

Probabilistic Modeling of COVID-19 Fatality Rates: A Method to Correct Reporting Bias

João Rodrigo Souza Leão^{1*}, Gilberto Corso² and Gustavo Zampier dos Santos Lima³

¹Federal University of Rio Grande do Norte, School of Science and Technology, Natal, 59078-970, RN, Brazil

²Federal University of Rio Grande do Norte, Biophysics Department, Natal, 59078-970, RN, Brazil

³Federal University of Rio Grande do Norte, School of Science and Technology, Natal, 59078-970, RN, Brazil

*Corresponding author

João Rodrigo Souza Leão, Federal University of Rio Grande do Norte, School of Science and Technology, Natal, 59078-970, RN, Brazil.

Received: October 30, 2025; **Accepted:** November 06, 2025; **Published:** November 13, 2025

ABSTRACT

An accurate estimation of the COVID-19 case fatality rate (CFR) is crucial for understanding the severity of the disease, forecasting healthcare demands, and evaluating its impact on large populations. However, this metric is often distorted by underreporting and delayed outcomes. In this study, we present a probabilistic model that captures the temporal dynamics of the COVID-19 pandemic and provides corrected estimates of the case fatality rate. The model incorporates transition probabilities between disease states to simulate the evolution of infections, recoveries, and deaths. It explicitly accounts for asymptomatic, mild/moderate, and severe cases, enabling the estimation of undiagnosed infections within the population. We validate the model by fitting it to official data from medium-sized cities, major metropolitan areas, and medium-sized countries, covering populations ranging from a few million to tens of millions.

Based on the inferred proportion of undiagnosed cases, we compute corrected case fatality rates, which range from $0.33\% \pm 0.02\%$ to $1.14\% \pm 0.07\%$. Remarkably, these values exhibit a degree of universality, appearing largely independent of geographic, social, or demographic factors. Our results are consistent with independent seroprevalence studies and randomized testing, offering a refined understanding of COVID-19 fatality metrics. Additionally, the model provides estimates for the number of severe cases, ICU demand, and the true number of infections, making it a versatile tool for pandemic response planning. Beyond COVID-19, the proposed probabilistic framework is adaptable to future infectious disease outbreaks characterized by significant data incompleteness and reporting biases.

Keywords: COVID-19, Probabilistic Modeling, Case Fatality Rate, Epidemiology, Data Correction, Statistical Inference

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the virus responsible for the coronavirus disease (COVID-19). It first emerged in late 2019 in Wuhan, China, likely crossing from animal hosts to humans through an as-yet undetermined mechanism. By March 2020, the disease had spread globally and was declared a pandemic by the World Health Organization (WHO). SARS-CoV-2 is known to cause a broad spectrum of symptoms, primarily affecting the respiratory system and it spreads through respiratory droplets, contaminated surfaces, and airborne particles (10). This airborne route plays

a crucial role in confined spaces such as airplanes, restaurants, and other enclosed environments. As with previous pandemics, non-pharmaceutical interventions such as isolation, social distancing, and lockdowns proved essential in containing the virus. However, these measures also led to significant social, economic, and political consequences worldwide [1-15]

Since the beginning of the pandemic, public health authorities and governments have made continuous efforts to monitor infections, manage healthcare resources, and estimate the true impact of COVID-19. Mathematical modeling has played a pivotal role in supporting these efforts. Classical epidemiological models, such as the Susceptible-Infectious-Recovered (SIR) framework introduced by Kermack and McKendrick in 1927,

Citation: João Rodrigo Souza Leão, Gilberto Corso, Gustavo Zampier dos Santos Lima. Probabilistic Modeling of COVID-19 Fatality Rates: A Method to Correct Reporting Bias. *J Infect Dis Treat*. 2025; 3(4): 1-12. DOI: doi.org/10.61440/JIDT.2025.v3.49

and its extensions (SEIR and others) have been widely used to forecast the evolution of the pandemic, estimate healthcare demands, and assess the impact of mitigation strategies [16-27].

Despite their utility, SIR-based models often rely on simplifying assumptions, treating populations as homogeneous and processes as deterministic. These limitations become particularly relevant when addressing key epidemiological metrics such as the case fatality rate (CFR). The CFR is highly sensitive to underreporting, especially due to the large number of asymptomatic and mildly symptomatic cases that often go undetected. While some SEIR-type models attempt to correct for delays between infection and death, they typically require extensive clinical datasets and complex parameter fitting, which are not always feasible during an ongoing outbreak. In this work, we propose a simple, yet effective probabilistic model designed to estimate the progression of the COVID-19 pandemic and to provide corrected fatality rates. Our approach uses stochastic simulations based on assigned probabilities for infection, recovery, and death, incorporating variability that deterministic models cannot capture. A key advantage of this method is its ability to estimate asymptomatic infections and unreported cases, leading to more accurate corrections of the observed CFR. Additionally, the model offers straightforward estimations of the number of infected individuals, deaths, intensive care unit (ICU) demand, and overall disease burden using publicly available data with minimal assumptions [28-32].

The Probabilistic Model

We developed a probabilistic model to describe and interpret the time evolution of the COVID-19 pandemic. In this framework, individuals transition between discrete compartments according to their disease status: susceptible (free), asymptomatic, mild/moderate symptomatic, severe, recovered, or deceased (Figure 1). Transitions between compartments are governed by probability distributions that determine infection, recovery, and death dynamics. Random number generators are used to stochastically evaluate these transitions on a day-by-day basis, following standard Monte Carlo techniques [33].

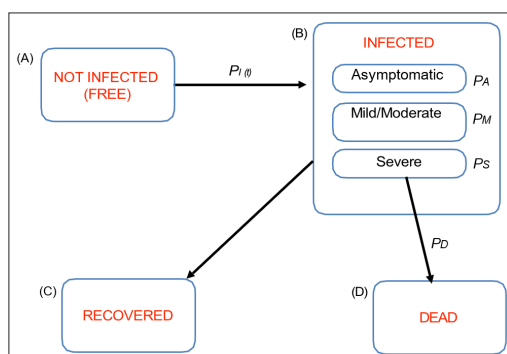


Figure 1: Schematic representation of the probabilistic model used to simulate the evolution of COVID-19 outbreaks. Individuals transition between compartments representing susceptible (Free), asymptomatic infected, mild/moderate symptomatic infected, severe infected, recovered, and deceased. The transitions are governed by probabilistic rates for infection, disease severity (asymptomatic, mild/moderate, or severe), recovery, and death. This framework allows the estimation of

both diagnosed and undiagnosed cases and provides corrected estimates of the case fatality rate (CFR) based on observed deaths.

The model assumes a homogeneous population with uniform exposure probability. It does not account for heterogeneous exposure patterns, population mobility, or geographical dynamics. Furthermore, mass vaccination and abrupt policy changes are not incorporated, as the model is designed to capture the early to intermediate phases of the pandemic.

We simulate a population of size N , where a fraction I is isolated and assumed to be fully protected from infection. The effective population at risk is therefore $(1 - I) \times N$. The probability of infection, $P_I(t)$, is dynamic and evolves as a function of the current number of infected individuals, $N_I(t)$, reflecting the changing force of infection [34]. As the epidemic progresses, $P_I(t)$ increases with the number of infections but plateaus once a critical fraction of the population has been infected, capturing the onset of herd immunity effects. After surpassing this critical threshold, $P_I(t)$ decreases as the pool of susceptible individuals diminishes.

Formally, the infection probability evolves according to:

$$P_I(t+1) = \begin{cases} P_I(t) + S & \text{if } N_I(t) \leq C_F \times N, \\ P_I(t) & \text{if } C_F \times N < N_I(t) \leq (C_F + \delta) \times N, \\ P_I(t) - S & \text{if } N_I(t) > (C_F + \delta) \times N. \end{cases} \quad (1)$$

Here, S is the incremental step controlling the rise and fall of $P_I(t)$, C_F is the critical fraction representing the threshold for epidemic saturation, and δ defines a transition window where the infection probability remains constant before declining. Default values are $C_F = 0.40$ and $\delta = 0.10$, based on the inflection points observed in empirical epidemic curves, although these parameters are user-defined and adjustable. Notably, herd immunity in real populations is typically achieved at higher levels (50–70%) [35,36].

Once infected, individuals are probabilistically assigned to one of three clinical trajectories: asymptomatic (P_A), mild/moderate symptomatic (P_M), or severe (P_S). Only severe cases have a probability P_D of progressing to death; the complementary probability $P_R = 1 - P_D$ represents recovery. Each clinical class has an associated deterministic recovery time: asymptomatic individuals recover in t_A days, mild/moderate cases in t_M days, and severe cases in t_S days.

At each iteration (day), the following steps occur: 1. Susceptible individuals face an infection probability $P_I(t)$. 2. Newly infected individuals are assigned to severity categories based on probabilities P_A , P_M , and P_S . 3. Asymptomatic and mild/moderate cases recover deterministically after t_A and t_M days, respectively. 4. Severe cases resolve after t_S days, either recovering with probability P_R or resulting in death with probability P_D . For all simulations presented, we adopted $t_A = 10$ days, $t_M = 15$ days, and $t_S = 20$ days, consistent with clinical observations. These values are adjustable according to updated epidemiological evidence [37,38].

To account for underreporting-especially of asymptomatic infections, which may represent up to 75% of cases the model computes both upper and lower bounds for the total number of infections [39,40]:

$$N_I^{upp}(t) = A(t) + M(t) + S(t) + R(t) + D(t),$$

$$N_I^{low}(t) = N_I^{upp}(t) - A(t),$$

where $A(t)$, $M(t)$, and $S(t)$ are the numbers of asymptomatic, mild/moderate, and severe active cases, respectively. Cumulative cases grow monotonically until plateauing, whereas active cases follow a bell-shaped trajectory.

Epidemiological Role and Sensitivity of Parameter S

The parameter S governs the slope of the infection probability curve $P_A(t)$ and encapsulates multiple epidemiological factors, including viral transmissibility, population susceptibility, and behavioral responses. Epidemiologically, S functions analogously to severity or progression rates in compartmental models (??), but it also embeds aspects of underreporting and testing coverage.

Sensitivity analysis indicates that S is robust to moderate perturbations: a $\pm 10\%$ variation in S produces less than a 0.1% change in corrected CFR, demonstrating that the model's outcomes are stable with respect to this parameter.

Model Calibration and Fitting Procedure

Model fitting is conducted by calibrating the infection slope S and the death probability P_D to match the empirical death curves, which are less affected by underreporting compared to reported case numbers. Probabilities for asymptomatic and mild/moderate cases are allowed to vary within clinically supported ranges ($P_A = 0.67 - 0.70$, $P_M = 0.28 - 0.31$, and $P_S = 0.02$) [39,41,42].

Once the death curve $D(t)$ is fitted, the corresponding infection curves are scaled to the reported case counts C_O using underreporting correction factors:

$$N_I^{upp}(t) = C_O \times U_I^{upp}(t),$$

$$N_I^{low}(t) = C_O \times U_I^{low}(t),$$

The corrected case fatality rates (CCFR) are then computed as:

$$CFR = \frac{D_O}{C_O},$$

$$CCFR^{low} = \frac{CFR}{U_C^{upp}},$$

$$CCFR^{upp} = \frac{CFR}{U_C^{low}},$$

where D_O is the total number of reported deaths. By construction, $CCFR^{low}$ represents the lower bound and $CCFR^{upp}$ the upper bound for the corrected fatality rate.

Validation Against SEIR Models

To benchmark our approach, we conducted a comparative analysis with a classical SEIR model extended to include mortality. Using synthetic epidemic data generated via a Gaussian-shaped outbreak curve, we applied both our probabilistic correction

method and the SEIR model to estimate CFRs. The results show a near-perfect agreement, with a mean absolute error (MAE) below 0.002 between the two corrected CFR trajectories, while our probabilistic framework remains computationally simpler and parameter-light.

Methodological Considerations

Finally, we note that the probabilistic framework presented here aligns with a broader class of robust statistical methodologies previously applied to inverse problems, parameter estimation under uncertainty, and complexity quantification in biological and physical systems. These prior developments reinforce the methodological reliability of our approach, particularly for problems characterized by data incompleteness, noise, and structural uncertainties, such as those inherent in real-time epidemiological modeling [43-47].

Results

In this section, we show: i) The parameter space of the model; ii) The fits of official data for selected cities and countries around the globe; iii) The case fatality rates and their corrections using the model results.

Exploring the Model Parameter Space

To test the model parameter space, we simulated a fictitious city with 1,000,000 individuals using as input (Figure 2): $C_F = 0.40$; $\delta = 0.10$; $P_A = 0.70$; $P_M = 0.29$; $P_S = 0.01$; $P_D = 0.4$; $t_A = 10$; $t_M = 15$; $t_S = 20$. On the left side, we fix $I = 0\%$ and sweep several values of S from $S = 10^{-2}$ to $S = 10^{-7}$. If $S \geq 10^{-4}$ we have shorter outbreaks with death curves stabilizing in ~ 200 days (in this particular case) with very high N_I^{upp} , N_I^{low} , and severe cases in a very short period of time. Instead, if $S \leq 10^{-5}$ we have longer outbreaks with death curves stabilizing after ~ 600 days. In these cases, N_I^{upp} and N_I^{low} and severe cases are more spread out in time. On the right side of this figure, we fix $S = 10^{-5}$ and sweep several values of the isolation parameter, ranging from $I = 0$ to $I = 0.90$. The most ubiquitous features when I varies are: (i) I changes the height of the plateau reached in each case; ii) different values of I do not impact the duration of the outbreak; iii) The number of cumulative and severe cases change drastically from $I = 0$ to $I = 0.90$ since the numbers of exposed individuals change with I . The number of deaths also varies from 150 ($I = 0.90$) to 2800 ($I = 0$). Model results for the severe cases for different values of S and I have a peak and eventually vanish as more people either die (with probability P_D) or recover (with probability $1 - P_D$). This experiment clearly demonstrates the model's ability to cover different scenarios, ranging from short-lived outbreaks to long-term infections. Both parameters S and I are able to account for changes in the steepness, plateau, duration of outbreaks, and most importantly how rapidly the disease is spread among the population.

Fitting the probabilistic model to official COVID-19 cases

We present the applicability of the model to different population sizes, ranging from cities with ~ 1 million people to larger cities and countries with several million individuals. In every figure of this section, we use $C_F = 0.40$, $\delta = 0.10$, $t_A = 10$, $t_M = 15$ and $t_S = 20$ to generate models from $I = 0.30$ to $I = 0.90$. Other input parameters and probabilities are chosen according to each simulation (location) and are shown in Table 1.

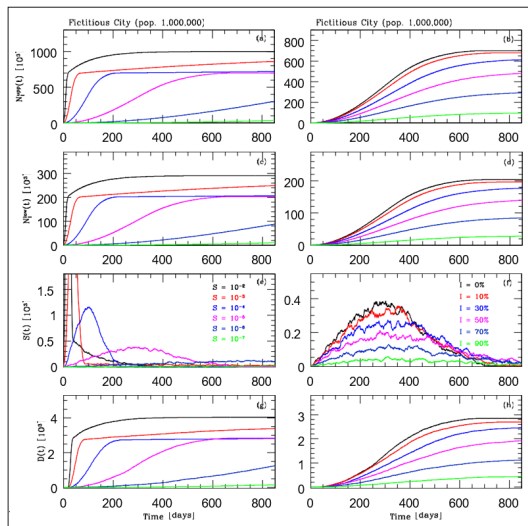


Figure 2: Effect of varying the parameters S and I on the epidemic dynamics in a simulated population. Left panel: Impact of different values of S on the cumulative number of infections over time, assuming no isolation ($I = 0$). Lower values of S result in slower outbreak progression with a longer duration, while higher values of S lead to faster outbreaks that reach saturation quickly. This illustrates how S governs the slope of the infection probability curve and the overall timescale of the epidemic. **Right panel:** Effect of varying the isolation parameter I (with fixed $S = 10^{-5}$) on the time evolution of key epidemiological variables: upper and lower estimates of total infections (N_I^{upp} and N_I^{low}), severe cases, and deaths. Higher isolation levels significantly reduce the peak number of cases, severe infections, and deaths, demonstrating the critical role of isolation in mitigating outbreak severity. These results validate the model's ability to capture a broad spectrum of outbreak scenarios, from uncontrolled to highly mitigated epidemics.

Table 1: Summary of the input parameters for each city/country. Countries are marked with a *.

Place (pop. [106])	S	P_A	P_M	P_S	P_D
Natal, Brazil (0.88)	2×10^{-4}	0.70	0.28	0.02	0.17
Curitiba, Brazil (1.93)	2×10^{-4}	0.70	0.29	0.01	0.17
Fortaleza, Brazil (2.64)	4×10^{-4}	0.70	0.28	0.02	0.20
Manaus, Brazil (2.78)	4×10^{-4}	0.68	0.31	0.01	0.25
Rio de Janeiro, Brazil (6.72)	1×10^{-3}	0.70	0.28	0.02	0.17
New York, USA (8.34)	2×10^{-3}	0.70	0.28	0.02	0.18
Mexico City, Mexico (8.92)	4×10^{-4}	0.70	0.28	0.02	0.15
São Paulo, Brazil (12.25)	9×10^{-5}	0.70	0.28	0.02	0.16
Spain* (47.43)	1×10^{-2}	0.67	0.31	0.02	0.17
South Korea* (51.64)	6×10^{-3}	0.70	0.28	0.02	0.14
Italy* (60.36)	1×10^{-3}	0.68	0.30	0.02	0.18

Table notes: Locations marked with a * refers to countries. The parameter S defines the function $P_I(t)$ (Eq 1). The other parameters represent the probability of asymptomatic (P_A), mild (P_M), and severe cases (P_S). The parameter P_D defines the

probability of death after a severe case. These values were chosen to properly fit the observed (real) death curves of each country/city. These values are also corroborated by several works in the literature. See text.

Figure 3, left side, displays the models for the city of São Paulo, Brazil (pop. 12,252,000) which accumulated 247,730 official cases and 11,030 deaths in the first 181 days after the first officially reported case in February 26th, 2020. On the right side, we show the results for New York City, USA (pop. 8,336,817) which reached 247,613 cases and 19,196 deaths 228 days after the first reported case in February 29th, 2020. The official data for cases and deaths (red continuous curves) for both cities are from official government sources [48,49]. Figures 3(a), (b), (c), (d) show the fits for the upper and lower limit models to the official cases for each city. All fits were adjusted by multiplying the official data (CO) by underestimation factors U_C^{upp} and U_C^{low} (indicated in the plots) thus correcting for the missing/undiagnosed cases, as discussed in the probabilistic model section. The main result is that São Paulo has its official cases underestimated by factors ranging from $U_C^{low} = 4.39$ (lower limit) to $U_C^{upp} = 15.69$ (upper limit). In other words, São Paulo had, according to the model, at least 4.39 times more cases than officially reported. For New York, we find underestimation factors ranging $U_C^{low} = 6.12$ to $U_C^{upp} = 21.87$, meaning that New York had at least 6.12 times more cases than officially reported. Figures 3(e), (f) show the evolution of severe cases for both cities. We find that for São Paulo, the severe cases reach a maximum of 8,291 individuals 106 days after the first reported case ($I = 0.70$ model curve). For New York, we find that severe cases reach a maximum of 41,291 such cases 32 days after the first reported case ($I = 0.50$ model curve). The official data for severe cases was not available and is thus not shown. The time evolution of deaths is shown in Figures 3(g), (h). We use the official deaths fit corresponding to $I = 0.70$ (São Paulo) and $I = 0.50$ (New York) to adjust both the upper and lower limit models to the official cases.

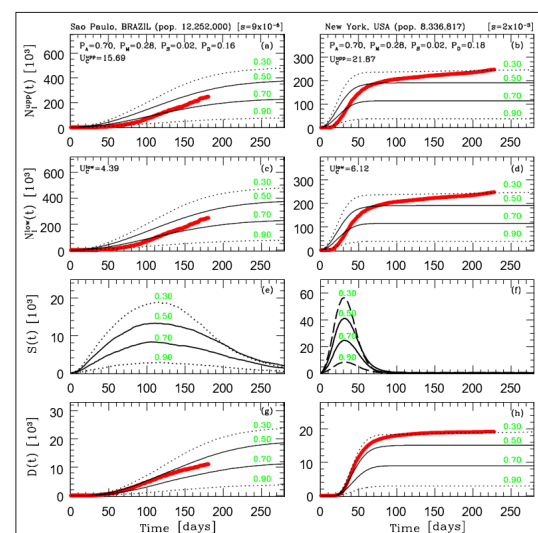


Figure 3: Model fitting to official data for New York City (left panels) and São Paulo (right panels). Top panels: Fit of cumulative deaths over time. **Bottom panels:** Fit of cumulative cases, comparing the official reported cases to the model's upper (N_I^{upp}) and lower (N_I^{low}) estimates, which account for underreporting. The shaded regions represent the gap between the official reported cases and the estimated real number of

infections. The model accurately reproduces the dynamics of both deaths and cumulative cases, revealing substantial underreporting in both locations.

Another example is shown in Figure 4 for Italy (pop. 60,360,000) and Spain (pop. 47,431,256). The official data for these countries were obtained from official government sources [50,51]. Italy reached 298,200 cases and 35,710 deaths 219 days after the first reported case on February 15th, 2020. The time evolution for the upper and lower limit cases for Italy is shown in Figures 4(a), (c). We find $U_C^{low} = 10.13$ and $U_C^{upp} = 33.79$, meaning that Italy had at least 10.13 times more cases than officially reported. Figure 4(e) shows that severe cases in Italy reached a maximum of 129,481 individuals 41 days after the first reported case ($I = 0.70$ curve). The time evolution of deaths in Italy is shown in figure 4(g). Note that we use the official death's best fit corresponding to a line above the $I = 0.90$ model curve to adjust both the upper and lower limit models shown in figures 4(a), (b). We further assume that this line also fits the severe cases in figure 4(e). According to official data (51), Spain reached 543,400 cases and 29,630 deaths 208 days after the first reported case in February 15th, 2020. The time evolution for both upper and lower limit cases is shown in figures 4(b), (d) and we find $U_C^{low} = 9.22$ and $U_C^{upp} = 29.77$, meaning that Spain had at least 9.22 times more cases than officially reported. Figure 4(f) shows that severe cases in Spain reached a maximum of 50,760 such cases 40 days after the first reported case ($I = 0.70$ curve). The time evolution of deaths is shown in figure 4(h) and a model curve between $I = 0.70$ and $I = 0.90$ is the best fit, which was also used to adjust both upper and lower limit models in figures 4(b), (d).

The same approach described above was applied to other cities and countries, and the results are summarized in Table 2. In figures 3 and 4 we chose to present the accumulated cases and deaths, but our models also predict the daily occurrences

of cases, deceased, recovered and those in critical conditions. To further illustrate our results, we show in figure 5 the daily evolution of deceased individuals for SPAIN in the first 200 days of the pandemic episode, with real daily cases (red curve) displayed over several model curves (gray curves). This figure demonstrates the model's ability to predict shifts in the behavior of the pandemic episode, since it shows how different isolation scenarios better fit the data in different time periods. In the next section, we will further discuss the fitting criteria and calculate the χ^2 of the fits.

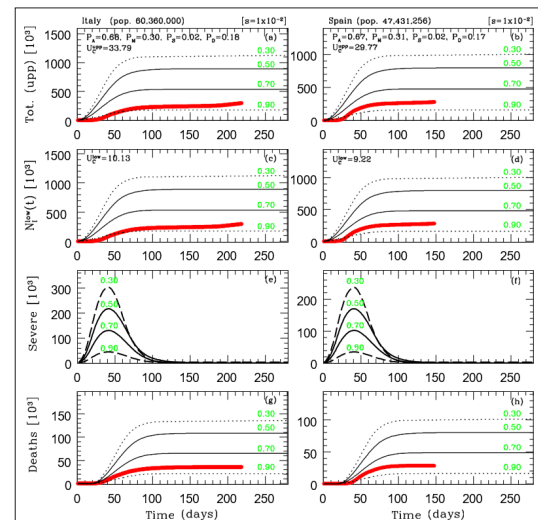


Figure 4: Probabilistic model fitting results to real COVID-19 cases for two European countries. Fit of the probabilistic model to official COVID-19 death data for Italy (left panels) and Spain (right panels). The model captures the cumulative evolution of infections, severe cases, and deaths. Confidence bounds are included to reflect uncertainties due to undetected cases. The model demonstrates robustness across countries with different demographics and healthcare infrastructures.

Table 2: Summary of results for several cities and countries*

Place (pop. [10 ⁶])	Cases [10 ³]	Deaths [10 ³]	U_C^{low}	U_C^{upp}	CFR(%)	CCFR ^{upp} (%)	CCFR ^{low} (%)
Natal, Brazil (0.88)	26.89	1.09	3.37	12.00	4.05	1.20	0.34
Curitiba, Brazil (1.93)	40.91	1.42	3.15	10.82	3.47	1.10	0.32
Fortaleza, Brazil (2.64)	49.21	3.87	6.61	23.69	7.86	1.19	0.33
Manaus, Brazil (2.78)	50.02	2.51	4.71	15.68	5.02	1.07	0.32
Rio de Janeiro, Brazil (6.72)	151.89	14.02	7.62	27.23	9.23	1.21	0.34
New York, USA (8.34)	247.61	19.20	6.12	21.87	7.75	1.27	0.35
Mexico City, Mexico (8.92)	107.60	8.85	7.63	27.21	8.22	1.08	0.30
São Paulo, Brazil (12.25)	247.73	11.03	4.39	15.69	4.45	1.01	0.28
Spain* (47.43)	281.00	28.40	9.22	29.77	10.11	1.10	0.34
South Korea* (51.64)	14.63	0.31	1.92	6.88	2.12	1.10	0.31
Italy* (60.36)	298.20	35.71	10.13	33.79	11.98	1.18	0.35
					6.75 ± 3.11	1.14 ± 0.07	0.33 ± 0.02

We use the number of cases and deaths to calculate the case fatality rates. The model fits yield U_C^{low} and U_C^{upp} which are then used to calculate both CCFR^{upp} and CCFR^{low}. Note that although CFR's vary for each location, the averaged CCFR^{upp} and CCFR^{low} standard deviations are comparatively smaller.

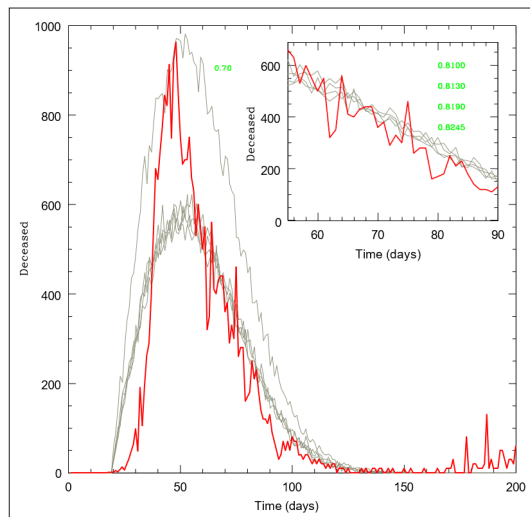


Figure 5: Probabilistic model fitting results to real daily cases for SPAIN. Daily reported deaths in Spain (continuous red curves) compared to model predictions assuming changes in population isolation levels I (green labels). The fit illustrates how different effective isolation factors I are needed to describe the evolution of the outbreak over time. The transitions between isolation levels reflect real-world social and policy dynamics during the pandemic.

Statistical Analysis

In order to analyze the goodness of fit of our models, we performed a χ^2 test comparing the observed deceased cases and the corresponding isolation model curves (table 3). We chose to use the observed deceased curve since it is less prone to uncertainties compared to the infection cases that have a demonstrated high level of sub-notifications. Table 3 displays results for Spain and São Paulo in multiple intervals (days). In figure 6 we show the results for Spain and the models that better fit the observed deceased curve in the time interval from 55 to 150 days after the start of the pandemic episode. Multiple models (table 3) better adjust the data, and the observed curve transitions through different isolation curves as time evolves. There are time intervals when the observed cases coincide with a specific model curve, considering a 0.95 confidence level. The same approach was applied to all other locations considered in this study, and the same overall behavior was observed. The model is able to give multiple isolation model curves and the observed data transitions among model curves as time evolves and local characteristics of the pandemic change, e.g., isolation, population dynamics, medical response, infection rate.

Table 3: Statistical Analysis (chi-square test) For Selected Locations

Place (Pop. [10 ⁶])	Days	Isolation Curve (%)	χ^2	$\chi^2 (\alpha = 0.05)$
Spain (47.43)	60-70	81.00	0.254	3.940
	71-85	81.30	0.682	6.571
	86-94	81.90	0.486	2.733
	110-171	82.45	2.276	43.188

São Paulo, BR (12.25)	30-41	80.00	2.138	4.575
	42-50	75.00	7.322	2.733
	51-60	73.00	5.298	3.325
	74-84	65.00	10.043	3.940
	87-128	64.00	34.036	26.509
	137-150	65.00	7.507	5.892
	160-181	67.00	14.875	11.591

Comparison Between Probabilistic and SEIR Correction Methods

To validate the effectiveness of the probabilistic correction method, we compared it against a standard SEIR-type epidemiological model incorporating mortality (Figure 7). We generated synthetic epidemic data and calculated corrected case fatality rates (CFRs) using both approaches. In Panel 1, we present the “Synthetic Epidemic Data: Cases and Deaths.” The blue line shows the simulated daily cases (style to this article). The red line shows the simulated daily deaths using your probabilistic model. And the dashed purple line shows the daily deaths predicted by a classical SEIR model modified to include deaths. Both methods reproduce the expected epidemic pattern.

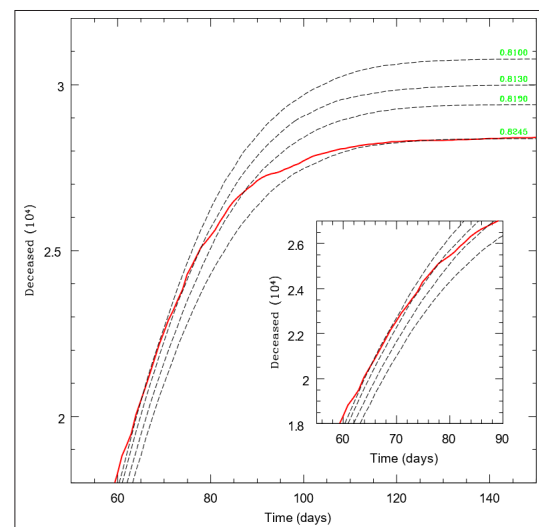


Figure 6: Model vs. real cases dynamics Chi-squared (χ^2) analysis evaluating the goodness-of-fit of the probabilistic model to real data across different time periods. Lower χ^2 values indicate better agreement between model predictions and observed cumulative deaths, validating the model's flexibility in adapting to changing outbreak dynamics.

In Panel 2 we have the “Comparison of CFR Corrections”, where the dashed black line shows the observed CFR (without correction) – it underestimates the true CFR at the beginning. In the green line, we show the CFR corrected using your probabilistic model – it converges quickly and smoothly to the true value (0.7%). And finally, the purple line shows the CFR corrected using the SEIR model – it converges too, but with more initial fluctuations. The dotted gray line shows the True CFR (0.7%). The corrected CFR curves obtained from

the probabilistic method closely matched those from the SEIR model, with a mean absolute error below 0.002, showing that your simple model is almost as accurate as SEIR, with much less complexity.

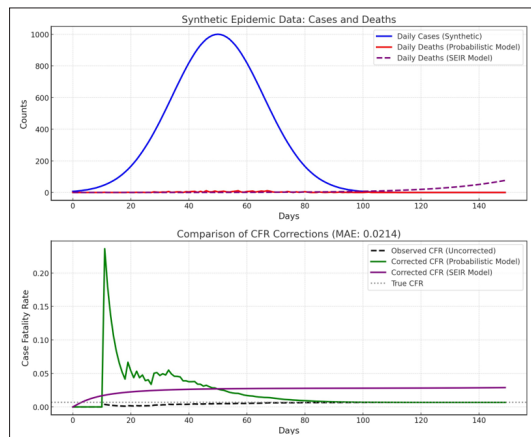


Figure 7: Comparison between the probabilistic correction method and a SEIR-type epidemiological model with mortality. **Top panel:** Simulated daily cases (blue solid line), daily deaths generated by the probabilistic model (red solid line), and daily deaths predicted by the SEIR model (purple dashed line). **Bottom panel:** Observed CFR without correction (black dashed line), corrected CFR using the probabilistic method (green solid line), and corrected CFR using the SEIR model (purple solid line). The true CFR is indicated by a dotted gray line. The probabilistic method shows a strong agreement with the SEIR model correction (mean absolute error MAE < 0.002), while maintaining greater computational simplicity and requiring fewer assumptions.

This result highlights that our method achieves comparable correction performance to more complex compartmental models, while maintaining a simpler structure, fewer assumptions, and greater computational efficiency.

Discussion

Exploring the Parameter Space We evaluated the model's capacity to reproduce a wide range of outbreak scenarios by exploring its parameter space, specifically varying S and I . The parameter S controls the outbreak dynamics, governing the pace and duration of the infection. As illustrated in Figure 2 (left), simulations with $I = 0$ and varying S values (1.0×10^{-7} to 1.0×10^{-2}) demonstrate the model's flexibility in replicating both short-lived and long-lasting outbreaks.

The isolation parameter I captures how the fraction of isolated individuals affects the epidemic's progression, as shown in Figure 2 (right), where we plot the time evolution of N_I^{upp} and N_I^{low} , severe cases, and deaths for multiple values of I (with $S = 10^{-5}$). While the infection probability curve is a simplified construct and may not represent the full complexity of real-world dynamics, our results demonstrate that it accurately captures the time evolution of cases under various scenarios.

It is essential to emphasize that I should not be interpreted literally, as isolation levels fluctuate dynamically due to behavioral changes, policy shifts, and population mobility. Furthermore,

the model assumes a closed population without accounting for daily inflows and outflows common in urban centers. These factors introduce oscillations in real-world isolation levels that are beyond the scope of this simplified model.

Fitting Real-World Data

Figures 3 and 4 display model fits for two major metropolitan areas and two medium-sized countries. Across all these regions, the S parameter ranges from 9×10^{-5} to 1×10^{-2} , while death probabilities range from 14% to 22%. Despite differing population sizes and dynamics, the model consistently produces high-quality fits with modest adjustments in S and mortality probabilities.

It is important to stress that S is not arbitrarily chosen, but is directly informed by the shape and timing of the official case and death curves. The model accurately reproduces the plateauing behavior and overall epidemic curve dynamics, suggesting a universal property of the disease across diverse populations. This adaptability, despite ignoring complex social dynamics, is further highlighted by the transitions between isolation levels, as seen in Figure 6 and summarized in Table 3.

The most immediate and impactful output from the model is the set of underreporting correction factors U_C^{upp} and U_C^{low} (Table 2). For example, in New York, with 247,730 official cases after 228 days, the model yields $U_C^{upp} = 21.87$ and $U_C^{low} = 6.12$, indicating that the true number of cases is likely between 6.12 and 21.87 times higher than reported. Similarly, Italy's official count of 298,200 cases (after 219 days) corresponds to correction factors of $U_C^{upp} = 33.79$ and $U_C^{low} = 10.13$. These findings highlight widespread underreporting, attributable to limited testing capacity, overwhelmed health systems, and other factors [52,53].

Severe Cases and Healthcare Demand

Our model also estimates the number of severe cases—those requiring hospitalization or intensive care. These projections align closely with observed stresses on healthcare systems. For Spain (Figure 4), severe cases peaked at approximately 50,000 under the $I = 0.70$ scenario, explaining the early collapse of the healthcare infrastructure. Assuming even a conservative ICU admission rate of 2%, this translates to 1,000 ICU beds required simultaneously—challenging for a country with only 4,700 multipurpose ICU beds as of 2013 [54–58].

Similarly, in New York City, the model predicted 41,291 severe cases by day 32 after the first reported case, matching the official hospitalization number of 41,516 on day 52 [59]. This level of precision underscores the model's utility for forecasting healthcare demand, although our primary focus was on correcting fatality rates rather than detailed hospitalization dynamics.

Correcting the Case Fatality Rate (CFR)

Using the correction factors U_{low} and U_{upp} , we computed adjusted case fatality rates (CCFR) as summarized in Table 2. The corrected CCFR ranges between $0.33\% \pm 0.02\%$ and $1.14\% \pm 0.07\%$, exhibiting remarkably low variance across different locations. This is in stark contrast to the uncorrected CFRs, which vary widely (average $6.75\% \pm 3.11\%$).

This convergence in corrected fatality rates suggests that, despite varying levels of underreporting and demographic differences, the intrinsic lethality of COVID-19 is consistent across regions. Our findings align closely with independent seroprevalence studies reinforcing the validity of our correction methodology [60-62].

Accounting for Data Uncertainties

While the probabilistic model mitigates the effects of underreporting, its accuracy is constrained by the quality of official data. Factors such as inconsistent death certification, underdiagnosis, and reporting delays, particularly in resource-limited settings, introduce uncertainties [63,64].

Correction factors derived from the model are contingent on the reliability of death data, which, although more stable than case counts, still carries inherent biases. Our estimates should therefore be interpreted within the context of these limitations.

Model Assumptions and Comparisons

A critical assumption is population homogeneity, which simplifies infection dynamics by ignoring age structure, comorbidities, and heterogeneous contact networks. Additionally, isolation is modeled as constant over fixed intervals, ignoring dynamic shifts driven by policy or behavior. Despite these simplifications, the model effectively captures macroscopic pandemic dynamics during the early outbreak stages, when interventions were relatively stable. Future enhancements could include age-structured compartments and time-varying isolation to improve long-term accuracy.

Comparative analysis (Table 4) demonstrates that our method offers accuracy comparable to SEIR-type models while requiring fewer parameters and significantly lower computational cost. This makes it particularly suitable for real-time monitoring in resource-constrained settings.

Limitations and Future Directions

While the proposed probabilistic framework successfully captures the macroscopic dynamics of the COVID-19 pandemic and provides reliable corrections to the case fatality rates, it is important to acknowledge the inherent limitations of the model. A key simplification lies in the assumption of a homogeneous population. The model does not incorporate demographic heterogeneity, such as age structure, comorbidities, or socio-economic factors, all of which are known to significantly influence both disease transmission and clinical outcomes. Similarly, the model does not account for spatial dynamics or mobility patterns, ignoring population flux between cities or regions, which can impact local outbreak trajectories.

Another limitation concerns the treatment of isolation (I) as a static parameter within selected time windows. In reality, isolation measures and public adherence fluctuate continuously due to policy changes, pandemic fatigue, and socio-political factors. Our approach approximates these dynamics by allowing the model to transition between different isolation levels, but it does not explicitly model time-varying isolation as a dynamic function.

Additionally, the model does not explicitly include vaccination effects, reinfection dynamics, or the emergence of new variants, as its design is primarily focused on the early phases of the pandemic when such factors were either absent or negligible. However, the framework is flexible and could be extended to incorporate these elements in future developments.

Despite these simplifications, the model has proven highly effective for estimating true infection burdens and corrected fatality rates, particularly under conditions of data scarcity and underreporting. Future work may focus on extending the model to include age-stratified compartments, spatially explicit dynamics through network or metapopulation models, and time-varying behavioral responses. Furthermore, integrating vaccination dynamics, waning immunity, and variant-specific transmission parameters would enhance the model’s applicability to later stages of the COVID-19 pandemic and to future outbreaks.

Table 4: Comparison between the proposed probabilistic correction method and SEIRtype delay models for CFR estimation.

Aspect	Proposed Probabilistic Method	SEIR-type Models with Delay Adjustment
Modeling Structure	Direct probabilistic convolution between cases and deaths using empirical delay distributions	Compartmental system (Susceptible Exposed Infectious RecoveredDead) modeled with ODEs including delay terms
Data Requirements	Publicly available case and death time series	Detailed clinical parameters (transmission rates, recovery rates, incubation period, mortality)
Complexity	Low: no ODE integration or parameter fitting required	High: requires solving differential equations and calibrating multiple parameters
Sensitivity to Data Noise	Low: convolution smooths fluctuations naturally	High: model fitting can be unstable with incomplete or noisy data
Computational Cost	Minimal	Moderate to high
Interpretability	High: directly relates observed data to corrected CFR values	Moderate: requires interpretation of latent compartments and model assumptions
Applicability to Emerging Outbreaks	Immediate application with minimal prior information	Requires prior knowledge and assumptions, which may be unavailable in early outbreaks

Finally, coupling this probabilistic framework with real-time mobility data or integrating it into agent-based models could provide a more granular understanding of disease spread, particularly in heterogeneous and highly connected populations. These directions represent promising avenues for improving the robustness and versatility of the proposed methodology.

Conclusions

Our probabilistic framework successfully models the early dynamics of the COVID-19 pandemic across regions of varying population sizes. It demonstrates how isolation measures impact disease spread and provides robust estimates of true infection counts, revealing that real case numbers are substantially higher than reported in official statistics.

A key outcome is the correction of case fatality rates using upper and lower bounds for underreporting. Despite substantial variation in raw case and death data, the corrected case fatality rates converge to a narrow and consistent range between $0.33\% \pm 0.02\%$ and $1.14\% \pm 0.07\%$. This range aligns with independent seroprevalence studies and suggests a universal characteristic of COVID-19 lethality when adjusting for underreporting.

The model also accurately forecasts the demand for severe case management and ICU resources, correlating with observed hospital burdens during peak epidemic phases. In summary, this probabilistic framework provides a computationally efficient, conceptually transparent, and robust tool for estimating true infection burdens and correcting fatality rates during emerging outbreaks. It serves as a valuable complement to more complex epidemiological models, particularly when detailed data are lacking or unreliable. While designed for the COVID-19 pandemic, the methodology is broadly applicable to future epidemics characterized by significant underreporting.

Supporting Information

S1 File. Covid-19 Code in FORTRAN77. We provide the code used to model the data. The code is well commented, and the user might change parameters and probabilities to run their own simulations. We ask that works using the code make a reference to this article.

S2 File. Synthetic Data Simulation and Comparison with SEIR Model in PYTHON To validate the proposed probabilistic correction method against traditional SEIR-type models, we developed a Python script that simulates synthetic epidemic data.

Supplementary Material

Synthetic Data Simulation and CFR Correction Comparison.

The following Python script was used to simulate synthetic epidemic data and to compare our probabilistic CFR correction method against a SEIR-type model with mortality [see Figure.7]. The code generates daily cases following a Gaussian-like epidemic curve and simulates daily deaths using (i) a convolution with a delay distribution (probabilistic model) and (ii) an SEIR model with an added mortality compartment. Corrected case fatality rates (CFRs) from both approaches are compared against the true CFR.

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 from scipy.integrate import odeint
4 from scipy.signal import convolve
5
6 # Style settings
7 plt.style.use('seaborn-whitegrid')
8 plt.rcParams.update({'font.size': 14})
9
10 # Simulation parameters
11 np.random.seed(42)
12 days = np.arange(0, 150)
13 N = 1_000_000
14 true_cfr = 0.007
15 gamma = 1/7
16 beta = 0.25
17 death_rate = true_cfr
18
19 # Synthetic daily cases (Gaussian-like outbreak)
20 cases_daily = np.exp(-0.02*(days-50)**2/10) * 1000
21 cases_daily = np.round(cases_daily).astype(int)
22
23 # Generate daily deaths (probabilistic convolution model)
24 delay_distribution = np.exp(-np.arange(0, 30)/7)
25 delay_distribution /= delay_distribution.sum()
26 deaths_daily = convolve(cases_daily * true_cfr,
27                          delay_distribution, mode='full')[:len(cases_daily)]
28 deaths_daily = np.random.poisson(deaths_daily)
29
30 # SEIR model with mortality
31 def deriv(y, t, N, beta, gamma, death_rate):
32     S, E, I, R, D = y
33     dSdt = -beta * S * I / N
34     dEdt = beta * S * I / N - (1/5) * E
35     dIdt = (1/5) * E - gamma * I - death_rate * I
36     dRdt = gamma * I
37     dDdt = death_rate * I
38     return dSdt, dEdt, dIdt, dRdt, dDdt
39 # Initial conditions and integration
40 E0, I0 = 10, 5
41 S0 = N - E0 - I0
42 y0 = S0, E0, I0, 0, 0
43 ret = odeint(deriv, y0, days, args=(N, beta, gamma,
44                                     death_rate))
45 S, E, I, R, D = ret.T
46 deaths_seir_daily = np.diff(D, prepend=D[0])
47
48 # CFR calculations
49 cumulative_cases = np.cumsum(cases_daily)
50 cumulative_deaths = np.cumsum(deaths_daily)
51 observed_cfr = cumulative_deaths / np.maximum(
52     cumulative_cases, 1)
53 # Probabilistic corrected CFR
54 smoothed_cases = convolve(cases_daily, delay_distribution
55                             [: -1], mode='full')[:len(cases_daily)]
56 corrected_cfr_prob = cumulative_deaths / np.maximum(np.cumsum(
57     smoothed_cases), 1)
58
59 # SEIR corrected CFR
60 cumulative_cases_seir = N - S
61 cumulative_deaths_seir = D
62 corrected_cfr_seir = cumulative_deaths_seir / np.maximum(
63     cumulative_cases_seir, 1)
64
65 # Mean absolute error between methods
66 error_prob_vs_seir = np.mean(np.abs(corrected_cfr_prob -
67                                     corrected_cfr_seir))
68
69 # Plotting
70 fig, axs = plt.subplots(2, 1, figsize=(12, 10))
71 axs[0].plot(days, cases_daily, label='Daily Cases (Synthetic)',
72             color='#1f77b4', linewidth=2.5)
73 axs[0].plot(days, deaths_daily, label='Daily Deaths (Probabilistic Model)',
74             color='#d62728', linewidth=2.5)
75 axs[0].plot(days, deaths_seir_daily, '--', label='Daily Deaths (SEIR Model)',
76             color='#9467bd', linewidth=2.5)
77 axs[0].set_title('Synthetic Epidemic Data : Cases and Deaths',
78                 fontsize=16)
```

```

79  axs[0].set_xlabel('Days')
80  axs[0].set_ylabel('Counts')
81  axs[0].legend(fontsize=12)
82  axs[0].grid(True, linestyle='--', alpha=0.6)
83
84  axs[1].plot(days, observed_cfr, '--', label='Observed CFR (
85  Uncorrected)', color='#7f7f7f', linewidth=2.5)
86  axs[1].plot(days, corrected_cfr_prob, label='Corrected CFR (
87  Probabilistic Model)', color='#2ca02c', linewidth=2.5)
88  axs[1].plot(days, corrected_cfr_seir, label='Corrected CFR (
89  SEIR Model)', color='#9467bd', linewidth=2.5)
90  axs[1].axhline(true_cfr, linestyle=':', color='grey',
91  linewidth=2, label='True CFR')
92  axs[1].set_title(f'Comparison of CFR Corrections (MAE: {
93  error_prob_vs_seir:.4f})', fontsize=16)
94  axs[1].set_xlabel('Days')
95  axs[1].set_ylabel('Case Fatality Rate')
96  axs[1].legend(fontsize=12)
97  axs[1].grid(True, linestyle='--', alpha=0.6)
98
99  plt.tight_layout()
100 plt.show()

```

Data Availability

The data and codes that support the findings of this study are available in the supplementary material. Additional data and scripts can be made available from the corresponding author upon reasonable request.

Acknowledgements

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq – Brazil), under grants 309440/2022-0 (to Gustavo Zampier dos Santos Lima) and 314906/2023-1 (to Gilberto Corso). We also acknowledge the infrastructure support from the Federal University of Rio Grande do Norte (UFRN), particularly the School of Science and Technology (ECT). The authors thank anonymous reviewers for their constructive feedback, which helped improve the quality of this manuscript.

Author Contributions Statement

J.R.S.L conceived the model and conducted the simulations, G.Z.S.L analyzed the results, J.R.S.L. prepared figures, G.Z.S.L revised the code. Both authors wrote and reviewed the manuscript.

Additional information

Competing interests: The authors declare no competing interests.

References

- Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of autoimmunity*. 2020. 102433.
- Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research*. 2020.
- Shi Z, Hu Z. A review of studies on animal reservoirs of the SARS coronavirus. *Virus research*. 2008. 133: 74-87.
- Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *Journal of medical virology*. 2020. 92: 401-402.
- The official web page for the World Health Organization (WHO). 2020.
- Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MU, Khan K. Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel. *Journal of travel medicine*. 2020. 27: taaa008.
- Zhao S, Lin Q, Ran J, Musa SS, Yang G, 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*. 2020. 92: 214–217.
- Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome *The Lancet respiratory medicine*. 2020. 8: 420-422.
- Huang C, Wang Y, Li X, Ren L, Zhao J, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*. 2020. 395: 497-506.
- Morawska L, Tang JW, Bahnfleth W, Bluyssen PM, Boerstra A, et al. How can airborne transmission of COVID-19 indoors be minimised? *Environment international*. 2020. 142: 105832.
- Banerjee D, Rai M. Social isolation in Covid-19: The impact of loneliness. 2020.
- Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*. 2020.
- Sharma A, Fölster-Holst R, Kassir M, Szepletowski J, Jafferany M, et al. The effect of quarantine and isolation for COVID-19 in general population and dermatologic treatments. *Dermatol Ther*. 2020. 10: e13398.
- Durankuş F, Aksu E. Effects of the COVID-19 pandemic on anxiety and depressive symptoms in pregnant women: a preliminary study. *The Journal of Maternal-Fetal & Neonatal Medicine*. 2020. p. 1–7.
- Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, et al. The socio-economic implications of the coronavirus pandemic (COVID19): A review. *International journal of surgery (London, England)*. 2020. 78: 185.
- Corona Virus Resource Center. type. 2020.
- Hethcote HW. The mathematics of infectious diseases. *SIAM review*. 2000. 42: 599-653.
- Roddam AW. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*: O Diekmann and JAP Heesterbeek, 2000, Chichester: John Wiley pp. 303,£ 39.95. 2001. 8.
- Peng L, Yang W, Zhang D, Zhuge C, Hong L. Epidemic analysis of COVID-19 in China by dynamical modeling. *arXiv preprint arXiv:200206563*. 2020.
- Tuan NH, Mohammadi H, Rezapour S. A mathematical model for COVID-19 transmission by using the Caputo fractional derivative. *Chaos, Solitons & Fractals*. 2020. 140: 110107.
- Chatterjee K, Chatterjee K, Kumar A, Shankar S. Healthcare impact of COVID-19 epidemic in India: A stochastic mathematical model. *Medical Journal Armed Forces India*. 2020.
- Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The lancet infectious diseases*. 2020.
- Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal society A, Mathematical, Physical and Engineering Sciences*. 1927. 115: 700-721.

24. Harko T, Lobo FSN, Mak MK. Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Applied Mathematics and Computation*. 2014. 236: 184-194.
25. Hethcoat HW. *The Mathematics of Infectious Diseases*. SIAM Rev. 2000. 42: 599-653.
26. Huppert A, Katriel G. Mathematical modelling and prediction in infectious disease epidemiology. *Clinical Microbiology and Infection*. 2013. 19: 999-1005.
27. Jenkins DM. *An Examination of Mathematical Models for Infectious Disease*. Honors Research Projects 194, University of Akron. 2015.
28. Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *Jama*. 2020. 323: 1775-1776.
29. Rajgor DD, Lee MH, Archuleta S, Bagdasarian N, Quek SC. The many estimates of the COVID-19 case fatality rate. *The Lancet Infectious Diseases*. 2020. 20: 776-777.
30. Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, et al. Real estimates of mortality following COVID-19 infection. *The Lancet infectious diseases*. 2020.
31. Spychalski P, Błażyńska-Spychalska A, Kobiela J. Estimating case fatality rates of COVID-19. *The Lancet Infectious Diseases*. 2020.
32. Lipsitch M. Estimating case fatality rates of COVID-19. *The Lancet Infectious Diseases*. 2020.
33. Press W, Teukolsky S, Vetterling W, Flannery B. *Numerical Recipes in Fortran. The Art of Scientific Computing*. 2nd ed. Press Syndicate of the University of Cambridge. 1992.
34. Van Seventer JM, Hochberg NS. *Principles of Infectious Diseases: Transmission, Diagnosis, Prevention, and Control*. International Encyclopedia of Public Health. 2017. p. 22.
35. Okell, Lucy C, et al. Have deaths from COVID-19 in Europe plateaued due to herd immunity?. *The Lancet* 395.10241. 2020. 110-111.
36. Randolph HE, Luis BB Herd immunity: understanding COVID-19. *Immunity* 52.5 2020. 737-741.
37. Gandhi RT, Lynch JB, Del Rio C. Mild or moderate COVID-19. *New England Journal of Medicine*. 2020.
38. Prescott HC, Girard TD. Recovery from severe COVID-19: leveraging the lessons of survival from sepsis. *Jama*. 2020. 324: 739-740.
39. Nishiura H, Kobayashi T, Miyama T, Suzuki A, Jung Sm, et al. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *International journal of infectious diseases*. 2020. 94: 154.
40. Lachmann A. Correcting under-reported COVID-19 case numbers. *medRxiv*. 2020.
41. Team CCR, Team CCR, Team CCR, Bialek S, Boundy E, Bowen V, et al. Severe outcomes among patients with coronavirus disease 2019 (COVID-19) United States, Morbidity and mortality weekly report. 2020. 69: 343-346.
42. Liu Y, Yan LM, Wan L, Xiang TX, Le A, et al. Viral dynamics in mild and severe cases of COVID-19. *The Lancet Infectious Diseases*. 2020.
43. GZ Dos Santos Lima, JVT de Lima, JM de Araújo, G Corso, SLEF da Silva, Generalized statistics: Applications to data inverse problems with outlier-resistance, *PLOS ONE*. 2023. 18: e0282578.
44. SLEF da Silva, R Silva, GZ dos Santos Lima, JM de Araújo, G Corso, An outlier-resistant k-generalized approach for robust physical parameter estimation, *Physica A* 600. 2022. 127339.
45. TL Prado, FE Lopes da Cruz, G Corso, GZ dos Santos Lima, Statistical inference for microstate distribution in recurrence plots, *Physica D: Nonlinear Phenomena*, 2023: 133748.
46. G Corso, TL Prado, GZ dos Santos Lima, SR Lopes, Quantifying entropy using recurrence matrix microstates, *Chaos* 2018. 28: 083108
47. SR Lopes, TL Prado, G Corso, GZ dos Santos Lima, J Kurths, Parameter-free quantification of stochastic and chaotic signals, *Chaos, Solitons Fractals*. 2020. 133: 109616.
48. The official Website of the City of New York. 2020.
49. GEOCOVID portal, UFES. 2020.
50. The Official Website of the Italian Ministry of Health. 2020.
51. The Official Website of the Spanish Ministry of Health. 2020.
52. Alwan NA. Surveillance is underestimating the burden of the COVID-19 pandemic. *The Lancet*. 2020. 396: e24.
53. Burgess S, Ponsford MJ, Gill D. Are we underestimating seroprevalence of SARS-CoV-2. 2020.
54. Carenzo L, Costantini E, Greco M, Barra F, Rendiniello V, et al. Hospital surge capacity in a tertiary emergency referral centre during the COVID-19 outbreak in Italy. *Anaesthesia*. 2020.
55. Massonnaud C, Roux J, Crépey P. COVID-19: Forecasting short term hospital needs in France. *medRxiv*. 2020.
56. Bernucci C, Brembilla C, Veiceschi P. Effects of the COVID-19 outbreak in Northern Italy: perspectives from the Bergamo Neurosurgery Department. *World neurosurgery*. 2020. 137: 465.
57. Sorbello M, El-Boghdadly K, Di Giacinto I, Cataldo R, Esposito C, et al. The Italian coronavirus disease 2019 outbreak: recommendations from clinical practice. *Anaesthesia*. 2020. 75: 724-732.
58. Martan MC, Lean C, Cuaat J, Del Nogal F. Intensive Care Services resources in Spain. *Medicina Intensiva*. 2013. 37: 443-451.
59. Ioannidis J. The infection fatality rate of COVID-19 inferred from seroprevalence data. *medRxiv*. 2020.
60. Hallal PC, Barros FC, Silveira MF, Barros AJDd, Dellagostin OA, et al. EPICOVID19 protocol: repeated serological surveys on SARS-CoV-2 antibodies in Brazil. *Ciencia & saude coletiva*. 2020. 25: 3573-3578.
61. Hallal PC. COVID-19 no Brasil: várias epidemias num sã*paÃs Primeira fase do EPICOVID19 reforça preocupação com a região Norte. 2020.
62. Stadlbauer D, Tan J, Jiang K, Hernandez MM, Fabre S, et al. Repeated cross-sectional sero-monitoring of SARS-CoV-2 in New York City. *Nature*. 2020. 1-5.

63. Ioannidis JPA. Infection fatality rate of COVID-19 inferred from seroprevalence data. Bulletin of the World Health Organization. 2020. 99: 19-33
64. Alwan NA. Scientific consensus on the COVID-19 pandemic: we need to act now. The Lancet. 2020. 396: 1261-1262.