



OPEN

Addressing hospitalisations with non-error-free data by generalised SEIR modelling of COVID-19 pandemic

Jorge M. Mendes✉ & Pedro S. Coelho

Successive generalisations of the basic SEIR model have been proposed to accommodate the different needs of the organisations handling the SARS-CoV-2 epidemic. These generalisations have not been able until today to represent the potential of the epidemic to overwhelm hospital capacity until today. This work builds on previous generalisations, including a new compartment for hospital occupancy that allows accounting for the infected patients that need specialised medical attention. Consequently, a deeper understanding of the hospitalisations rate and probability as well as of the recovery rates for hospitalised and non-hospitalised individuals is achieved, offering new information and predictions of crucial importance for the planning of the health systems and global epidemic response. Additionally, a new methodology to calibrate epidemic flows between compartments is proposed. We conclude that the two-step calibration procedure is able to recalibrate non-error-free data and showed crucial to reconstruct the series in a specific situation characterised by significant errors over the official recovery cases. The performed modelling also allowed us to understand how effective the several interventions (lockdown or other mobility restriction measures) were, offering insight for helping public authorities to set the timing and intensity of the measures in order to avoid the implosion of the health systems.

In December 2019, a new coronavirus named severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2), causing severe acute respiratory disease emerged in the region of Wuhan, China^{1,2}. SARS-CoV-2 is an acute respiratory infectious disease that spreads through the respiratory tract by droplets, respiratory secretions, and person-to-person contact^{3,4}.

At the time of the writing of this introduction, a “second wave” of new COVID-19 infections is striking some European countries, while in South America and North America a first wave is spreading at an alarming rate. Its transmissibility is high (reproductive number, R_0 , estimated to be between 2 and 3^{5–8}), indicating that a large proportion of the world population can be infected. As a significant number of patients have severe symptoms and need dedicated medical care (around one in every five people who are infected), the potential to overwhelm our health care systems and intensive care units is a real threat⁹. In Portugal, after an incidence spike in early April (cf. Fig. 1a), the lockdown measures had been easing in early May, resulting in an increase of the daily incidence of new cases (cf. Fig. 1a). Although the Portuguese national health system limits (dashed horizontal lines in Fig. 1b,c) have never been reached, the current course of the epidemic raises critical management issues, in particular, because the more significant growth of the second wave is threatening to overcome hospitalisation capacity in the short term.

Much work has been done so far on the COVID-19 outbreak course using the so-called compartmental models^{10–12}. The origin of compartmental models dates back to the early 20th century with the seminal work of Kermack and McKendrick (in 1927)¹³. Compartmental models simplify the mathematical modelling of infectious diseases. The population is assigned to compartments and may progress between compartments. Compartmental order usually shows the flow patterns between the compartments; for example, SIS means susceptible, infectious, and then susceptible again¹⁴. The models are most often run with ordinary differential equations (which are deterministic), but can also be used with a stochastic framework, as it is the case here¹⁴.

Compartmental models allow predicting how a disease spreads out through the total number infected, the duration of an epidemic, and infection reproductive number (R_0), among other important epidemic course

NOVAIMS, Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisbon, Portugal. ✉email: jmm@novaims.unl.pt

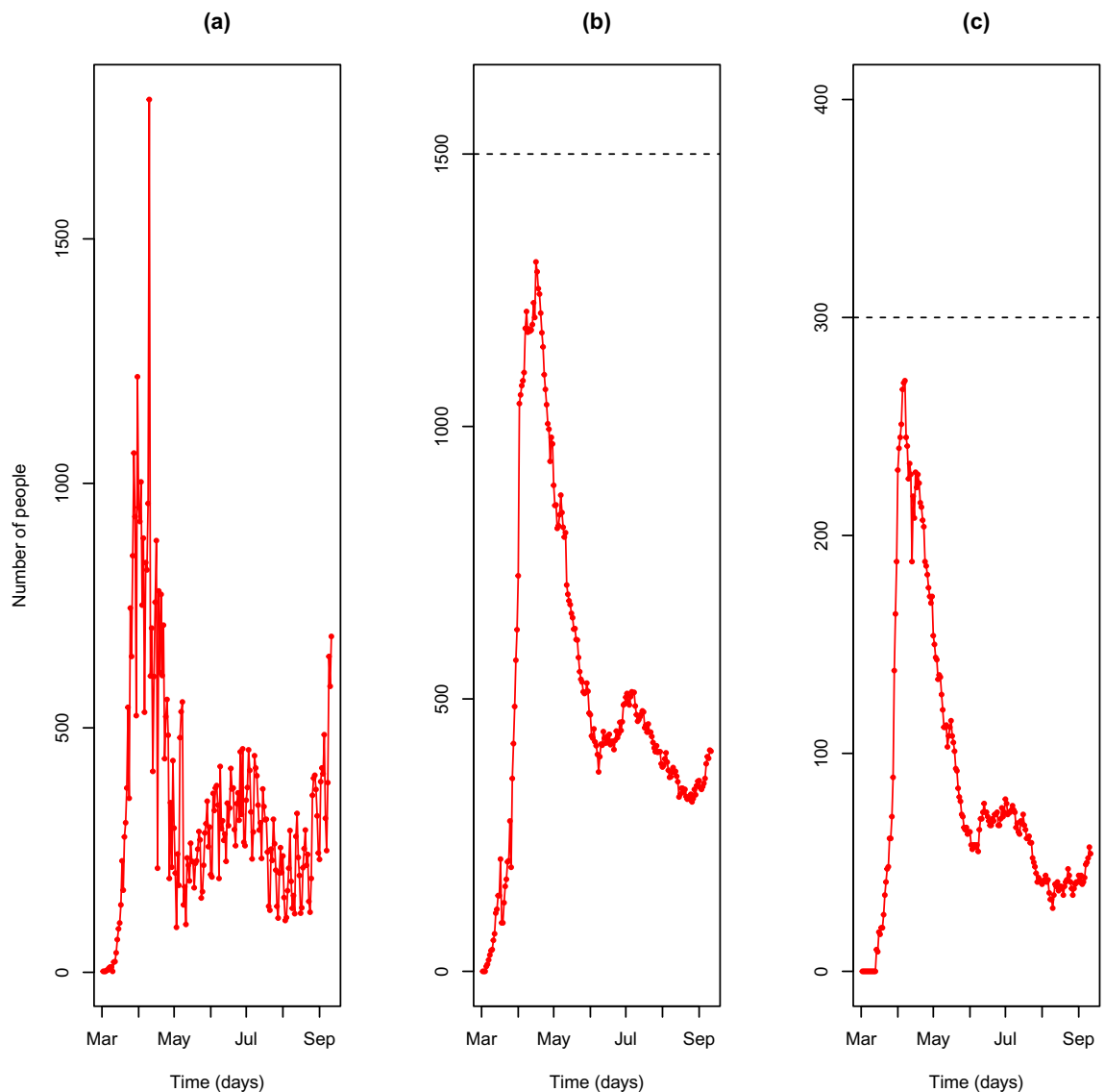


Figure 1. Course of the COVID-19 epidemic in Portugal. (a) Daily incidence of COVID-19 cases in Portugal, (b) Daily hospital occupancy due to COVID-19 in Portugal, (c) Daily ICU occupancy due to COVID-19 in Portugal. Data collected from the daily bulletins issued by *Direção-Geral de Saúde* (Directorate-General of Health, DGS), the Portuguese health authority. The dashed horizontal lines show the generally accepted national health system limits.

parameters. Moreover, such models can show how different public health interventions (e.g. vaccination, limited social contacts, lockdown) may affect the outcome of the epidemic.

The basic compartmental model is the SIR model, and other more complex variations are derivatives of the basic SIR model¹³. The SIR model comprises three compartments: *S* for the number of susceptible, *I* for the number of infectious, and *R* for the number of removed (recovered, deceased, or immune) individuals at a particular time. This model assumes incidence grows exponentially, which is not in agreement with the observed epidemic course, as the measures adopted by public health authorities at different moments, as well as the change of human behaviours tend to flatten the parabolic incidence curve. Therefore, its predictive value for infectious diseases that are transmitted from humans to humans, without any change on the constant transmission rate is of limited usefulness.

To represent that the number of susceptible, infectious and removed individuals may vary over time (even if the total population size remains constant) a time index, *t*, might be added: S_t , I_t and R_t . For a specific disease within a particular population, these functions may be worked out to predict possible outbreaks and bring them under control.

For infections with the characteristics of SARS-CoV-2, there is a significant incubation period during which individuals have been infected but are not yet infectious themselves^{9,15–19}. During this period the individual is in compartment *E* (for exposed), resulting in the so-called SEIR model. The SEIR model's basic version is still short for our current needs, especially regarding the burden put in the health system. Indeed, the basic version of the SEIR model is unable to distinguish between deceased and recovered people and provide any helpful



Figure 2. Basic SEIR model.

information on the daily burden put on a health system by individuals in need of medical care. Therefore successive generalisations of the basic SEIR model have been proposed to accommodate either the different needs of the different organisations handling epidemic-related problems or the need to account for public health data not specifically used in the basic SEIR model. This work seeks to pursue these needs and extend such generalisations. For public health officers, the number of confirmed cases is the part of the iceberg that is visible as many more people are infected but still not infectious because the virus is still in the replication phase (incubation period). Indeed, these people would come up as confirmed cases a few days ahead. Knowing in advance how the number of confirmed cases might evolve is of crucial importance for planning reasons.

Moreover, it is of unquestionable value for health system officers to know the share of confirmed COVID-19 confirmed cases who are going to need for specialised medical care and how many of them will recover or die. Additionally, for researchers with no access to micro-data files (due to unavailability or lack of data quality) it is impossible to know some of the flows between some compartments of the SEIR model's generalisations. This work also addresses this problem, establishing a methodology to calibrate epidemic flows between compartments, that is relevant to recalibrate non-error-free data. To accomplish these goals, we build on the work of²⁰ and decided to propose several extensions and restrictions, including the consideration of a new compartment representing hospital occupancy.

This work is organised as follows: “[The SEIQRHD model](#)” section describes the classic SEIR model summarily and its proposed generalisation to meet our aims. It also describes our proposals for accommodating some of the characteristics already mentioned as well as others suggested by the empirical analysis of the Portuguese epidemic course; “[Data and empirical analysis](#)” section describes the used data and the major data processing tasks prior to modelling using a Bayesian framework and, finally, “[Bayesian hierarchical modelling](#)” section presents the hierarchical Bayesian framework used for parameters estimation. “[Results](#)” and “[Discussion](#)” sections respectively present and discuss, respectively, the achieved results for the case of Portuguese epidemic course and the main implications for managing and controlling the course of the outbreak.

Data and model

The SEIQRHD model. Assuming the vital dynamics (birth, deaths and migration) may be neglected (which is not a hard assumption given the short period of the pandemic) the basic SEIR model is defined as in Fig. 2.

The boxes represent the compartments (or states) and arrows represent transitions from one compartment to another. S_t is the number of people susceptible at time t , E_t is the number of people exposed to the infection at time t (people that become infected but not infective), I_t is the number of infective people infected at time t and R_t : number of people removed (recovered or deceased) at time t . The parameter β denotes the transmission rate (the expected amount of people an infected person infects per day and is the result of the contact rate—the number of people an average person enters into contact with each day—and the probability that a contact provokes the transmission of the disease). It is multiplied by the ratio S/N to avoid counting contacts between two people who cannot infect each other (e.g., because one of them has already recovered and immune, or because both are infective²¹). The parameter δ is the disease incubation rate (the inverse of the incubation period in days), and γ is the removal rate (the inverse of the removal period in days). When some of these three parameters are considered fixed, they do not, in general, fit the observed course of the current epidemic. For example, the social distancing measures or even the lockdown damp the transmission rate down by reducing the daily contacts one person regularly has; or the improvement and gained experience of the health care system pull the removal rate up by lowering the recovery period or the death rate. Making these two parameters, β_t and γ_t , functions of time, with user-defined functional forms, is the solution to describe the course of the disease spread appropriately. As already mentioned, for epidemics that last for a short period (some months, for instance) it does not worth to account for demographics dynamics (births, migration, deaths by other causes) as the population can be considered fixed. Nonetheless, for the cases whenever the demographics may impact other transitions might also be considered also to account for those dynamics.

Based on the SEIR model, many measures of the epidemic development measures might be assessed^{15,21–24}. We refer to²⁵ as an excellent review of the recent work done in this area. The SEIR model was also used to compare the effects of the Hubei province's lockdown on the transmission dynamics in Wuhan and Beijing¹⁶.

As the SEIR model can generate interpretable results, wave derivatives are being developed²⁵ (a) to simulate the processes of transmission from infection source; (b) to assess the transmission risk and prediction of patient number based into two subpopulations of infected people, those the quarantined and unquarantined; (c) to simulate the incubation period and the period before recovery.

Many of these works consider different generalisations of the basic SEIR model as those four compartments are often insufficient to describe a much more complex reality. First, considering only one compartment to account for removed people, R , is not enough to describe the evolution of patients that recover from the infection and the share that deceases. Secondly, the compartment I does not match the observed amount of the infected people in many situations as many infected people are asymptomatic or only experience mild symptoms (and they are not accounted for by the figures released out by the health authorities) or it does not account for the

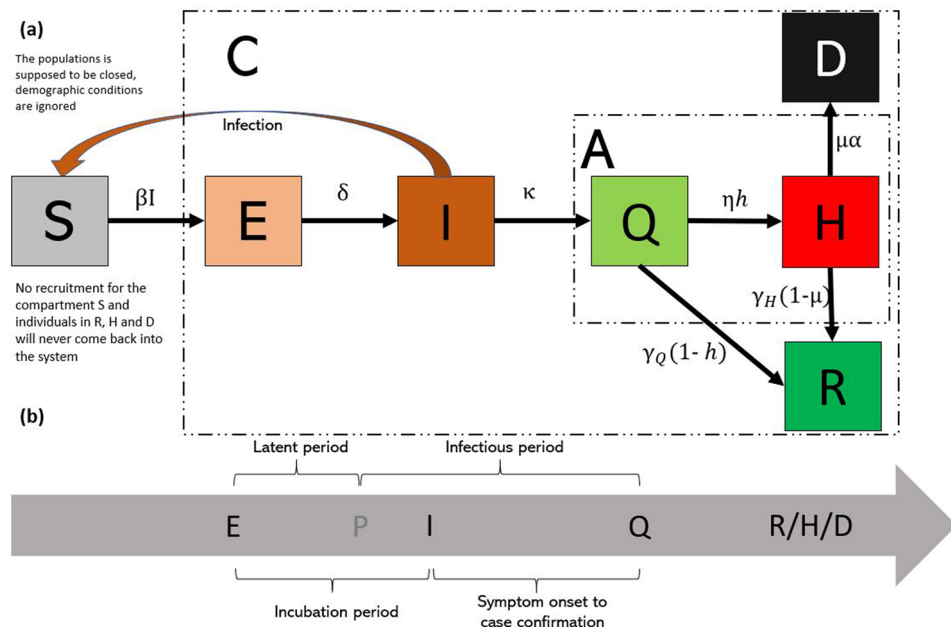


Figure 3. The SEIQRHD model used in this work. (a) SEIQRHD model. (b) Schematic representation of temporal evolution between compartments E, I and Q.

isolation (quarantine) usually imposed by health authorities making these people no longer infectious. Indeed, the daily number of “confirmed infections” corresponds to people who are isolated (quarantined) to avoid new infection chains. As a solution, many authors^{20,25} have been proposing generalisations of this basic SEIR model by splitting the compartment, R into two compartments, one to account for the infected people that genuinely recover from the disease, R, and a second one to account for the mortality induced by the infection, D. Moreover, to precisely match the observed number of confirmed cases, one additional compartment, Q, has been considered immediately after the I compartment²⁰. Building on the work of²⁵ we propose an additional compartment to account for hospital occupancy, H. Figure 3a reproduces the generalisation of the SEIR model used in this work. Figure 3b describes schematically the temporal evolution between compartments E, I and Q.

Figure 3a contains two additional compartments to keep track of the cumulative number of COVID-19 infections, C, and the current number of ill individuals, the so-called “active cases”, A. This models includes other not yet described parameters. Further ahead, we elaborate on the potential functional forms to describe their temporal evolution and use the opportunity to define them.

The following differential equations describe this is generalisation of the SEIR model (SEIQRHD):

$$\begin{aligned}
 \partial S_t / \partial t &= -\beta S_t \frac{I_t}{N} \\
 \partial E_t / \partial t &= \beta S_t \frac{I_t}{N} - \delta E_t \\
 \partial I_t / \partial t &= \delta E_t - \kappa I_t \\
 \partial Q_t / \partial t &= \kappa I_t - \eta h Q_t - \gamma_Q (1-h) Q_t \\
 \partial H_t / \partial t &= \eta h Q_t - \gamma_H (1-\mu_t) H_t - \mu \alpha H_t \\
 \partial R_t / \partial t &= \gamma_Q (1-h) Q_t + \gamma_H (1-\mu) H_t \\
 \partial D_t / \partial t &= \mu \alpha H_t \\
 \partial C_t / \partial t &= \delta E_t \\
 \partial A_t / \partial t &= \kappa I_t - \gamma_Q (1-h) Q_t - \gamma_H (1-\mu) H_t - \mu \alpha H_t \\
 N &= S_t + E_t + I_t + Q_t + R_t + H_t + D_t
 \end{aligned} \tag{1}$$

The discussion on whether the parameters are fixed is done further ahead; therefore, the subscript t in the parameters is omitted for the time being. Some of the model coefficients might be considered fixed (δ, κ, α), but others (β, η, h, γ and μ) vary in time according to some specific functional form.

The parameter δ is the inverse of the average incubation period and governs the lag between having undergone an infectious contact and showing symptoms. It is generally considered fixed because it depends mainly on the the infectious agent’s characteristics.

The parameter κ is the inverse of the average period taken to isolate a symptomatic person, which usually happens after the person has been tested positive. Therefore, the quarantined compartment matches the “active confirmed cases” in many official data sources in most of the developed countries and the quarantined individuals are excluded from the infectious compartment (I) because they are supposed to be isolated. The time required to isolate a person with symptoms depends mainly on the protocol followed by the health authorities

which with some minor adjustments is constant. As explained further ahead, we consider this parameter fixed after an initial period of adaptation, and use an empirical estimate, jointly with a literature-based value of δ to estimate the flows between compartment S and E and E and I . Within the hierarchical Bayesian model they are even considered fixed.

As the lockdown and social distancing measures impact on the number of individuals a person comes in contact with on a daily basis, the parameter β should not be considered constant.

The parameters γ_Q and γ_H are the recovery rate for hospitalised and quarantined people. They correspond to the inverse of the average time required for an active case (quarantined/hospitalised) to be classified as recovered. They provide precious information about how fast the people may recover from the disease (in days, in general). We believe that both γ_Q and γ_H are related to how a health system can improve its capability to treat people over time (e.g., with the introduction of a new therapeutic strategies). However, γ_Q is much more dependent on recovery confirmation protocols and therefore, much more susceptible to increase over time.

The outer flows from compartment Q depend on γ_Q , already discussed, η and h . The rate at which quarantined individuals are hospitalised, η (η^{-1} is the average period spent in compartment Q), mainly depends on the illness severity and health system capacity to accept ill individuals, whilst the share of people in need of specialised medical care, h , mainly depends on the population health conditions, as well as the mix of the quarantine compartments regarding demographic characteristics as gender and age. Whilst the former might be considered fixed to some extent, the latter, due to the mentioned reasons cannot be. Indeed, recent empirical analysis shows that after the epidemic spike in late April 2020, the share of COVID-19 patients admitted to a hospital decreased from almost 100%, in the early days of the epidemic, to a tiny constant value, in late July. This phenomenon might be explained by the uncertainty and lack of knowledge of the disease course in the early days of the epidemic. However, as time goes by, the medical practitioners learn and become more selective about the patients to be admitted to a hospital. Moreover, as the prevalence curve increases, the health systems tend to avoid the intolerable burden and become more demanding on the hospitalisation requirements. Therefore, there is enough evidence to consider η a constant parameter and h_t a time-varying parameter.

The parameter μ is the fatality rate and provides information on the proportion of hospitalised individuals who, unfortunately, die. It depends either on the resilience of the patients and the severity of the disease. As this variation is not time-related we consider it to be fixed. It is worth mentioning here that any transition from the quarantined Q to the death D compartments is not considered at all, as any ill and quarantined individual that gets worse is, at some point in time, admitted to a hospital and eventually recovers or dies. The parameter α denotes the transition rate between H and D . Its inverse corresponds to the time a hospitalised individual takes to die. We anticipate these parameters to be highly variable as, in the ultimate analysis, it depends on the individuals themselves, more than the quality of the health care provided, but we do not see any reason to consider it a time-varying parameter.

Many infected people experience mild or no symptoms at all and are not accounted for (ascertained). Therefore, the Infection Fatality Rate (IFR), i.e. the number of deaths as a proportion of all persons infected with the SARS-CoV-2 novel coronavirus is expected to be much lower than the Case Fatality Rate (CFR), the proportion among confirmed cases. Some works have reported values for the ascertainment rate between 0.4 and 14% in Wuhan^{7,26–30}, 28.4% in Italy¹⁶ and just 0.23% in Iran³¹. Unfortunately, the available data does not allow the estimation of this ascertained rate²⁴, that account for an increased number of people in each compartment than the actual figures.

For modelling and empirical data analysis performed hereafter one uses a discrete-time approximation to the stochastic continuous-time SEIQRHD model defined in Eq. (1). Consider a time interval $(t, t + h)$, where h represents the length between the time points at which measurements are taken, here $h = 1$ day. Let dSE_t denote the daily number of susceptible individuals who become infected, dSI_t the number of cases by date of symptom onset, dIQ_t the daily number of confirmed cases, dQR_t the daily number of quarantined cases who recover, dHR_t the daily number of hospitalised cases who recover, dQH_t the daily number of quarantined cases admitted to a hospital for more specialised care and dHD_t the daily number of cases who decease. Given initial conditions $S_0 = s_0$, $E_0 = e_0$, $I_0 = i_0$, $Q_t = q_0$, $R_0 = 0$, $H_t = 0$ and $D_0 = 0$, and the population size N , the discretised stochastic SEIQRHD model is specified by:

$$\begin{aligned}
S_{t+1} &= S_t - dSE_t \\
E_{t+1} &= E_t + dSE_t - dEI_t \\
I_{t+1} &= I_t + dEI_t - dIQ_t \\
Q_{t+1} &= Q_t + dIQ_t - dQH_t - dQR_t \\
H_{t+1} &= H_t + dQH_t - dHR_t - dHD_t \\
R_{t+1} &= R_t + dQR_t + dHR_t \\
D_{t+1} &= D_t + dHD_t \\
C_{t+1} &= C_t + dSE_t \\
A_{t+1} &= A_t + dIQ_t - dHD_t - dHR_t - dQR_t \\
dSE_t &= \beta S_t \frac{I_t}{N} \\
dEI_t &= \delta E_t \\
dIQ_t &= \kappa I_t \\
dQH_t &= \eta h_t Q_t \\
dQR_t &= \gamma_Q (1 - h_t) Q_t \\
dHR_t &= \gamma_H (1 - \mu) H_t \\
dHD_t &= \mu \alpha H_t \\
N &= S_t + E_t + I_t + Q_t + R_t + H_t + D_t,
\end{aligned} \tag{2}$$

This set of equations, jointly with the initial conditions, define the SEIQRHD model used in this work.

Data and empirical analysis. We collected the epidemic data from the *Direção-Geral de Saúde* (Directorate-General of Health, DGS), the Portuguese health authority. Using the SEIQRHD model described in the previous section we propose to disclose either the course of the Portuguese SARS-CoV-2 epidemic curves at the country level on a daily basis as long as the analysis of the information provided by model parameters. The results of the work based on a simplified version of the model can be found on <https://insights.cotec.pt/> on the mosaic *Modelos epidemiológicos*. Data collection and analysis has taken into consideration guidelines of good practices (e.g.³²).

Since the beginning of the pandemic, DGS releases daily data on its course (<https://covid19.min-saude.pt/ponto-de-situacao-atual-em-portugal/>). The key released variables are number of confirmed cases, number of deaths (nationwide and by health region), number of recovered cases, hospital occupancy, intensive care unit occupancy (nationwide), as well as the characterisation of the positive cases by origin, gender and age group. The figures released by DGS are used for modelling purposes as explained below. By September 11th, 2020, the course of the epidemic is as Figs. 1 and 4 describe.

The epidemic started on March 2nd, 2020 with the first two cases. Some unexpected spikes of daily recovered cases can be observed between May 14th and 24th 2020. This information corresponds to the release of recovered cases than have recovering during the previous weeks but have not been accounted for before as the recover notifications issued by the primary health care system were not included in the released figures. Moreover, it is globally clear that the released data about recoveries does not reflect the actual daily number of cases as some of these cases tend to be accumulated and only released later on. Therefore, we decided to proceed to a calibration on the number of recovered cases. This calibration impacts the hospital occupancy (H_t) and quarantined individuals (Q_t) and the two flows of recovered cases coming into the compartment R_t . Only the series of daily new quarantined (dIQ_t), and the daily states of the active ($A_t = Q_t + H_t$), recovered (R_t), hospitalised (H_t), and death (D_t) cases are released by DGS, but the fact that the number of recoveries is released with some delay, also leads to inaccuracies over the stock of quarantined cases and consequently on the active cases. To account for the mentioned dependencies we decided to calibrate the unknown flows (dSE_t , dEI_t , dQH_t , dQR_t , dHR_t), using the available data: (a) the daily number of new confirmed cases, dIQ_t , (b) the daily number of new deaths, dHD_t , and its daily stock D_t , (c) the daily number of new recovered cases, $dHR_t + dQR_t$ and is stock R_t (d) the daily number of hospitalisations, H_t and (e) the daily number of quarantined cases $Q_t = A_t - H_t$.

The calibration methodology is a two-step procedure. In the first step we calibrate the series of exposed (dSE_t) and infected (dEI_t) based on the incubation period of COVID-19 and the time from symptom onset to laboratory confirmation (based on the analysis of COVID-19 micro-data file provided by DGS). Table 1 summarises the distribution parameters used to calibrate the daily incidences of dSE_t and dEI_t as described in Figs. 5 and 6. The median incubation period of SARS-Cov-2 has been estimated at approximately 5–6 days, with a range between 1 and 21 days^{9,15–18,33–35}. We used the value provided by³⁵ for the gamma distribution. We assume that the mean time from symptom onset to diagnosis confirmation in Portugal decreased from 5.34 days (with a standard deviation of 4.8 days) in the early period of the epidemic (early March) to 4.74 days (standard deviation of 6.5 days) in early April. These values are based on empirical analysis of microdata file provided by DGS. The file provides information for positive cases on the dates of symptom onset and test results. The latter is used to start counting the isolation period. The time necessary to isolate an individual has decreased steeply in the early days of the epidemic to a steady value, achieved one April 2nd, around 1 month after the epidemic started. To reflect the above, we modelled the time between case reporting and symptom onset with a Gamma distribution with μ_X/σ_X parameter of 5.34/4.80 on March 2, decreasing linearly to 4.72/6.45 on April 2 and after (Fig. 5a). Similarly the time between case exposure (infection with SARS-CoV-2) and diagnosis confirmation (based on empirical analysis DGS provided microdata) was modelled with a Gamma distribution with μ_X/σ_X parameter of 11.84/7.4 on March 2, decreasing linearly to 11.23/9.1 on April 2 and after (Fig. 6a).

The calibrated series dEI_t and dSE_t are shown in Figs. 5b and 6b.

In the second step, as the available data provide no information on the daily flows from the Q compartment to the H and R compartments (dQH_t , dQR_t), and from the H compartment to the R compartment, (dHR_t), we

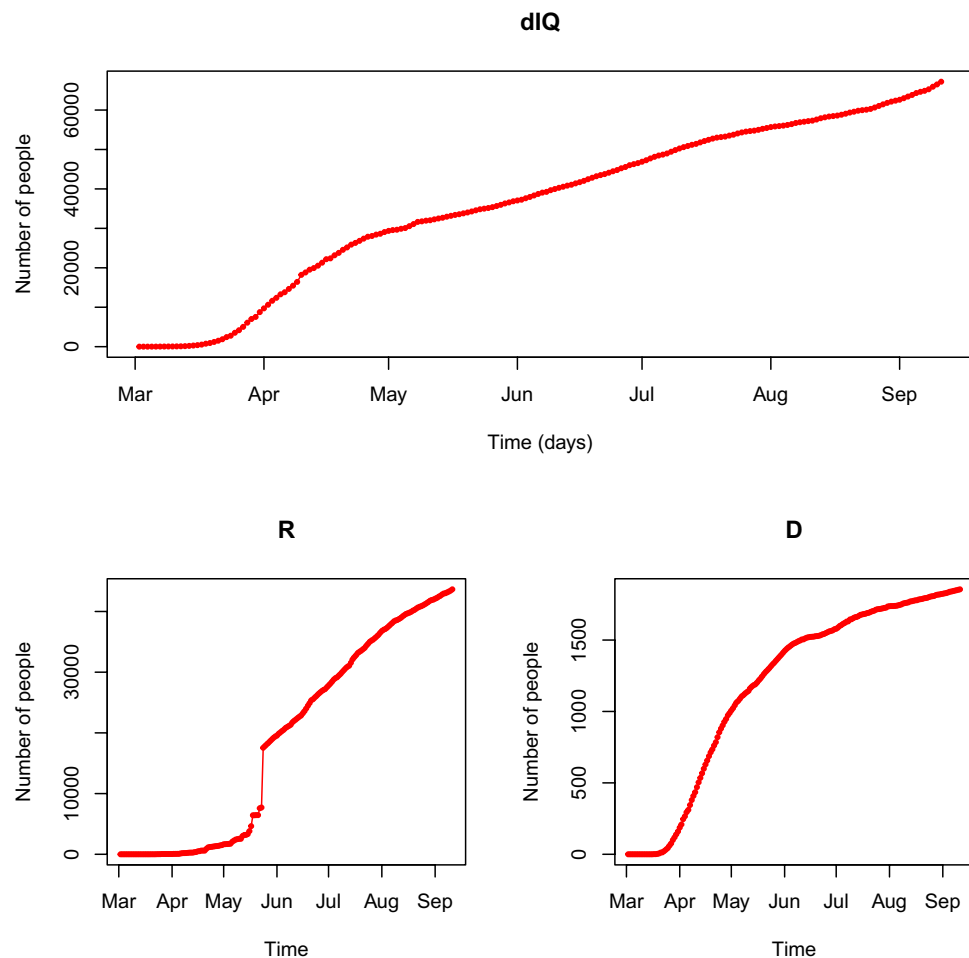


Figure 4. Course of the COVID-19 epidemic in Portugal. (a) Confirmed infected cases (cumulative). (b) Recovered cases (cumulative). (c) Deaths (cumulative).

Estimate	Mean	Std deviation
Incubation period (in days) of SARS-Cov-2	6.50	2.60
Time (in days) from symptom onset to case confirmation (prior April, 2nd)	5.34	4.80
Time (in days) from symptom onset to case confirmation (post April, 2nd)	4.73	6.45

Table 1. Mean time and corresponding standard deviation for calibration of distribution of daily incidences of non-observed series.

decided to use the discrete-time approximation to the stochastic continuous-time SEIQRHD model defined in Eq. (1) with the following assumptions:

1. $dQR_t = \gamma_Q Q_{t-1}, t \geq 12$
2. $dHR_t = \frac{1-\omega}{1-\omega} dQR_t, t \geq 12$
3. $dQH_t = (H_t - H_{t-1} + dHD_t + dHR_t) I_{(H_t - H_{t-1} + dHD_t + dHR_t \geq 0)}, t \geq 13$;

where $I_{(H_t - H_{t-1} + dHD_t + dHR_t \geq 0)}$ is an indicator function taking the value 1, if the argument $H_t - H_{t-1} + dHD_t + dHR_t \geq 0$ is true, and zero otherwise, $\gamma_Q > 0$ and $0 < \omega < 1$.

The first assumption originates from the fact, already mentioned, that the recovery rate is roughly constant for small time intervals. The second assumption originates from the empirical evidence that about 80% of COVID-19 patients recover from the disease without needing special treatment, and for the majority—especially for children and young adults—illness due to COVID-19 is generally minor. However, for some people it can cause serious illness (20%)—difficulty in breathing requiring hospital care—which is particularly true for people who are aged

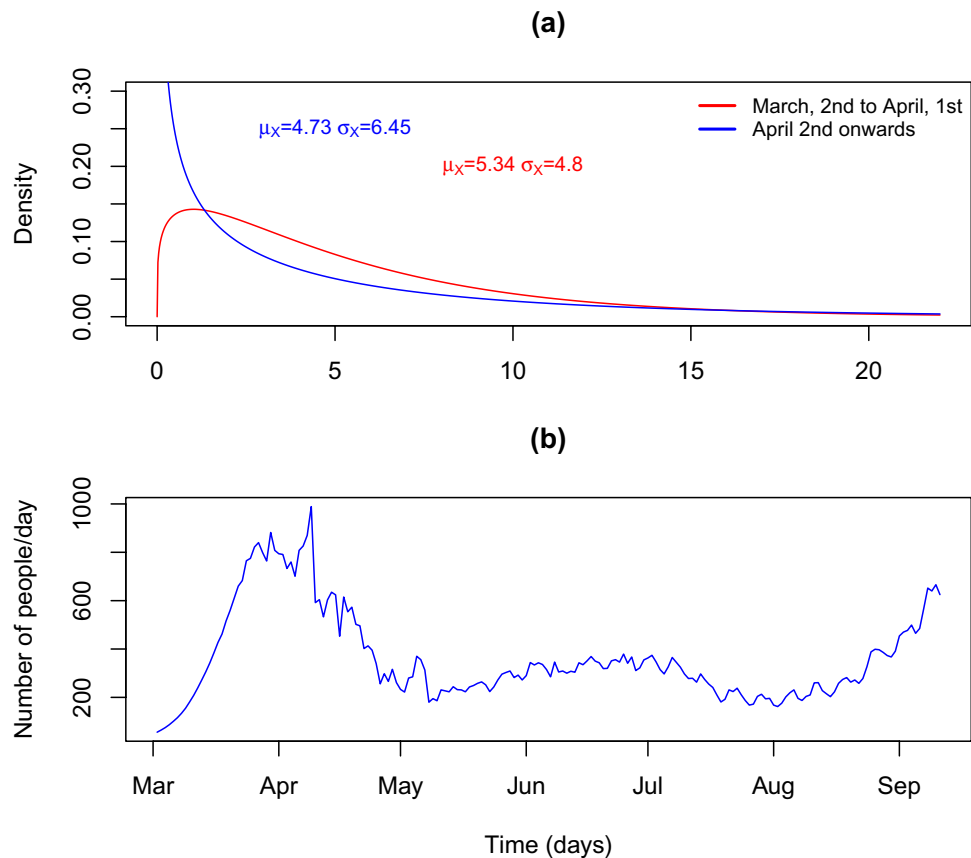


Figure 5. Distribution of time from symptom onset to case confirmation and reporting and the resulting curve of the estimated curve of daily infective people. **(a)** Distribution of time from symptom onset to case confirmation and reporting. **(b)** Estimated infectious (dEI_t) cases in Portugal.

over 60 years and people who have underlying medical conditions such as diabetes, heart disease, respiratory disease or hypertension. Therefore, assuming that $0 < \omega < 1$ of the total outgoing flows from Q_t go to H_t we have:

$$\begin{aligned} dHR_t &= \omega dR_t, \quad 0 < \omega < 1 \Leftrightarrow \\ dHR_t &= \omega (dHR_t + dQR_t) \Leftrightarrow \\ dHR_t &= \frac{\omega}{1 - \omega} dQR_t, \end{aligned} \quad (3)$$

The third assumption relies on the fact that, having calibrated the values of dHR_t and using the observed values of H_t , dQH is calibrated by difference, provided that results in a non-negative flow.

The flows described in the assumptions are only used for $t \geq 13$, and $t \geq 12$, because the number of hospitalised individuals, H_t , is equal to the observed cumulative number of quarantined cases in the first 12 days of the epidemic, $h_t = \sum_{i=1}^t dIQ_i$, $t = 1, \dots, 12$ and we assume no recoveries in the first 11 days of the epidemic.

Day 84th of the epidemic corresponds to May 24th, 2020, the time point where 9844 recovered cases were released regarding recoveries that happened prior to May 24th, 2020. To calibrate the aforementioned flows the following algorithm is implemented as:

1. For the period $t = 1$ to $t = 84$ (May 24th 2020):

- Initialise $\hat{S}_1 = s_1$, $\hat{E}_1 = e_1$, $\hat{I}_1 = i_1$, $\hat{Q}_1 = q_1$, $\hat{H}_1 = 0$, $\hat{R}_1 = 0$, $\hat{D}_1 = 0$.
- for $t = 2$ to T

1. $\hat{S}_t = \hat{S}_{t-1} - \widehat{dSE}_t$
2. $\hat{E}_t = \hat{E}_{t-1} + \widehat{dSE}_t - \widehat{dEI}_t$
3. $\hat{I}_t = \hat{I}_{t-1} + \widehat{dEI}_t - \widehat{dIQ}_t$
4. $\widehat{dQR}_t = \text{round}(\gamma_{0Q}(1 - \exp(-\rho_{yQ}t))\hat{Q}_{t-1})I_{(t \geq 13)} + 0$
5. $\widehat{dHR}_t = \text{round}(\omega \exp(-\rho_{\omega}t)dQR_t)I_{(t \geq 12)} + 0$
6. $\widehat{dHD}_t = dD_t$
7. $\widehat{dQH}_t^{(1)} = (H_t - H_{t-1} + \widehat{dHD}_t + \widehat{dHR}_t)I_{(H_t - H_{t-1} + \widehat{dHD}_t + \widehat{dHR}_t \geq 0)}I_{(t \geq 13)} + 0$
8. $\hat{Q}_t = \hat{Q}_{t-1} + \widehat{dIQ}_t - \widehat{dQH}_t - \widehat{dQR}_t$

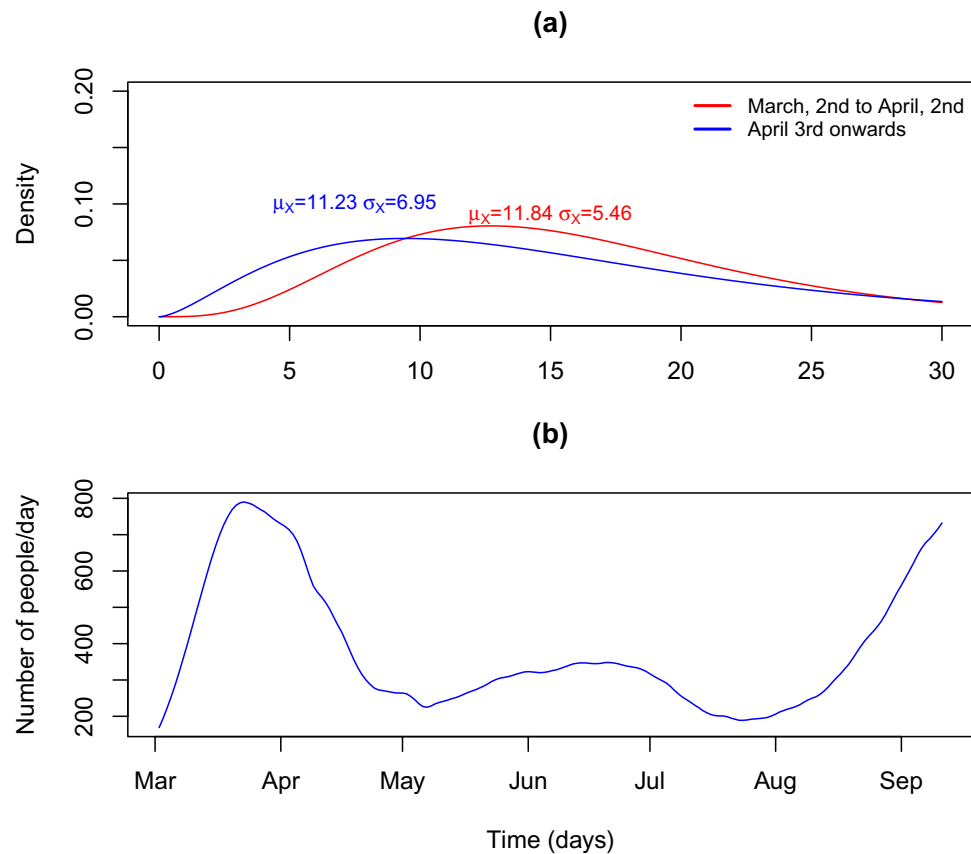


Figure 6. Distribution of time from infection to case confirmation and reporting and resulting curve of estimated curve of daily infective people. (a) Distribution of time from infection to case confirmation and reporting. (b) Estimated infections (dSE_t) in Portugal.

$$\begin{aligned}
 9. \quad \hat{H}_t &= \hat{H}_{t-1} + \widehat{dQH}_t - \widehat{dHR}_t - \widehat{dHD}_t \\
 10. \quad \widehat{dQH}_t^{(2)} &= (\widehat{dQH}_t - (\hat{H}_t - H_t)I_{(\hat{H}_t - H_t > 0)})I_{(\widehat{dQH}_t - (\hat{H}_t - H_t)I_{(\hat{H}_t - H_t > 0)}) > 0)}I_{(t \geq 13)} + 0 \\
 11. \quad \hat{R}_t &= \hat{R}_{t-1} + \widehat{dQR}_t + \widehat{dHR}_t \\
 12. \quad \hat{D}_t &= \hat{D}_{t-1} + \widehat{dHD}_t \\
 13. \quad \hat{C}_t &= \hat{C}_{t-1} + dSE_t \\
 14. \quad \hat{A}_t &= \hat{A}_{t-1} + \widehat{dIQ}_t - \widehat{dHD}_t - \widehat{dHR}_t - \widehat{dQR}_t
 \end{aligned}$$

2. For the remaining period ($t = 85$ to $t = T$) we run the previous algorithm for every disjoint time frame of 7 consecutive days, initialising the values of each compartment with the final values of the previous period, allowing for different estimates of γ_Q and ω across time.

The “hat” in the above notation denotes calibrated figures (either on the first or the second procedure steps) whilst the figure with no “hat” are currently observed. The calibrated value \widehat{dQH}_t (7th step) is submitted to a second verification in the 10th step of the algorithm to avoid flows into compartment H that are higher than actually observed. Indeed, if the calibrated value \hat{H}_t is higher than the observed value H_t we reduce the inflow into H_t , \widehat{dQH}_t , by the difference $\hat{H}_t - H_t$, provided this reduction produces a non-negative value, \widehat{dQH}_t .

In evolutionary computation, differential evolution (DE)³⁶ is a method that optimises a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. Such methods are commonly known as meta-heuristics as they make few or no assumptions about the problem being optimised and can search very large spaces of candidate solutions. DE is used for multidimensional real-valued functions but does not use the gradient of the problem being optimised, which means DE does not require the optimisation problem to be differentiable, as is required by classic optimisation methods such as gradient descent and quasi-Newton methods. Therefore, DE can also be used on optimisation problems that are not even continuous, are noisy, change over time, etc.. DE optimises a problem by maintaining a population of candidate solutions and creating new candidate solutions by combining existing ones according to its simple formulae, and then keeping whichever candidate solution has the best score on the optimisation problem at hand. Using a DE algorithm (provided by “R” package “DEoptimR”), for the first period ($t = 1$ to $t = 84$) we minimise the loss function:

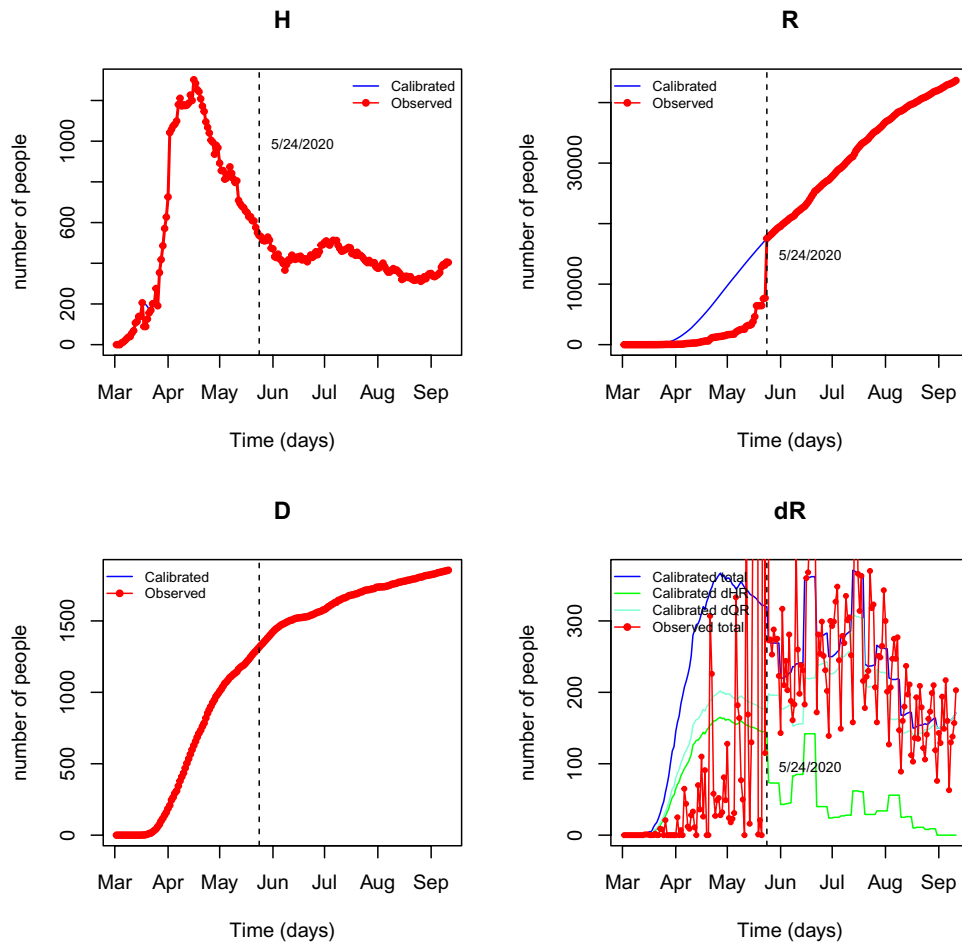


Figure 7. Observed and calibrated course of compartments H, R and D, and flows coming into compartment R.

$$\mathcal{L}_1(\theta) = \sum_{t=1}^{t^f} (H_t - \hat{H}_t)^2 + (R_{84} - \hat{R}_{84})^2 + (R_{84}^{(Q)} - 9652)^2. \quad (4)$$

and for the remaining sets of 7-day periods we minimise the loss function:

$$\mathcal{L}_2(\theta) = \sum_{t=1}^{t^f} \left((H_t - \hat{H}_t)^2 + (R_t - \hat{R}_t)^2 \right) \quad (5)$$

The first row of loss function (4) ensures the calibrated daily hospital occupancy is equal to the observed daily hospital occupancy. The second row ensures that the calibrated number of recovered individuals on the correction (day 84th) matches the observed and reliable figure as closely as possible. The third row ensures the accumulated number of recovered individuals originated on the quarantined compartment, $R_t^{(Q)} = \sum_{i=1}^t dQR_i$, is equal to the number of recovered individuals recorded by the Portuguese primary health care system, but released at once on May 24th, 2020 (9652 individuals). The second loss function (5) ensures the calibrated daily hospital occupancy is equal to the observed daily hospital occupancy and the daily accumulated number of recovered individuals is equal to the daily accumulated number of recovered individuals.

This procedure allows the calibrations of the flows dQH_t , dQR_t and dHR_t , conditional to the flows dSE_t and dEI_t calibrated in the first place, as described above.

The upper and lower limits of the hypersquare where the best solution is searched for are:

- $(\gamma_Q^{(u)}, \eta^{(u)}, h_0^{(u)}, h_1^{(u)}, \rho_h^{(u)}, \tau^{(u)}) = (1/10, 0.999)$ and
- $(\gamma_Q^{(l)}, \eta^{(l)}, h_0^{(l)}, h_1^{(l)}, \rho_h^{(l)}, \tau^{(l)}) = (1/120, 0.001)$.

The initial conditions are the two calibrated values in step one, $e_1 = dSE_1$ and $i_1 = dEI_1$ and the observed value $q_1 = dIQ_1$ and $s_1 = N - e_1 - i_1 - q_1$.

The calibrated values H , R , and D compartments resulting from the second step of the calibration procedure along with the original observed data are represented in panels (H), (R) and (D) of Fig. 7. The calibrated flows

\widehat{dQR}_t , \widehat{dHR}_t and $\widehat{dR} = \widehat{dQR}_t + \widehat{dHR}_t$ are represented in Fig. 7 (dR) along with the observed total flow dR (the deviation between the observed and calibrated dR on 2020-09-27, the last day under analysis, is red 0.3%). As expected, the accumulated number recovered individuals on May 25th (day 85) is interpolated, back-casting the excess of recovered cases released on May 24th (day 84), accordingly to the interdependence model dynamics. As such, the curve of calibrated quarantined individuals reproduces almost exactly its course after May 24th and flattens before this day, reproducing the calibrated recovered flows (dQR_t). This last flow, jointly with the flow of recovered individuals from the hospitalised compartment, dHR_t are calibrated resulting in reliable courses both for accumulated recovered individuals and daily recovered individuals.

Bayesian hierarchical modelling. Using the discrete-time approximation to the stochastic continuous-time SEIQRHD model (2) defined above we set up a Bayesian hierarchical model where the incidence variables are assumed stochastic. The transitions of individuals from the one compartment to the next one of the SEIQRHD model are considered stochastic movements between the corresponding population compartments. In each period an individual either stays in or moves on to the next compartment. In reliability analysis life-time is usually considered to follow an exponential distribution. By analogy, here the time length that an individual spends in a compartment is exponentially distributed with some compartment-specific rate λ_t . Therefore, the probability of extending the stay by a further period of length h is $\exp(-\lambda_t h)$ and the probability of leaving is therefore $1 - \exp(-\lambda_t h)$. The summation over the individual Bernoulli trials assuming they are independent and identical for all members of a compartment would result in binomial distributions³⁷. Due to the scale we are working with, we found it useful to take advantage of the approximation of the binomial to the Poisson distribution. Therefore incidences are assumed to follow the distributions:

$$\begin{aligned} dSE_t &\sim \text{Pois}(S_t \times p_{dSE}(t)), \\ dEI_t &\sim \text{Pois}(E_t \times p_{dEI}), \\ dIQ_t &\sim \text{Pois}(I_t \times p_{dIQ}), \\ dQH_t &\sim \text{Pois}(Q_t \times h_t \times p_{dQH}), \\ dQR_t &\sim \text{Pois}(Q_t \times (1 - h_t) \times p_{dQR}(t)), \\ dHR_t &\sim \text{Pois}(H_t \times (1 - \mu) \times p_{dHR}), \\ dHD_t &\sim \text{Pois}(H_t \times \mu \times p_{dHD}), \end{aligned} \quad (6)$$

where the transition probabilities are given by:

$$\begin{aligned} p_{dSE}(t) &= 1 - \exp\left(-\beta \frac{I_t}{N}\right), \\ p_{dEI} &= 1 - \exp(-\delta), \\ p_{dIQ} &= 1 - \exp(-\kappa), \\ p_{dQH} &= 1 - \exp(-\eta), \\ p_{dQR}(t) &= 1 - \exp(-\gamma_Q(t)), \\ p_{dHR} &= 1 - \exp(-\gamma_H), \\ p_{dHD} &= 1 - \exp(-\alpha) \end{aligned} \quad (7)$$

assuming the time unit h is 1 day.

Conditional on all information up to time t , the Poisson random variables dSE_t , dEI_t , dIQ_t , dQH_t , dQR_t , dHR_t and dHD_t are independent. The model further assumes that the population size N remains constant and that individuals mix homogeneously. In order to account for the effects of control measures such as lockdown and social distancing, we assume that the transmission parameter β is constant up to the time point when the control measures are introduced and after that decays exponentially at a constant rate. During the SARS-CoV-2 epidemic in Portugal four distinct periods can be distinguished. The first one, between the beginning of the epidemic and the declaration of the *state of emergency*, the first hard lockdown control measures (on March 18th 2020) where the transmission parameter is assumed to be constant. The second one, between March 18th, 2020, and the declaration of the *situation of calamity* when the lockdown measures started easing (May 4th, 2020). A third period from May 4th onwards and until August 18th, corresponding to the traditional Portuguese holidays time and the beginning of the school year and finally, from August 18th onwards. This is formulated as:

$$\beta(t) = \begin{cases} \beta_0, & t < \tau_0 \\ \beta_0 e^{-\rho_{\beta_0}(t-\tau_0)}, & \tau_0 \leq t < \tau_1, \\ \beta_1 e^{-\rho_{\beta_1}(t-\tau_1)}, & \tau_1 \leq t < \tau_2, \\ \beta_2 e^{-\rho_{\beta_2}(t-\tau_2)}, & t \geq \tau_2 \end{cases} \quad (8)$$

where τ_0 , τ_1 and τ_2 correspond to those three intervention dates (March 18th, May 4th, and August), β_0 is the initial transmission rate, and ρ_{β_0} , ρ_{β_1} , $\rho_{\beta_2} > 0$ are the rates at which β_0 , β_1 and β_2 decay on $\tau_0 \leq t < \tau_1$, $\tau_1 \leq t < \tau_2$ and $t \geq \tau_2$, respectively.

The basic reproduction number R_0 is defined as the average number of secondary cases generated by a primary case over his/her infectious period when introduced into a large population of susceptible individuals³⁸. The constant R_0 thus measures the initial growth rate of the epidemic and for the model above is $R_0 = \frac{\beta_0}{\kappa}$ ³⁹.

Furthermore⁴⁰, define the time dependent effective reproductive number $R_0(t) = \frac{\beta_t S_t}{\kappa N}$ as the number of secondary cases per infectious case at time t . Because $S_t \approx N$, it follows that $R_0(t) = \frac{\beta_t}{\kappa}$ is a function proportional to

the time-varying transmission rate in (8). The time point at which $R_0(t)$ assumes values smaller than 1 indicates when control measures have become effective in controlling the epidemic.

The parameter κ governs the passage from the infective to the quarantined compartment, as explained before. An infected individual is not automatically quarantined, because the authorities are often unable to test enough people while keeping pace with the spread. This measure is especially difficult because many people do not develop symptoms at all, but can transmit the infection to others. So, we believe that κ also contains some information about the percentage of the detected infective people. A study by⁴⁰ proposed the introduction of an additional parameter to understand this issue better. Despite all these considerations, we used values estimated from a microdata file provided by DGS in the first step of data calibration (as explained before). Therefore, we decided to move on a parsimonious parameterisation for κ , considering it fixed and equal to 1/4.73 days (see Table 1). Similarly, the incubation period, δ^{-1} is considered fixed and equal to 6.5 days (see Table 1). Accordingly to what we have discussed before, we assume η is constant, and h_t follows the model:

$$h(t) = h_0 \exp(-\rho_h t). \quad (9)$$

The parameters γ_Q and γ_H govern the passage from the Q and H compartments to the R compartment. As already mentioned, we assume γ_Q is much more dependent on protocols for recovery confirmation and therefore much more susceptible to increase over time than γ_H . Therefore, γ_Q follows an exponential trend, whilst γ_H is considered fixed. The assumption is that the recovery rate, $\gamma_Q(t)$ converges towards an asymptotic value:

$$\gamma_Q(t) = \gamma_{0Q}(1 - \exp(-\rho_\gamma t)). \quad (10)$$

The value of γ_{0Q} is the final asymptotic value of the recovery rate. It depends on ρ_γ which represents how fast the health system learns to respond to the disease and the protocol on recovery confirmation evolves.

Finally, the parameter μ , which governs the passage from the H compartment to the D compartment, corresponds to the share of the hospitalised people who dies due to the disease. Again, we believe μ is related to how a health system can improve its capability to treat people over time. Nevertheless, for parsimonious reasons and given the residual hospitalised fatality rate of COVID-19 we decided to assume μ is fixed. Evoking the aforementioned reasons, the parameter α is also considered fixed, but highly variable between individuals.

The epidemic model specified in (2), (6) and (7), together with the transmission rate model (8) and models (9) and (10) and the parameters δ , κ , η , γ_H , μ and α has parameter vector $\theta = (\beta_0, \beta_1, \rho_{\beta_0}, \rho_{\beta_1}, \delta, \kappa, \eta, h_0, h_1, \rho_h, \tau, \gamma_{0Q}, \rho_\gamma, \gamma_H, \mu, \alpha)$, which we would like to estimate from the knowledge of initial conditions $S_0 = s_0$, $E_0 = e_0$, $I_0 = i_0$, $Q_t = q_0$, $R_0 = 0$, $H_t = 0$ and $D_0 = 0$, population size, N , and from observed values of dSE_t , dEI_t , dIQ_t , dQH_t , dQR_t , dHR_t , dHD_t .

Gamma and uniform priors are assigned to each of the parameters in θ as

$$\begin{aligned} \beta_0 &\sim \text{Ga}(2, 10), & \beta_1 &\sim \text{Ga}(2, 10), \\ \beta_2 &\sim \text{Ga}(2, 10), & \rho_0 &\sim \text{Ga}(10, 100), \\ \rho_1 &\sim \text{Ga}(10, 100), & \rho_2 &\sim \text{Ga}(10, 100), \\ h_0 &\sim \text{Unif}(0, 1), & \rho_h &\sim \text{Ga}(1, 10), \\ \gamma_{0Q} &\sim \text{Ga}(10, 100), & \rho_\gamma &\sim \text{Ga}(1, 10), \\ \eta &\sim \text{Ga}(2, 10), & \gamma_H &\sim \text{Ga}(10, 100), \\ \mu &\sim \text{Unif}(0, 1), & \alpha &\sim \text{Ga}(10, 100) \end{aligned} \quad (11)$$

where $\text{Ga}(a, b)$ refers to a gamma distribution with parameters shape a and rate b , mean a/b , and variance a/b^2 and $\text{Unif}(a, b)$ to a uniform distribution between a and b with a mean $(a + b)/2$ and variance $1/(12(b - a))$.

Let $\mathbf{dSE} = \{dSE_t, t = 1, 2, \dots, T\}$ be the daily observed (calibrated) counts of susceptible individuals who become infected. We define similarly the other vectors: \mathbf{dEI} , \mathbf{dIQ} , \mathbf{dQH} , \mathbf{dQR} , \mathbf{dHR} and \mathbf{dHD} . Because the series are conditionally independent the likelihood of the data is:

$$\begin{aligned} \mathcal{L} &= p(\mathbf{dSE}|\cdot)p(\mathbf{dEI}|\cdot)p(\mathbf{dIQ}|\cdot)p(\mathbf{dQH}|\cdot)p(\mathbf{dQR}|\cdot)p(\mathbf{dHR}|\cdot)p(\mathbf{dHD}|\cdot) \\ &= \prod_{t=1}^T p(dSE_t|\cdot)p(dEI_t|\cdot)p(dIQ_t|\cdot)p(dQH_t|\cdot) \times \\ &\quad p(dQR_t|\cdot)p(dHR_t|\cdot)p(dHD_t|\cdot), \end{aligned} \quad (12)$$

where $p(\cdot|\cdot)$ stands for the Poisson transition probabilities specified in (6) conditioned on θ and on all the information up to time t . Given the hierarchical representation presented above and using the same notation, one can evaluate the posterior distribution of all of the parameters, given the observed counts:

$$\Pi(\theta|\mathbf{dSE}, \mathbf{dEI}, \mathbf{dIQ}, \mathbf{dQH}, \mathbf{dQR}, \mathbf{dHR}, \mathbf{dHD}). \quad (13)$$

One cannot evaluate this posterior distribution analytically and must resort to numeric simulation methods. We use the special case of MCMC known as Gibbs sampling⁴¹ and implement the algorithm using the “R” package “JAGS” (all code used in this paper can be obtained from authors upon request).

Results

After a burn-in period of 10,000 iterations by which we believe convergence has been achieved, a sample of size 250 taken every 500 iterations (to avoid serial correlation, especially for parameters where identification problem may arise, such as η and h and μ and α) of the chain was used to obtain marginal posterior distributions for all model parameters.

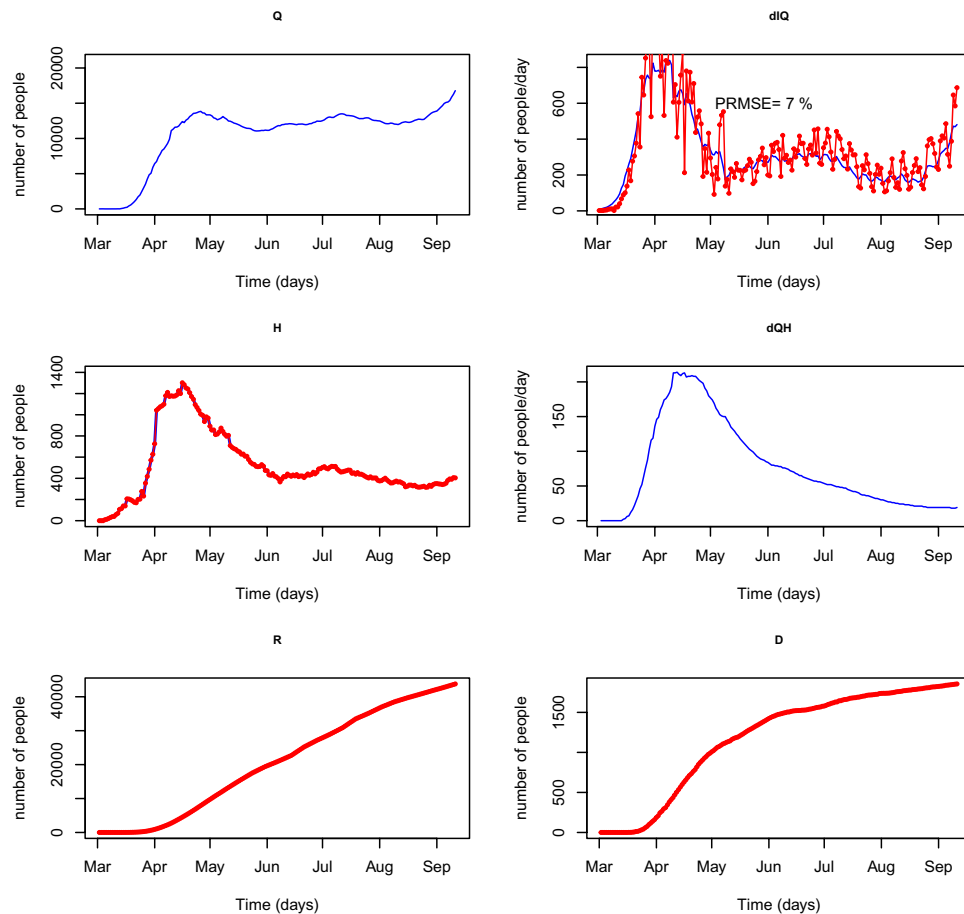


Figure 8. Results of SEIQRHD model. Means of the posterior predictive distributions of the most important flows for pandemic course management are represented in blue lines. Red lines represent compartments or flows for which observed values exist. The percentual root mean square error (PRMSE) is indicated in panel *dIQ* for the central flow that drives the course of the epidemic.

Convergence of the Markov chain was assessed using a series of runs for four different chains with different starting values and also inspecting the autocorrelation function (cf. Figures 15, 16, 17 and 18 in Supplementary Material). In all cases the Markov chain appeared to have converged after the burn-in period.

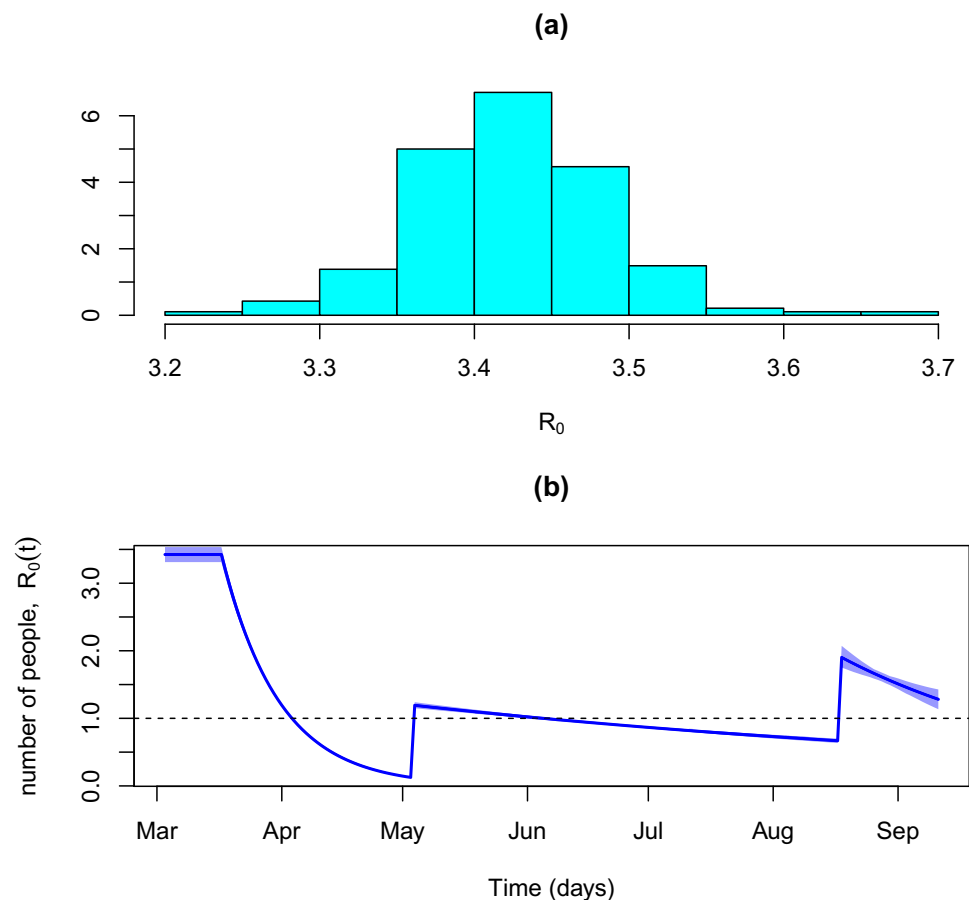
Analysing the fit of the model against real data is relevant as a tool of external validation. Therefore, the model fit is assessed by plotting the available data against the estimated expected values of the compartments. The fit of the model is evident from Fig. 8 for the period corresponding to the used data. Among the model variables represented, the flow of positive cases that are isolated, *dIQ_t*, is of essential importance, as is the one that drives the course of the epidemic and captures most of the attention of the health authorities. Thus, we have assessed the predictive relevance of the model by calculating the root mean square error of prediction (represented in the percentage of the mean of *dIQ_t*). The resulting value of 7% allows to confirm the validity of the model. Moreover, it is possible to assess the daily number of people in need of medical care through the flow *dQH* which is not released by the health authorities.

Table 2 presents the main results for the model parameters.

The mean of the basic reproduction number (R_0) is 3.434 with a 95% credible interval of (3.312–3.535) and varies between 3.250 and 4.165. The distribution of the basic reproduction number is depicted in Fig. 9a. The temporal evolution of the effective reproduction number is represented in Fig. 9b. The imposition of the lockdown and social distancing measures were effective in pulling the reproduction number down to a value below the epidemic control of one. Nevertheless, easing those measures implied a significant increase on the infection transmission in early May. Still, the level that was reached kept the disease's spread under control and additional restrictions made it possible to bring it to levels close to the threshold, although the end of the epidemic cannot be anticipated yet.

The posterior mean of transmission rate before first intervention (β_0) is 0.724 with a 95% credible interval of (0.700–0.747). These values are consistent with the values published in the literature (see, for example,²⁰). By the time the second intervention takes place, the value of $\beta(t)$ is only 0.027 (0.025–0.029). This explains the benefits of imposing lockdown measures to bring the transmission rate down from 0.724 to 0.027. As already mentioned, on May 4th the lockdown measures started to be eased and the transmission rate raised to 0.252 (0.243–0.262).

Parameter	Description	Units	Prior mean (std)	Posterior mean (std)	Days (95% CrI)
R_0	Basic infection reproduction number	No.	–	3.424 (0.0610)	–
β_0	Contact rate before first intervention	days ⁻¹	0.2 (0.1414)	0.724 (0.0130)	–
β_1	Contact rate on second intervention	days ⁻¹	0.2 (0.1414)	0.252 (0.0051)	–
β_2	Contact rate on third intervention	days ⁻¹	0.2 (0.1414)	0.402 (0.0174)	–
ρ_{β_0}	Rate of decay of β_0	days	0.1 (0.0316)	0.007 (0.0012)	–
ρ_{β_1}	Rate of decay of β_1	days	0.1 (0.0316)	0.006 (0.0004)	–
ρ_{β_2}	Rate of decay of β_2	days	0.1 (0.0316)	0.016 (0.0038)	–
η	Hospitalisation rate	days ⁻¹	0.2 (0.1414)	0.636 (0.1349)	1.13–2.69
h_0	Initial probability of hospitalisation	%	0.5 (0.2887)	0.067 (0.0116)	–
ρ_h	Rate at which probability of hospitalisation decreases	%	0.1(0.1)	0.018 (0.0002)	–
γ_0	Final (asymptotic) recovery rate of quarantined people	days ⁻¹	0.01 (0.0032)	0.016 (0.0001)	62.89–64.43
ρ_γ	Rate at which recovery rate of quarantined people increases	days	0.1 (0.0316)	0.081 (0.0106)	–
γ_H	Recovery rate of hospitalised individuals	days ⁻¹	0.01 (0.0032)	0.180 (0.0051)	5.24–5.85
μ	Hospitalised case fatality rate	%	0.5 (0.2887)	0.26 (0.0177)	–
α	Rate of death	days ⁻¹	0.01 (0.0032)	0.07 (0.0052)	12.12–16.31

Table 2. Posterior statistics of SEIQRHD model parameters.**Figure 9.** Infection reproduction number. (a) Basic Reproduction Number distribution, R_0 . (b) Effective reproduction number, $R_0(t)$.

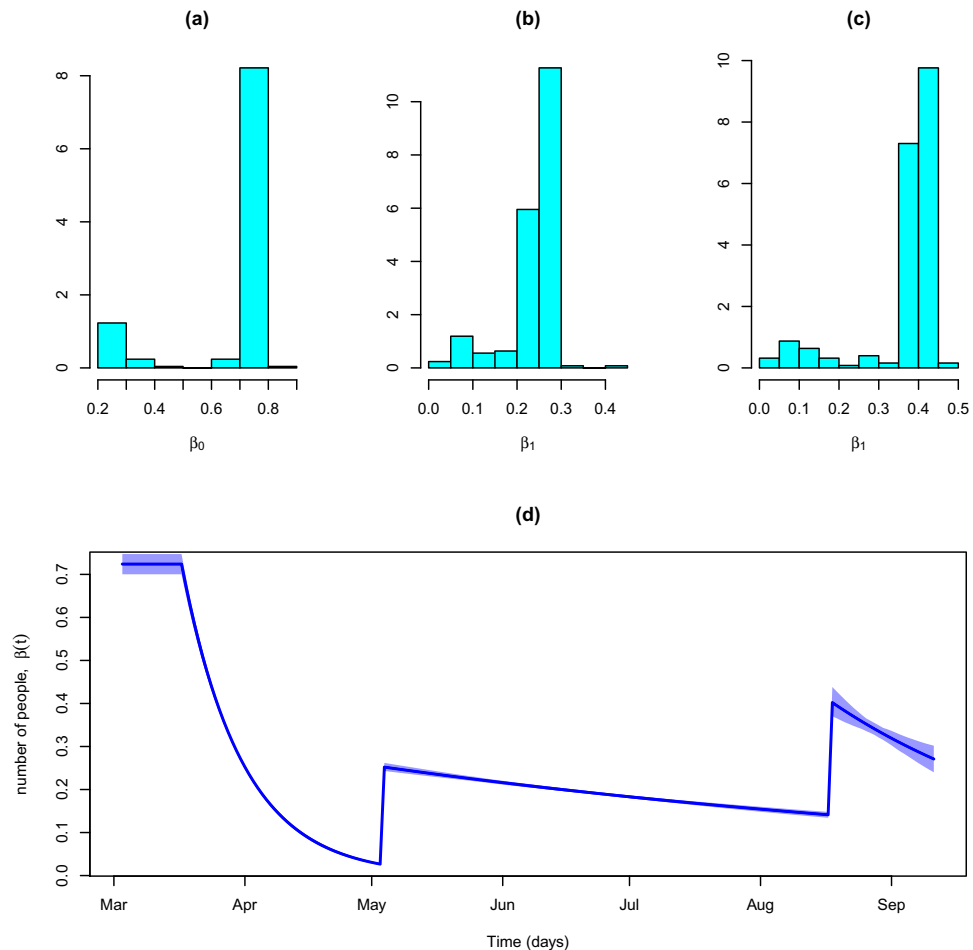


Figure 10. Transmission rate, β_t . (a) Initial transmission rate distribution, β_0 . (b) Transmission rate distribution after the second intervention, β_1 . (c) Transmission rate distribution after the third intervention, β_2 . (d) Course of the transmission rate, $\beta(t)$.

starting a new downward path. Nevertheless, the end of the Portuguese traditional holidays period and (second half of August) and the beginning of the academic year push the transmission rate up again from 0.141 to 0.402 due to the natural increase of daily contacts. The distribution of ρ_{β_0} , ρ_{β_1} , and ρ_{β_2} are Gaussian-like with posterior means 0.07 (0.068–0.072), 0.006 (0.005–0.006), and 0.016 (0.009–0.024). The rate of decay of β_1 is very minimal. Indeed, the value of $\beta(\tau_2 - 1)$ (the day before of the third intervention) is 0.14 indicating the transmission rate after the second intervention (0.24) has a small decay. Actually, although the current mobility levels (cf. *Google Mobility Reports*) are below the pre-pandemic recorded figures, from early May onwards they started to catch up and are currently close to the pre-pandemic levels from early May onwards, potentially inducing a large number of daily infections. The posterior distribution of the transmission rates, β_0 , β_1 and β_2 are represented in Fig. 10a–c. The temporal evolution of the transmission rate is represented in Fig. 10d.

The posterior mean of the hospitalisation rate (η) is between 0.636 (0.371–0.888) corresponding to a mean period of 1.13–2.69 days between case confirmation and hospital admittance. The posterior mean of the initial hospitalisation probability h_0 is 6.71% (5.12–9.89%). Confirming what is currently known about the severity of the disease this figure indicates that only a small proportion of the quarantined individuals need specialised medical care. The posterior distributions of the hospitalisation rate, η , and initial hospitalisation fraction, h_0 are represented in Fig. 11a,b. The temporal evolution of the hospitalisation fraction is represented in Fig. 11c.

Regarding recovery, we observe a mean asymptotic recovery rate of 0.016 (0.0155–0.0159). This rate translates in a mean asymptotic recovery period of 63.6 (62.89–64.43) days. Although it seems to be larger than expected, this value is affected by the speed of the reporting of recovered individuals. Indeed, it is consensual between epidemiologists in Portugal that recovered individuals figures released by DGS have been permanently below the expert expectations, given recovery time reported in the literature (around 14 days, on average) and admitted by clinical practitioners. Indeed, by the time we of finishing this manuscript DGS has already released a batch of additional 13529 recovered individuals (referring to the previous period with unspecified allocation) that would result in a reduction of the recovery time consistent with the usual clinical mean recovery period. The posterior distribution of the asymptotic recovery rate is depicted in Fig. 12a. The temporal evolution of the recovery rate is represented in Fig. 12c.

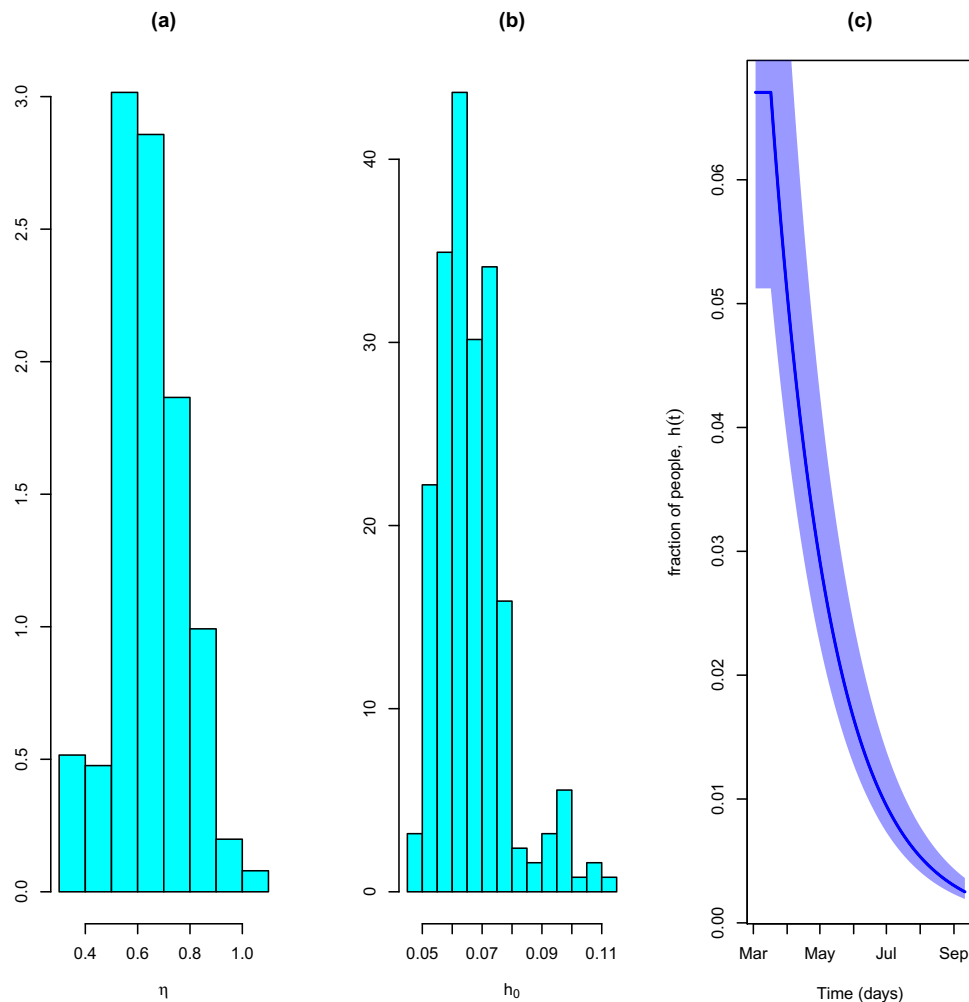


Figure 11. Hospitalisation rate. **(a)** Hospitalisation rate distribution, η . **(b)** Initial hospitalisation fraction, h_0 . **(c)** Hospitalisation fraction course, $h(t)$.

On the other hand, the posterior mean of the recovery rate of hospitalised individuals is 0.18 (0.171–0.191) which corresponds to a period of 5.5 (5.24–5.85) days in hospital before recovery confirmation. That is a perfectly reasonable period despite its dependence on the health system promptness. The posterior distribution of the mean recovery rate of hospitalised individuals is depicted in Fig. 12b.

Finally, the posterior mean of the hospitalised case fatality rate (HCFR) (μ) is 26.12% (22.40–29.53%). We must remark that this HCFR concerns only the hospitalised individuals and should not be compared with the case fatality rate (CFR) only concerns the recorded number of infections. Indeed, in our model the case fatality rate is given by the ratio of D_t (accumulated number of deaths) to C_t (accumulated number of infected individuals). The course of the CFR is represented in Fig. 13. It follows a convergent path to a maximum value around 3.9% which is in line with the values reported by the literature, for countries that have adopted a lockdown policy for a significant period^{42,43}. The posterior mean of the death rate, α , is 0.071 (0.061–0.082) which corresponds to a period of 14.2 (12.12–16.31) days in hospital before death. As anticipated this is highly dispersed value denoting a significant heterogeneity of the length of stay in hospital among individuals.

Discussion

This work adopts a generalised SEIR model, the so-called SEIQRHD model, to offer a quantitative overview of the complex structural analysis and prediction of the SARS-CoV-2 epidemic. The work was developed during the first wave of course of the epidemic. By the time we completed the manuscript, a new wave of increased infections was striking all over Europe.

The proposed SEIQRHD model is, to the best of our knowledge, the first considering a specific compartment to follows the hospitalised individuals who might put an intolerable burden on the national health system, especially now that a second wave is starting and no one can anticipates the course of the epidemic when the Winter arrives. This model offers to health authorities the possibility to predict of the number of individuals who are going to require hospitalisation, thus allowing for better planning of resources. It must be emphasized that the data available to calibrate the model and discover some of the necessary flows includes infirmary and

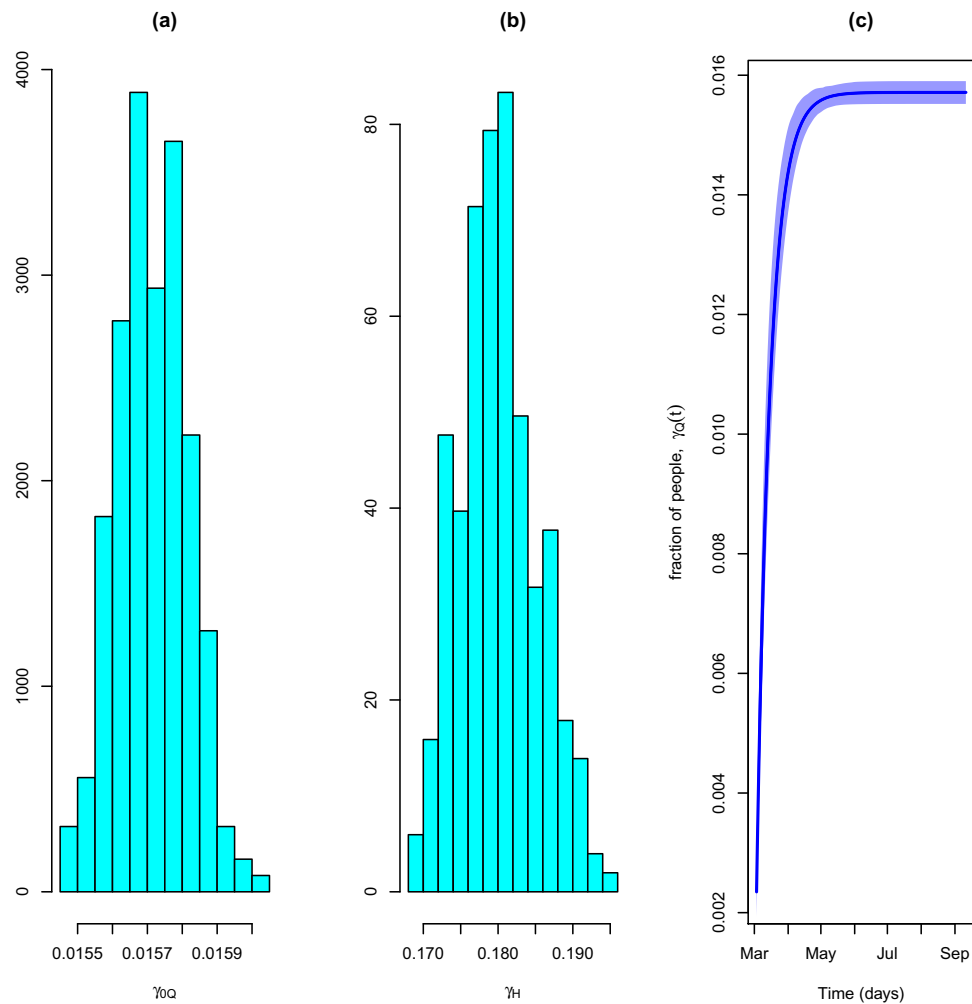


Figure 12. Recovery rate. (a) Asymptotic recovery rate for quarantined individuals' distribution, γ_{0Q} . (b) Recovery rate for hospitalised individuals' distribution, γ_H . (c) Recovery rate course for quarantined individuals, $\gamma_Q(t)$.

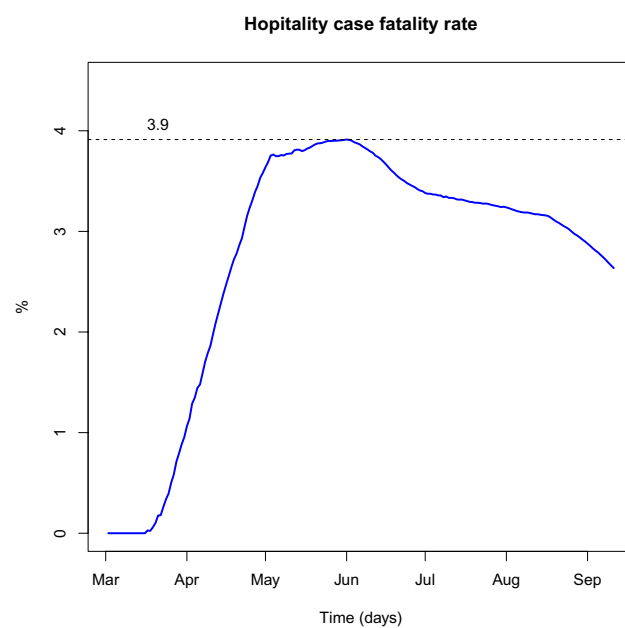


Figure 13. Case fatality rate.

ICU occupancy, that is, the daily number of occupied beds. The calibration procedure allows to estimate the in and outflows of compartment H, thus allowing to approximate these unobserved values, whose knowledge is of extreme importance for the health system to allocate the necessary resources. Moreover, the set of model parameters (some of them ordinary in SEIR model specification), either time-fixed or time-varying, provide useful insights on the health authorities reaction to the daily challenges. Namely, it is possible to monitor (a) the time between case symptom onset and subsequently isolation, (b) the time spent in isolation, (c) the hospitalisation rate and the probability of hospitalisation and its evolution over time which provides information on the near future burden to the health system among other important insights as discussed in the section devoted to the analysis of results. It is also possible to distinguish between recovery rates and duration for hospitalised and non-hospitalised individuals, allowing a much deeper understanding on the recovery evolution and the efficiency of hospital care. Daily estimation of the SEIQRHD along with the predictions for stocks and flows between compartments may have strong implications on public health policy and resources allocation to control the pandemic. Results from the model may help health authorities to adapt their strategies and plan their resources within at least three action vectors: (a) monitor how fast the public health teams in the field identify and isolate positive cases turning possible to anticipate the number of professionals that need to be allocated to these tasks; (b) by following and predicting the hospitalization needs, allowing to anticipate possible stress levels in the healthcare provided and the reallocation of beds devoted to SARS-CoV-2 and other pathologies; (c) by better predicting the course of the pandemic in terms of incidence and prevalence, allowing a better planning of mitigation measures, including more informed decisions about the timing and intensity of confinement measures.

An additional future step over of this line of modelling is the separation between hospitalised noncritical patients from the ones in intensive care, which may of course offer authorities with additional insight for planning purposes.

Other researchers might use the methodology developed here to either calibrate non-error free or missing epidemic data. Indeed, as⁴⁴ points out, frequently the official databases are not exempt from measurement errors and/or missing data that must be accounted for and fixed before any further analysis. Estimating missing flows or the calibrated version of observed series can be done using the rationale behind the two-step calibration procedure proposed in this paper. For Portuguese data the calibration procedure has shown to be crucial to cope with errors on the official recovery cases, with impact over virtually all the transition rates between compartments and particularly the ones related with hospital occupancy: hospitalisation rate and recovery rate from hospitalised patients.

Both the calibration methodology and the Bayesian model were adapted to the Portuguese case and corresponding available data. However, we believe the data released in Portugal, with some minor differences, are widely available in most of the developed countries. Therefore, the methodologies developed here might be successfully replicated in many other countries.

The numbers retrieved from the Portuguese data show that Portugal has been dealing with the epidemic promptly and the measures taken to flatten the incidence curve have proved to produce the desirable results. In particular, the performed modeling allowed to understand how the contact rate has evolved over time and more importantly, how effective several interventions were (lockdown or other mobility restriction measures). For example, the first Portuguese intervention allowed the transmission rate to reduce from 0.724 to 0.027, but the second intervention was much less effective only allowing a reduction from 0.24 to 0.14. It is important to notice that the first intervention corresponded to a full lockdown, while the second one corresponded to setting limited mobility restrictions. It also allows us to understand the price to pay when restrictions are removed or eased. In fact, we have observed a rapid increase of the transmission rate from 0.027 to 0.252 right before the second intervention and from 0.141 to 0.402 before the third one.

These results may help public authorities to set the timing and intensity of the measures in order to achieve the planned results and particularly to avoid the implosion of the health system.

Received: 22 February 2021; Accepted: 14 September 2021

Published online: 04 October 2021

References

1. Zhu, N. *et al.* A novel coronavirus from patients with pneumonia in China. *N. Engl. J. Med.* **382**, 727–733. <https://doi.org/10.1056/NEJMoa2001017> (2020).
2. Chan, J. F. *et al.* A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *Lancet* **395**, 514–523. [https://doi.org/10.1016/S0140-6736\(20\)30154-9](https://doi.org/10.1016/S0140-6736(20)30154-9) (2020).
3. Chen, J. Pathogenicity and transmissibility of 2019-nCoV—A quick overview and comparison with other emerging viruses. *Microbes Infect.* **22**, 69–71 (2020).
4. Wang, D. W. *et al.* Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* **323**, 1061–1069. <https://doi.org/10.1001/jama.2020.1585> (2020).
5. Chinese Center for Disease Control and Prevention. Epidemic update and risk assessment of 2019 Novel Coronavirus 2020. <http://www.chinacdc.cn/yyrdgz/202001/P020200128523354919292.pdf> (2020).
6. Riou, J. & Althaus, C. L. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance* **25**(4), 2000058. <https://doi.org/10.2807/1560-7917> (2020).
7. Wu, J. T. *et al.* Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* <http://www.nature.com/articles/s41591-020-0822-7> (2020).
8. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *Lancet* **395**–10225, 689–697. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9) (2020).

9. European Centre for Disease Prevention and Control (ECDC). Outbreak of novel coronavirus disease 2019 (COVID-19): increased transmission globally—fifth update. <https://www.ecdc.europa.eu/sites/default/files/documents/RRA-outbreak-novel-coronavirus-disease-2019-increase-transmission-globally-COVID-19.pdf>.
10. Abou-Ismaïl A. Compartmental Models of the COVID-19 Pandemic for Physicians and Physician-Scientists [published online ahead of print, 2020 Jun 4]. *SN Compr. Clin. Med.* 1–7. <https://doi.org/10.1007/s42399-020-00330-z> (2020).
11. Ndaïrou, F., Area, I., Nieto, J. J. & Torres, D. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos Solitons Fractals* **135**, 109846. <https://doi.org/10.1016/j.chaos.2020.109846> (2020).
12. Tianbing, W. *et al.* A four-compartment model for the COVID-19 infection-implications on infection kinetics, control measures, and lockdown exit strategies. *Precision Clin. Med.* 3–2, 104–112. <https://doi.org/10.1093/pcmedi/pbaa018> (2020).
13. Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. A* **115**(772), 700–721. <https://doi.org/10.1098/rspa.1927.0118> (1927).
14. Hethcote, H. W. The mathematics of infectious diseases. *SIAM Rev.* **42**–4, 599–653 (2000).
15. Biao, T. *et al.* An updated estimation of the risk of transmission of the novel coronavirus (2019-ncov). *Infect. Dis. Model.* **5**, 248–255. <https://doi.org/10.1016/j.idm.2020.02.001> (2020).
16. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* **368**–6490, 489–493. <https://doi.org/10.1126/science.abb3221> (2020).
17. Lauer, S. A. *et al.* The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Intern. Med.* **172**(9), 577–582. <https://doi.org/10.7326/M20-0504> (2020) (Epub 2020 Mar 10).
18. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *Lancet*. **395**(10229), 1054–1062. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3) (2020).
19. Xinhai, L., Xumao, Z. & Yuehua, S. The lockdown of Hubei province causing different transmission dynamics of the novel coronavirus (2019-ncov) in Wuhan and Beijing. *medRxiv* <https://doi.org/10.1101/2020.02.09.20021477> (2020).
20. Godio, A., Pace, F. & Vergnano, A. SEIR modeling of the Italian epidemic of SARS-CoV-2 using computational swarm intelligence. *Int. J. Environ. Res. Public Health* **17**(10), 3535 (2020).
21. Samuel, J. C. *et al.* Interventions targeting air travelers early in the pandemic may delay local outbreaks of sars-cov-2. *MedRxiv* <https://doi.org/10.1101/2020.02.12.20022426> (2020).
22. Tang, B. *et al.* Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *J. Clin. Med.* **9**(2), 462. <https://doi.org/10.3390/jcm9020462> (2020).
23. Mingwang, S. *et al.* Novel coronavirus in Hubei province, China. *medRxiv* <https://doi.org/10.1101/2020.02.11.20022236> (2019).
24. Hao, X. & Huili, Y. Simulating the infected population and spread trend of 2019-nCoV under different policy by SEIR model. *MedRxiv* <https://doi.org/10.1101/2020.02.10.20021519> (2020).
25. Peng, L., Yang, W., Zhang, D., Zhuge, C. & Hong, L. Epidemic analysis of COVID-19 in China by dynamical modeling. *MedRxiv Epidemiol.* <https://doi.org/10.1101/2020.02.16.20023465> (2020).
26. Kobayashi, T. *et al.* Communicating the risk of death from novel coronavirus disease (COVID-19). *J. Clin. Med.* **9**(2), 580 (2020).
27. Read, J. M., Bridgen, J. R., Cummings, D. A., Ho, A. & Jewell, C. P. Novel coronavirus 2019-nCoV: Early estimation of epidemiological parameters and epidemic predictions. *medRxiv* <https://doi.org/10.1101/2020.01.23.20018549> (2020).
28. Lytras, T., Panagiotakopoulos, G. & Tsiodras, S. Estimating the ascertainment rate of SARS-CoV-2 infection in Wuhan, China: Implications for management of the global outbreak. *MedRxiv* <https://doi.org/10.1101/2020.03.24.20042218> (2020).
29. Nishiura, H. *et al.* The rate of underascertainment of novel coronavirus (2019-nCoV) infection: Estimation using Japanese passengers data on evacuation flights. *J. Clin. Med.* **9**(2), 419 (2020).
30. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **369**–6490, 489–493. <https://doi.org/10.1126/science.abb3221> (2020).
31. Tuite, A. R., Ng, V., Rees, E. & Fisman, D. Estimation of COVID-19 outbreak size in Italy based on international case exportations. *medRxiv* <https://doi.org/10.1101/2020.03.02.20030049> (2020).
32. Swart, E. *et al.* Gute Praxis Sekundärdatenanalyse (GPS): Leitlinien und Empfehlungen [Good Practice of Secondary Data Analysis (GPS): guidelines and recommendations]. *Gesundheitswesen*. **77**(2):120–126. <https://doi.org/10.1055/s-0034-1396815> (2015) (Epub 2015 Jan 26, , German).
33. Tuite, A. R. *et al.* Estimation of COVID-19 burden and potential for international dissemination of infection from Iran. *medRxiv* <https://doi.org/10.1101/2020.02.24.20027375> (2020).
34. Backer, J. A., Klinkenberg, D. & Wallinga, J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance*. **25**(5), 2000062. <https://doi.org/10.2807/1560-7917.ES.2020.25.5.2000062> (2020).
35. Mode, C. J. & Sleeman, C. K. *Stochastic Processes in Epidemiology* (World Scientific, 2020).
36. Storn, R. & Price, K. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.* **11**–4, 341–359. <https://doi.org/10.1023/A:1008202821328> (1997).
37. Diekmann, O., Heesterbeek, J. & Metz, J. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382 (1990).
38. van den Driessche, P. Reproduction numbers of infectious disease models. *Infect. Dis. Model.* **2**(3), 288–303. <https://doi.org/10.1016/j.idm.2017.06.002> (2017).
39. Chowell, G., Hengartner, N. W., Castillo-Chavez, C., Fenimore, P. W. & Hyman, J. M. The basic reproductive number of Ebola and the effects of public health measures: The cases of Congo and Uganda. *J. Theor. Biol.* **229**(1), 119–26. <https://doi.org/10.1016/j.jtbi.2004.03.006> (2004).
40. Calafiore GC, Novara C, Possieri C. A Modified SIR Model for the COVID-19 Contagion in Italy, [arXiv:physics/2003.14391](https://arxiv.org/abs/2003.14391) (2020).
41. Gilks, W. R., Richardson, S., Spiegelhalter, D. J. (eds) *Markov Chain Monte Carlo in Practice. Interdisciplinary Statistics* (Chapman & Hall, 1996).
42. Wu, Z. & McGoogan, J. M. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. *J. Am. Med. Assoc.* **323**(13), 1239–42. <https://doi.org/10.1001/jama.2020.2648> (2020).
43. Stokes, E. K. *et al.* Coronavirus disease 2019 case surveillance—United States, January 22–May 30, 2020. *Morb. Mortal. Wkly. Rep.* **69**(24), 759–765. <https://doi.org/10.15585/mmwr.mm6924e2> (2020).
44. Ashofteh, A. & Bravo, J. M. A study on the quality of novel coronavirus (COVID-19) official datasets. *Stat. J. IAOS* **36**–2, 291–301. <https://doi.org/10.3233/SJI-200674> (2020).

Author contributions

J.M.M. wrote the manuscript and prepared tables and figures. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-98975-w>.

Correspondence and requests for materials should be addressed to J.M.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021