medRxiv preprint doi: https://doi.org/10.1101/2020.05.03.20052779; this version posted May 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Estimative of real number of infections by COVID-19 on Brazil and possible scenarios

Cintra, H. P. C[1] and Fontinele, F. N.[1]

***Abstract*— This paper attempts to provide methods to estimate the real scenario of the novel coronavirus pandemic crisis on Brazil and the states of Distrito Federal, Sao Paulo, Pernambuco, Espirito Santo and Amazonas. By the use of a SEIRD mathematical model with age division, we predict the infection and death curve, stating the peak date for Brazil and these states. We also carry out a prediction for the ICU demand on these states for a visualization of the size of a possible collapse on the local health system. By the end, we establish some future scenarios including the stopping of social isolation and the introduction of vaccines and efficient medicine against the virus.**

I. INTRODUCTION

On December 2019, the city of Wuhan on mainland China started experiencing an outbreak of unknown pneumonia cases. Later, the cause of this outbreak was identified as a virus belonging to the *Orthocoronavidae* subfamiliy and the *Betacoronavirus* genus [1], similar to the SARS-CoV virus that caused the SARS crisis on 2003 [2]. That similarity suggested the name SARS-CoV-2 to the novel coronavirus, and COVID-19 to the disease.

The virus quickly spread to other countries, reaching several countries by the end of February and being declared as a pandemic crisis by the World Health Organization (WHO) at 11th March, being classified as a threat of high risk for the world [3]. Since then, several mathematical models were used to predict the dynamics of the pandemic crisis on other countries. One of those models with the biggest impact was developed by the Imperial College London [4].

On Brazil, the first case registered dates back to 25th February, but on this study we suggest evidence that the infection might have started 19 to 24 days before the official record. We then proceed to simulate the crisis on specific states and attempt to estimate the real scale of the outbreak, predicting when the infections peak might occur as well as the curve for ICU demand for each of those states. Finally, we present some future scenarios based on how the stop of the intervention might affect the curve and how the introduction of vaccines or available medicine might also change the infection curve since there are several studies being made to evaluate possible use of pharmaceutical drugs to cure the disease [5], [6] and [7].

II. MODEL DESCRIPTION

We make use of a SEIRD model, dividing the population into 5 groups: Susceptible, Exposed, Infected, Recovered and Dead. The exposed population differs from the infected population on the development of symptoms; an individual with the virus enters first the exposed group, carrying the virus during the incubation period; then, with the development of symptoms, the individual passes to the infected group. The rate of infection is proportional to the number of infected and a contact constant β given by the average number of contacts between individuals times the probability of contracting the virus on each contact. The rate of symptoms development is proportional to the incubation period c^{-1}. The rate of recovery is proportional to the percentage of people who recovers divided by the average time taken from symptoms onset to recovery, similarly to the death rate. Another consideration is that people on the exposed group might infect susceptible people with an infection rate k which is a small percentage of β.

The following diagram represents the dynamics of these populations:

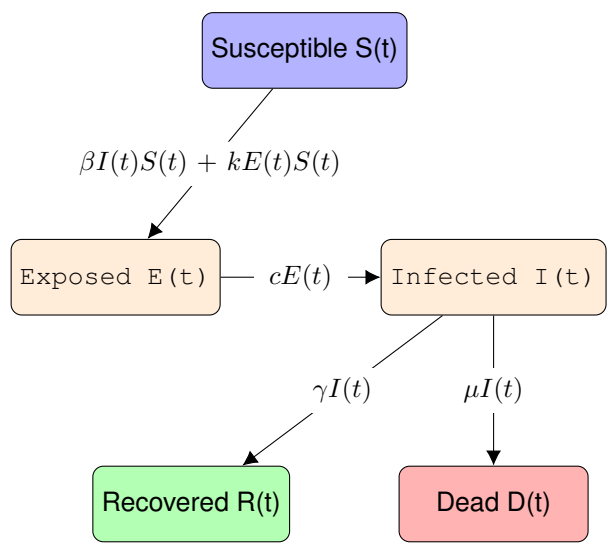

Fig. 1: Representation of a SEIRD model, a susceptible person gets exposed to the virus, being infected afterwards and either dies or recovers from the disease.

This model is represented by the following set of differential equations

[1]Physics Institute, University of Brasilia, Brasilia, Brazil

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.03.20052779; this version posted May 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

$$\frac{dS}{dt} = -\frac{\beta}{N}I(t)S(t) - \frac{k}{N}E(t)S(t) \quad (1)$$

$$\frac{dE}{dt} = \frac{\beta}{N}I(t)S(t) + \frac{k}{N}E(t)S(t) - cE(t) \quad (2)$$

$$\frac{dI}{dt} = cE(t) - \gamma I(t) - \mu I(t) \quad (3)$$

$$\frac{dR}{dt} = \gamma I(t) \quad (4)$$

$$\frac{dD}{dt} = \mu I(t) \quad (5)$$

where the recovery rate γ and death rate μ are represented in terms of the Case Fatality Rate (CFR) $P_{:(}$ and the average time from symptoms onset to recovery τ_r and death τ_d.

$$\mu = \frac{P_{:(}}{\tau_d} \quad (6)$$

$$\gamma = \frac{1 - P_{:(}}{\tau_r} \quad (7)$$

All equations described conserve the total population N, which is assumed constant and homogeneous for the model to be valid. This, of course, presents a limitation of the model, since in reality N is not homogeneous. Therefore, here N carries the role of effective population, being equivalent to the population in which the virus might get to under the interval of some months. Estimating the real N is not an easy task, on next sessions we discuss how we decided to estimate this number.

We then, divided the population into age groups to better describe how these rates vary from group to group. With that, we suggest the following changes already proposed by [8]:

$$P_{:(} \rightarrow P_{:(_i}, \qquad i = 1, 2, \cdots, M \quad (8)$$

$$\beta \rightarrow \beta_i = \sum_{j=1}^{M} P_{inf} C_{ij} \quad (9)$$

where M is the number of age groups, C_{ij} is the social contact matrix, representing the average contacts between a member of the i-th group with all other j-th groups and P_{inf} is the probability of being infected at each contact.

With these definitions, we represent non-pharmaceutical interventions such as social isolation and lock-down with a decrease of β given by a logistic function of the type

$$\beta = \frac{P_d \beta_i}{1 + \tau e^{t - t_c}} + (1 - P_d)\beta_i \quad (10)$$

here, β_i is the infection rate before the intervention, t_c is the time when the intervention starts, P_d is the percentage of reduction achieved and τ is a constant related to the time taken from the start of the intervention until P_d is reached.

When simulating the curve for infections and deaths from Brazil and the states of Pernambuco, Espirito Santo, Distrito Federal, Sao Paulo and Amazonas, we used the model described above. Meanwhile, when simulating the ICU demand, we do not apply the age division for lack of specific data for each age group, thus, we apply the simple SEIRD model with β extracted from the fitting of data of each state and $P_{:(}$, τ_r and τ_d appropriate for ICU patients by COVID-19.

III. ESTIMATING THE PERCENTAGE OF LOST CASES

A. *Number of hospitalizations by SARS*

According to [9], 14% of COVID-19 cases are severe and require hospitalization and 5% of are critical and require an ICU unit, another study found similar percentages, stating that 19% of the infections resulted on hospitalizations [10].

With the emergence of the novel coronavirus, the number of hospitalizations by SARS per week increased when compared to the years of 2019, 2018 and 2017. Using the number of hospitalizations by SARS on those years, we construct an background behavior, that is, the expected number of hospitalizations due to other respiratory diseases (Figure 2). The number released by the Health Ministry per week is subjected to alterations due to new results on the following weeks regarding the one released. For example, by the end of the 6th week of the year 2018, the official report estimates a number of hospitalizations around 50, but on later on, this number was corrected to be close to 200. Because of this uncertainty on the most recent data, for this estimation we use the values available a few weeks before the most recent released (Figure 3).

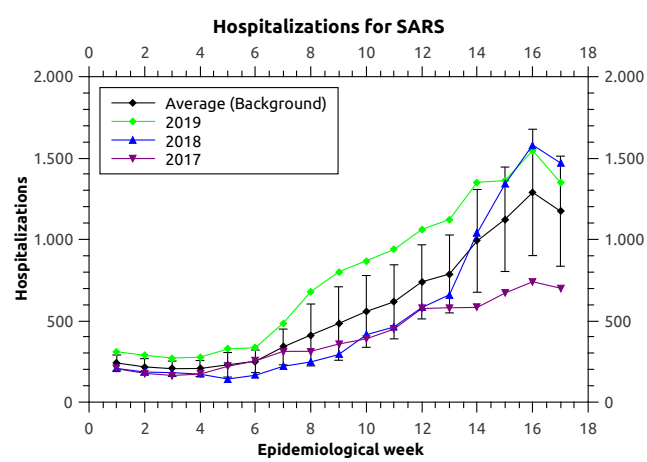

Fig. 2: Hospitalizations by SARS on Brazil on the years of 2017, 2018 and 2019.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.03.20052779; this version posted May 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

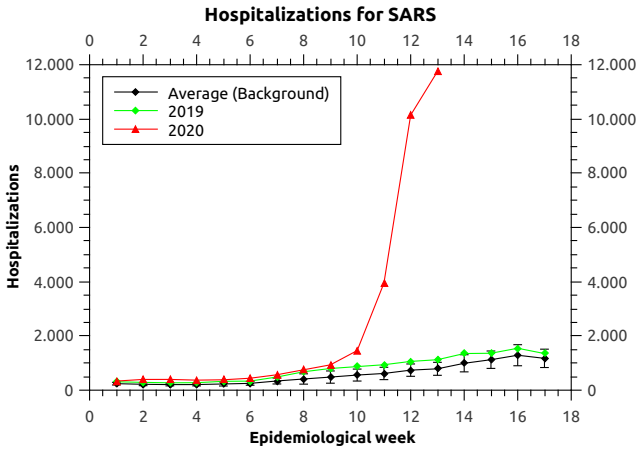

Fig. 3: Number of hospitalizations by SARS on Brazil on the year of 2020, 2019 and the background average.

From the data, at the 6th week, the number of hospitalizations was higher than the upper error bar of the background by 121 hospitalizations, and higher than the year of 2019 by 106, an increase of 31%. According to a study done on COVID-19 patients on Shanghai, the hospitalization occurs on average 4 days after symptoms onset, ranging from 2 to 7 [11], this study, together with the increase of SARS hospitalizations by the 6th week of 2020 suggests the possible existence of COVID-19 cases on Brazil around February 1st to February 6th, 19 to 24 days before the official record of the first case on 25th February.

Following the increase of cases, by the end of the 13th epidemiological week of 2020 (28/03/2020), the number of hospitalizations by SARS on Brazil was already, 12260, while the upper error bar of the background is 1028, and the year of 2019 registered 1123. Supposing that, 90% of the excess of hospitalizations is due to COVID-19; based on the observation that the year of 2019 is about 10% bigger than the background, meaning we could see this behavior on 2020 as well; that marks around 10023 to 10108 hospitalizations by infections of the SARS-CoV-2 virus, which reflects on 52752 to 53200 infections between 21/03/2020 and 26/03/2020 (According to the average time taken to hospitalization). The comparison with the official numbers reported on this period gives a real number 18 to 46 times bigger than the one released (24, using the average). That represents a lost of 96% (94 - 97.8) of the infections. By comparison, a study done on China found 86% of undocumented infections prior to 23th january [12].

B. Number of tests performed

A study of the Imperial College London estimated the number of infections on 11 European countries until 28th March, based on the basic reproduction number of the disease, found to be between 2 and 3 [13], [14], [15] and [16], and the type of non-pharmaceutical intervention done by the countries on specific dates [17]. With these estimations, we may find the percentage of lost cases on these countries until 28th March by comparing the estimate number of people infected with the official data available at 28th March. Relating these percentages with the number of tests done per 1000 habitants and the number of tests done per day per 1000 habitants, we got an linear relation between the number of total tests done per 1000 habitants and per day per 1000 habitants on a country and the percentage of lost cases (Figure 4 and 5).

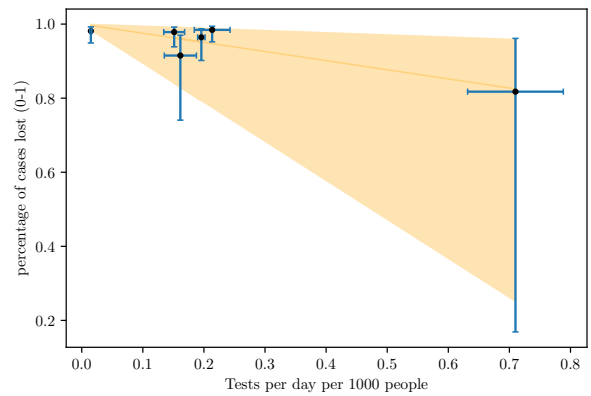

Fig. 4: Relation between the number of tests performed per day per 1000 habitants and the percentage of lost cases.

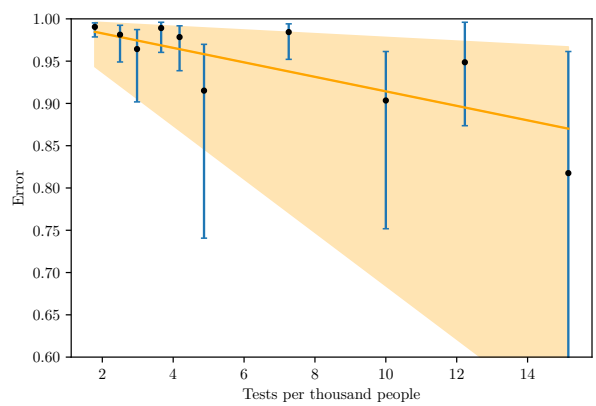

Fig. 5: Relation between the number of tests performed on total per 1000 habitants and the percentage of lost cases.

The number of points on each graph is different because, although the study considered 11 countries, not all of them had data of tests per day available on [18]. We also compared the undocumented cases with the progression of the outbreak on each country and the day on which the non-pharmaceutical interventions were imposed, but found no correlation. We evaluated the effect of the increasing rate of testing as well, but it had no observable effect. From this relation, a country needs to perform 4 (0.94 - 17) tests per day per 1000 habitants. Here, the large margin for the higher values of testing arises from the low density of data points on the bigger values of the x-axis on figure 4.

The last official register of the total number of tests done per 1000 habitants on Brazil was 1.37, which corresponds

medRxiv preprint doi: https://doi.org/10.1101/2020.05.03.20052779; this version posted May 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

to 98.8% of cases lost (95.6 - 99.7). A more precise number could be achieved with the data of tests per day per 1000 habitants, allowing a 2-dimensional regression, unfortunately, we found no record of this information, still, when fitting the data to a 2-dimensional regression algorithm, the resulting function states that the most important factor controlling the uncertainty of cases is tests per day per 1000 habitants. That could also be observed by looking at the graphs individually, the number of total tests performed per 1000 habitants decreases the percentage of undocumented infections in a much lower rate than the number of tests per day per 1000 habitants.

Both methods found an region of agreement (95.6% to 97.8%) of undocumented infections on Brazil. With the agreement of both methods, we decided to accept the estimate for undocumented infections on Brazil and moved on to the simulations of the country and some specific regions.

IV. SIMULATIONS

For the simulation of Brazil, we used the *World Population Prospects* from the United Nations (UN) to evaluate the age distribution on Brazil on the year of 2020 [19]. We found no study measuring the social contact matrix for the country, but the study [20] evaluated the high levels of social contact on Brazil as an important factor for the spreading of leprosy. Therefore, we decided to use the social contact matrix found with the highest entries among those available (Poland) due to Brazilian culture of proximity.

For the values of γ and μ we choose to use the ones found on South Korea, Germany, Iceland and Taiwan data, since these countries are performing more tests per 1000 habitants than Brazil, making their data more reliable (Figures 6 and 7). For each country, we acquired the average values for τ_d and τ_r, knowing the CFR.

Country	τ_d	τ_r
Iceland	14.3 $\pm$ 4.3 days	11.5 $\pm$ 4.5 days
South Korea	13.5 $\pm$ 5.6 days	21 $\pm$ 10 days
Germany	13.6 $\pm$ 5.8 days	17.5 $\pm$ 8 days
Taiwan	–	–
Average	13.8 $\pm$ 5.2 days	16 $\pm$ 7.5 days

TABLE I: Average values of τ_d and τ_r acquired from data.

Data from Taiwan presented large fluctuations on the behavior of μ and γ, even with a almost constant Case Fatality Rate (CFR) $1.3\% \pm 0.2\%$, making the values for τ_d and τ_r inconclusive. That might be explained by the early intervention made by the local government, drastically changing the values for the parameters. Clinical studies performed on Wuhan patients found τ_d on average 18 days (6-32) [21], and 20 days (17-24) [22].

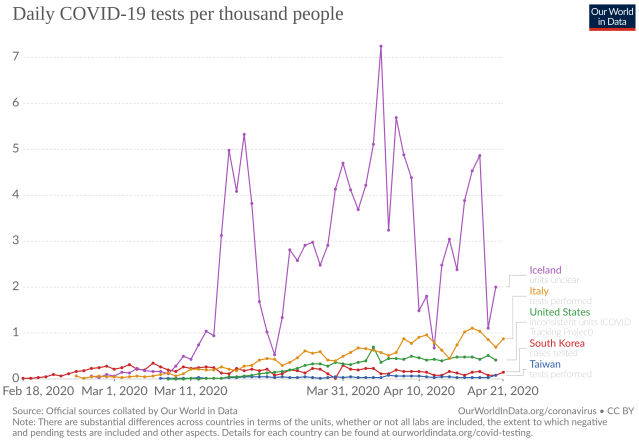

Fig. 6: Tests Per Day Per 1000 habitants. Taken from [18].

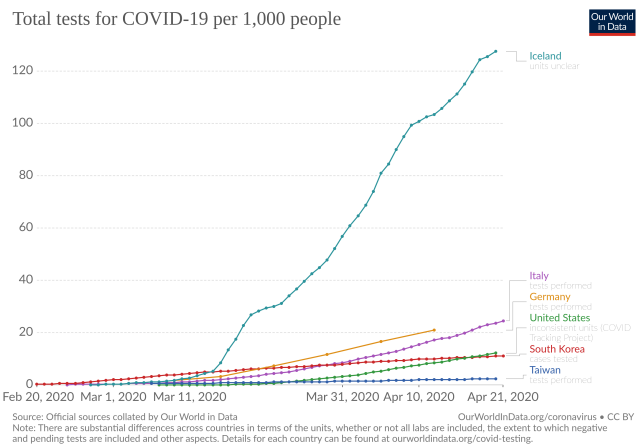

Fig. 7: Total number of tests per 1000 habitants. Graph taken from [18].

When fitting the data of those countries with the model to extract β (Table II), we took into consideration on the simulations the non-pharmaceutical intervention on each country in order to better describe β. The value of β was used to set an reference to compare with the ones found with the fitting of data from each state.

Country	β
Taiwan	0.427 $\pm$ 0.066
Germany	0.483 $\pm$ 0.025
South Korea	0.534 $\pm$ 0.040
Iceland	0.685 $\pm$ 0.121
Average	0.532 $\pm$ 0.063

TABLE II: Values of β.

For the incubation period c^{-1}, we took an average found of previous studies (Table III)

medRxiv preprint doi: https://doi.org/10.1101/2020.05.03.20052779; this version posted May 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

incubation time	95% confidence	Reference
6.4 days	5.6-7.7	[23]
5.2 days	4.1-7	[24]
5 days	–	[25]
4 days	–	[26]
5.1 days	4.5 - 5.8	[27]

TABLE III: Incubation time of the disease according to other studies with an average of 5.1 days.

The value for k was set to 9% of β based on the findings that asymptomatic cases were responsible for 9% of the infections [28]. The parameter $P_{:(}$ for each age group was set according to the international average [29] (Table IV), while $P_{:)} = 1 - P_{:(}$.

Age (years)	Case Fatality Rate
0-9	0
10-19	0.2%
20-29	0.2%
30-39	0.2%
40-49	0.4%
50-59	1.3%
60-69	3.6%
70-79	8.0%
+80	14.8%

TABLE IV: CFR of COVID-19 for different ages.

V. RESULTS

On the simulation for the whole country, we considered N as 1% of the total population based on an international behavior for the total number of infections on other countries. We also selected $P_{inf} = 14\%$ according to [8].

In order to input on the simulation, the effect of the use of masks by a large number of individuals we use a logistic function to decrease the value of P_{infec} on 50% based on [30], the slope of the decreasing region was set to be 10x slower than the one simulated for the social distancing.

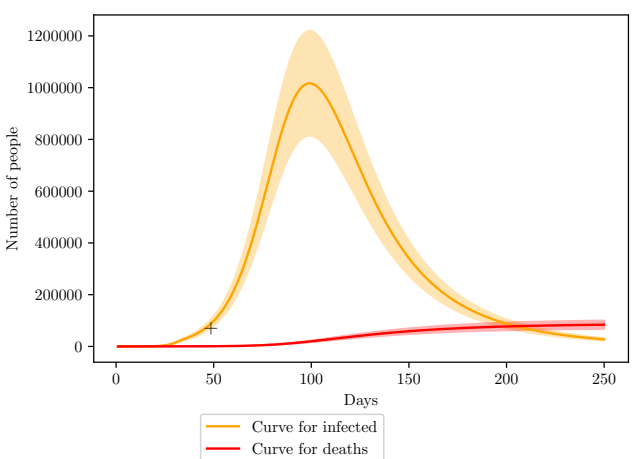

Fig. 8: Simulation of the COVID-19 pandemic crisis on Brazil.

The curve shows a good agreement with the estimated values by the number of SARS hospitalizations on the last weeks of March, shown by the + mark on the graph. We also predict that the peak of the infection curve on Brazil should be 100 days after the first case, in which we considered to be the beginning of February. Therefore, the peak should be on the middle to end of May with a million of infections, ranging from 800 hundred thousand to 1.2 million. The number of deaths is estimated to be around 80000, ranging from 60000 to 100000.

The shading areas represent a 20% deviation from the simulated curve. The high value of deviation was chosen as a reflection of the uncertainty on the value for effective population N.

A. Pernambuco

Online data available from the local government on [31] states a total of 0.84 tests per 1000 habitants and an average of 0.05 tests per day per 1000 habitants, placing it on more than 90% of infections being undocumented. For the simulation, we acquired data regarding the age and geographical distribution of the population from the last census from IBGE [32].

The official record for the first case dates to 12th March, however, data from [31] now shows a ICU entry of a 71 year old man on the capital Recife, diagnosed with the virus SARS-CoV-2. The patient started with the symptoms on March 1st. We choose to set this date as the starting point of the simulation.

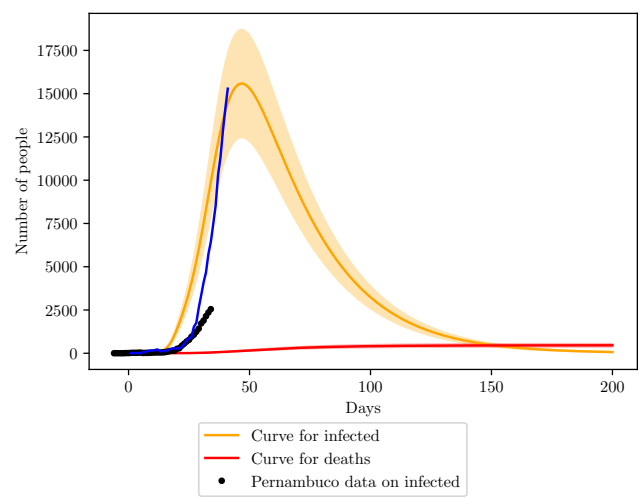

Fig. 9: Simulation of the COVID-19 pandemic crisis on Pernambuco. The blue curve shows the behavior of the data considering 90% loss of infections.

The simulation shows a peak close to the 50th day, on the beginning of May, with 15000 infections, ranging from 12500 to 18000. The number of deaths estimated is 400 (320 - 480).

Despite the large number of cases lost, when fitting the data with a simulated curve, the value of β is 0.460 ± 0.050, which agrees to international standards. That indicates a good tracking of the rate of change of the infection curve on Pernambuco. The state might not have the precise values of the real infections, but it has a good knowledge of their growth. That is an important feature for the state to be able

medRxiv preprint doi: https://doi.org/10.1101/2020.05.03.20052779; this version posted May 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

to say that it's data might represent a small scale of the real scenario.

To simulate the ICU population, we changed the parameters of τ_r and τ_d to those corresponding to ICUs, which were $\tau_r = 16 \pm 4$ days [11] and $\tau_d = 7$ (3-11) days [33]. The CFR was also changed to 52% [34], and β was set proportional to τ_{ICU}, the mean time taken from symptoms onset to ICU entry, which is 3.5 days [34].

The state of Pernambuco has a total of 1315 ICU beds according to a census carried by the Brazilian Association of Intensive Medicine (AMIB) on the year 2016 [35]. However, recent news point to 80% of these beds already being occupied, bringing the available number of ICU beds to 263.

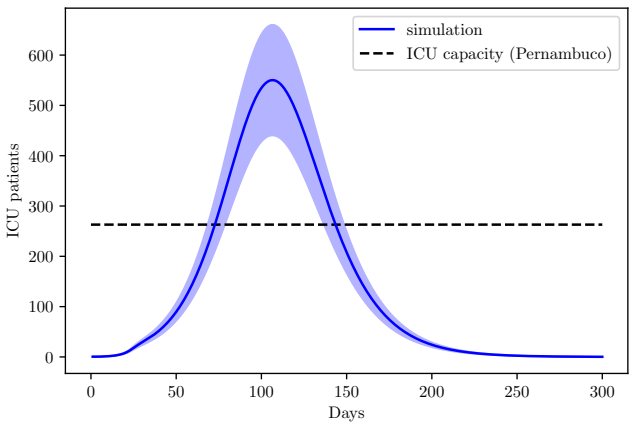

Fig. 10: Simulation of the ICU population on Pernambuco due to COVID-19.

B. Espirito Santo

On Espirito Santo, the online data provided by the government states a total of 1.8 tests per 1000 habitants realized, 0.04 tests per day per 1000 habitants, placing the uncertainty percentage close to 90%. There are also 161 ICU units available for COVID-19 cases [36]. The population data for the simulations was retrieved from a local census done by IBGE [37].

Like Pernambuco, the fitting of Espirito Santo data reveals a good agreement of β with international parameters, $\beta = 0.436 \pm 0.199$, however, the large margin of error shows low confidence on that data.

We found no record of previous hospitalizations due to COVID-19 prior the first case announced on 6th March, like the one found on Pernambuco, therefore, we choose the official day as the starting point of the disease. The first infectious individual was on the age group between 30-39 years.

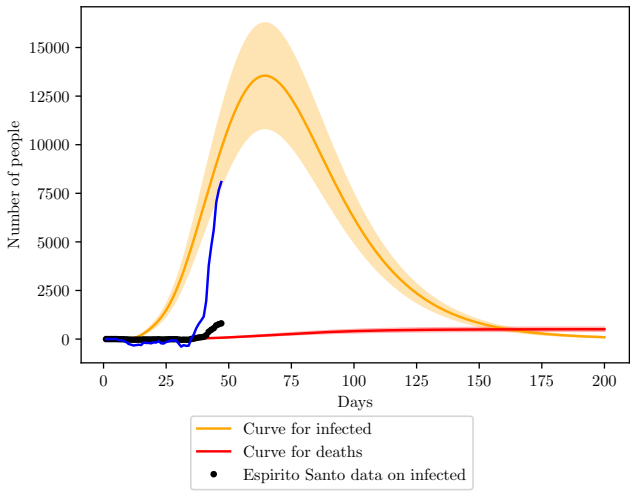

Fig. 11: Simulation of the COVID-19 pandemic crisis on Espirito Santo.

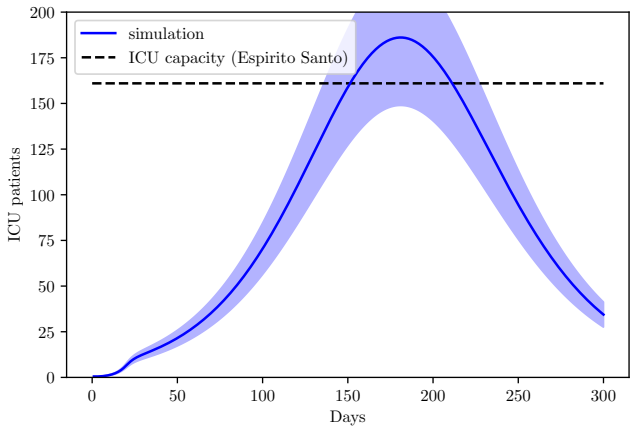

Fig. 12: Simulation of the ICU population on Espirito Santo due to COVID-19. Blue curve represents data following the 90% of undocumented infections.

The peak of Espirito Santo is close to 70 days before the start, being this date close to 15th May, with a maximum infection number around 130000 (105000 - 1560000). The number of deaths is estimated to 500 (400-600).

C. Distrito Federal

Recent data from the government reveals 20716 tests, meaning 6.8 tests per 1000 habitants, placing the state on most likely 86% (78.5 - 94.2) of undocumented infections, unfortunately, no record of tests per day was found, so a better accuracy of lost cases was not possible. The first register of COVID-19 on the state is from 5th March, with non-pharmaceutical interventions starting at 10th March [38].

Like previous states, the IBGE census was used to extract population distribution [39].

The fit of data with the simulations returns an efficiency of 88% of the social isolation, but β and τ_d are off the margin of acceptance, indicating that the state is not tracking well

medRxiv preprint doi: https://doi.org/10.1101/2020.05.03.20052779; this version posted May 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

the rate of increase of deaths and cases, possibly invalidating the estimate percentage of efficiency of the social isolation.

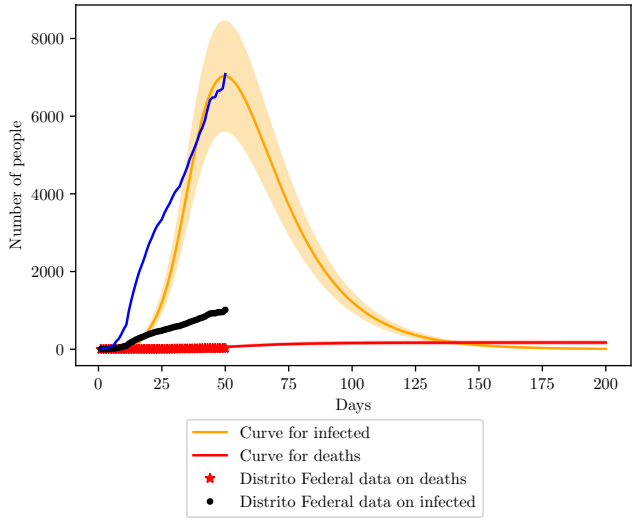

Fig. 13: COVID-19 scenario for Distrito Federal. Blue curve represents the behavior of data considering 86% loss of infections.

The simulation shows that the Distrito Federal is currently at it's higher number of infections, around 7000 (5600 - 8400). The maximum number of deaths is projected to 160 (128 - 192). Also, with the current number of infections, Distrito Federal is losing 85% of cases (82 - 88), in agreement with the margin estimated by the number of tests performed.

From the AMIB census, the state posses 659 ICU beds, we assume 70% of occupation before the disease reached the state.

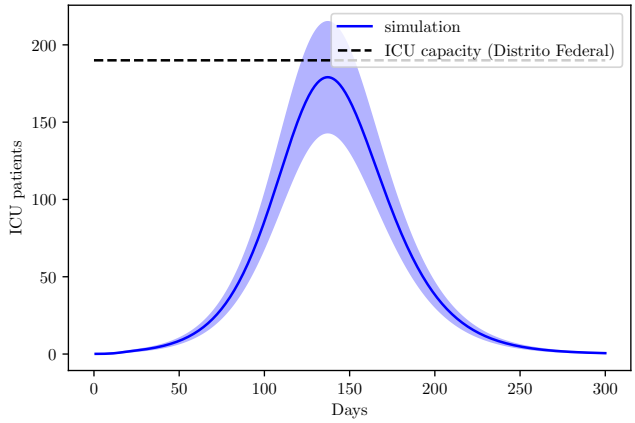

Fig. 14: ICU demand on Distrito Federal due to COVID-19.

D. Sao Paulo

The state of Sao Paulo also provided online data gathered by the government [40]. The first infection notified dates from 26th February. Studies done with cellphone data from Sao Paulo habitants saw an average of $53.6\% \pm 3.4\%$ of the population is respecting the social isolation imposed by the local government on 24th March [40].

When fitting the data with the model, considering a non-pharmaceutical intervention starting 27 days after the first case, we found an quarantine efficiency of $58.3\% \pm 7\%$, with agreement of the study. We also found $\beta = 0.454 \pm 0.52$, indicating that Sao Paulo is also on good track of the increasing rate of the outbreak. Unfortunately, the government did not display data on infections, but with such a high mortality, around 8%, the number of infections is probably 4x bigger than the official number (meaning 75% of undocumented infections), assuming that the number of deaths is in good agreement with the real scenario. However, given the behavior of previous states, and the general scenario of Brazil, it is most likely that Sao Paulo founds itself on a 90% loss scenario.

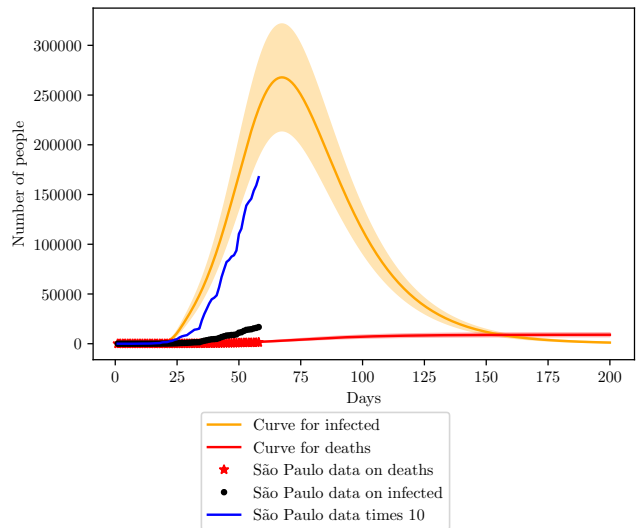

Fig. 15: Simulation of the COVID-19 pandemic crisis on Sao Paulo. The blue curve was shown to represent the 90% loss of data on Sao Paulo considering a constant testing rate.

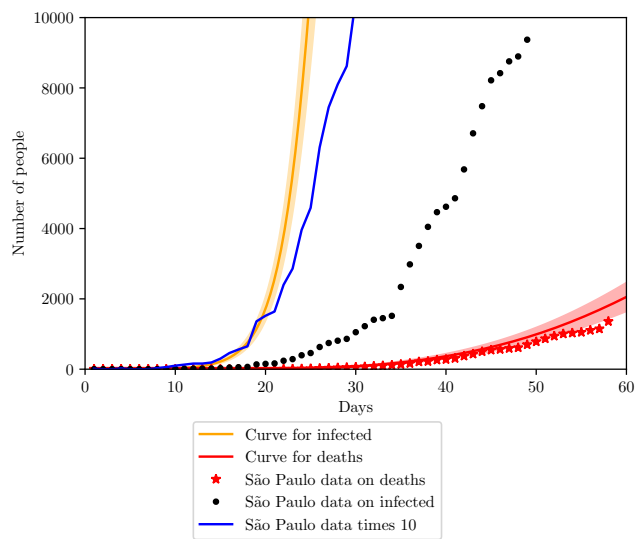

Fig. 16: Simulation of the COVID-19 pandemic crisis on Sao Paulo on most recent days.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.03.20052779; this version posted May 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

The state is with it's peak projected to be around the 70th day of infection, namely close to 7th May. The peak number of infections should be 260000 (208000 - 312000). For the number of deaths, the estimate is close to 6500 (5200 - 7800).

From the AMIB census, the state of Sao Paulo has a total of 7312 ICU beds and recent news point to 53% of them already occupied, leaving around 3400 ICU beds available for COVID-19 treatment.

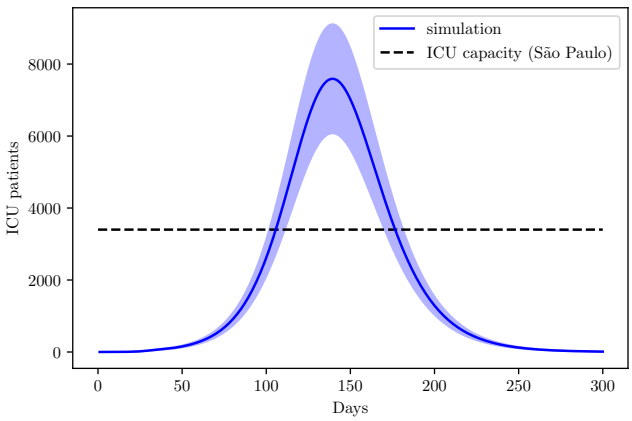

Fig. 17: Simulation of the ICU demand on the state of Sao Paulo.

E. Amazonas

For Amazonas, the fitting of data acquired from the Health Ministry yields $\beta = 0.406 \pm 0.096$ and $\tau_d = 16 \pm 6$, showing that, despite the high number of undocumented infections, the state is on the same situation found on other states. Knowing the behavior of the curve, but not the true number of each point on the curve. The difference from previous states is that the value of τ_d is also in agreement with international values.

The census from IBGE [41] was also used here to acquire population data for the state.

From the AMIB census, Amazonas posses 249 ICU beds, with 55% of them occupied before the outbreak. Unfortunately, no data on tests was found for Amazonas, therefore we consider a 90% loss of infections.

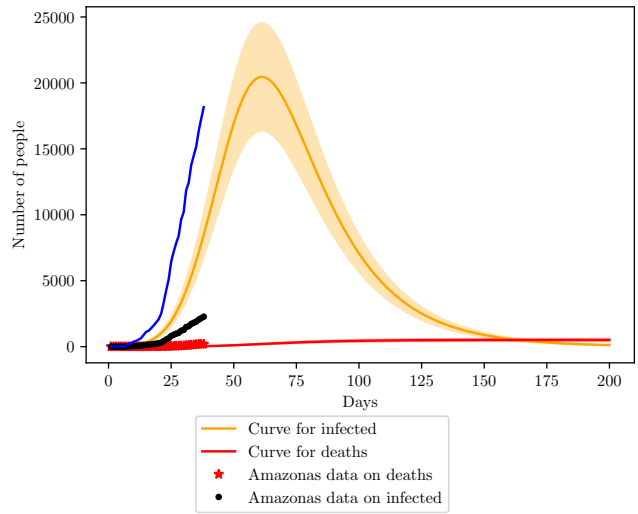

Fig. 18: Simulation of the COVID-19 pandemic crisis on Amazonas.

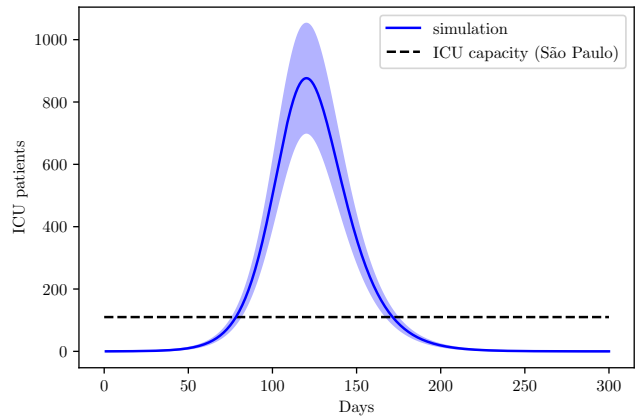

Fig. 19: Simulation of the ICU demand on the state of Amazonas.

Amazonas peak is estimated to May 16th, with a total of 20000 infections peak (16000 - 24000). Deaths are estimated to reach 500 (400 - 600).

VI. FUTURE SCENARIOS

Simulating the stopping of non-pharmaceutical interventions is equivalent to make β increase back to it's starting value. By making such simulations, we observe a increase of cases, that is, a second peak of the disease right after the stop.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.03.20052779; this version posted May 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

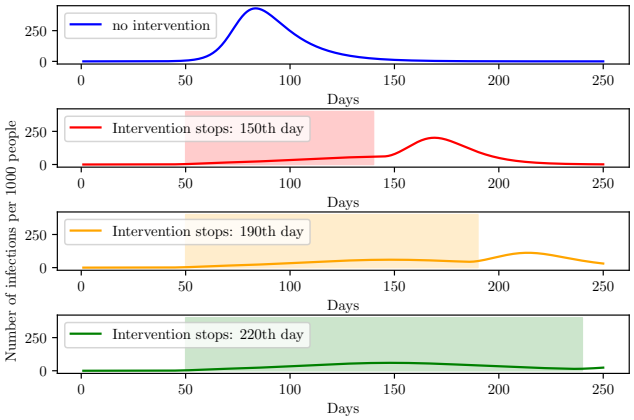

Fig. 20: Height of the second peak for different times on social isolation.

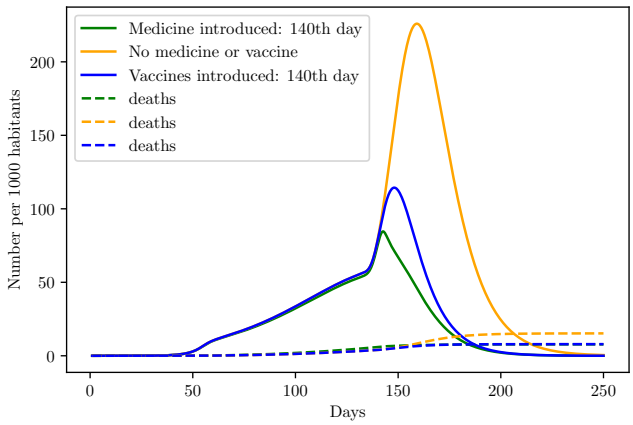

Fig. 21: Behave of the infection curve if the vaccination/medication occurs at the same time of intervention stopping.

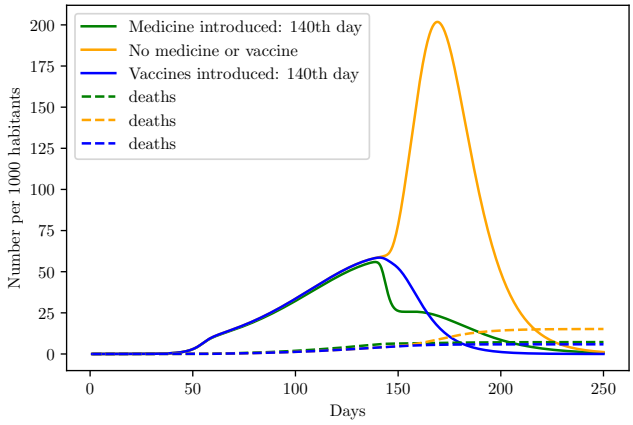

Fig. 22: Behave of the infection curve if the vaccination/medication occurs 10 days before time of intervention stopping.

Figure 20 shows that to drastically diminish the second peak, the social isolation must endure about 220 days supposing an efficiency of 70%, it is equivalent to state that on Brazil, quarantine should hold until October, while for a total prevention of the second peak, social isolation must take place until December. That is expected and agrees to other simulations made by different groups, another group from the University of Harvard projected that, to prevent a second peak on the world the possible re-incidence of the virus, social isolation must hold until the beginning of 2021 and social distancing until 2022 or 2024 [42].

However, that scenario might drastically change with the introduction of vaccines or efficient medicine on the population. As show on simulations, such pharmaceutical interventions are able to decrease rapidly the infection curve. In order to simulate the effect of medicine on the population, we started decreasing the death probability $P_{:(}$ and time taken from symptoms onset to recovery τ_r from a specific date, until it reaches a maximum value. We supposed that the introduction of medicine decreased both $P_{:(}$ and τ_r by half on the period of 10 days after the introduction on the population.

For the vaccines, we added the term $-vS(t)$ on (1), which takes out individuals from the susceptible group at a rate v called vaccination rate, and added the term $vS(t)$ on (4), adding those individuals on the recovery group, granting them immunity against the virus. The vaccination rate v was chosen to behave according to a logistic function starting on 0, and gradually increasing to 0.2 after an specific time.

From the simulations, the most safe method is not to stop the intervention and introduce the vaccines or drugs into the population, but to wait a small period of 10 days before stopping the intervention.

VII. CONCLUSION

Simulations of the COVID-19 outbreak vary from model to model, here we try to find balance on the most precise model, which could be achieved considering also a group of asymptomatic infections and hospitalized, and the availability of data. In doing so, we decided to simulate the behavior of the disease on Brazil based on international parameters under the assumption that the virus would not be much different on Brazil and the main aspects regarding the transmission would be intervention efficiency, population demographics and social contact. This assumption might prove limited if later should be found that climate effects strongly alters the spread.

Another limitation of the model is on the assumption of homogeneous population. We tried here to counter-attack this limitation by estimating the effective population N according to international parameters and by widening the error margin

medRxiv preprint doi: https://doi.org/10.1101/2020.05.03.20052779; this version posted May 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

of the predictions. A better estimate of the outbreak could be done by assessing cities individually, however that would represent a loss of data, since demographics available by IBGE regard mainly states and major cities. Another outtake would be the testing data, the states which provided testing data, did only for the whole state but not for individual cities. We did not consider comorbidities on the population such as diabetes and cancer, however the age of the individual seems to be the most important factor on determining mortality factors [43].

We also state here that the nature of the process is stochastic, allowing fluctuations from the deterministic model used to run the simulations. Thus, this study present an estimate of the real situation and expected behavior given the parameters associated with the disease and the efficiency of the intervention. The above results present the dimension of the real scenario, but due to possible initial fluctuations on the stochastic behavior of reality, we might find some deviations from the expectancy.

Even with limitations, the model has proven efficient on generating curves that agree with the estimated loss of cases for each state. From the states studied here, Sao Paulo, Amazonas and Pernambuco present the highest risk of collapse on the health system, while Espirito Santo and Distrito Federal should have minor issues with system collapse. The blue curve representing the behavior of the official data considering the error percentage for Amazonas exhibited a growth far from the simulation region, however, it falls perfectly inside this region when data is translated by 10 days, meaning that if the infection on Amazonas begun 10 days earlier than previously thought, data fits the simulated curve.

On the duration of social isolation, the safer situation is to hold the isolation for as long as possible in order to decrease the second peak height, while increasing the number of tests performed. All simulations considered here did not assume the end of the intervention, therefore, numbers of deaths may be higher. Should any efficient drugs on combating the virus come along, the simulations shows the safer way is to first introduce them on the population without breaking the social isolation, and about 10 days later start the process of reopening.

References

[1] J. Cui, F. Li, and Z.-L. Shi, "Origin and evolution of pathogenic coronaviruses," *Nature reviews Microbiology*, vol. 17, no. 3, pp. 181–192, 2019.

[2] K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, and R. F. Garry, "The proximal origin of sars-cov-2," *Nature Medicine*, vol. 26, no. 4, pp. 450–452, 2020.

[3] W. H. Organization *et al.*, "Coronavirus disease 2019 (covid-19): situation report, 85," 2020.

[4] N. Ferguson, D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg, *et al.*, "Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand," 2020.

[5] E. K. McCreary and J. M. Pogue, "Coronavirus disease 2019 treatment: a review of early and emerging options," in *Open Forum Infectious Diseases*, vol. 7, p. ofaa105, Oxford University Press US, 2020.

[6] E. M. Negri, B. Piloto, L. K. Morinaga, C. V. P. Jardim, S. A. E.-D. Lamy, M. A. Ferreira, E. A. D'Amico, and D. Deheinzelin, "Heparin therapy improving hypoxia in covid-19 patients-a case series," *medRxiv*, 2020.

[7] N. Tang, H. Bai, X. Chen, J. Gong, D. Li, and Z. Sun, "Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy," *Journal of Thrombosis and Haemostasis*, 2020.

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

[9] Z. Wu and J. M. McGoogan, "Characteristics of and important lessons from the coronavirus disease 2019 (covid-19) outbreak in china: summary of a report of 72 314 cases from the chinese center for disease control and prevention," *Jama*, 2020.

[10] C. COVID and R. Team, "Severe outcomes among patients with coronavirus disease 2019 (covid-19)—united states, february 12–march 16, 2020," *MMWR Morb Mortal Wkly Rep*, vol. 69, no. 12, pp. 343–346, 2020.

[11] J. Chen, T. Qi, L. Liu, Y. Ling, Z. Qian, T. Li, F. Li, Q. Xu, Y. Zhang, S. Xu, *et al.*, "Clinical progression of patients with covid-19 in shanghai, china," *Journal of Infection*, 2020.

[12] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman, "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov2)," *Science*, 2020.

[13] S. Zhang, M. Diao, W. Yu, L. Pei, Z. Lin, and D. Chen, "Estimation of the reproductive number of novel coronavirus (covid-19) and the probable outbreak size on the diamond princess cruise ship: A data-driven analysis," *International Journal of Infectious Diseases*, vol. 93, pp. 201–204, 2020.

[14] S. Zhao, Q. Lin, J. Ran, S. S. Musa, G. Yang, W. Wang, Y. Lou, D. Gao, L. Yang, D. He, *et al.*, "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in china, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak," *International journal of infectious diseases*, vol. 92, pp. 214–217, 2020.

[15] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv, "The reproductive number of covid-19 is higher compared to sars coronavirus," *Journal of travel medicine*, 2020.

[16] J. M. Read, J. R. Bridgen, D. A. Cummings, A. Ho, and C. P. Jewell, "Novel coronavirus 2019-ncov: early estimation of epidemiological parameters and epidemic predictions," *MedRxiv*, 2020.

[17] S. Flaxman, S. Mishra, A. Gandy, H. Unwin, H. Coupland, T. Mellan, H. Zhu, T. Berah, J. Eaton, P. Perez Guzman, *et al.*, "Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on covid-19 in 11 european countries," 2020.

[18] H. R. Max Roser and E. Ortiz-Ospina, "https://ourworldindata.org/coronavirus," 2020.

[19] D. o. E. United Nations and S. Affairs, "World population prospects 2019: Highlights," 2019.

[20] P. D. Deps, B. Guedes, J. Bucker Filho, M. K. Andreatta, R. S. Marcari, and L. C. Rodrigues, "Characteristics of known leprosy contact in a high endemic area in brazil," *Leprosy review*, vol. 77, no. 1, pp. 34–40, 2006.

[21] Q. Ruan, K. Yang, W. Wang, L. Jiang, and J. Song, "Clinical predictors of mortality due to covid-19 based on an analysis of data of 150 patients from wuhan, china," *Intensive care medicine*, pp. 1–3, 2020.

[22] J. T. Wu, K. Leung, M. Bushman, N. Kishore, R. Niehus, P. M. de Salazar, B. J. Cowling, M. Lipsitch, and G. M. Leung, "Estimating clinical severity of covid-19 from the transmission dynamics in wuhan, china," *Nature Medicine*, pp. 1–5, 2020.

[23] J. A. Backer, D. Klinkenberg, and J. Wallinga, "Incubation period of 2019 novel coronavirus (2019-ncov) infections among travellers from wuhan, china, 20–28 january 2020," *Eurosurveillance*, vol. 25, no. 5, 2020.

[24] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, *et al.*, "Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia," *New England Journal of Medicine*, 2020.

[25] N. M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S.-m. Jung, B. Yuan, R. Kinoshita, and H. Nishiura, "Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis

medRxiv preprint doi: https://doi.org/10.1101/2020.05.03.20052779; this version posted May 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

of publicly available case data," *Journal of Clinical Medicine*, vol. 9, no. 2, p. 538, 2020.

[26] W.-j. Guan, Z.-y. Ni, Y. Hu, W.-h. Liang, C.-q. Ou, J.-x. He, L. Liu, H. Shan, C.-l. Lei, D. S. Hui, *et al.*, "Clinical characteristics of coronavirus disease 2019 in china," *New England Journal of Medicine*, 2020.

[27] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler, "The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application," *Annals of internal medicine*, 2020.

[28] H. Nishiura, T. Kobayashi, Y. Yang, K. Hayashi, T. Miyama, R. Kinoshita, N. M. Linton, S.-m. Jung, B. Yuan, A. Suzuki, *et al.*, "The rate of underascertainment of novel coronavirus (2019-ncov) infection: Estimation using japanese passengers data on evacuation flights," 2020.

[29] WorldMeters, "https://www.worldometers.info/coronavirus/coronavirus-cases/," 2020.

[30] C. R. MacIntyre, Q. Wang, S. Cauchemez, H. Seale, D. E. Dwyer, P. Yang, W. Shi, Z. Gao, X. Pang, Y. Zhang, *et al.*, "A cluster randomized clinical trial comparing fit-tested and non-fit-tested n95 respirators to medical masks to prevent respiratory virus infection in health care workers," *Influenza and other respiratory viruses*, vol. 5, no. 3, pp. 170–179, 2011.

[31] CIEVSPE, "http://dados.seplag.pe.gov.br/apps/corona.html," 2020.

[32] IBGE, "https://cidades.ibge.gov.br/brasil/pe/panorama," 2017.

[33] X. Yang, Y. Yu, J. Xu, H. Shu, H. Liu, Y. Wu, L. Zhang, Z. Yu, M. Fang, T. Yu, *et al.*, "Clinical course and outcomes of critically ill patients with sars-cov-2 pneumonia in wuhan, china: a single-centered, retrospective, observational study," *The Lancet Respiratory Medicine*, 2020.

[34] M. Arentz, E. Yim, L. Klaff, S. Lokhandwala, F. X. Riedo, M. Chong, and M. Lee, "Characteristics and outcomes of 21 critically ill patients with covid-19 in washington state," *Jama*, 2020.

[35] A. de Medicina Intensiva Brasileira, "Censo amib 2016," 2016.

[36] G. do Estado do Espírito Santo, "https://coronavirus.es.gov.br/painel-covid-19-es," 2020.

[37] IBGE, "https://cidades.ibge.gov.br/brasil/es/panorama," 2017.

[38] S. de Saúde do Distríto Federal, "http://www.saude.df.gov.br/," 2020.

[39] IBGE, "https://cidades.ibge.gov.br/brasil/df/panorama," 2017.

[40] G. do Estado de São Paulo, "https://www.saopaulo.sp.gov.br/coronavirus/," 2020.

[41] IBGE, "https://cidades.ibge.gov.br/brasil/am/panorama," 2017.

[42] S. M. Kissler, C. Tedijanto, E. Goldstein, Y. H. Grad, and M. Lipsitch, "Projecting the transmission dynamics of sars-cov-2 through the postpandemic period," *Science*, 2020.

[43] N. U. Kingdom, "https://www.covid19survivalcalculator.com/research," 2020.

VIII. SUPPLEMENTARY MATERIAL

Source code used for some simulations and with didatic example of predictions available at https://github.com/PedroHPCintra/Coronavirus.