Alexandria Engineering Journal (2021) 60, 4363–4380

HOSTED BY

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

The COVID-19 (SARS-CoV-2) uncertainty tripod in Brazil: Assessments on model-based predictions with large under-reporting

Saulo B. Bastos [a,b], **Marcelo M. Morato** [c,*], **Daniel O. Cajueiro** [a,b,d], **Julio E. Normey-Rico** [c]

[a] *Departamento de Economia, FACE, Universidade de Brasília (UnB), Campus Universitário Darcy Ribeiro, 70910-900 Brasília, Brazil*
[b] *Machine Learning Laboratory in Finance and Organizations (**LAMFO**), FACE, Universidade de Brasília (UnB), Campus Universitário Darcy Ribeiro, 70910-900 Brasília, Brazil*
[c] *Renewable Energy Research Group (**GPER**), Departamento de Automação e Sistemas (**DAS**), Universidade Federal de Santa Catarina (UFSC), Florianópolis, Brazil*
[d] *Nacional Institute of Science and Technology for Complex Systems (INCT-SC), Brazil*

Received 26 June 2020; revised 25 February 2021; accepted 3 March 2021
Available online 18 March 2021

KEYWORDS

COVID-19;
Under-reporting;
SIR model;
Uncertainty;
Brazil

Abstract The COVID-19 pandemic (SARS-CoV-2 virus) is the global crisis of our time. The absence of mass testing and the relevant presence of asymptomatic individuals causes the available data of the COVID-19 pandemic in Brazil to be largely under-reported regarding the number of infected individuals and deaths. We develop an adapted Susceptible-Infected-Recovered (SIR) model, which explicitly incorporates the under-reporting and the response of the population to public health policies (confinement measures, widespread use of masks, etc). Large amounts of uncertainty could provide misleading predictions of the COVID-19 spread. In this paper, we discuss the role of uncertainty in these model-based predictions, which is illustrated regarding three key aspects: (i) Assuming that the number of infected individuals is under-reported, we demonstrate anticipation regarding the infection peak. Furthermore, while a model with a single class of infected individuals yields forecasts with increased peaks, a model that considers both symptomatic and asymptomatic infected individuals suggests a decrease of the peak of symptomatic cases. (ii) Considering that the actual amount of deaths is larger than what is being registered, we demonstrate an increase of the mortality rates. (iii) When we consider generally under-reported data, we demonstrate how the transmission and recovery rate model parameters change qualitatively and quantitatively. We also investigate the "the uncertainty tripod": under-reporting level in terms of cases, deaths, and the true mortality rate of the disease. We demonstrate that if two of these factors

* Corresponding author.
E-mail addresses: marcelomnzm@gmail.com (M.M. Morato), danielcajueiro@unb.br (D.O. Cajueiro), julio.normey@ufsc.br (J.E. Normey-Rico).
Peer review under responsibility of Faculty of Engineering, Alexandria University.

https://doi.org/10.1016/j.aej.2021.03.004

are known, the remainder can be inferred, as long as proportions are kept constant. The proposed approach allows one to determine the margins of uncertainty by assessments on the observed and true mortality rates.

1. Introduction

The COVID-19 pandemic is definitely the global crisis of our time. A Chinese scientist first identified the SARS-CoV-2 virus in humans in Wuhan, in the province of Hubei, China by December 2019. This virus causes severe acute respiratory syndromes which can become potentially fatal. By the end of June 2020, the World Health Organization (WHO) estimated that the number of confirmed cases was already reaching the order $10,000,000$, with over $490,000$ confirmed deaths.

Much more than presenting drastic effects on health systems, social and economical backlashes are already felt by many countries. This phenomenon is especially evident in countries with larger social inequalities, such as Brazil. The effects of the virus on populations with poorer access to health systems and sanitation facilities are strikingly harder [21,45,25]. The city of São Paulo shows a very illustrative example of these differences: the city hall released a technical note by the end of April[1] stating that the observed mortality rate is 10 times larger in neighborhoods of the city with worse social conditions and precarious housing.

In this paper, we consider the context of the spread of the SARS-CoV-2 virus in Brazil, as detailed in [51]. Brazil is facing many issues since the beginning of the contagion, such as the advance of the virus to farthest western cities, away from urban areas, where medical care is somehow less present. The country has 26 federated states, which have been choosing different social distancing measures since mid-March.[2] Even though a strong public health system is available in Brazil, many states have been exhibiting near-collapsing conditions since May, with over 95% of Intense Care Unit (ICU) hospital beds occupied with COVID-19 patients [9]. Furthermore, we note that the SARS-CoV-2 is currently posing a great threat to indigenous communities, such as the Yanomami and Ye'kwana ethnicities[3]. Clearly, the situation is borderlining.

The first official death due to the SARS-CoV-2 virus in Brazil was registered in March 17, while the first case was officially notified in February 26, 2020. Through inferential statistics, [12] acknowledge the fact that community transmission has been ongoing in the state of São Paulo since the beginning of February (over one month before the first official reports). This points to empirical evidence that the true amount of infected individuals, and possibly registered deaths, are actually very under-reported.

Due to the absence of mass testing in the majority of cities, Brazil has basically only accounted for moderate to severe COVID-19 cases. People with mild or no symptoms are being oriented by the Ministry of Health to stay home. Added to this fact, the SARS-CoV-2 virus shows itself as an asymptomatic contagion for a large number of individuals. For these reasons, the scientific community has been warning for a possibly huge margin of underestimated cases in Brazil [47,42,12]. Some studies, such as [4], point out to the presence of over 700% of under-reported cases. Furthermore, the daily reports ("datasets") disclosed by the Brazilian Ministry of Health,[4] only give an impression of the virus contagion in the past, since, on average, a person exhibits acute symptoms only 20 days of the moment of infection. Through statistical procedures, [35] have recently confirmed the empirical evidence that the margin of under-reported cases is quite large in the majority of Brazilian states. The state of São Paulo seems to be the one with less uncertainty regarding the number of deaths because the data also partially incorporates those deceased due to severe acute respiratory syndromes even without COVID-19 testing.

Therefore, it seems evident that uncertainty plays a significant role in this contagion in Brazil. This issue should be directly taking into account in the formulation of nationwide public health policies. With respect to this context, this paper investigates the role of uncertainty in such a way that decision-makers are able to plan more coherent, and adherent to reality, policies. It is worth mentioning that propositions to address the pandemic through recurrent social isolation periods has been recently assessed through optimal control in [31,32].

Table 1 summarizes the estimates available in the literature regarding the under-reporting levels and true mortality rates for Brazil. In this Table, we evidence the percentage increase of under-reports w.r.t. to reported deaths and cases, as evaluated by prior studies. Most of these works show that, on average, the number of infected individuals could be 3 to 14 times higher. Some studies from other countries point out that this number could reach up to 30.[5] According to technical news-pieces disclosed by [49,19], the amount of under-reports in terms of deaths due to the SARS-CoV-2 virus ranges from

[1] Secretaria Municipal da Saúde de São Paulo. **Boletim Quinzenal Covid-19**. April 30, 2020. Available at: https://www.prefeitura.sp.gov.br/cidade/secretarias/upload/saude/PMSP_SMS_COVID19_Boletim%20Quinzenal_20200430.pdf.

[2] Throughout this paper, the Year/Month/Day notation is used.

[3] The Brazilian Socioenvironmental Institute (ISA, *Instituto Socioambiental*, see https://www.socioambiental.org/en) has released a technical note [16] which warns for the contagion of COVID-19 of up to 40% of Yanomami Indigenous Lands, amid the states of Amazonas and Roraima and along the border between Brazil and Venezuela, due to the presence of approximately 20000 illegal mining prospectors. Datasets regarding the COVID-19 spread amid indigenous communities are available in https://covid19.socioambiental.org.

[4] These datasets comprise the number of infected and deceased patients on the given day.

[5] Naomi Martin. **Mass. official coronavirus count is 218, but experts say that the true number could be as high as 6,500**. The Boston Globe. March 17, 2020. Available at: https://www.bostonglobe.com/2020/03/17/metro/mass-official-coronavirus-count-is-197-experts-say-true-number-could-now-be-high-6000/.

Table 1 Estimates of COVID-19 sub-report levels in Brazil and true mortality rates.

	Infected		Deaths		True
Source	Reported	× more	Reported	× more	Mortality rate
[1,17]	6.55%	14.2	100%	0	1.08% to 1.11%
[38]	7.8% to 8.1%	11.8 to 12.3	100%	0	1.3%
[34]			26.9% to 37.5%	1.67 to 2.72	
[19]			85.5%	0.17	
[49]			67.5%	0.48	
[15]	10% (5.9% to 20%)	9 (4 to 16)			0.42% (0.23% to 0.87%)
[14]	25% (13.9% to 41.7%)	3 (1.4 to 6.2)			0.97% (0.84% to 1.12%)
[4]	12.5%	7			
[10]	7.4%	12.5			

17% to 122%. [1,17] estimate the real mortality rate estimated for Brazil to be approximately 1.08% to 1.11%, but estimates from the research group [15,14] from random samples of the Brazilian population point out that this mortality could be ranging from 0.42% to 0.97%. Studies from other countries, such as [26,6] show that this true mortality rate varies roughly from 1 to 5%, but [30] states that recent research is converging to the estimation of the true mortality rate to be between 0.5 to 1%. [37] find the true mortality rate to be approximately 0.64% for Geneva, Switzerland.

Bearing in mind the previous discussion, the main motivation of this paper is to present an adapted Susceptible-Infected-Recovered (SIR) model which inherently takes into account these uncertainty levels, considering the Brazilian COVID-19 context. Furthermore, the motivation is also to assess the role of uncertainty in the predictions cast with such models. We denote uncertainty as the amount of sub-notification with respect to infected and deceased individuals. Our approach comprises the following ingredients:

- Firstly, we propose a new modeling scheme that incorporates a dynamic decaying parameter for the viral transmission rate. The dynamic decaying parameter for the transmission rate, adapted from [31], considers that the government applies contagion mitigation measures (such as social isolation and incentives to use masks, which we refer to as "pandemic policies" henceforth), which decreases the contagion spread dynamics.
- Secondly, we develop an uncertainty measure with respect to infected and deceased individuals. The uncertainty is embedded in the optimization procedures used to determine the epidemiological parameters, in order to correct underestimates of infected and deceased individuals.
- Thirdly, we use these adapted SIR models to make predictions for the Brazilian scenario regarding several different uncertainty sets, with short and long-term forecast spans. By this, we are able to illustrate the role of under-reporting in the model response curves. Specifically, we study its effect upon the peak of infections (in terms of amplitude and time shift), upon the total number of deaths, observed mortality rate, and model epidemiological parameters.
- Finally, considering the uncertainty tripod, i.e., the strong link between under-reporting of infected and deaths and the true mortality rate, we extrapolate and suggest an alignment of the observed and the true mortality rate to infer on the level of uncertainty present on the measurements.

This paper relates to a previous paper by the Authors [5], wherein SIR-like models present short and long-term outlooks for Brazil. The previous work covered only the data from an early stage of the contagion, until March 30, 2020. Furthermore, the Authors used parametric variations of the parameters to introduce the uncertainty in the identified model, differing from the approach proposed herein. In this paper, we consider the data available until June 2020.

Fig. 1 shows the SIR model forecasts back on March 17, with respect to real data, considering variations for the transmission parameter β. We note that parametrically changing the transmission parameter is not enough to make the model fit the real data. Therefore, a time-varying parameter is added to model the effect of pandemic policies, which meddle with the viral transmission rates. Such an adapted model is able to account for the population response to governmental enacted policies. In practice, the pandemic policy can be understood as feedback of endogenous variables, also depending on non-observed time-varying factors (people can simply decide to relax quarantine measures, even if the social isolation policy is still enacted). Through our simulations, we are able to replicate the real data with a fair amount of similarity, as shown in the sequel.

We must note that one could argue that the large amount of data from other countries could serve as a potential thread to quantify the level of uncertainty regarding the datasets available in Brazil. Nevertheless, this is not exactly true, since many social and local factors play a significant role in the spread of COVID-19. We point to the recent spread event in the city of Manaus, at the beginning of 2021, where the spread drastically increases despite local seroprevalence, which is unaccounted for elsewhere [44].

This paper is organized as follows. Section 2 presents the proposed SIR-adapted epidemiological model, which incorporates a new dynamic variable that describes the response of the population to enacted pandemic policies. In this section, we also discuss uncertainty modeling, in terms of the available datasets. Section 3 details the parameter estimation procedure, which inherently includes uncertainty. Section 4 shows the results in terms of parameter estimation and forecasts for the COVID-19 pandemic in Brazil. We present a thorough discussion on the achieved results and concluding remarks in Section 5.

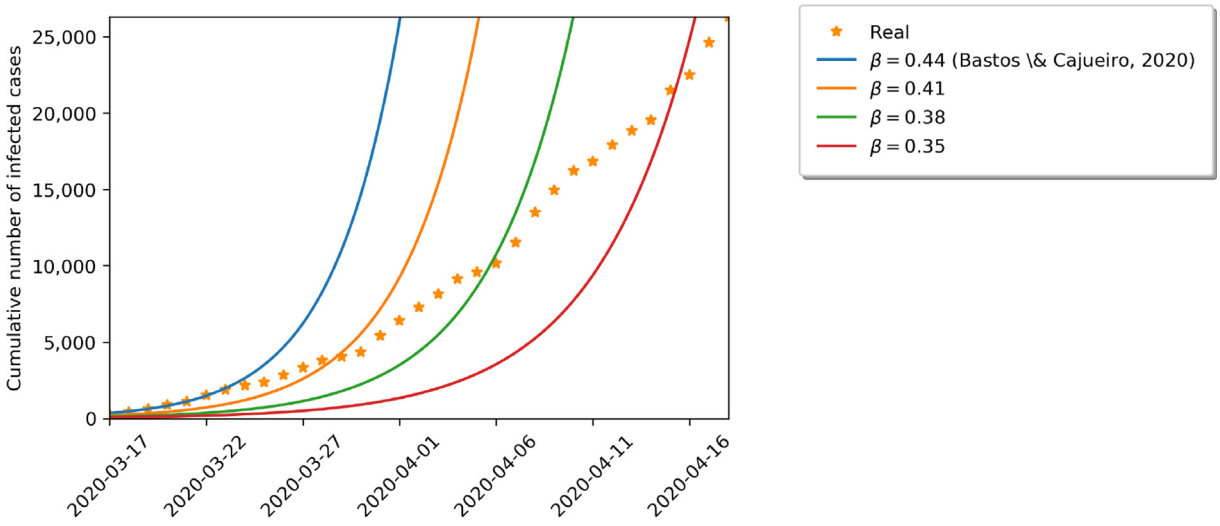

Fig. 1 Cumulative number of infected.

2. SIR epidemiological models

Recent literature [36,28] has demonstrated how the infection rate and evolution dynamics of the SARS-CoV-2 virus can be adequately described by Susceptible-Infected-Recovered kind models. In this Section, we present the classical SIR model due to [24], the new dynamic variable which models the population's response to isolation policies (enacted by local governments) and discuss some remarks on data uncertainty.

2.1. Epidemiological model

The SIR describes the spread of a given disease with respect to a population split into three non-intersecting classes, which stand for:

- The total amount of susceptible individuals, that are prone to contract the disease at a given moment of time t, denoted through the dynamic variable $S(t)$;
- The individuals that are currently infected with the disease (active infections at a given moment of time t), denoted through the dynamic variable $I(t)$;
- The total amount of recovered individuals, that have already recovered from the disease, from an initial time instant 0 until the current time t, denoted through the dynamic variable $R(t)$

Due to the evolution of the spread of the disease, the size of each of these classes change over time and the total population size N is the sum of these three classes, as follows:

$$N(t) = S(t) + I(t) + R(t). \tag{1}$$

In the SIR model, the parameter β stands for the average number of contacts that are sufficient for transmission of the virus from one individual, per unit of time t. Therefore, $\beta I(t)/N(t)$ determines the average number of contacts that are sufficient for transmission from infected individuals, per unit of time, to one susceptible individual; and $(\beta I(t)/N(t))S(t)$ determines the number of new cases per unit of time due to the amount of $S(t)$ susceptible individuals (they are "available for infection").

Furthermore, the parameter γ stands for the recovery rate, which is the rate that each infected individual recovers (or dies). This parameter characterizes the number of individuals that "leaves" the infected class, considering a constant probability quota per unit of time. We model the number of deceased individuals due to a SARS-CoV-2 infection following the lines of [22], where $D(t)$ is the dynamic variable that describes the number of deaths and ρ is the observed mortality rate.

The "**SIRD**" (Susceptible-Infected-Recovered-Dead) model is expressed as follows:

$$\begin{aligned} \frac{dS(t)}{dt} &= -(1-\psi(t))\frac{\beta I(t)S(t)}{N(t)} \\ \frac{dI(t)}{dt} &= (1-\psi(t))\frac{\beta I(t)S(t)}{N(t)} - \frac{\gamma I(t)}{1-\rho} \\ \frac{dR(t)}{dt} &= \gamma I(t) \\ \frac{dD(t)}{dt} &= \frac{\rho}{1-\rho}\gamma I(t) \end{aligned} \quad [\textbf{SIRD}] \tag{2}$$

In this model, ψ represents a transmission rate mitigation factor[6]: for $\psi = 0$, there is "no control" of the viral spread, while for $\psi = 1$, the contagion is completely controlled, with no more social interactions (a complete lockdown with no social mobility would represent this scenario, which is impracticable in reality). It holds that $N(t) = N_0 - D(t)$, where N_0 is the initial population size. Remark that, in SIR kind models, $I(t)$ represents the active infections at a given moment, while $D(t)$ represents the total amount of deaths until this given moment; for this reason, it follows that $\frac{dD(t)}{dt}$ is proportionally dependent to $I(t)$.Since the SIR model is used herein to

[6] We note, with respect to previous works [5], that the ψ parameter used in this work represents for $1-\psi$ in the prior. We believe the current representation is easier to grasp. Furthermore, in [32,31], this parameter is taken as a "control input" of the system, since model-based social distancing policies are synthesized. In this paper, we consider "open-loop" data, meaning that ψ is known or estimated, but no based on any feedback action.

describe a short-term pandemic outbreak, we do not consider the effects of demographic variations. Despite recent discussion regarding the possibilities of reinfection [11], we assume that a recovered individual does not contract the disease twice in the short period of time of this pandemic. Since, in the case of the SARS-CoV-2 virus, there is a relevant percentage of the infected individuals that are asymptomatic, we progress by splitting the class of infected individuals into the classes of symptomatic (I_S) and asymptomatic individuals (I_A), as suggested by [41,2,29]. This yields the "**SIRASD**" model[7]:

$$\begin{array}{rcl}
\frac{dS}{dt} & = & -(1-\psi)(\beta_A I_A(t) + \beta_S I_S(t))\frac{S}{N(t)} \\
\frac{dI_A(t)}{dt} & = & (1-\psi)(1-p)(\beta_A I_A(t) + \beta_S I_S(t))\frac{S}{N(t)} - \gamma_A I_A(t) \\
\frac{dI_S(t)}{dt} & = & (1-\psi)p(\beta_A I_A(t) + \beta_S I_S(t))\frac{S}{N(t)} - \gamma_S I_S(t) - \frac{\rho_S}{1-\rho_S}\gamma_S I_S(t) \\
\frac{dR_A(t)}{dt} & = & \gamma_A I_A(t) \\
\frac{dR_S(t)}{dt} & = & \gamma_S I_S(t) \\
\frac{dD(t)}{dt} & = & \frac{\rho_S}{1-\rho_S}\gamma_S I_S(t)
\end{array} \quad \text{[SIRASD]} \tag{3}$$

In this model, ρ_S denotes the actual observed mortality rate for the symptomatic class only, while ρ stands for the total observed mortality rate (for both symptomatic and asymptomatic classes).

Remark 1. There are alternative, more attractive description models to describe the COVID-19 contagion, rather than the SIRASD model. The "SIDARTHE" model, as proposed by [18], provides much more detailed dynamics.[8] Nevertheless, these "complex" models cannot be considered in the Brazilian setting, due to insufficient data. The country-wide data disclosed by the Ministry of Health only represents the total amount of infections and the total amount of deaths, per day. Since there is no pool sample testing in the country, there is a lack of data regarding detected asymptomatic individuals, for instance, as it is available in Italy, where the SIDARTHE model was conceived. If more complex models than SIRASD were considered in this work, the truthfulness/validity of the identification and forecast results could be largely over-corrupted, which would result in worse estimates for parameters and uncertainty assessments.

Two important issues should be highlighted: *i*) the mortality for asymptomatic individuals is extremely low; for simplicity, it is taken as null in the SIRASD model; and *ii*) the observed (real) mortality rates also includes all those individuals that died from the contagion, while not being accounted for in the available data sets. This is, if there is a number of sub-reported deaths, these should also influence the mortality rate parameters.The instantaneous values for the morality rates are evaluated as follows:

$$\rho_S(t) = \frac{D(t)}{I_S(t) + R_S(t) + D(t)} \tag{4}$$

$$\rho(t) = \frac{D(t)}{I_A(t) + R_A(t) + I_S(t) + R_S(t) + D(t)} \tag{5}$$

Therefore, when the COVID-19 spread ceases, it follows that $\rho_S = \lim_{t\to\infty}\rho_S(t) = D(\infty)/(R_S(\infty) + D(\infty))$ and $\rho = \lim_{t\to\infty}\rho(t) = D(\infty)/(R_A(\infty) + R_S(\infty) + D(\infty))$. Due to this fact, it holds that $\frac{\rho_S}{\rho} = \frac{1}{p}$, i.e. $\rho = p\,\rho_S$. The parameter p is included to represent the percentage of infected individuals who present symptoms; $(1-p)$ denotes the percentage of those without symptoms.[9] In order to take into account the effect of public health policies (that are enacted by local governments to "control" and mitigate the effects of the COVID-19 pandemic), such as social isolation, incentives to use of masks, etc., we include the dynamic equation for $\psi(t)$ to the SIRD and SIRASD models:

$$\frac{d\psi(t)}{dt} = \begin{cases} \alpha(\psi_\infty - \psi(t)) & \text{if under the pandemic policies effect,} \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Note that $\psi(t)$ converges to ψ_∞ with a settling time[10] of $\frac{3}{\alpha}$. It follows that ψ_∞ is a factor which stands for the maximal observed effect of pandemic policies. Note that for larger values of ψ_∞ (closer to 1), the spread of the SARS-CoV-2 virus gets slower, with smaller peak of infections and number of deaths. The condition of $\psi = 1$ represents a total isolation condition, where the amount of contacts is reduced to zero. In the recent papers by the Authors [31,32], these new models were used to conceive model-based optimal control policies, under a Predictive Control optimization formalism. Through the sequel, we refer to these extended models as SIRD+ψ and SIRASD+ψ models, respectively.

Remark 2. The time-varying social distancing parameter ψ is theoretically conceived within the interval [0, 1], being $\psi = 1$ the case of a total lockdown. Nevertheless, we stress that very high values of ψ (near 1, close to total lockdown with no social mobility) are not possible in practice. In Brazil, for instance, the maximal social distancing factor observed was near 50% (e.g. Fig. 2, in the sequel). We also note that the identification procedures regarding ψ_∞ (upper-bound to ψ) set this parameter freely within [0, 1], but the results coincide with practical observations, i.e. near 50%. As a result, we do not consider situations of complete lockdown. More importantly, we emphasize the time varying nature of the transmission rate, which opens opportunities for identification procedures or even the proposal of other epidemiological models for the COVID-19 dynamics.

2.2. Uncertainty

As discussed in the Introduction, the sub-notification regarding the number of infected and deceased individuals, due to COVID-19, has been pointed out as rather elevated in Brazil, as brought to focus by a number of recent papers [7,48,47,42,43,31]. These studies discuss that the amount of under-reported cases regarding infected individuals could be up to 30 times the reported values. Furthermore, the sub-

[7] SIRASD stands for Susceptible-Infected-Recovered-Asympto matic-Symptomatic-Deaths.

[8] This model splits the infections into (symptomatic, asymptomatic) detected, undetected, recovered, threatened and extinct classes.

[9] We note that $R_S(\infty) + D(\infty)$ result from all the symptomatic infections, while $R_A(\infty)$ results from the asymptomatic cases, since these do not lead to deaths.

[10] Considering that $\psi(t) = \psi_\infty(1 - e^{-\alpha t})$ and $\frac{\psi_\infty - \psi_\tau}{\psi_\infty} = e^{-\alpha\tau} = \delta \Rightarrow \tau = -\frac{\ln\delta}{\alpha}$, if $\delta = 0.05$ then $\tau \approx \frac{3}{\alpha}$.

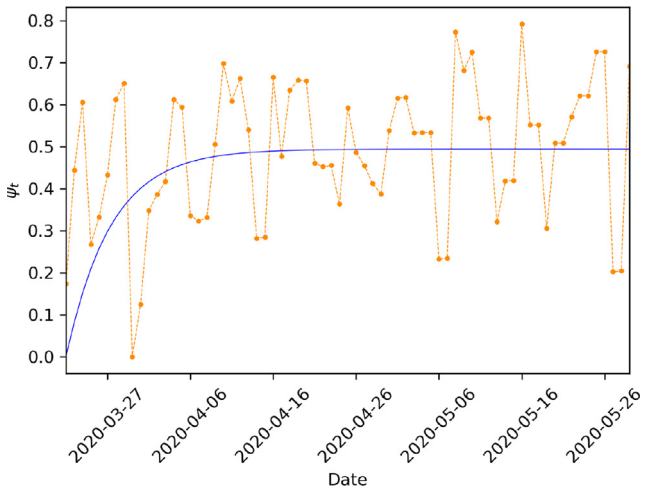

Fig. 2 Estimation of $\psi(t)$ for the SIRD+ψ model using the data provided by the Ministry of Health ($\tilde{q}_I = \tilde{q}_D = 0$), which we get $\alpha = 0.186353$ and $\psi_\infty = 0.494027$.

notification in terms of deaths due to the SARS-CoV-2 virus might be over 120 % of the mortality disclosed by the Ministry of Health.Therefore, we proceed by modeling these uncertainty margins as follows:

$$I_S^{\text{real}}(t) = I_S^{\text{measured}}(t) + I_S^{\text{unknown}}(t), \tag{7}$$

$$D^{\text{real}}(t) = D^{\text{measured}}(t) + D^{\text{unknown}}(t), \tag{8}$$

where the super-index "measured" denotes the values as prescribed by the Ministry of Health (data), "unknown" as the sub-notified values, and "real" for total value, including uncertainty.It follows that the amount of uncertainty in any of these variables, generically referred to as X, can be described by a concentrated multiplicative parameter, this is:

$$\left.\begin{array}{rcl} X^{\text{measured}}(t) & = & q_X X^{\text{real}}(t) \\ X^{\text{unknown}}(t) & = & (1-q_X)X^{\text{real}}(t) \end{array}\right\} \Rightarrow \frac{X^{\text{measured}}(t)}{X^{\text{unknown}}(t)} = \frac{q_X}{1-q_X}, \tag{9}$$

with X representing either D or I_S and q_X denoting the respective uncertainty parameter. Such concentrated parameter $q_j \in [0,1]$ gives a measure for the amount of sub-notification. For instance, if we consider $q_D = 1$, it means that there is no sub-notification with respect to the disclosed datasets for the number of deaths. For $q_D = 0.5$, we observe that the actual number of deaths is twice the reported amount.Using the same notation, through the sequel, we present our results using the following transformation, for understanding simplicity:

$$\tilde{q}_X = \frac{1}{q_X} - 1, \tag{10}$$

which holds for $X^{\text{unknown}}(t) = \tilde{q}_X X^{\text{measured}}(t) \Rightarrow X^{\text{real}}(t) = (1+\tilde{q}_X)X^{\text{measured}}(t)$. This notation concatenates the following idea: as an example, if we consider $\tilde{q}_I = 0$, it means that the reported amount of infections is equal to the real amount. Consequently, for $\tilde{q}_I = 0.25$, for instance, it means that the real amount of infections is 25% bigger than the number reported cases.With respect to the uncertainty modeling, we are concerned with the exposure of the effects simulation/prediction procedure when using uncertainty-embedded identification applied to SIR-like models in order to forecast the COVID-19 pandemic dynamics in Brazil. The discussion resides in the following key points:

(a) the influence of uncertainty in these forecasts;
(b) the temporal shift of the peak of infections according to the level of uncertainty (which would also impact directly in the enacted pandemic policies);
(c) the temporal aspect of the mortality rate;
(d) the under-reporting tripod.Furthermore, we must emphasize that the predictions for the number of deaths is very dependent on the level of uncertainty. This happens since the uncertainty parameters q_D and q_I directly impact on the dynamics of $D(t)$ in the SIRASD+ψ model. It seems reasonable that, as time progresses and the pandemic contagion gets "controlled" (i.e. stabilizes), the amount of uncertainty tends to decrease. Anyhow, we are concerned with forecasting phenomena that will irrevocably take place (the peak of infections and a possible second peak), if the enacted pandemic policies remain the same.

3. Estimation procedure

The Brazilian Ministry of Health provides (daily) two useful data time-series which are used to evaluate (i.e. identify) the SIR model parameters: *i*) the total number of infected individuals, denoted

$$I^{\text{total}}(t) = I(t) + R(t) + D(t)$$

and *ii*) the total number of deaths, which is $D(t)$. In this paper, we use data from February 25, 2020, to May 31, 2020.

Parameters β, γ, and ρ are computed according to the procedure adapted from [5], considering that the period for which pandemic policies had not yet been formally implemented (this period comprises the weeks from February 25, 2020, until March 22, 2020). We find that the recovery rate γ is found to be roughly constant at 0.150876, which is consistent with other findings in the literature [40,33]. We maintain this value for γ throughout the following identification steps.

In order to estimate the parameters of the SIRD+ψ and SIRASD+ψ models, considering the remainder of the available data (from March 22, 202 onward), we proceed by performing a Least-Squares identification procedure, with respect to different levels of uncertainty ($\tilde{q}_D$ and $\tilde{q}_I$). These uncertainty levels are directly embedded in the datasets prior to the actual identification. Minimizing the square-error between the integrated variables and their real values is in accordance with regular identification methodologies [3,8]. Then, we proceed by following a hierarchical procedure, as done in [5]: firstly, we estimate parameters of the SIRD model, then, assuming that the infected individuals in the SIRD model are the symptomatic individuals in the SIRASD model, we estimate the remaining parameters of the SIRASD model.

Our identification procedure also includes constraints[11] to the possible parameter values: $\beta, \beta_S, \beta_A \in [1/20, 2]$, $\gamma, \gamma_S, \gamma_A \in [1/14, 1/2]$, $\rho, \rho_S \in [0.001, 0.2]$, $\alpha \in [0,1]$ and $\psi_\infty \in [0.0, 1.0]$. These limits are in accordance with those presented by [51,5].

Globally, the identification procedure for the SIRD+ψ model resides in solving the following minimization problem:

[11] If more restrictive bounds are used, we mention them explicitly.

$$\min_{\beta,\rho,\alpha,\psi_\infty} \frac{1}{2}\sum_t \left(\mathscr{L}_I^2 + \mathscr{L}_D^2\right) \tag{11}$$

$$\mathscr{L}_I = \log\left[1 + \left(\frac{I^{\text{total}}(t) - D(t)/q_D}{q_I}\right)\frac{1}{I_{\max}}\right] - \log\left(1 + \frac{\widehat{I}(t) + \widehat{R}(t)}{I_{\max}}\right) \tag{12}$$

$$\mathscr{L}_D = \log\left[1 + \left(\frac{D(t)}{q_D}\right)\frac{1}{D_{\max}}\right] - \log\left(1 + \frac{\widehat{D}(t)}{D_{\max}}\right) \tag{13}$$

where $\widehat{I}(t)$, $\widehat{R}(t)$ and $\widehat{D}(t)$ are estimated values of the infected, recovered and deaths, respectively, and $I_{\max} = \max_t\left(\frac{I^{\text{total}}(t) - D(t)/q_D}{q_I}\right)$, $D_{\max} = \max_t\left(\frac{D(t)}{q_D}\right)$. We decided to normalize the each series by its maximum value and take the logarithm as an approach to balance different exponential characteristics of the infected and deaths series. The initial conditions for the identification procedures are $D_0 = D(0)/q_D$, $I(0) = I^{\text{total}}(0) - D(0)$, $R(0) = 0, S(0) = N_0 - I(0) - R(0) - D(0)$, where $N_0 = 210147125$ is the initial population according to Brazilian Institute of Geography and Statistics (IBGE). We estimate α and ψ_∞ only in the simulation without any uncertainty, and use these values on the other minimization procedures because we believe that these parameters (a response of the society) should be the same, regardless of the uncertainty level.

We stress that the nonlinear map $\mathscr{L}_a$ is increasing in its domain and normalized within $[0, 1]$. This function is used to further weight the last values of the identification data series w.r.t. to the first data steps. This logarithmic correction is implied to compensate for the exponential characteristics of the COVID-19 spread. Moreover, the cost formulations in Eqs. (12) and (13) are coherent with regular Least-Square identification procedures, minimizing the squared difference between model and data. It is implied that $I^{\text{total}}(t) - D(t) = I(t) + R(t)$.

Equivalently, the identification procedure for the $SIRASD + \psi$ follows through the minimization problem below:

$$\min_{\beta_A,\gamma_A,\rho} \frac{1}{2}\sum_t \left(\mathscr{L}_{I_S}^2 + \mathscr{L}_D^2\right) \tag{14}$$

$$\mathscr{L}_{I_S} = \log\left[1 + \left(\frac{I^{\text{total}}(t) - D(t)/q_D}{q_I}\right)\frac{1}{I_{\max}}\right] - \log\left(1 + \frac{\widehat{I}_S(t) + \widehat{R}_S(t)}{I_{\max}}\right) \tag{15}$$

where $\widehat{I}_S(t)$ and $\widehat{R}_S(t)$ are estimated values of the symptomatic infected and recovered, respectively, and the initial conditions are $D(0) = D(0)/q_D, I_S(0) = I^{\text{total}}(0) - D(0), I_A(0) = I_S(0)(1 - q_I)/q_I, R_S(0) = R_A(0) = 0, S(0) = N_0 - D(0) - I_S(0) - I_A(0) - R_S(0) - R_A(0)$. For this model, we begin the identification procedure with $\beta_S = \beta, \gamma_S = \gamma$, where β and γ are the values directly taken from the identification procedure regarding the $SIRD + \psi$. Hence, we denote the procedure as hierarchical.

We stress, once again, that a large number of different SIRD$+\psi$ and SIRASD$+\psi$ models are obtained through the identification procedures detailed in the prequel. Each one of these models is identified for different levels of uncertainty (q_D and q_I).

4. Results

In this Section, we present the main results of our paper, which comprise the role that sub-notification uncertainty plays in the model-based predictions of the COVID-19 contagion, harshly affecting the outlooks for its evolution spread in Brazil. To have the available data with such large amounts of uncertainty is a very troublesome issue, since public health policies are currently derived through the available datasets, meaning that these policies may be equivocating, such as reducing social isolation policies before adequate time, and thus lead to possibly unwanted phenomena.

The following results comprise predicted forecasts with the SIRD$+\psi$ and SIRASD$+\psi$ models; they have been obtained using Python software.

4.1. $SIRD + \psi$ model

We begin by depicting the results achieved with the SIRD$+\psi$ model.

Fig. 2 shows the estimation of the $\psi(t)$ variable, which models the population's response to pandemic policies using Eq. (6), considering no uncertainty ($q_I = q_D = 1$, only for the moment). In this Figure, the orange dots represent the identification results found directly through the *SIRD* model, as if ψ was a parameter estimated for each sample. The solid blue line represents the time-varying $\psi(t)$ curve. Evidently, we can notice that the estimation of the proposed $\psi(t)$ model absorbs a great deal of error. Consequently, the forecasts derived from such models (SIRD$+\psi$, SIRASD$+\psi$) can vary quite significantly from one day to the next (considering the identification procedure performed daily), while the main trend is followed. Some recent studies [31,27] discuss that the re-estimation (re-identification) should be performed each week, in such a way that these daily errors get dynamically absorbed and are thus represented through the mean behavior. We must re-stress that the following forecasts are extremely sensitive to the available datasets (and to the uncertainties) and, thus, to re-perform the identification as time progresses is essential. This is: one cannot use the derived models as if their parameters would not change over time. The forecasts only offer qualitative insights.

We note that the social distancing parameter ψ is embedded in the first-order dynamics of Eq. (6). Consequently, the identification procedure is performed to estimate α and ψ_∞ in order to best match the "average" tendency of social distancing measures, as presented in Fig. 2. In previous works (see [31,32], we consider that ψ is analytically related to a control input which models the social distancing measures.

We consider several different values for the uncertainty parameters $\tilde{q}_I$ and $\tilde{q}_D$ in the following simulation runs.[12] Anyhow, for exhibition lightness purposes, we will only explicitly explore the ones that found to be more likely scenarios for Brazil (coherent with prior discussion in the literature). Our baseline uncertainty parameters are $\tilde{q}_I = \tilde{q}_D = 0$ (nominal condition), corresponding to the Ministry of Health datasets. We explore $\tilde{q}_I \in \{7, 14, 30\}$ and $\tilde{q}_D \in \{0.25, 0.5, 1\}$, in consonance with previous references. These uncertainty margins relate to the percentage increase regarding infections and

[12] To ease the Reader's interpretation, we transformed q_I and q_D in $\tilde{q}_I$ and $\tilde{q}_D$, respectively, by the use of the transformation in Eq. (10).

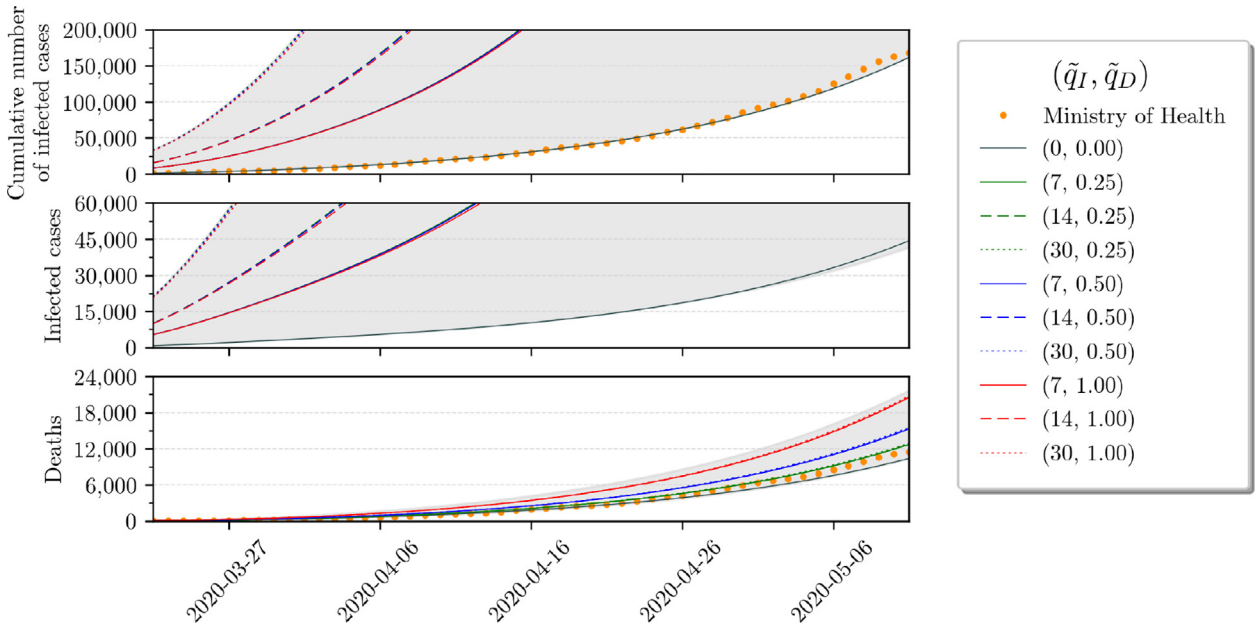

Fig. 3 Short-term simulation for the SIRD+ψ model using different values of uncertainty.

deaths, respectively. The gray area presented in the short and long term predictions of Figs. 3 and 4 correspond to uncertainty over the region $\{(\tilde{q}_I, \tilde{q}_D) | \tilde{q}_I \in [0, 30]$ and $\tilde{q}_D \in [0, 1]\}$.

Based on these uncertainty values, we simulate the SIRD+ψ model along time. We show short and long-term perspectives in Figs. 3 and 4. Detailed values that arise in these simulations are concatenated in Table 2. With respect to these Figures, we remark the following key points:

1. The long-term forecast of the infected cases suggests that greater uncertainty of the number of infected has the effect of anticipating the peak in time and also increase its amplitude. That is if we have 7, 14, and 30 times more infected, then the peak on September 4 will be anticipated to July 28, 17, and 4, respectively, and also the peak amplitude will increase from 3.9% to approximately 4.6%, 4.7%, and 4.8%, respectively. We note that this peak percentage is given w.r.t. to the total population size.

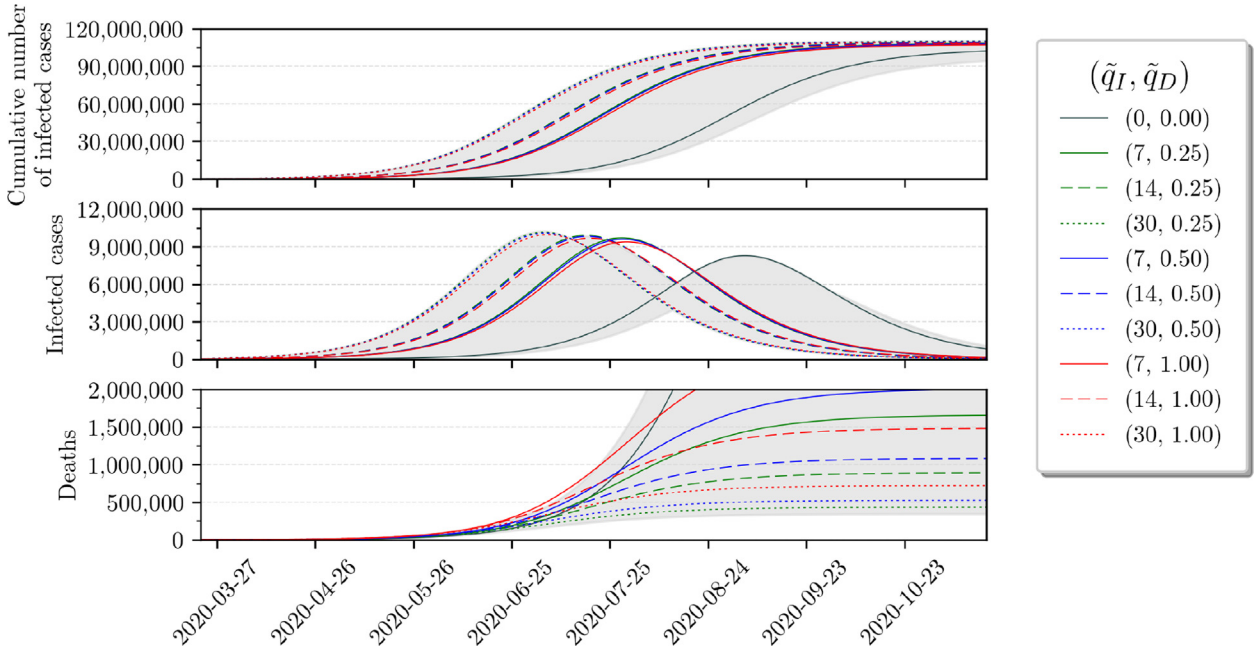

Fig. 4 Long-term simulation for the SIRD+ψ model using different values of uncertainty.

Table 2 Observed mortality, peak percentage and its respective date of occurrence and number of deaths after 360 days for the SIRD+ψ model for different values of uncertainty.

$\tilde{q}_D$	$\tilde{q}_I$	ρ	Peak (%)	Peak Forecast	Number of Deaths
0	0	8.8569%	3.9405%	September 4, 2020	9,226,000
0	7	1.2086%	4.6651%	July 28, 2020	1,316,778
0	14	0.6447%	4.7508%	July 17, 2020	706,295
0	30	0.3090%	4.8573%	July 4, 2020	341,003
0.5	7	1.8597%	4.5681%	July 29, 2020	2,012,722
0.5	14	0.9958%	4.6793%	July 18, 2020	1,085,310
0.5	30	0.4785%	4.7992%	July 5, 2020	525,776
1	7	2.5450%	4.4673%	July 30, 2020	2,735,068
1	14	1.3683%	4.6049%	July 18, 2020	1,483,225
1	30	0.6594%	4.7390%	July 5, 2020	721,160

2. The peak of infections gets shifted for at least **one month** for larger uncertainties. This could be quite troublesome, since public health policies concerning ICU beds, for instance, maybe accounting for erroneous data and hospitals may be surprised by a larger demand of ICUs than what is expected when disregarding uncertainties. Further analysis is necessary since we cannot separate who are asymptomatic from the ones that are symptomatic and, therefore, would actually demand health care.
3. The amplitude of the peak of infections shows small variations, despite uncertainties. This suggests that the SARS-CoV-2 will infect the same amount of people. Moreover, the virus does not distinguish between asymptomatic or symptomatic individuals, only aiming to spread its genetic material.
4. The mortality rate obtained from the Brazilian Ministry of Health datasets does not reflect reality because it differs significantly from the true mortality rate (see Table 1). The long-term deaths forecast suggests that the number of infected uncertainty drastically decreases the true mortality rate, while increases in the death uncertainty increase the true mortality rate.

We proceed by exploring how the infected uncertainty influences the mortality rate for a given death uncertainty level, as shows Fig. 5. In this Figure, we present the estimated values for the mortality rate (ρ) from Eq. (2) as a function of the infected uncertainty ($\tilde{q}_I$) for a given (fixed) death uncertainty ($\tilde{q}_D$) amount. The values used in this graph and some estimated epidemiological parameters are in Tables 4 and 6 in Appendix A.1. We must stress that the observed mortality rate decreases with respect to the increase of infected uncertainty $\tilde{q}_I$; this is shown directly in Fig. 5. The reason for this phenomenon resides in the fact that more uncertainty over the infected individuals while keeping the number of deaths constant, implies that the mortality rate will decrease see Eq. (5).

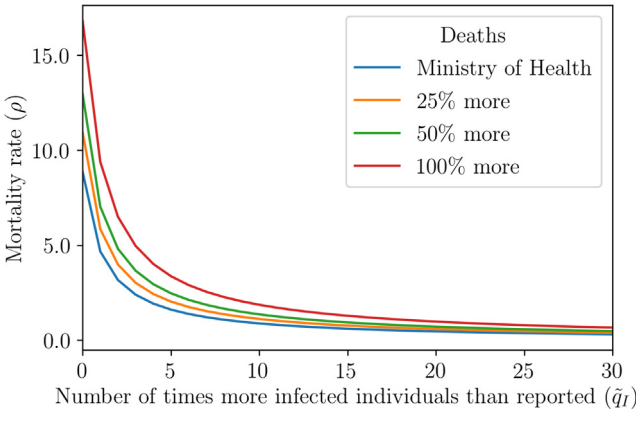

Fig. 5 The number of times more infected than reported ($\tilde{q}_I$) vs. the observed mortality rate (ρ) for the SIRD+ψ model. The blue, orange and green curves represent simulations using data from the Ministry of Health ($\tilde{q}_D = 0$), 50% more deaths ($\tilde{q}_D = 0.5$) and 100% more deaths ($\tilde{q}_D = 1$).

Fig. 6 provides the uncertainty paths for infected and deaths considering specific mortality rates.

4.2. SIRASD+ψ model

With respect to the uncertainty-embedded identification procedure in Section 3, Figs. 7 and 8 show, respectively, short and long term predictions using the SIRASD+ψ model. Table 3 collects the essential information of these forecasts. Regarding the long-term forecast, we call attention to the infected cases. An increase in the infected uncertainty anticipates the peak in time, just as in the SIRD+ψ model, but the symptomatic peak amplitude decreases.

With respect to the simulated mortality rate, Fig. 9 shows the relationship between ρ and the uncertainties q_D and q_I; these values are detailed in Table 5 of Appendix A.1. Results differ only slightly from those obtained with the SIRD+ψ model and, thus, the same conclusions can be inferred. Furthermore, Fig. 10 gives an insight on the possible "trajectories" of ρ with respect to these uncertainty margins, i.e. showing the static gain between q_D, q_I and ρ, for different levels of mortality.

The identified epidemiological parameters for the SIRASD+ψ model are presented in Table 7, in the Appendix. There are two interesting phenomena regarding β_A: (i) in all simulations we have $\beta_A > \beta_S$, meaning that asymptomatic individuals are more likely to transmit the disease since they probably do not know they are infected; and (ii) increasing the infected uncertainty causes a decrease in the parameter β_A, meaning that the probability to transmit the disease also decreases. As for γ_A, we notice that: (i) it is greater than γ_S in all cases, meaning that asymptomatic individuals have a smaller infectious period ($1/\gamma_A$); and (ii) it decreases as the infected uncertainty increases, meaning that the average infectious period increases with the increases of asymptomatic individuals, i.e. it is more likely to find more asymptomatic individuals spreading the disease for a longer period.

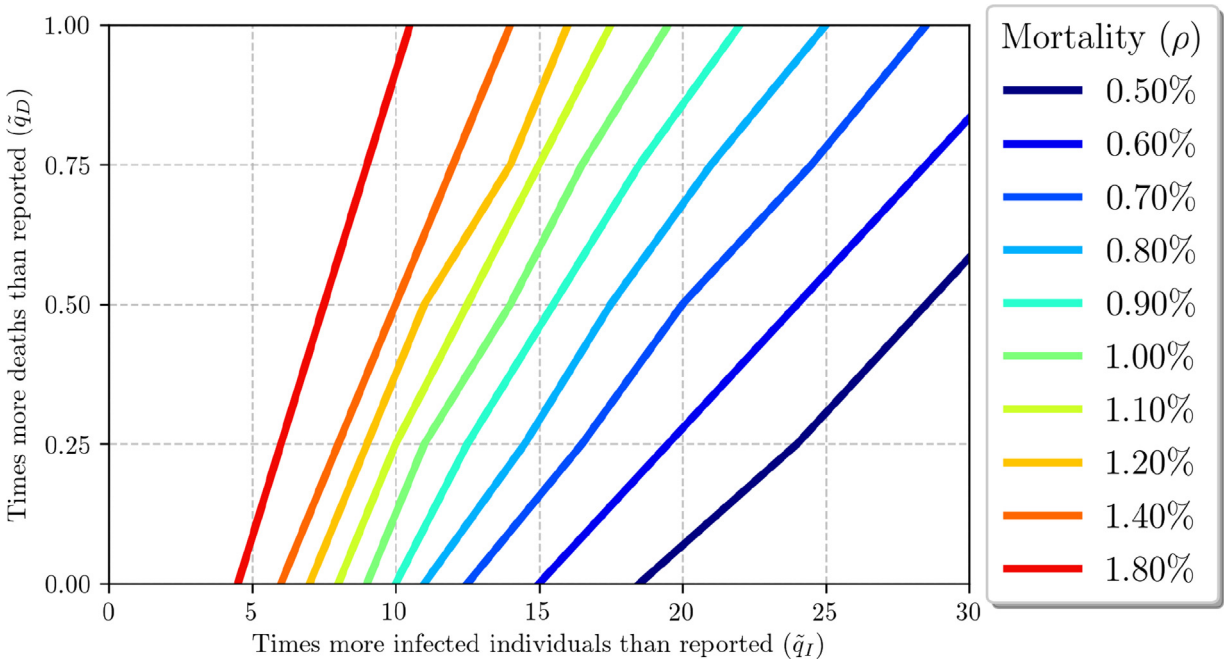

Fig. 6 Mortality rate paths using the number of times more infected than reported ($\tilde{q}_I$) versus the number of times more deaths than reported ($\tilde{q}_D$) for the SIRD+ψ model.

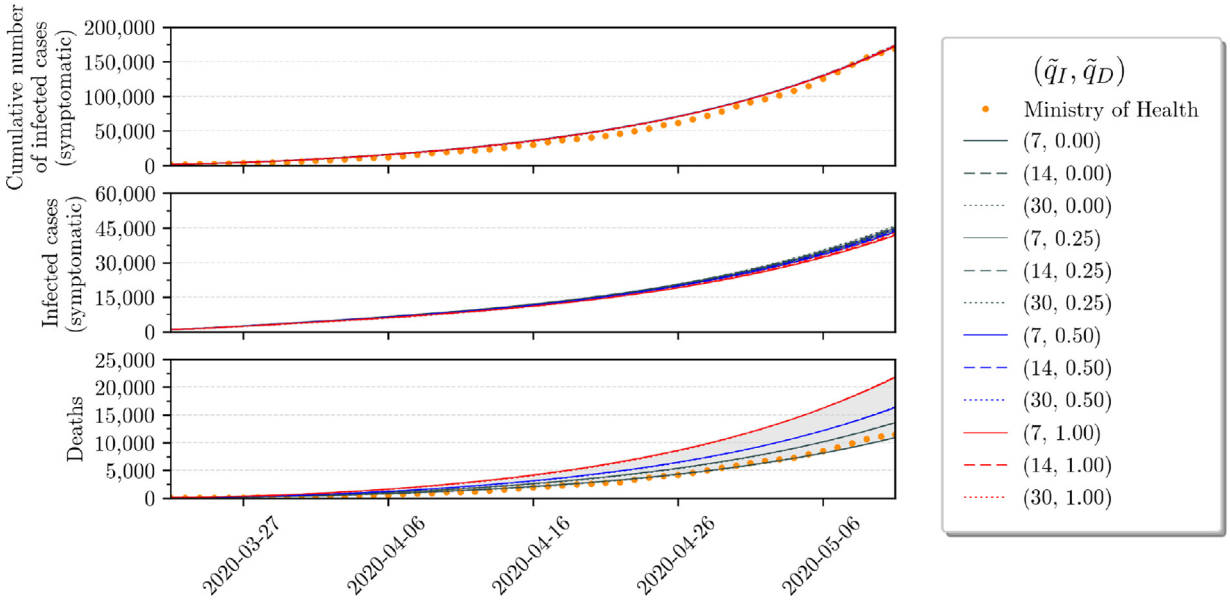

Fig. 7 Short-term simulation for the SIRASD+ψ model using different values of uncertainty.

We must point out that analogous results to those achieved with the SIRD+ψ model are found, as expected due to the equivalence between the nonlinear differential equations of these two models. We stress some key points:

- An increase in the sub-notification with respect to infected individuals directly implies in the reduction of the peak of symptomatic infections: essentially, the virus "does not care" if it causes symptoms or not on the infected individual (its only goal is to infect and replicate);
- Assuming that the virus infects the same amount of people (despite the margins of the sub-reports), and as it infects (faster) more individuals without causing symptoms, we observe a smaller percentage peak of symptomatic infections for larger sub-notification.
- The uncertainty margins upon the infected (q_I) and the death count (q_D) influence both infected and death curves, $I(t)$ and $D(t)$, respectively. The infected curve is directly and mostly influenced by the infected uncertainty, but the death uncertainty affects this curve more modestly through

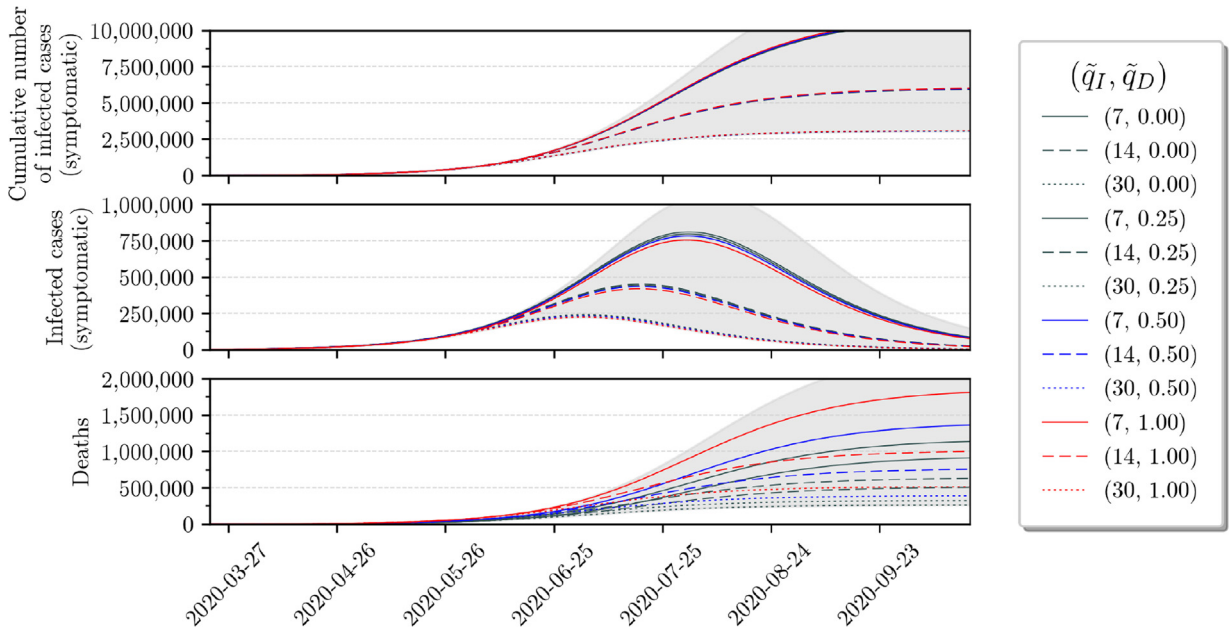

Fig. 8 Long-term simulation for the SIRASD+ψ model using different values of uncertainty.

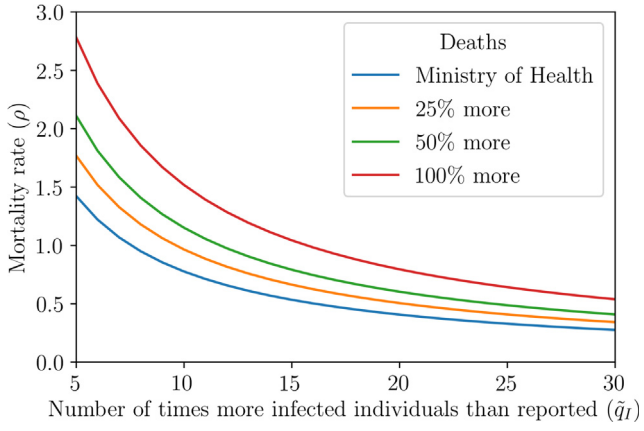

Fig. 9 The number of times more infected than reported ($\tilde{q}_I$) vs. the observed mortality rate (ρ) for the SIRD+ψ model. The blue, orange and green curves represent simulations using data from the Ministry of Health ($\tilde{q}_D = 0$), 50% more deaths ($\tilde{q}_D = 0.5$) and 100% more deaths ($\tilde{q}_D = 1$).

the total number of the population, $N(t)$, on the transmission channel (see Eq. (2) and Fig. 4). Additionally, we emphasize that the death curve also depends on both uncertainty parameters, q_D and q_I. The smallest death count forecast is found with an elevated under-reporting of asymptomatic infections and small sub-report margins regarding deaths. This fact seems empirically reasonable, once an unaccounted increase in the number of deceased individuals should increase the number of deaths.

- The instantaneous mortality rates, as of Eqs. (4) and (5), are show in Fig. 11. These instantaneous values indeed converge to those presented in Table 3. Anyhow, it is imperious to recall that the mortality rate, in practice, varies according to the amount of testing. Since the dynamics for symptomatic and asymptomatic infections evolve differently over time, the mortality rate also depends on the stage of the pandemic evolution. If the local epidemic scenario is an ending stage, the mortality rate tends to increase and stabilize. Note that if the margins of uncertainty are known (or roughly estimated), we can forecast quite accurately what will be the observed rates of mortality in the country.
- The real mortality rate, measured with population samples, can be quite misleading since it shows only an instantaneous snippet of the pandemic at a given moment (much like a "photograph"). Fig. 11 shows that the real mortality rate evolves in an asymptotic-like behavior, converging to some steady-state value. Therefore, if one computes the mortality rate of a sample population in a country with mass testing (unlike Brazil), and this country is not yet in an ending stage of its COVID-19 epidemic, one can observe values that are not the steady-state ones. In other words, there would still be symptomatic and asymptomatic to-be individuals which would alter the real mortality rate. And, as shows Fig. 11, this real rate tends to be greater than in the beginning stages of the SARS-CoV-2 spread.

5. Discussion and conclusions

In this paper, we discussed the Brazilian COVID-19 pandemic scenario, the effects of under-reporting regarding the number of infections and deaths due to the SARS-CoV-2 virus in the country. Through two SIR-like adapted models, which include the population's response to measures of control of the pandemic (such as social isolation, the use of masks, etc.) and deaths due to the disease. Furthermore, we make a set of possible forecasts and evaluate how uncertainty meddles with the COVID-19 pandemic prediction curves. Specifically, we analyzed how uncertainty affects the infection peak displacement

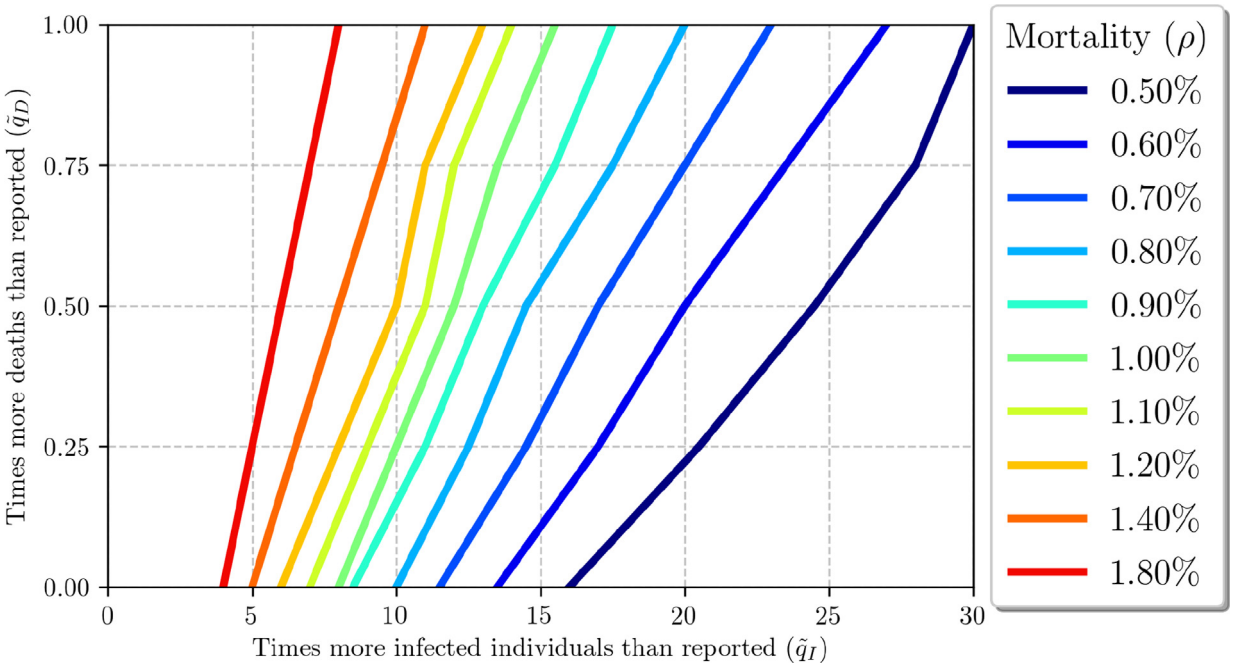

Fig. 10 Mortality rate paths using the number of times more infected than reported ($\tilde{q}_I$) versus the number of times more deaths than reported ($\tilde{q}_D$) for the SIRASD+ψ model.

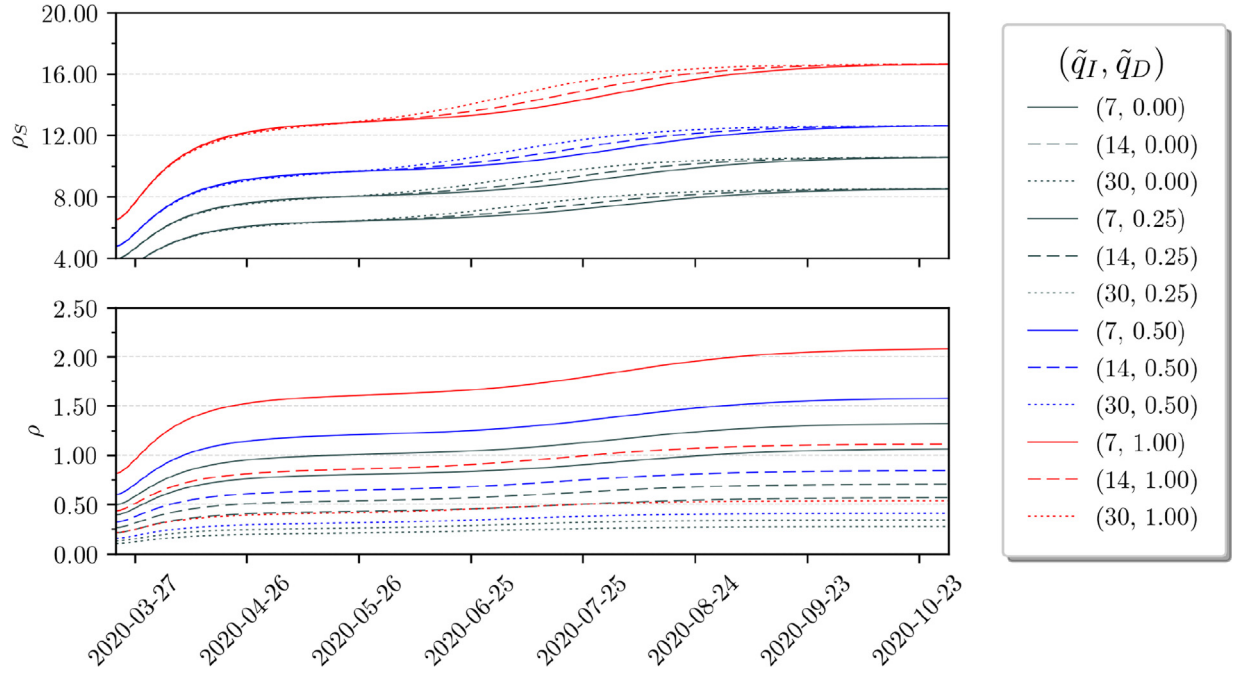

Fig. 11 Symptomatic and overall mortality rate over time for the SIRASD+ψ model.

in time and its amplitude, epidemiological parameters, the observed and real mortality rate evolution, and how infection and death uncertainty and the true mortality rate could be related.

We used two SIR-like models, SIRD+ψ and SIRASD+ψ, which differ in two aspects. First, the SIRD+ψ model considers that asymptomatic and symptomatic parameters are equal, that is, $\beta_S = \beta_A = \beta$ and $\gamma_S = \gamma_A = \gamma$ and that the mortality rate is applied in the whole infected population, while the SIRASD+ψ model distinguishes symptomatic and asymptomatic parameters and applies the mortality rate to the symptomatic only. Second, the initial condition of the SIRD+ψ model suggests that the infected class of individuals is actually split as $100q_I\%$ symptomatic and $100(1-q_I)\%$ asymptomatic,

Table 3 Observed mortality (ρ_S and ρ), peak percentage and its respective date of occurrence and number of deaths after 360 days for the SIRD+ψ model for different values of uncertainty.

				Both		Symptomatic		Asymptomatic		
$\tilde{q}_D$	$\tilde{q}_I$	ρ_S	ρ	Peak (%)	Peak Forecast	Peak (%)	Peak Forecast	Peak (%)	Peak Forecast	Deaths
0	7	8.5408%	1.0676%	3.0667%	2020–07-31	0.3856%	2020-08-01	2.4267%	2020-07-31	939,784
0	14	8.5335%	0.5689%	4.2421%	2020-07-18	0.2149%	2020-07-18	2.7655%	2020-07-18	512,176
0	30	8.5208%	0.2749%	4.7137%	2020-07-03	0.1155%	2020-07-03	3.3167%	2020-07-03	262,242
0.25	7	10.6162%	1.3270%	3.0667%	2020-07-31	0.3791%	2020-08-01	2.4505%	2020-07-31	1,172,135
0.25	14	10.6073%	0.7072%	4.2421%	2020-07-18	0.2111%	2020-07-18	2.7858%	2020-07-18	638,423
0.25	30	10.5882%	0.3416%	4.7137%	2020-07-03	0.1130%	2020-07-03	3.3205%	2020-07-03	325,994
0.5	7	12.6686%	1.5836%	3.0761%	2020-07-31	0.3726%	2020-07-31	2.4751%	2020-07-31	1,403,689
0.5	14	12.6580%	0.8439%	4.2421%	2020-07-18	0.2073%	2020-07-18	2.8063%	2020-07-18	764,006
0.5	30	12.6358%	0.4076%	4.7117%	2020-07-03	0.1109%	2020-07-03	3.3414%	2020-07-03	390,023
1	7	16.7050%	2.0881%	3.0761%	2020-07-31	0.3596%	2020-07-31	2.5249%	2020-07-31	1,864,283
1	14	16.6919%	1.1128%	4.2494%	2020-07-18	0.1998%	2020-07-18	2.8485%	2020-07-18	1,013,367
1	30	16.6639%	0.5375%	4.7117%	2020-07-03	0.1068%	2020-07-03	3.3830%	2020-07-03	516,959

Table 4 Observed mortality rates (ρ) for different simulations of the SIRD+ψ model under infected ($\tilde{q}_I$) and death ($\tilde{q}_D$) uncertainty.

Times more infected ($\tilde{q}_I$)	Mortality rate (ρ)			
	$\tilde{q}_D = 0$	$\tilde{q}_D = 0.25$	$\tilde{q}_D = 0.5$	$\tilde{q}_D = 1$
0	8.8569%	10.9324%	12.9481%	16.7947%
1	4.6676%	5.8437%	7.0219%	9.3818%
2	3.1651%	3.9819%	4.8088%	6.4925%
3	2.3932%	3.0183%	3.6543%	4.9602%
4	1.9233%	2.4292%	2.9457%	4.0113%
5	1.6070%	2.0318%	2.4663%	3.3658%
6	1.3798%	1.7458%	2.1207%	2.8988%
7	1.2086%	1.5300%	1.8597%	2.5450%
8	1.0750%	1.3614%	1.6555%	2.2677%
9	0.9678%	1.2261%	1.4915%	2.0446%
10	0.8799%	1.1151%	1.3568%	1.8612%
11	0.8066%	1.0224%	1.2444%	1.7079%
12	0.7444%	0.9438%	1.1490%	1.5777%
13	0.6911%	0.8764%	1.0671%	1.4658%
14	0.6447%	0.8177%	0.9958%	1.3683%
15	0.6042%	0.7664%	0.9335%	1.2830%
16	0.5684%	0.7211%	0.8784%	1.2077%
17	0.5365%	0.6807%	0.8293%	1.1405%
18	0.5080%	0.6446%	0.7854%	1.0803%
19	0.4824%	0.6121%	0.7459%	1.0263%
20	0.4591%	0.5826%	0.7100%	0.9771%
21	0.4380%	0.5559%	0.6775%	0.9325%
22	0.4187%	0.5315%	0.6478%	0.8917%
23	0.4010%	0.5090%	0.6204%	0.8542%
24	0.3847%	0.4884%	0.5953%	0.8198%
25	0.3697%	0.4693%	0.5721%	0.7879%
26	0.3557%	0.4517%	0.5507%	0.7585%
27	0.3428%	0.4353%	0.5307%	0.7311%
28	0.3307%	0.4200%	0.5121%	0.7055%
29	0.3195%	0.4057%	0.4948%	0.6817%
30	0.3090%	0.3924%	0.4785%	0.6594%

Table 5 Observed mortality rates ($\rho = p\rho_S$) for different simulations of the SIRASD+ψ model under infected ($\tilde{q}_I$) and death ($\tilde{q}_D$) uncertainty.

Times more infected ($\tilde{q}_I$)	Mortality rate (ρ)			
	$\tilde{q}_D = 0$	$\tilde{q}_D = 0.25$	$\tilde{q}_D = 0.5$	$\tilde{q}_D = 1$
0.1	7.9411%	9.8027%	11.6102%	15.2003%
1	4.2743%	5.3126%	6.3393%	8.3585%
2	2.8489%	3.5411%	4.2255%	5.5716%
3	2.1364%	2.6554%	3.1687%	4.1782%
4	1.7089%	2.1241%	2.5346%	3.3422%
5	1.4239%	1.7698%	2.1119%	2.7848%
6	1.2203%	1.5168%	1.8100%	2.3867%
7	1.0676%	1.3270%	1.5836%	2.0881%
8	0.9489%	1.1794%	1.4074%	1.8559%
9	0.8539%	1.0614%	1.2665%	1.6701%
10	0.7761%	0.9647%	1.1513%	1.5181%
11	0.7114%	0.8842%	1.0552%	1.3915%
12	0.6566%	0.8161%	0.9739%	1.2843%
13	0.6096%	0.7578%	0.9043%	1.1924%
14	0.5689%	0.7072%	0.8439%	1.1128%
15	0.5333%	0.6629%	0.7910%	1.0431%
16	0.5018%	0.6238%	0.7444%	0.9817%
17	0.4739%	0.5891%	0.7030%	0.9270%
18	0.4489%	0.5580%	0.6659%	0.8781%
19	0.4264%	0.5300%	0.6325%	0.8341%
20	0.4061%	0.5048%	0.6024%	0.7943%
21	0.3876%	0.4818%	0.5749%	0.7581%
22	0.3707%	0.4608%	0.5499%	0.7251%
23	0.3552%	0.4415%	0.5269%	0.6948%
24	0.3409%	0.4238%	0.5058%	0.6670%
25	0.3278%	0.4075%	0.4862%	0.6412%
26	0.3156%	0.3923%	0.4682%	0.6174%
27	0.3043%	0.3783%	0.4514%	0.5953%
28	0.2938%	0.3652%	0.4358%	0.5747%
29	0.2840%	0.3530%	0.4212%	0.5555%
30	0.2749%	0.3416%	0.4076%	0.5375%

while the initial condition of the SIRASD+ψ model considers that the proportion of one symptomatic per $\frac{(1-q_I)}{q_I}$ asymptomatic individual (as gives Eq. (9)).

It is a fact (and heavily discussed by recent literature) that the available datasets disclosed by the Brazilian Ministry of Health have large sub-notification margins. Brazil currently conducts roughly 4000 tests per million inhabitants, which is one of the lowest rates in the world [50]. Likewise, the number of deaths may be underestimated for the same reason. According to [10], only 8% of infections are reported; the real number of infected individuals is possibly up to 30 times bigger than what is being disclosed by the authorities [46,50].

Herein, we have tried to expose some essential insights regarding sub-notification and how to incorporate them into pandemic models; below, we summarize the main findings of this paper, enlightening the key points:

- Since the spread of the SARS-CoV-2 virus is inherently complex and varies according to multiple factors (some of which are possibly unmodelled and external), exact forecasts of the pandemic dynamics are not viable. Therefore, the correct procedure should be based on a recurrent model tuning (via identification), always taking into account the uncertainty margins. This measure would allow presenting more coherent forecasts as time goes since the uncertainty margins tend to decrease as the pandemic ceases (and as more testing is done). We remark that, in this paper, the uncertainty margins are assumed constant along the forecast horizons in order to tune/estimate the model parameters. In future works, the Authors plan in exploring the possible differences obtained when implying a dynamic behavior for the uncertainty margins (asymptotic and decreasing).
- The simulation forecasts indicate that the amount of uncertainty influences directly on the date of the infection peak, on the number of deaths, and on the mortality rate of the decease.Higher levels of infection uncertainty anticipate the infected peak in time. Considering no death uncertainty, the symptomatic peak at July 31, 2020, for $\tilde{q}_I = 7$ is anticipated to July 3, 2020, for $\tilde{q}_I = 30$ (see Table 3), almost a month difference. This is followed by the corresponding decrease (increase) in the peak amplitude of symptomatic (asymptomatic) individuals.Higher levels of death uncertainty cause increase in the observed and true mortality rates. Considering a fixed amount of infected uncertainty (q_I or $\tilde{q}_I$ is constant), the number of deaths and the mortality rate are higher if the death uncertainty is higher (see Figs. 8 and 11). Also, considering a fixed amount of death uncertainty (q_D or $\tilde{q}_D$ is constant), the number of deaths and the mortality rate are lower if the infected uncertainty is higher.Furthermore, there is a direct relationship between the uncertainty level and the observed mortality rates, which are in fact time-varying (Fig. 11). The instant mortality rates are calculated using the number of deaths and the cumulative number of infected, while the true mortality rate is its asymptotic value. So depending on which stage of the epidemic we are at, estimations using population samples could vary, not only by the method itself but also due to the natural evolution of the proportions of infected and deaths.
- Higher uncertainty levels of asymptomatic individuals cause a decrease in the epidemiological parameters β_A and γ_A (Table 7). There are two possible interpretations for the transmission parameter: (i) according to [23], asymptomatic individuals presumably have more contacts, since they do not have any symptoms and, therefore, do not do a self-induced quarantine; and (ii) according to [41], the transmission of the infection is more readily on symptomatic individuals due to physical signs of illness (coughing, sneezing, etc.), which outweighs this first factor. Since $\beta_A > \beta_S$, our results support the first hypothesis.That said, the transmission parameter β_A is approximately 0.47 for $\tilde{q}_I = 5$ and 0.45 for $\tilde{q}_I = 30$, corresponding to 2.3% to 6.8% higher than the symptomatic transmission parameter,[13] meaning that asymptomatic individuals could possibly transmit the disease in a greater extent, since the greater number of contacts may over-weights the probability of transmitting the disease due to physical signs of illness. Of course, the viral load in these cases is smaller, which is a relevant transmission factor not accounted for in our analysis.The duration of the infection $1/\gamma_A$ is approximately $1/0.21 \approx 4.8$ days for $\tilde{q}_I = 5$ and $1/0.17 \approx 5.9$ days for $\tilde{q}_I = 30$, corresponding to 12% to 28.4% less than the average time to recovery of symptomatic individuals ($1/\gamma_S = 1/0.15 \approx 6.7$ days), meaning that the infectious period of the asymptomatic is smaller than the symptomatic one. This is consistent with other pandemic, such as the H1N1 pandemic (see Table 1 from [41] for γ_A and γ_S, and corresponding references).
- We find that the effect of under-reporting of the number of infected and deaths is related to the true mortality rate, which we call "COVID-19 under-reporting tripod". Considering that there are three variables with uncertainty (the under-reporting of infected individuals, under-reporting of deaths, and the true mortality rate) if two of them are known (or measured empirically), the other can be inferred assuming a constant amount of uncertainty. Alternatively, if one is known, it is possible to infer a path for the other two. This is shown in Fig. 10.
  This approach allows aligning the observed to the true mortality rate in order to find the population uncertainty. [30] mentions that many studies are estimating the true mortality rate to be in the range of 0.5–1%. Assuming that this is the true mortality rate, the number of times more infected than reported is about: 8–16 considering the deaths reported by the Ministry of Health; 10–20.5 considering 25% more deaths than reported; 12–24.5 considering 50% more deaths than reported; and 15.5–30 considering 100% more deaths than reported. Alternatively, considering 9 times more infected than reported, the true mortality rate varies from 0.87% to 1%, resulting in 0 to 13% more deaths than reported; if we consider 12 times more deaths than reported, the true mortality rate range is 0.67%-1%, and the death uncertainty range is 0–50%. Other analyses can be inferred directly from Fig. 10. These results are consistent with estimated and empirical findings presented in Table 1.

[13] Considering $\beta_S = 0.44$, then $\beta_A = 0.47$ corresponds to $0.03/0.44 \approx 6.8\%$ and $\beta_A = 0.45$ corresponds to $0.01/0.44 \approx 2.3\%$.

- We note that the level of uncertainty plays a significant role in shifting the Susceptible-Infected transmission curves of the contagion. Uncertainty not only influences the effective transmission rate per infectious contact, but also the recovery and mortality rates associated with the virus. It is possible to make a parallel between this paper and the recent study by [20], which investigates how macroscopic patterns and social reinforcement of interacting contagions through SIR-like models; it is shown how social reinforcement can directly meddle with the contagion transmission rate along with the spreading regime. The results point to very similar behaviors of the SIR curves as to the influence of the uncertainty levels. We note, nonetheless, that uncertainty also shifts and increases the peak of these curves, which does not happen in the demonstrations presented in [20].
- We provide different relationships between the "uncertainty level" w.r.t. infections and deaths (namely parameters q_I and q_D, respectively) with the epidemiological status of the COVID-19 disease spread in Brazil. We correlate these margins of sub-reports with mortality rates, asymptomatic/symptomatic ratios, and so forth. One key issue is that to estimate and quantify the actual level of sub-reports (through q_I and q_D), pool sample testing and statistical procedures should be performed. Such analyses have been done for some regions of the country [35,13], and also for other nations [39].
- A possible guide to plan adequate public health policies is the following: perform tests in statistically chosen groups of the population and, then, conclude on the level of uncertainty of the available datasets (regarding the number infections and deaths). With the level of uncertainty at hand, one can better quantify the dynamics of the COVID-19 spread within that sampled group with the framework provided in our paper. Knowing the ongoing stage/transmission rate of the contagion, better containment measures can be planned.
- Lastly, we mention that we are not necessarily suggesting to perform mass testing in Brazil, which can be impractical for many reasons. The use of pool sampling (and the tests available from some regions of the country) could serve to infer/estimate the uncertainty level in the datasets disclosed by the Ministry of Health. Then, relying on this, we can better quantify and interpret the effects of the ongoing pandemic. The effects of the divergence in forecasts with regard to different levels of uncertainty are very significant and must be taken into account, as we have been shown in this work.

Table 6 Estimated values of the epidemiological parameters for the SIRD+ψ model. We used $\gamma = 0.150876$ provided by [5], and $\alpha = 0.186353$ and $\psi_\infty = 0.494027$ (calculated with the data from the Ministry of Health, without any uncertainty) in all simulations.

Number of deaths provided by the Ministry of Health ($\tilde{q}_D = 0$)		
$\tilde{q}_I$	β	ρ
0	0.441881	0.088569
5	0.424779	0.016070
10	0.423526	0.008799
15	0.423275	0.006042
20	0.423254	0.004591
25	0.423415	0.003697
30	0.423607	0.003090
25% more deaths ($\tilde{q}_D = 0.25$)		
$\tilde{q}_I$	β	ρ
0	0.447062	0.109324
5	0.425398	0.020318
10	0.423706	0.011151
15	0.423286	0.007664
20	0.423174	0.005826
25	0.423276	0.004693
30	0.423427	0.003924
50% more deaths ($\tilde{q}_D = 0.5$)		
$\tilde{q}_I$	β	ρ
0	0.452330	0.129481
5	0.426041	0.024663
10	0.423897	0.013568
15	0.423304	0.009335
20	0.423098	0.007100
25	0.423139	0.005721
30	0.423247	0.004785
100% more deaths ($\tilde{q}_D = 1$)		
$\tilde{q}_I$	β	ρ
0	0.463086	0.167947
5	0.427405	0.033658
10	0.424317	0.018612
15	0.423359	0.012830
20	0.422956	0.009771
25	0.422871	0.007879
30	0.422889	0.006594

We must stress, once again, that the used models have a series of limitations (such as unmodelled phenomena and disregarded transient behaviors), the available datasets are very imprecise, and also that the forecasting/ model-based prediction problem has a lot of associated sensibility: it is composed of several coupled nonlinear differential equations, which heavily rely on initial conditions (and also contour factors due to the time-series behavior implied via $\psi(t)$). Furthermore, the pandemic dynamics may vary abruptly if more intense health policies are adopted (or dropped) in the future. Therefore, we must recall that the results presented in this paper are qualitative. Our intention in showing long-term predictions is not to provide perfect accurateness regarding the number of infections and deaths, but to show relevant phenomena regarding the levels of uncertainty. If more testing is performed, for example, the uncertainty levels tend to decrease and, thus, the forecast should also change. Also, if the elderly people were to be isolated from the virus in a more effective way, we would expect to see a decrease in death rate, even if the number of infected is increasing, since the mortality rate for this group of people is considerably higher than for people younger than 65 years, as shown by [37].

Table 7 Estimated values of the epidemiological parameters for the SIRASD+ψ model. We used $\alpha = 0.186353$ and $\psi_\infty = 0.494027$ in all simulations.

Number of deaths provided by the Ministry of Health ($\tilde{q}_D = 0$)						
$\tilde{q}_I$	β_A	β_S	γ_A	γ_S	ρ	p
0.1	0.596899	0.441881	0.286665	0.150876	0.087352	0.909091
1	0.495901	0.441881	0.206414	0.150876	0.085485	0.500000
5	0.469934	0.441881	0.186263	0.150876	0.085431	0.166667
10	0.464844	0.441881	0.182116	0.150876	0.085377	0.090909
15	0.462002	0.441881	0.179598	0.150876	0.085324	0.062500
20	0.459022	0.441881	0.177151	0.150876	0.085274	0.047619
25	0.456489	0.441881	0.174976	0.150876	0.085225	0.038462
30	0.453604	0.441881	0.172562	0.150876	0.085208	0.032258
25% more deaths ($\tilde{q}_D = 0.5$)						
$\tilde{q}_I$	β_A	β_S	γ_A	γ_S	ρ	p
0.1	0.595876	0.447062	0.286665	0.150876	0.107830	0.909091
1	0.488607	0.447062	0.200907	0.150876	0.106252	0.500000
5	0.468361	0.447062	0.185085	0.150876	0.106190	0.166667
10	0.463912	0.447062	0.181421	0.150876	0.106122	0.090909
15	0.461257	0.447062	0.179043	0.150876	0.106060	0.062500
20	0.458375	0.447062	0.176669	0.150876	0.105999	0.047619
25	0.455899	0.447062	0.174537	0.150876	0.105939	0.038462
30	0.453410	0.447062	0.172425	0.150876	0.105882	0.032258
50% more deaths ($\tilde{q}_D = 0.5$)						
$\tilde{q}_I$	β_A	β_S	γ_A	γ_S	ρ	p
0.1	0.594797	0.452330	0.286665	0.150876	0.127712	0.909091
1	0.481588	0.452330	0.195552	0.150876	0.126786	0.500000
5	0.466812	0.452330	0.183913	0.150876	0.126715	0.166667
10	0.462976	0.452330	0.180717	0.150876	0.126638	0.090909
15	0.460515	0.452330	0.178486	0.150876	0.126566	0.062500
20	0.457731	0.452330	0.176187	0.150876	0.126494	0.047619
25	0.455307	0.452330	0.174095	0.150876	0.126425	0.038462
30	0.452822	0.452330	0.171994	0.150876	0.126358	0.032258
100% more deaths ($\tilde{q}_D = 1$)						
$\tilde{q}_I$	β_A	β_S	γ_A	γ_S	ρ	p
0.1	0.496673	0.463086	0.208421	0.150876	0.167203	0.909091
1	0.468433	0.463086	0.185298	0.150876	0.167170	0.500000
5	0.463762	0.463086	0.181556	0.150876	0.167088	0.166667
10	0.461252	0.463086	0.179352	0.150876	0.166994	0.090909
15	0.459041	0.463086	0.177363	0.150876	0.166901	0.062500
20	0.456445	0.463086	0.175211	0.150876	0.166811	0.047619
25	0.454125	0.463086	0.173201	0.150876	0.166724	0.038462
30	0.451705	0.463086	0.171152	0.150876	0.166639	0.032258

The Authors truly hope that the proposition herein formalized can serve to help to determine adequate public health policies for Brazil.

Notes

The authors report no financial disclosure nor any potential conflict of interests.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

D. O. C. and J. E. N. acknowledge the financial support of CNPQ under respective grants 302629/2019 − 0 and 304032/2019 − 0.

Appendix A. *A.1. Observed mortality rates*

Tables 4 and 5.

A.2. Simulation parameters

Tables 6 and 7.

References

[1] D. Alves, R. Gaete, N. Miyoshi, B. Carciofi, L. Olveira, T. Sanchez, Estimativa de casos de covid-19, 2020. https://ciis.fmrp.usp.br/covid19-subnotificacao/.

[2] R.M. Anderson, H. Heesterbeek, D. Klinkenberg, T.D. Hollingsworth, How will country-based mitigation measures influence the course of the covid-19 epidemic?, Lancet 395 (2020) 931–934.

[3] Y. Bard, Nonlinear Parameter Estimation, 1974.

[4] A.J. Barros, C.G. Victora, A. Menezes, B.L. Horta, F. Hartwig, G. Victora, L.C. Pellanda, O.A. Dellagostin, C.J. Struchiner, M. N. Burattini, et al, Social distancing patterns in nine municipalities of Rio Grande do Sul, Brazil: the EPICOVID19/RS study, Revista de saude publica 54 (2020) 75.

[5] S.B. Bastos, D.O. Cajueiro, Modeling and forecasting the early evolution of the Covid-19 pandemic in brazil, Sci. Rep. 10 (2020), https://doi.org/10.1038/s41598-020-76257-1.

[6] D. Baud, X. Qi, K. Nielsen-Saines, D. Musso, L. Pomar, G. Favre, Real estimates of mortality following COVID-19 infection, Lancet Infect. Diseases (2020).

[7] S. Bhatia, A. Cori, K.V. Parag, S. Mishra, L.V. Cooper, K.E.C. Ainslie, M. Baguelin, S. Bhatt, A. Boonyasiri, O. Boyd, L. Cattarino, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, S. Dorigatti, Ilaria van-Elsland, R. FitzJohn, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, D. Haw, S. Hayes, W. Hinsley, N. Imai, D. Jorgensen, E. Knock, D. Laydon, G. Nedjati-Gilani, L.C. Okell, S. Riley, H. Thompson, J. Unwin, R. Verity, M. Vollmer, C. Walters, H.W. Wang, P.G. Walker, O. Watson, C. Whittaker, Y. Wang, P. Winskill, X. Xi, A.C. Ghani, C.A. Donnelly, N.M. Ferguson, P. Nouvellet, Short-term forecasts of COVID-19 deaths in multiple countries, 2020. https://mrc-ide.github.io/covid19-short-term-forecasts/index.html#authors [Online; Accessed 29-April-2012].

[8] F. Brauer, C. Castillo-Chavez, Z. Feng. Mathematical Models in Epidemiology, 2019.

[9] Brazilian Federal Medicine Council. Icu beds in Brazil, 2020. https://portal.cfm.org.br/images/PDF/leitosdeutiestados2018.pdf. Accessed: 2020-03-30.

[10] COVID-19 Brasil Research Team, Subnotification analysis (text in portugues). https://ciis.fmrp.usp.br/covid19/analise-subnotificacao/.

[11] C. Del Rio, P.N. Malani, Covid-19–new insights on a rapidly changing epidemic, Jama (2020).

[12] E. Delatorre, D. Mir, T. Graf, G. Bello, Tracking the onset date of the community spread of SARS-CoV-2 in western countries, medRxiv, 2020.

[13] M.F. Do Prado, B.B. de Paula Antunes, L.D.S.L. Bastos, I.T. Peres, A.D.A.B. Da Silva, L.F. Dantas, F.A. Baião, P. Maçaira, S. Hamacher, F.A. Bozza, Analysis of COVID-19 under-reporting in Brazil, Revista Brasileira de terapia intensiva 32 (2020) 224.

[14] EPICOVID19, Epidemiologia da covid-19 no rio grande do sul: Resultado do quarto inquérito, 2020. https://wp.ufpel.edu.br/covid19/files/2020/05/Coletiva-Fase-4-20200526.pdf.

[15] EPICOVID19, Epidemiologia da covid-19 no rio grande do sul: Resultado do terceiro inquérito, 2020. https://wp.ufpel.edu.br/covid19/files/2020/05/Epicovid-resultados-do-3%C2%BA-inqu%C3%A9rito.pdf.

[16] L. Ferrante, P.M. Fearnside, Protect indigenous peoples from COVID-19, Science 368 (2020) 251.

[17] R. Gaete, Análise subnotificação, 2020. https://ciis.fmrp.usp.br/covid19/analise-subnotificacao/.

[18] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, M. Colaneri, Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy, Nature Med. 26 (2020) 855–860.

[19] Globo, Mortes por covid-19 no brasil podem ser ainda maiores em função dos atrasos nas estatísticas da saúde, diz pesquisador, 2020. https://oglobo.globo.com/sociedade/coronavirus/mortes-por-covid-19-no-brasil-podem-ser-ainda-maiores-em-funcao-dos-atrasos-nas-estatisticas-da-saude-diz-pesquisador-24381446.

[20] L. Hébert-Dufresne, S.V. Scarpino, J.-G. Young, Macroscopic patterns of interacting contagions are indistinguishable from social reinforcement, Nat. Phys. 16 (2020) 426–431, https://doi.org/10.1038/s41567-020-0791-2.

[21] M. Kabir, M.S. Afzal, A. Khan, H. Ahmed, COVID-19 economic cost; impact on forcibly displaced people, Travel Med. Infect. Disease (2020).

[22] M.J. Keeling, P. Rohani, Modeling Infectious Diseases in Humans and Animals, 2011.

[23] J.T. Kemper, The effects of asymptomatic attacks on the spread of infectious disease: A deterministic model, Bull. Math. Biol. 40 (1978) 707–718, https://doi.org/10.1016/S0092-8240(78)80003-2, http://www.sciencedirect.com/science/article/pii/S0092824078800032.

[24] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. Roy. Soc. A 115 (1927) 700–721.

[25] H.H. Khachfe, M. Chahrour, J. Sammouri, H. Salhab, B.E. Makki, M. Fares, An epidemiological study on COVID-19: a rapidly spreading disease, Cureus 12 (2020).

[26] S. Khan, The true case fatality of covid19: An analytical solution, medRxiv, 2020, https://www.medrxiv.org/content/early/2020/05/22/2020.05.17.20104554. doi:10.1101/2020.05.17.20104554. arXiv:https://www.medrxiv.org/content/early/2020/05/22/2020.05.17.20104554.full.pdf.

[27] A.J. Kucharski, T.W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R.M. Eggo, Early dynamics of transmission and control of covid-19: a mathematical modelling study, Lancet. Infect. Dis March (2020) 1–7.

[28] A.J. Kucharski, T.W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R.M. Eggo, F. Sun, M. Jit, J.D. Munday, et al, Early dynamics of transmission and control of COVID-19: a mathematical modelling study, Lancet Infect Diseases (2020).

[29] I.M. Longini Jr., M.E. Halloran, A. Nizam, Y. Yang, Containing pandemic influenza with antiviral agents, Am. J. Epidemiol. 159 (2004) 623–633.

[30] S. Mallapaty, How deadly is the coronavirus? scientists are close to an answer, Nature (2020), https://www.nature.com/articles/d41586-020-01738-2.

[31] M.M. Morato, S.B. Bastos, D.O. Cajueiro, J.E. Normey-Rico, An optimal predictive control strategy for COVID-19 (SARS-CoV-2) social distancing policies in Brazil, Annu. Rev. Control 50 (2020) 417–431.

[32] M.M. Morato, I.M.L. Pataro, M.V.A. da Costa, J.E. Normey-Rico, A parametrized nonlinear predictive control strategy for relaxing COVID-19 social distancing measures in brazil, ISA Trans. (2020).

[33] H. Nishiura, T. Kobayashi, A. Suzuki, S.-M. Jung, K. Hayashi, R. Kinoshita, Y. Yang, B. Yuan, A.R. Akhmetzhanov, N.M. Linton, Estimation of the asymptomatic ratio of novel coronavirus infections. Forthcoming, Int. J. Infect. Diseases (2020).

[34] Observatório Covid-19 BR, 2020. Óbitos por COVID-19. Technical Report Observatório Covid-19 BR. https://covid19br.github.io/analises.html?aba=aba2.

[35] B. Paixão, L. Baroni, R. Salles, L. Escobar, C. de Sousa, M. Pedroso, R. Saldanha, R. Coutinho, F. Porto, E. Ogasawara, Estimation of COVID-19 under-reporting in brazilian states through SARI, 2020. arXiv:2006.12759.

[36] L. Peng, W. Yang, D. Zhang, C. Zhuge, L. Hong, Epidemic analysis of COVID-19 in China by dynamical modeling, 2020. arXiv preprint arXiv:2002.06563.

[37] J. Perez-Saez, S.A. Lauer, L. Kaiser, S. Regard, E. Delaporte, I. Guessous, S. Stringhini, A.S. Azman, D. Alioucha, I. Arm-Vernez, et al, Serology-informed estimates of SARS-CoV-2 infection fatality risk in Geneva, Switzerland, Lancet Infect. Diseases (2020).

[38] M. Prado, L. Bastos, A. Batista, B. Antunes, F.B. ao, P. Maçaira, S. Hamacher, F. Bozza, Análise de subnotificação do número de casos confirmados da COVID-19 no Brasil. Technical Report Núcleo de Operações e Inteligência em Saúde (NOIS), 2020. https://sites.google.com/view/nois-pucrio/publica%C3%A7%C3%B5es#h.au6kdlkrdotc.

[39] H. Rahmandad, T.Y. Lim, J. Sterman, Estimating COVID-19 under-reporting across 86 nations: implications for projections and control, 2020. Available at SSRN 3635047.

[40] J.M. Read, J.R.E. Bridgen, D.A.T. Cummings, A. Ho, C.P. Jewell, Novel coronavirus 2019-ncov: early estimation of epidemiological parameters and epidemic predictions. medRxiv, 2020. https://www.medrxiv.org/content/early/2020/01/28/2020.01.23.20018549. doi:10.1101/2020.01.23.20018549. arXiv:https://www.medrxiv.org/content/early/2020/01/28/2020.01.23.20018549.full.pdf.

[41] M. Robinson, N.I. Stilianakis, A model for the emergence of drug resistance in the presence of asymptomatic infections, Math. Biosci. 243 (2013) 163–177.

[42] T.M. Rocha Filho, F.S.G. dos Santos, V.B. Gomes, T.A. Rocha, J.H. Croda, W.M. Ramalho, W.N. Araujo, Expected impact of COVID-19 outbreak in a major metropolitan area in Brazil, medRxiv, 2020.

[43] A.J. Rodriguez-Morales, V. Gallego, J.P. Escalera-Antezana, C. A. Mendez, L.I. Zambrano, C. Franco-Paredes, J.A. Suárez, H. D. Rodriguez-Enciso, G.J. Balbin-Ramon, E. Savio-Larriera, et al., COVID-19 in latin america: the implications of the first confirmed case in Brazil, Travel Med. Infect. Disease (2020).

[44] E.C. Sabino, L.F. Buss, M.P. Carvalho, C.A. Prete, M.A. Crispim, N.A. Fraiji, R.H. Pereira, K.V. Parag, P. da Silva Peixoto, M.U. Kraemer, et al, Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence, Lancet 397 (2021) 452–455.

[45] L. San Lau, G. Samari, R.T. Moresky, S.E. Casey, S.P. Kachur, L.F. Roberts, M. Zard, COVID-19 in humanitarian settings and lessons learned from past epidemics, Nat. Med. 26 (2020) 647–648.

[46] L.F. Scabini, L.C. Ribas, M.B. Neiva, A.G. Junior, A.J. Farfán, O.M. Bruno, Social interaction layers in complex networks for the dynamical epidemic modeling of COVID-19 in Brazil, 2020, arXiv preprint arXiv:2005.08125.

[47] R.R. Silva, W.D. Velasco, W. da Silva Marques, C.A.G. Tibirica, A bayesian analysis of the total number of cases of the COVID-19 when only a few data is available. a case study in the state of Goias, Brazil, medRxiv, 2020.

[48] The Lancet, Covid-19 in brazil: "so what?", Lancet 395 (2020) 1461. http://www.sciencedirect.com/science/article/pii/S0140673620310953. doi: 10.1016/S0140-6736(20)31095-3.

[49] UOL, Covid-19: declarações de óbito apontam 48% mais mortes do que dado oficial, 2020. https://noticias.uol.com.br/saude/ultimas-noticias/redacao/2020/04/09/covid-19-declaracoes-de-obito-apontam-48-mais-mortes-do-que-dado-oficial.htm.

[50] D.T. Volpatto, A.C.M. Resende, L. Anjos, J.V.O. Silva, C.M. Dias, R.C. Almeida, S.M.C. Malta, Spreading of COVID-19 in brazil: Impacts and uncertainties in social distancing strategies, medRxiv, 2020, https://www.medrxiv.org/content/early/2020/06/03/2020.05.30.20117283.

[51] G.L. Werneck, M.S. Carvalho, The COVID-19 pandemic in brazil: chronicle of a health crisis foretold, 2020.