $1 - \chi_j$ enters into mild covid-19 class Q_{2j} . For symptomatic persons, after an average period $1/\gamma_{1j}$ in class D_{1j} , where γ_{1j} is the infection rate of pre-diseased individuals, pre-diseased individuals enter into severe covid-19 class D_{2j} (with probability $1 - m_j$) or class Q_{2j} (with probability m_j). Individuals in class D_2 acquire immunity after period $1/\gamma_{2j}$, where γ_{2j} is the recovery rate of severe covid-19, and enter into class I or die under the disease-induced (additional) mortality rate α_j . Individuals in class Q_{2j} acquire immunity after period $1/\gamma_{3j}$, where γ_{3j} is the recovery rate of mild covid-19, and enter into immune class I .

The new coronavirus transmission model is described by the system of ordinary differential equations, with $j = y, o$. Equations for susceptible individuals are

$$\begin{cases} \frac{d}{dt} S_y &= \phi N - (\varphi + \mu) S_y - \lambda S_y \\ \frac{d}{dt} S_o &= \varphi S_y - \mu S_o - \lambda \psi S_o, \end{cases} \quad (8)$$

for infectious individuals,

$$\left\{ \begin{array}{l} \frac{d}{dt} E_j = \lambda (\delta_{jy} + \psi \delta_{jo}) S_j - (\sigma_j + \mu) E_j \\ \frac{d}{dt} A_j = p_j \sigma_j E_j - (\gamma_j + \mu) A_j \\ \frac{d}{dt} D_{1j} = (1 - p_j) \sigma_j E_j - (\gamma_{1j} + \mu) D_{1j} \\ \frac{d}{dt} Q_{2j} = (1 - \chi_j) \gamma_j A_j + m_j \gamma_{1j} D_{1j} - (\gamma_{3j} + \mu) Q_{2j} \\ \frac{d}{dt} D_{2j} = (1 - m_j) \gamma_{1j} D_{1j} - (\gamma_{2j} + \mu + \alpha_j) D_{2j}, \end{array} \right. \quad (9)$$

and for immune individuals,

$$\frac{d}{dt} I = \chi_y \gamma_y A_y + \gamma_{3y} Q_{2y} + \gamma_{2y} D_{2y} + \chi_o \gamma_o A_o + \gamma_{3o} Q_{2o} + \gamma_{2o} D_{2o} - \mu I, \quad (10)$$

where $N_j = S_j + E_j + A_j + D_{1j} + Q_{2j} + D_{2j}$, and $N = N_y + N_o + I$ obeys

$$\frac{d}{dt} N = (\phi - \mu) N - \alpha_y D_{2y} - \alpha_o D_{2o}, \quad (11)$$

with the initial number of population at $t = 0$ being $N(0) = N_0 = N_{0y} + N_{0o}$, where N_{0y} and N_{0o} are the size of young and elder subpopulations at $t = 0$. If $\phi = \mu + (\alpha_y D_{2y} + \alpha_o D_{2o}) / N$, the total size of the population is constant. The accumulated severe covid-19 cases Ω is obtained from

$$\frac{d}{dt} \Omega = \gamma_{1y} D_{1y} + \gamma_{1o} D_{1o}, \quad \text{with } \Omega(0) = 0. \quad (12)$$

Table 2 summarizes the model parameters. The description of the assigned values can be found in [7]. The transmission and additional mortality rates are estimated.

The basic reproduction number R_0 for the system of equations (8), (9) and (10) written in terms of the corresponding fractions is

$$R_0 = (R_{1y} + R_{2y}) s_y^0 + (R_{1o} + R_{2o}) s_o^0, \quad (13)$$

where

$$\left\{ \begin{array}{l} s_y^0 = \frac{\phi}{\phi + \varphi} \\ s_o^0 = \frac{\varphi}{\phi + \varphi}, \end{array} \right.$$

Table 1: Summary of the model parameters ($j = y, o$) and values (rates in $days^{-1}$, and proportions are dimensionless). The values (*) correspond to São Paulo State. For Spain, $\phi = \mu = 1/(83.4 \times 365) days^{-1}$, $\varphi = 1.14 \times 10^{-5} days^{-1}$, and $\psi = 1.1$.

Symbol	Meaning	Value
μ^*	Natural mortality rate	$1/(78.4 \times 365)$
ϕ^*	Birth rate	$1/(78.4 \times 365)$
φ^*	Aging rate	6.7×10^{-6}
$\sigma_y (\sigma_o)$	Incubation rate	$1/5.8 (1/5.8)$
$\gamma_y (\gamma_o)$	Recovery rate of asymptomatic persons	$1/12 (1/14)$
$\gamma_{1y} (\gamma_{1o})$	Infection rate of pre-diseased persons	$1/4 (1/4)$
$\gamma_{2y} (\gamma_{2o})$	Recovery rate of severe covid-19	$1/12 (1/21)$
$\gamma_{3y} (\gamma_{3o})$	Infection rate of mild covid-19 persons	$1/13 (1/16)$
$\alpha_y (\alpha_o)$	Additional mortality rate	$0.00185 (0.0071)$
$z_y (z_o)$	Proportion circulating of mild covid-19 persons	$0.5 (0.2)$
ψ^*	Scaling factor of transmission among elder persons	1.15
$\chi_y (\chi_o)$	Proportion of remaining as asymptomatic persons	$0.98 (0.95)$
$p_y (p_o)$	Proportion of asymptomatic persons	$0.8 (0.8)$
$m_y (m_o)$	Proportion of mild (non-hospitalized) covid-19	$0.92 (0.75)$

and

$$\left\{ \begin{array}{l} R_{1y} = p_y \frac{\sigma_y}{\sigma_y + \phi} \frac{\beta_{1y}}{\gamma_y + \phi} + (1 - p_y) \frac{\sigma_y}{\sigma_y + \phi} \frac{\beta_{2y}}{\gamma_{1y} + \phi} \\ R_{1o} = p_o \frac{\sigma_o}{\sigma_o + \phi} \frac{\beta_{1o}\psi}{\gamma_o + \phi} + (1 - p_o) \frac{\sigma_o}{\sigma_o + \phi} \frac{\beta_{2o}\psi}{\gamma_{1o} + \phi} \\ R_{2y} = \left[p_y \frac{\sigma_y}{\sigma_y + \phi} (1 - \chi_y) \frac{\gamma_y}{\gamma_y + \phi} + (1 - p_y) \frac{\sigma_y}{\sigma_y + \phi} m_y \frac{\gamma_{1y}}{\gamma_{1y} + \phi} \right] \frac{z_y \beta_{3y}}{\gamma_{3y} + \phi} \\ R_{2o} = \left[p_o \frac{\sigma_o}{\sigma_o + \phi} (1 - \chi_o) \frac{\gamma_o}{\gamma_o + \phi} + (1 - p_o) \frac{\sigma_o}{\sigma_o + \phi} m_o \frac{\gamma_{1o}}{\gamma_{1o} + \phi} \right] \frac{z_o \beta_{3o}\psi}{\gamma_{3o} + \phi} \end{array} \right. \quad (14)$$

Details of calculations can be found in [7].

To estimate the transmission rates, we consider $\beta_y = \beta_{1y} = \beta_{2y} = \beta_{3y}$ and $\beta_o = \beta_{1o} = \beta_{2o} = \beta_{3o} = \psi\beta_y$, and we use equation (6) and Ω given by equation (12). The values for the model parameters are those given in Table 1. The basic reproduction number R_0 is calculated using equation (13).

The initial conditions supplied to the system of equations (8), (9) and (10) are, for young

and elder subpopulations,

$$\left\{ \begin{array}{l} \text{young} \left(S_y(0) = N_{0y}, \quad E_y(0) = 30, \quad A_y(0) = 24, \quad D_{1y}(0) = 6, \quad Q_{2y}(0) = 6, \quad D_{2y}(0) = 0 \right) \\ \text{and} \\ \text{elder} \left(S_o(0) = N_{0o}, \quad E_o(0) = 20, \quad A_o(0) = 16, \quad D_{1o}(0) = 4, \quad Q_{2o}(0) = 3, \quad D_{2o}(0) = 1, \right) \end{array} \right. \quad (15)$$

plus $I(0) = 0$, where the initial simulation time $t = 0$ corresponds to the calendar time when the first case was confirmed (February 26 for São Paulo State, and January 31 for Spain). For São Paulo State, $N_{0y} = 37.8$ million and $N_{0o} = 6.8$ million, and for Spain, $N_{0y} = 35.17$ million and $N_{0o} = 12.23$ million.

For the data collected from São Paulo State, we obtained $R_0 = 9.24$, with $Sum = 1.22 \times 10^6$, while for the data collected from Spain, we obtained $R_0 = 8.0$, with $Sum = 5.8 \times 10^8$. If we let $z_y = z_o = 0$ (mild covid-19 cases do not transmit) and $\chi_y = \chi_o = 1$ (asymptomatic individuals do not relapse to mild covid-19), the estimated basic reproduction number is $R_0 = 8.72$ for São Paulo State, with $Sum = 1.23 \times 10^6$, and $R_0 = 7.35$ for Spain, with $Sum = 1.59 \times 10^8$.

3 Discussion and conclusion

Firstly, in the elaborated model, there are several infectious classes but the severe covid-19 cases do not transmit the SARS-CoV-2, for this reason, R_0 does not depend on the additional mortality rates α_y and α_o (see equations (13) and (14)). On the other hand, in the SIR model, there is only one infectious class, and R_0 depends on the additional mortality rate α (see equation (A.2)).

As we have pointed out, at the beginning and also in the early phase of the covid-19 epidemic, only hospitalized severe covid-19 cases were registered after confirmed by serological or PCR test. These individuals are either isolated in hospitals (receiving treatment) or discharged from hospitals but recommend to be isolated in their home. Then, somehow, the majority of these individuals are not participating in the populational SARS-CoV-2 chain transmission. In the SIR model, the unique way to estimate the transmission rate is considering that severe covid-9 cases are forming the infective class I . Hence, the estimation of R_0 provided by the SIR model using severe covid-19 cases is unrealistic.

There are different manners to define an epidemic curve. For instance, one possible definition is the curve formed by those positive for serological and PCR tests. However, in the early phase of (natural) epidemic, the covid-19 epidemic curve must be defined by severe cases, which are the only available data. In the elaborated model, the covid-19 epidemic curve was retrieved by

estimating the transmission rates of asymptomatic, pre-diseased, and a fraction of mild classes. The estimated basic reproduction number was higher than that usually accepted: $R_0 = 9.24$ for São Paulo State, and $R_0 = 8.0$ for Spain. The usually accepted estimation of R_0 is around 3 obtained from the SIR model, which results from the estimated transmission rate of the severe covid-19 class. In [9], it was shown that the ratio between non-apparent and apparent covid-19 is around 24, showing that SARS-CoV-2 is being transmitted by a huge number of hidden cases.

In the elaborated model, the initial conditions $E(0) = 50$ and $I(0) = 1$ were supplied to the dynamical system, resulting in $R_0 = 9.24$ (São Paulo State) and $R_0 = 8.0$ (Spain). Instead of comparing with the SIR model, let us compare with the SEIR model considering the same initial conditions $E(0) = 50$ and $I(0) = 1$ supplied to the elaborated model (see Appendix B). The estimated basic reproduction number for São Paulo State was $R_0 = 3.92$, and for Spain, $R_0 = 4.41$.

Comparing SIR, SEIR, and elaborated model, as the model incorporates more aspects of the natural history of the infection, higher becomes the estimation of R_0 . For SIR and SEIR models, there is not any alternative except considering severe covid-19 cases as infective. However, in the elaborated model, the asymptomatic (A), pre-diseased (D_1) and a fraction of mild covid-19 (Q_2) individuals are transmitting SARS-CoV-2, but the severe covid-19 (D_2) individuals are isolated and do not contribute, except to the hospital staff [9].

As we have pointed out, the estimation of R_0 is important because this number determines the magnitude of effort to eradicate infection. In [7], describing the lockdown in Spain, they argued that if $R_0 = 3$, at least 67% of the population must be isolated to control the epidemic, while for $R_0 = 8$, at least 87% of the population must be isolated. They showed that 70% of the population in lockdown did not explain the covid-19 data in Spain, but 90% of the population in lockdown explained better the observed data.

With respect to the SIR model, Li *et al.* [10] explicitly cited that, by using data from January 10 to February 8, 2020, they estimated the effective reproduction number R_{ef} instead of R_0 , arguing that the most recent common ancestor could have occurred on November 17, 2019. The time elapsed from November 17, 2019 (the first case) to January 10, 2020 (the first day in the estimation) is 54 days. However, on January 23, 2020, Wuhan and other cities of Hubei province imposed a rigid lockdown, showing that they estimated R_0 taking also into account data recorded 16 days after the lockdown. In other words, they used the severe covid-19 data from 54 after the possible first case to 16 days after the implementation of a rigid lockdown (January 10 to February 8).

We conclude that the SIR model is not appropriate to estimate the basic reproduction number R_0 , specifically in the case of the covid-19 epidemic when the severe covid-19 cases

may not transmit SARS-CoV-2 populationally. Elaborated models taking into account important aspects related to the natural history of the infection estimate more accurately R_0 . For instance, the incorporation of the asymptomatic individuals, the pre-diseased individuals, different lethality according to age, for instance, to improve the mathematical model.

References

- [1] Koo JR, *et al.* Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. *Lancet Infect Dis.* 2020; 20 (6): 678-688. doi 10.1016/S1473-3099(20)30162-6.
- [2] Ferguson NM, *et al.* Impact of non-pharmaceutical interventions (NPIs) to reduce covid-19 mortality and healthcare demand. Imperial College covid-19 Response Team 2020; March 16, 2020. doi: <https://doi.org/10.25561/77482>.
- [3] Anderson RM, May, RM. *Infectious Diseases of Human. Dynamics and Control.* Oxford, New York, Tokyo: Oxford University Press; 1991: 757 p.
- [4] Yang HM. Modeling directly transmitted infections in a routinely vaccinated population – The force of infection described by Volterra integral equation. *Applied Mathematics and Computation* 2001; 122 (1): 27-58.
- [5] SEADE, SP contra o novo coronavírus – Boletim completo. Available from: https://www.seade.gov.br/coronavirus/?utm_source=portal&utm_medium=banner&utm_campaign=completo, (accessed on August 14, 2020).
- [6] CNE, Covid19. Available from: <https://cnecovid.isciii.es/covid19/>. (accessed on August 14, 2020).
- [7] Yang HM, Lombardi Junior LP, Campos AC, Castro FFM. Mathematical modeling of the transmission of SARS-CoV-2 – Evaluating the impacts of isolation in São Paulo State (Brazil) and lockdown in Spain associated with protective measures on the epidemic of covid-19. *PLOS One*, submitted (2020). [MedRxiv, <https://www.medrxiv.org/content/10.1101/2020.07.30.20165191v1>.]
- [8] Luo J. When will COVID-19 end? Data-driven prediction. Data-Driven Innovation Lab. 2020 [accessed on October 10, 2020]. Available from: <https://ddi.sutd.edu.sg/publications>.

- [9] Yang HM, Lombardi Junior LP, Castro FFM, Campos AC. Mathematical model describing covid-19 in São Paulo State, Brazil – Evaluating isolation as control mechanism and forecasting epidemiological scenarios of release. *Epidemiology and Infection* 2020 148: e155. doi: 10.1017/S0950268820001600.
- [10] Li RY, *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* 2020; 368(6490): 489-493. doi: 10.1126/science.abb3221.

A Steady-state analysis of the SIR model

The system of equations (1) does not reach steady state, except if $\phi = \mu + \alpha I/N$. However, the system (1) in term of fractions attains steady state [7]. Defining the fraction $x = X/N$, with $X = \{S, I, R\}$, we have

$$\frac{d}{dt} \frac{X}{N} = \frac{1}{N} \frac{d}{dt} X - \frac{X}{N} \frac{1}{N} \frac{d}{dt} N = \frac{1}{N} \frac{d}{dt} X - x(\phi - \mu - \alpha i),$$

using equation (2), and the system of equations (1) becomes

$$\begin{cases} \frac{d}{dt}s &= \phi - \beta is - \phi s + \alpha is \\ \frac{d}{dt}i &= \beta is - (\gamma + \phi + \alpha)i + \alpha i^2 \\ \frac{d}{dt}r &= \gamma i - \phi r + \alpha ir, \end{cases} \quad (\text{A.1})$$

with $s + i + r = 1$, hence, the equation for r can be decoupled from the system, through $r = 1 - s - i$. Notice that $d(s + i + r)/dt = 0$, and the system of equations in terms of fractions attain a steady state.

The system of equations (A.1), dropping out the decoupled equation for r , has two equilibrium points: The trivial (disease-free) equilibrium point $P^0 = (\bar{s} = 1, \bar{i} = 0)$ and non-trivial (epidemic) equilibrium point $P^* = (\bar{s} = s^*, \bar{i} = i^*)$.

The Jacobian matrix J evaluated at the trivial equilibrium point $P^0 = (\bar{s} = 1, \bar{i} = 0)$ is

$$J = \begin{bmatrix} -\phi & -\beta + \alpha \\ 0 & \beta - (\gamma + \phi + \alpha) \end{bmatrix},$$

with eigenvalues $\rho_1 = -\phi$ and $\rho_2 = (\gamma + \phi + \alpha)(R_0 - 1)$, where the basic reproduction number

R_0 is given by

$$R_0 = \frac{\beta}{\gamma + \phi + \alpha}. \quad (\text{A.2})$$

Hence, P^0 is locally asymptotically stable if $R_0 < 1$.

The non-trivial equilibrium point $P^* = (\bar{s} = s^*, \bar{i} = i^*)$ has the coordinates given by

$$\begin{cases} s^* = \frac{\phi}{(\beta - \alpha)i^* + \phi} \\ P_2(i) = 0, \end{cases} \quad (\text{A.3})$$

where i^* is the positive root but small than one of the second degree polynomial $P_2(i)$ given by

$$P_2(i) = \frac{\alpha}{\phi} \left[(R_0 - 1) + \frac{\gamma + \phi}{\gamma + \phi + \alpha} \right] i^2 - \left[\frac{\gamma + \phi + \alpha}{\phi} (R_0 - 1) + \frac{(\gamma + \phi)^2 + \gamma\alpha}{\phi(\gamma + \phi + \alpha)} \right] i + (R_0 - 1), \quad (\text{A.4})$$

which has the value, at $i = 1$,

$$P_2(1) = - \left[\frac{\gamma}{\phi} (R_0 - 1) + \frac{\gamma(\gamma + \phi) + \phi(\gamma + \phi - \alpha)}{\phi(\gamma + \phi + \alpha)} \right].$$

When $R_0 > 1$, we have $P_2(1) < 0$ (the condition $\gamma + \phi > \alpha$ is satisfied once $\gamma > \alpha$), and the two positive roots of $P_2(i)$ are such that $0 < i_1^* < 1 < i_2^*$. Hence the small root i_1^* is biologically feasible. When $R_0 = 1$, we have $i_1^* = 0$ and $i_2^* > 1$, hence $i_1^* = 0$ is biologically feasible. When $R_0 < 1$, we have $i_1^* < 0$ and $i_2^* > 1$, hence $i_1^* = 0$ is biologically feasible. Therefore, the small root i_1^* , which is biologically feasible, assumes a negative value for $R_0 < 1$, zero at $R_0 = 1$, and positive value but lower than 1 for $R_0 > 1$. The small root of $P_2(i)$ is given by

$$i_1^* = \frac{\left[\frac{\gamma + \phi + \alpha}{\phi} (R_0 - 1) + \frac{(\gamma + \phi)^2 + \gamma\alpha}{\phi(\gamma + \phi + \alpha)} \right] - \sqrt{\Delta}}{2 \frac{\alpha}{\phi} \left[(R_0 - 1) + \frac{\gamma + \phi}{\gamma + \phi + \alpha} \right]}, \quad (\text{A.5})$$

where Δ is

$$\Delta = \left[\frac{\gamma + \phi + \alpha}{\phi} (R_0 - 1) + \frac{(\gamma + \phi)^2 + \gamma\alpha}{\phi(\gamma + \phi + \alpha)} \right]^2 - 4 \frac{\alpha}{\phi} \left[(R_0 - 1) + \frac{\gamma + \phi}{\gamma + \phi + \alpha} \right] (R_0 - 1).$$

The complexity arises due to the non-constant population under the additional mortality

rate. Let us consider $\alpha = 0$. In this case, $P_2(i)$ has a unique positive solution

$$i^* = \frac{\phi(R_0 - 1)}{(\gamma + \phi)R_0}, \quad (\text{A.6})$$

and the fraction of susceptible individuals, from equation (A.3), is

$$s^* = \frac{1}{R_0}. \quad (\text{A.7})$$

When $\alpha > 0$, comparing equations (A.3) and (A.5), we notice that s^* has a complex dependency with R_0 , not simply $1/R_0$.

B The SEIR model

Including the exposed class E in the SIR model, we have the SEIR model

$$\begin{cases} \frac{d}{dt}S &= \phi N - \beta \frac{I}{N}S - \mu S \\ \frac{d}{dt}E &= \beta \frac{I}{N}S - (\sigma + \mu)E \\ \frac{d}{dt}I &= \sigma E - (\gamma + \mu + \alpha)I \\ \frac{d}{dt}R &= \gamma I - \mu R, \end{cases} \quad (\text{B.1})$$

where σ is the incubation rate. As the previous analysis for the SIR model, we can obtain the basic reproduction number R_0 , which is given by

$$R_0 = \frac{\sigma}{\sigma + \mu} \times \frac{\beta}{\gamma + \mu + \alpha}, \quad (\text{B.2})$$

when $\phi = \mu$, and the accumulated cases Ω obeys

$$\frac{d}{dt}\Omega = \sigma E, \quad \text{with} \quad \Omega(0) = 0.$$

To estimate R_0 for the elaborated model in the main text, we used as the initial conditions $E(0) = 50$ and $I(0) = 1$ among others. For SEIR model, let us consider as the initial conditions $E(0) = 50$ and $I(0) = 1$, but $S(0)$ and $R(0)$ are the same supplied to the SIR model. We let $\sigma = 1/5.8 \text{ days}^{-1}$ (elaborated model), and the values for other parameters are those provided in the SIR model. The value $\gamma = 1/10 \text{ days}^{-1}$ in somehow considered an average value among γ , γ_1 , γ_2 and γ_3 in the elaborated model. The estimated basic reproduction number for São Paulo State was $R_0 = 3.92$ with $Sum = 1.29 \times 10^6$, and for Spain, $R_0 = 4.41$ with $Sum = 2.24 \times 10^8$.

Notice that $\sigma/(\sigma + \mu) = 0.99999$, hence, R_0 given by equations (A.2) and (B.2) must be equal if β is the same. However, β is estimated by the dynamical systems (1) or (B.1) against the accumulated severe covid-19 cases. For this reason, the estimated β must be different between SIR and SEIR models. Let us consider SEIR model considering the initial conditions $S(0) = N_0$, $E(0) = 0$, $R(0) = 0$, and varying $I(0)$.

For the data collected from São Paulo State, with $S(0) = 44.6$ million, we obtained for $I(0) = 1$, $R_0 = 7.84$ with $Sum = 1.11 \times 10^6$, for $I(0) = 10$, $R_0 = 5.1$ with $Sum = 4.52 \times 10^5$, for $I(0) = 25$, $R_0 = 4.02$ with $Sum = 1.21 \times 10^6$, and for $I(0) = 100$, $R_0 = 2.74$ with $Sum = 3.51 \times 10^6$. The lowest Sum occurs when $R_0 = 5.1$.

For the data collected from Spain, with $S(0) = 47.4$ million, we obtained for $I(0) = 1$, $R_0 = 6.86$ with $Sum = 7.01 \times 10^8$, for $I(0) = 10$, $R_0 = 5.1$ with $Sum = 7.79 \times 10^8$, for $I(0) = 25$, $R_0 = 4.46$ with $Sum = 9.54 \times 10^8$, and for $I(0) = 100$, $R_0 = 3.58$ with $Sum = 1.33 \times 10^9$. The lowest Sum occurs when $R_0 = 6.86$.

Comparing with the estimated R_0 provided by the SIR model (see the main text), we observe that SEIR model estimated with higher value. Let us assess the role played by $E(0)$. We choose data from São Paulo State, letting $I(0) = 10$, $S(0) = 44.6$ million, $R(0) = 0$, and varying $E(0)$ (for the SIR model, $R_0 = 2.4$). for $E(0) = 1$, $R_0 = 5.03$ with $Sum = 4.57 \times 10^5$, for $E(0) = 10$, $R_0 = 4.61$ with $Sum = 6.19 \times 10^5$, for $E(0) = 25$, $R_0 = 4.13$ with $Sum = 9.86 \times 10^5$, and for $E(0) = 100$, $R_0 = 3.04$ with $Sum = 3.31 \times 10^6$. The lowest Sum occurs when $R_0 = 5.03$, however, as $E(0)$ increases, the estimated R_0 approaches to that provided by the SIR model.