PLOS ONE

RESEARCH ARTICLE

Current forecast of COVID-19 in Mexico: A Bayesian and machine learning approaches

Kernel Prieto *

Instituto de Matemáticas, Universidad Nacional Autónoma de México, Mexico City, México

* kernel@ciencias.unam.mx

Abstract

The COVID-19 pandemic has been widely spread and affected millions of people and caused hundreds of deaths worldwide, especially in patients with comorbilities and COVID-19. This manuscript aims to present models to predict, firstly, the number of coronavirus cases and secondly, the hospital care demand and mortality based on COVID-19 patients who have been diagnosed with other diseases. For the first part, I present a projection of the spread of coronavirus in Mexico, which is based on a contact tracing model using Bayesian inference. I investigate the health profile of individuals diagnosed with coronavirus to predict their type of patient care (inpatient or outpatient) and survival. Specifically, I analyze the comorbidity associated with coronavirus using Machine Learning. I have implemented two classifiers: I use the first classifier to predict the type of care procedure that a person diagnosed with coronavirus presenting chronic diseases will obtain (i.e. outpatient or hospitalised), in this way I estimate the hospital care demand; I use the second classifier to predict the survival or mortality of the patient (i.e. survived or deceased). I present two techniques to deal with these kinds of unbalanced datasets related to outpatient/hospitalised and survived/deceased cases (which occur in general for these types of coronavirus datasets) to obtain a better performance for the classification.

Citation: Prieto K (2022) Current forecast of COVID-19 in Mexico: A Bayesian and machine learning approaches. PLoS ONE 17(1): e0259958. https://doi.org/10.1371/journal.pone.0259958

Editor: Simone Lolli, Italian National Research Council (CNR), ITALY

Received: January 23, 2021

Accepted: October 29, 2021

Published: January 21, 2022

Data Availability Statement: All relevant data are within the paper and its Supporting information files.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

1 Introduction

Several mathematical models for disease transmission, and to predict and control disease spread have been proposed because emerging and re-emerging infectious diseases represent a major threat to public health, and may cause large economic and social losses. Vaccination is the principal control measure for reducing the spread of many infectious diseases [1, 2]. However, recent epidemics such as H1N1, Ebola, and MERS-CoV have required strong government interventions for fast eradication [3]. Based on previous pandemics, scientists have warned that another pandemic could strike at any moment. Therefore, a considerable effort has been made to study the impact of control measures to eradicate the outbreak of an epidemic, and in particular an immediate response for a possible influenza pandemic crisis [4]. Mathematical models include compartmental epidemic models, which are deterministic systems of ordinary and partial differential equations or stochastic difference equations [5]. For diseases such as influenza, typhoid fever, anthrax, diphtheria, tetanus, cholera, hepatitis B, pertussis, pneumonia, and coronavirus, the process of transmission between individuals takes

place because of an initial inoculation of a small amount of pathogen units. The pathogen then reproduces quickly within the host during a period of time, which is called the incubation time. During this period, the pathogen affluence is enough to activate transmission to other susceptible individuals [6]. Many mathematical models assume that the disease incubation period is negligible once an individual is infected; that is, the infected individual becomes infectious instantaneously. The compartmental model based on these assumptions is known as the Susceptible-Infectious-Removed (SIR or SIRS) model [7], depending on whether the acquired immunity is permanent or temporal. For viral infections such as rubella and measles, the infected individual acquires permanent immunity. However, many diseases have an incubation (latent) period of time before the hosts become infectious, such as influenza, typhoid fever, anthrax, diphtheria, tetanus, cholera, hepatitis B, pertussis, pneumonia and coronavirus [8]. Meanwhile, diseases with a long immune period include polio, chicken-pox, whooping cough, smallpox and dengue fever. To take this incubation period of the disease into account, another population compartment, named exposed class, E, is incorporated into this type of model (i.e. SIR or SIRS). A susceptible individual who has just been infected first goes through the exposed class during the incubation period of the disease and the exposed individual then becomes infectious. The resulting models are of Susceptible-Exposed-Infectious-Removed (SEIR or SEIRS) type. I note that there is more literature on SIR and SEIR models than SIRS and SEIRS models; that is, those which permanent immunity is not assumed. I refer the reader to [9–11] for references on SEIRS models and [6, 12–16] for references on SEIR models.

Numerous efforts to forecast and produce mathematical control models for disease transmission have been proposed since the re-emergence of the coronavirus named SARS-CoV-2 [17–23]. The first coronavirus outbreak was named SARS-CoV (where SARS stands for severe acute respiratory syndrome), which caused a pandemic with a variety of incidences in 29 countries around the world. A Bayesian compartment model (SEIR: Susceptible, Exposed, Infected and Removed) was presented to study the spread of the first coronavirus in 2002 [24]. The mean incubation period was 5.3 days (95% Credible Interval 4.2 − 6.8 days), which is close to the latter coronavirus mean incubation period, which is reported as 5.1 [25]. In addition, the reported mean recovery period, from symptom onset to recovery, was 21 days (%95 Credible Interval 16 − 26 days), which is higher when compared to the second coronavirus recovery period, which is reported to be around 14 days [26]. The use of social distance as a control strategy for SARS was explored in [27]. The basic and effective reproductive numbers of SARS-Cov were estimated in [28]. In addition, a spatiotemporal analysis of SARS-CoV was presented in [29]. Another type of coronavirus emerged in 2015 in the Republic of Korea, which was named Middle East Respiratory Syndrome Coronavirus (MERS-CoV). Seventeen years after the first appearance of SARS-CoV (November, 2002), another virus strain has emerged; which is called SARS-CoV-2 or COVID-19. Many attempts to predict the dynamics of the coronavirus pandemic have been presented since the start of the second coronavirus outbreak in Wuhan City in December of 2019, some with a Bayesian inference approach [22, 30, 31]. A wide range of predictions have been presented in model calibrations using confirmed-case data because of the nonidentifiability in these models [32].

The rest of this paper is organised as follows. Section 2 describes the mathematical formulation of the contact tracing model for coronavirus disease and it outlines the Bayesian inference framework to predict the dynamics of its spread. Besides predicting the coronavirus cases, mathematical methods are used to forecast hospital care demand and mortality among patients with COVID-19 who present comorbidities related with COVID-19. I aimed to develop models of COVID-19 using Machine Learning to accurately predict both hospital care demand and mortality based on patients who present diseases such as hypertension, obesity, diabetes and smoking. These models and methods are presented in Section 4 using the dataset

[33]. Each section presents the mathematical framework and numerical results. A discussion and some conclusions are presented in the last section 5.

2 Bayesian forecasting

2.1 Model formulation

Control strategies for infectious diseases include effective vaccination [34], early detection, proper treatment, isolation, quarantine, and educational campaigns. With the aim of studying the effect of contact tracing in the propagation of an infectious disease, I formulate a contact tracing model. Here, it is assumed that the disease transmits horizontally (i.e., vertical transmission is neglected). The horizontal transmission can occur either by direct contact (e.g., touching, licking, or biting) or by indirect contact with no physical contact (e.g., vectors or fomites).

The SIR and SEIR frameworks have been used in most current studies of COVID-19 transmission dynamics. Inspired by a full data-driven approach, I have tried to use all of the available reliable data to forecast the spread of the HIV-AIDS disease, keeping in mind that a simple model may fit better than complex models [32]. Next, I formulate a mathematical model considering isolation due to contact tracing as suggested in [6] and the models proposed in [27, 32, 35]. This model analyzes the significance of isolating the probable infected individuals. The total population, N, is divided into the following seven epidemiological classes SsEIQR: susceptible S, suspects (susceptible quarantined) s: people who have had contact with an infectious person or with someone who had contact with an infectious person), exposed E, people who have contracted the virus but are not yet infectious, the undetected infectives A, asymptomatic people, sick people reported in quarantine I (i.e., individuals are isolated at home or in the hospital), recovered people R, and the last state variable P denotes the deceased by coronavirus. I assume that the disease transmission rate, λ, is decomposed of two parts: the disease transmission rate by symptomatic people and by asymptomatic people; $\lambda = \beta_a + \beta_s$. I assume that a fraction q of the contacts whom infected individuals have had recently are sought and isolated. I model contact tracing by forcing a fraction q of those who have recently had contact with an infectious individual to be quarantined, where they will spend an average $1/\tau$ days. Importantly, I assume that these individuals are quarantined before they have a chance to generate any subsequent infection. Because of this latter assumption, contact tracing does not need to be recursive. The parameter α^{-1} and γ^{-1} represents the mean latent period and the recovery period, respectively. The parameter ρ represents the proportion between the symptomatic class and the asymptomatic class. Finally, the parameter σ denotes the death rate by the disease. The description of all the parameters of the contact tracing model proposes here is on Table 1. My

Table 1. Parameters of the contact tracing model (1).

Parameter	Description	Value
β_s	transmission rate of the disease by symptomatic individuals	to be estimated
β_a	transmission rate of the disease by asymptomatic individuals	to be estimated
ρ	the fraction of asymptomatics/symptomatics	to be estimated
$1/\gamma$	recovery period (days)	to be estimated
σ	death rate by the disease	to be estimated
q	proportion of quarantined individuals by contact tracing	to be estimated
$1/\tau$	period of quarantined (days)	14 [26]
$1/\alpha$	latent period (days)	5.1 [25]
E_0	initial condition for exposed class $E(0)$	to be estimated
A_0	initial condition for asymptomatic class $A(0)$	to be estimated
I_0	initial condition for symptomatic class $I(0)$	to be estimated

https://doi.org/10.1371/journal.pone.0259958.t001

suggested model reads as follows

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{((1-q)\beta_s I + \beta_a A)S}{N} - \frac{q\beta_s IS}{N} + \tau s \\
\frac{ds}{dt} &= \frac{q\beta_s IS}{N} - \tau s \\
\frac{dE}{dt} &= \frac{((1-q)\beta_s I + \beta_a A)S}{N} - \alpha E \\
\frac{dA}{dt} &= \rho\alpha E - \gamma A \\
\frac{dI}{dt} &= (1-\rho)\alpha E - (\gamma + \sigma)I \\
\frac{dR}{dt} &= \gamma(A + I) \\
\frac{dD}{dt} &= \sigma I
\end{aligned} \tag{1}$$

The total population $N(t)$ is determined by $N(t) = S(t) + s(t) + E(t) + A(t) + I(t) + R(t) + D(t)$. I note that a more complex model is suggested in [6], considering stages in the exposed and infectious compartments but considering to decompose the force of transmission λ. In model (1), I have assumed that the compartment of suspect people are unexposed people of the disease during the quarantine period similarly to the quarantined compartment, S_q, and the compartment, E_q, of model proposed in [36, 37], respectively. A less similar compartment, Q, to my proposed compartment s is proposed in [38]. A more realistic version would be to consider that people during the quarantine period are exposed to be infected as in [39, 40]. Actually, in [39] is considered a parameter which measures the efficacy of quarantine to prevent the acquisition of infection by quarantined-susceptible individuals during the quarantine period. Finally, these articles Reviews [41, 42] analyze and categorize studies of quarantine through contact tracing.

Future work may explore the contact tracing model in [43], which proposes a very interesting and robust force of transmission λ that is dependent of time and with a delay. A sensitivity analysis shows that λ is the highest sensible parameter in this kind of compartment model. Therefore, it is very important to select this parameter adequately. Further interesting options for contact tracing models can be found in [44]. A robust review of contact tracing models can be found in [45] and quarantine models can be found in [46]. A detailed mathematical analysis of this type of SEIR model can be found in [47, 48].

3 Data

All code and data used to complete these simulations and analyses presented in section 2 based on the Stan Package, the *t- walk* Package is publically available on https://github.com/kernelprieto/COVID_MEX2, and https://github.com/kernelprieto/COVID_MEX1, respectively. All code and data used to complete the simulations and analyses presented in section 4 based on Machine Learning methods is publically available on https://github.com/kernelprieto/COVID-19_Comorbidities.

3.1 Parameter estimation

I used the daily updated data for the parameter estimation [33]. From the mathematical point of view, the parameter estimation of a system of ordinary differential equations is regarded as

an inverse problem. The fitting curve or estimation of the parameters of a model is considered to be an inverse problem from the mathematical point of view. Typically, an optimisation method such as the Landweber in [49–53], or faster methods such as the Levenberg-Marquardt or Conjugate Gradient methods, and regularisation techniques, such as Tikhonov, Sparsity or Total Variation are used to solve this inverse problem. In this manuscript, I used Bayesian inference to solve the inverse problem because it is a tool that combines uncertainty propagation of measured data with available prior information of the model's parameters. It is also a numerically more stable approach than classical methods that rely on the starting parameter point being relatively close to the true one, otherwise the solution obtained corresponds to a local minimum. Moreover, the classical methods only give a point estimate solution instead of a band of solutions using Bayesian inference; that is, in a Bayesian framework, one works with credible intervals. Some studies that have used Bayesian inference include [5, 18, 30, 34, 35, 54–60]. A Bayesian framework to model the spread of the first coronavirus (i.e., SARS-CoV) was presented in [24]. Using Bayesian inference, solutions of the inverse are obtained from the posterior distribution of the parameters of interest, and a solution of interest is obtained using the Maximum a Posterior (MAP). This MAP gives the parameter value for which the posterior density is maximal. I can also calculate the median and quantiles from this posterior sample. As previously mentioned, the Bayesian framework provides a natural and formal way to quantify the uncertainty of the quantities of interest. By denoting the state variable $x = (S(t), s(t), E(t), I(t), Q(t), R(t), P(t)) \in (L^2([0, T])^n$ (i.e., n denotes the number of state variables, here $n = 7$) and the parameters $\theta = (\beta, q, \delta, \alpha, \gamma, \sigma, s(0), E(0), I(0), Q(0)) \in \mathbb{R}^m$ (i.e., m denotes the dimension number of parameters to estimate, here $m = 10$), I can write the model (1) as the following initial value problem

$$\begin{aligned} \dot{x} &= \varphi(x, \theta) \quad (2a) \\ x(0) &= x_0. \quad (2b) \end{aligned}$$

Problem (2), defines a mapping $\Phi(\theta) = x$ from parameters θ to state variables x, where $\Phi : \mathbb{R}_+^m \to (L^2([0, T])^n$, where $\mathbb{R}_+$ denotes the nonnegative real numbers. I assume that Φ has a Fréchet derivative; that is, the mapping $F'(\theta) : \mathbb{R}_+^m \to (L^2([0, T]))^n$, is injective. Thus, the forward problem (2) has a unique solution x for a given θ. The Fréchet derivative of Φ, denoted by Φ', results to be the usual derivative for the system (1) because the domain and range of Φ' are finite dimensional spaces. Usually, not all states of the system can actually be directed measured (i.e., the data consists of measurements of some state variables at a discrete set of points $t_1, \ldots, t_k$); for example, in epidemiology, these data consist of number of cases of confirmed infected people. This defines a linear observation mapping from state variables to data $\Psi : (L^2([0, T]))^n :\to \mathbb{R}^{s\times k}$, where $s \leq n$ is the number of observed variables and k is the number of sample points. Let $F : \mathbb{R}^m \to \mathbb{R}^{s\times k}$ be defined by $F(\theta) = \Psi(\Phi(\theta))$, which is called the forward problem. The inverse problem is formulated as a standard optimisation problem

$$\min_{\theta\in\mathbb{R}^m} \| F(\theta) - y_{\text{obs}} \|^2, \quad (3)$$

such that $x = \Phi(\theta)$ holds, with y_{obs} is the data that has error measurements of size η. Problem (2) may be solved using numerical tools to deal with a nonlinear least-squares problem or the Landweber method, or a combination of both. I implement Bayesian inference to solve the inverse problem (3) in this manuscript. From the Bayesian perspective, all of the state variables x and parameters θ are considered as random variables and the data y_{obs} is fixed. For random variables x, θ, the joint probability distribution density of data x and parameters θ, denoted by $\pi(\theta, x)$, is given by $\pi(\theta, x) = \pi(x|\theta)\pi(\theta)$, where $\pi(x|\theta)$ is the conditional probability distribution,

which is also called the likelihood function, and $\pi(\theta)$ is the prior distribution, which involves the prior information of parameters θ. Given $x = y_{\text{obs}}$, the conditional probability distribution $\pi(\theta|y_{\text{obs}})$, which is called the posterior distribution of θ, is given by the Bayes' theorem:

$$\pi(\theta|y_{\text{obs}}) \propto \pi(y_{\text{obs}}|\theta)\pi(\theta), \tag{4}$$

If additive noise is assumed:

$$y_{\text{obs}} = F(\theta) + \eta,$$

where η is the noise due to discretisation, model error and measurement error. If the noise probability distribution $\pi_H(\eta)$ is known, and θ and η are independent, then

$$\pi(y_{\text{obs}}|\theta) = \pi_H(y_{\text{obs}} - F(\theta)).$$

All of the available information regarding the unknown parameter θ is codified into the a prior distribution $\pi(\theta)$, which specifies our belief in a parameter before observing the data. All of the available information regarding how I obtained the measured data is codified into the likelihood distribution $\pi(y_{\text{obs}}|\theta)$. This likelihood can be seen as an objective or cost function because it punishes deviations of the model from the data. To solve the associated inverse problem (4), one may use the maximum a posterior (MAP)

$$\theta_{\text{MAP}} = \max_{\theta} \pi(\theta|x), \quad \theta_{\text{CM}} = \mathbb{E}[\pi(\theta|x)].$$

I used the dataset $y_{\text{obs}} = (\tilde{s}, \tilde{Q}, \tilde{P})$, which correspond to the suspects, diagnosed sick cases and the deceased, respectively. I note that I have not used the data column corresponding to the recovered here because this data was not been collected in a large range (from the beginning) of days. A Poisson distribution with respect to the time is typically used to account for the discrete nature of these counts. However, the variance of each component of the dataset y_{obs} is larger than its mean, which indicates that there is over-dispersion of the data. Thus, it is more appropriate to use the Negative Binomial likelihood distribution because it has an additional parameter that allows the variance to exceed the mean [34, 60, 61]. In fact, the Negative Binomial is a mixture of Poisson and Gamma distributions, where the rate parameter of the Poisson distribution itself follows a Gamma distribution [61, 62]. Here, I have used the following expression for the Negative Binomial distribution

$$\mathcal{NB}(y|\mu, \phi) = \frac{\Gamma(y+\phi)}{\Gamma(y)\Gamma(\phi)} \left(\frac{\mu}{\mu+\phi}\right)^{y} \left(\frac{\phi}{\mu+\phi}\right)^{\phi}, \tag{5}$$

where μ is the mean of the random variable $y \sim \mathcal{NB}(y|\mu, \phi)$ and ϕ is the overdispersion parameter; that is,

$$\mathbb{E}[Y] = \mu, \quad \text{Var}(Y) = \mu + \frac{\mu^2}{\phi}.$$

I recall that Poisson distribution has mean and variance equal to μ, so $\mu^2/\phi > 0$ is the additional variance of the negative binomial with respect to the Poisson distribution. Therefore, the inverse of the parameter ϕ controls the overdispersion. This is important when selecting its support for parameter estimation. In addition, there are alternative forms of the Negative Binomial distribution. In fact, I have used the first option *neg_bin* of the Negative Binomial distribution of Stan [63]. I acknowledge that some scientists have had success with the second alternative representation of the NB distribution [58]. I assume independent Negative Binomial distributed noise η; that is, all dependency in the data is codified into the contact tracing

model. In other words, the positive definite noise covariance matrix η is assumed to be diagonal. Therefore, using Bayes formula, the likelihood is

$$\pi(\theta|\tilde{s},\tilde{I},\tilde{D}) \propto \pi(\tilde{s}|\theta)\pi(\tilde{I}|\theta)\pi(\tilde{D}|\theta)\pi(\theta).$$

As mentioned earlier, I approximate the likelihood probability distribution corresponding to suspects, diagnosed cases, and deaths with a Negative Binomial distribution

$$\tilde{s}_i \sim \mathcal{NB}(s_i(\theta),\phi_0^2), \quad \tilde{I}_i \sim \mathcal{NB}(I_i(\theta),\phi_1^2), \quad \tilde{D}_i \sim \mathcal{NB}(D_i(\theta),\phi_2^2),$$

where the index i denotes the number time, in our case the number of days, and ϕ_0, ϕ_1 and ϕ_2 are the parameters corresponding to the overdispersion parameter of the Negative Binomial distribution (5), respectively, of each data component.

For independent observations, the likelihood distribution $\pi(y|\theta)$, is given by the product of the individual probability densities of the observations

$$\pi(y_{\text{obs}}|\theta) = \prod_{i=1}^{n} \pi(\tilde{s}_i|\theta)\pi(\tilde{I}_i|\theta)\pi(\tilde{D}_i|\theta),$$

where the mean μ of the negative binomial distribution $\mathcal{NB}(I_i(\theta),\phi_1^2)$, is given by the solution $I(t)$ of the model (1) at time $t = t_i$. Analogously, the mean for the negative binomial distributions $\mathcal{NB}(s_i(\theta),\phi_0^2)$ and $\mathcal{NB}(D_i(\theta),\phi_2^2)$ are the solutions $s(t)$ and $D(t)$ of (1) at time t_i, respectively. For the prior distribution, I select LogNormal distribution for the β parameter and Uniform distributions for the rest of parameters to estimate: q, δ, α, γ, σ, s_0, E_0, I_0, Q_0.

$$\pi(\theta) = \prod_{i=1}^{n} \mathcal{LN}(a_\beta,b_\beta)\mathcal{U}(a_q,b_q)\mathcal{U}(a_\delta,b_\delta)\mathcal{U}(a_\alpha,b_\alpha)\mathcal{U}(a_\gamma,b_\gamma) \tag{6}$$

$$\times\, \mathcal{U}(a_{s_0},b_{s_0})\mathcal{U}(a_{E_0},b_{E_0})\mathcal{U}(a_{I_0},b_{I_0})\mathcal{U}(a_{Q_0},b_{Q_0}). \tag{7}$$

The posterior distribution $\pi(\theta|y_{\text{obs}})$ given by (4) does not have an analytical closed form because the likelihood function, which depends on the solution of the nonlinear SsEAIRD model, does not have an explicit solution. Then, I explore the posterior distribution using the Stan Statistics package [63], general purpose Markov Chain Monte Carlo Metropolis-Hasting (MCMC-MH) algorithm to sample it, the package *t- walk* [64]. Both algorithms generate samples from the posterior distribution $\pi(\theta|y_{\text{obs}})$ that can be used to estimate marginal posterior densities, mean, credible intervals, percentiles, variances, and so on. I refer the reader to [65] for a more complex description of MCMC MH algorithms. The dataset in [33] contains the information regarding the number of diagnosed cases, deaths, and suspects. Figs 1–3 show the results of forecasting the disease using the Stan package [63]. Fig 1 show the credible intervals of parameters of model (1) within 95% Highest-Posterior Density. Fig 1 shows the results of six chains of the MCMC MH algorithm. Fig 2 shows the incidence analysis for Mexico considering data for the first 182 days of the pandemic. Left column corresponds to the inference analysis using the Stan Package. Right column corresponds to the inference analysis using the *t- walk* Package. Row from top to bottom correspond to the confirmed cases, deceases and suspects. Posterior uncertainty is illustrated with the blue shadow areas within the 95% Highest-Posterior Density. Red bars correspond to the data, i.e., the confirmed cases, deceases and suspects. Blue line denotes the median, and the purple line on the right column correspond to the mode. Fig 3 shows Joint probability density distributions of parameters of model (1) within 95% HPD using the Stan Package [63]. The blue lines represent the medians. Table 2 shows the parameter estimated using the Stan package with the quantiles 2.5%, 25%, 50%, 75%,

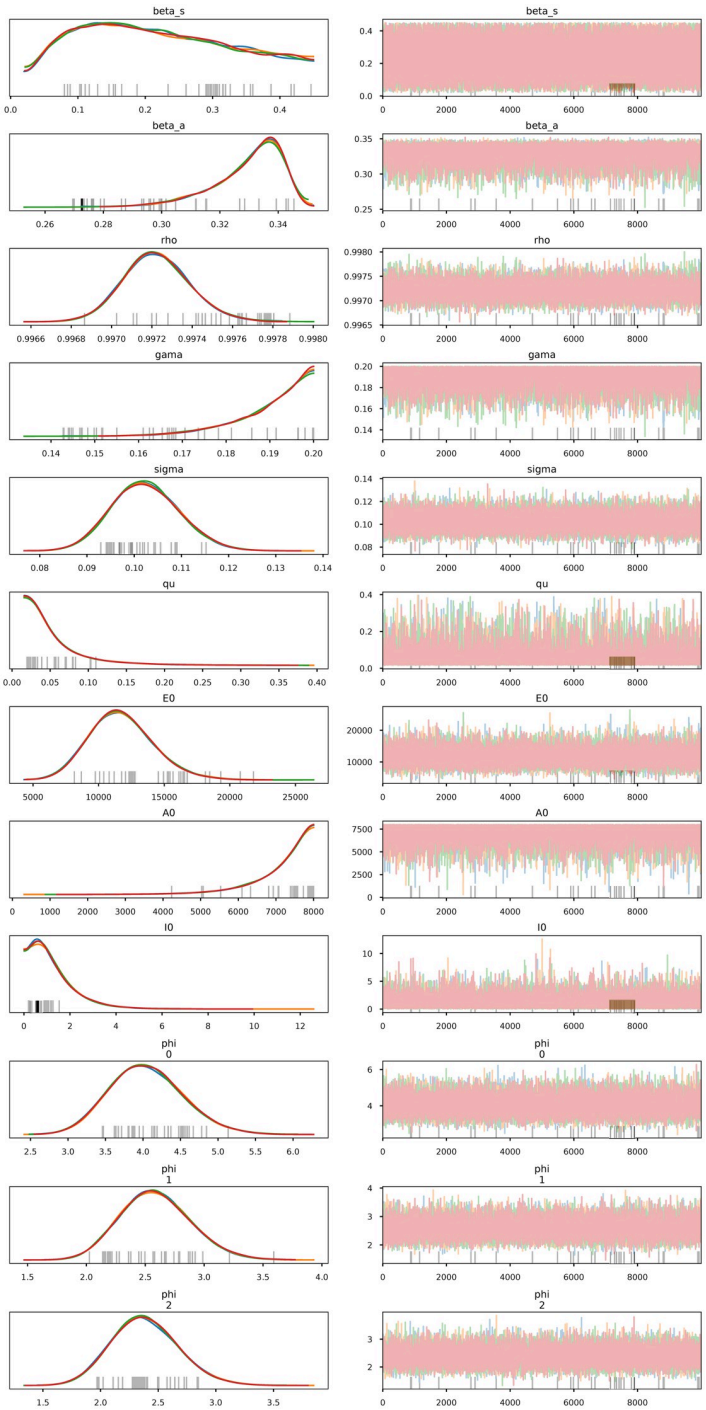

Fig 1. Credible intervals of parameters of model (1). Credible intervals within 95% Highest-Posterior Density (HPD).

https://doi.org/10.1371/journal.pone.0259958.g001

97.5%. I perform 20000 iterations, with 10000 of them as a burn-in. I have used the interface in Python (PyStan). I have used the Hamilton Monte Carlo and No-U-Turn Sampler (NUTS) algorithms, obtaining similar performance. I point out that using Automatic Differentiation Variational Inference (ADVI) is much faster than the previously mentioned algorithms, with

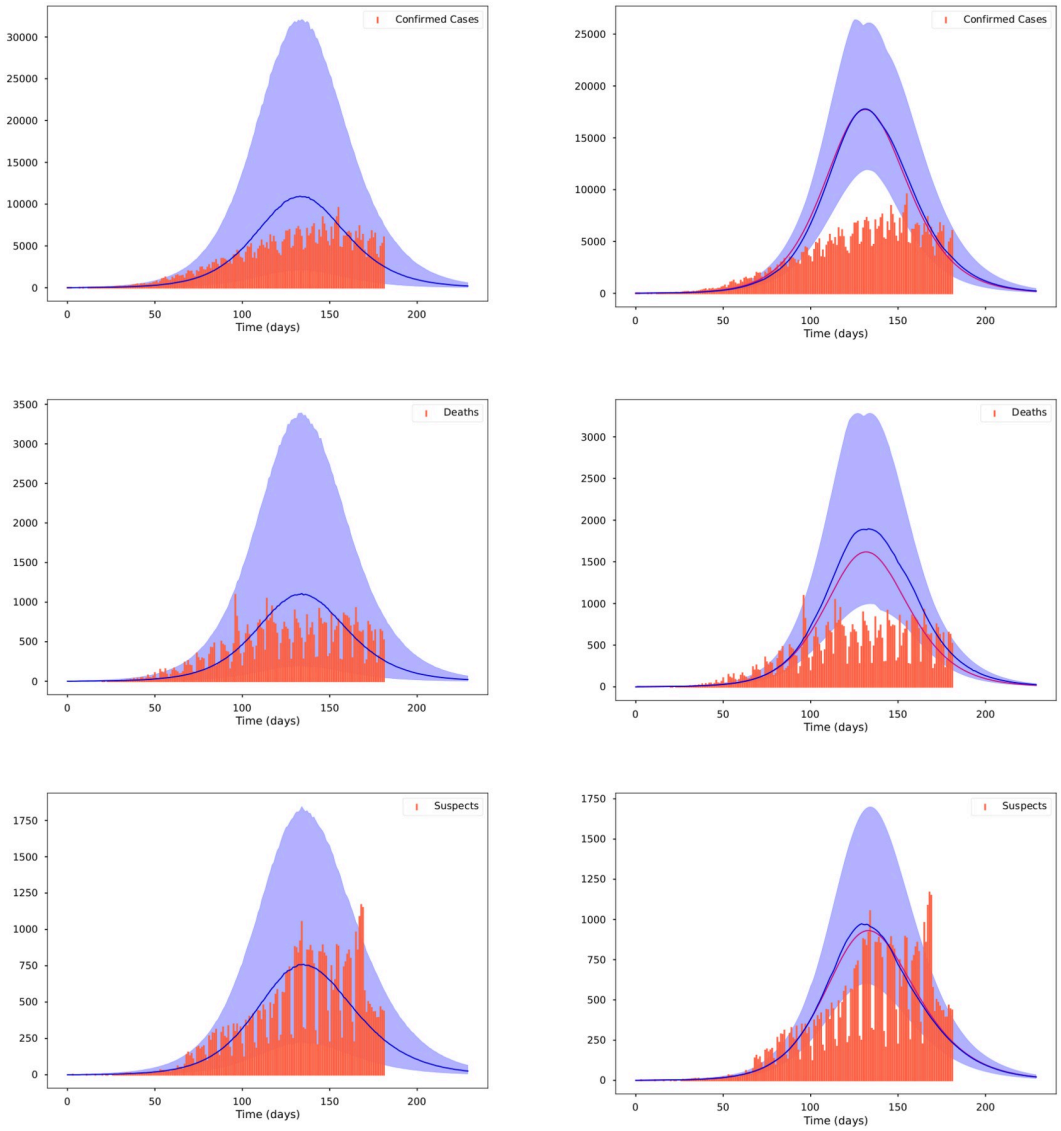

Fig 2. Incidence analysis for Mexico considering data for the first 182 days of the pandemic (until 6 August 2020). Left column corresponds to the inference analysis using the Stan Package. Right column corresponds to the inference analysis using the *t- walk* Package. Row from top to bottom correspond to the confirmed cases, deceases and suspects. Posterior uncertainty is illustrated with the blue shadow areas within the 95% Highest-Posterior Density. Red bars correspond to the data, i.e., the confirmed cases, deceases and suspects. Blue line denotes the median, and the purple line on the right column correspond to the mode.

https://doi.org/10.1371/journal.pone.0259958.g002

very similar results. Fig 4 and the right-hand column of Fig 2 show corresponding results using the *t- walk* package (the Python version). Fig 4 shows the credible intervals for the estimated parameters of model (1) within %95 of HPD using the *t- walk* Package [64]. Top row from left to right, the parameters: β_s, β_a, ρ. Middle row from left to right: γ, σ, q. Middle row from left to right: E_0, A_0, I_0. Bottom row from left to right: ϕ_0, ϕ_1, ϕ_2 I performed 600000 iterations with 300000 of them as burn-in. Using both packages, I have made predictions until the day 240, meaning 16 October. Future work will analyze the identifiability of the parameters of model (1), as suggested in [59, 66, 67], specifically the ρ parameter, because this parameter is multiplied by the period of incubation of the disease, α. Thus, estimating both parameters

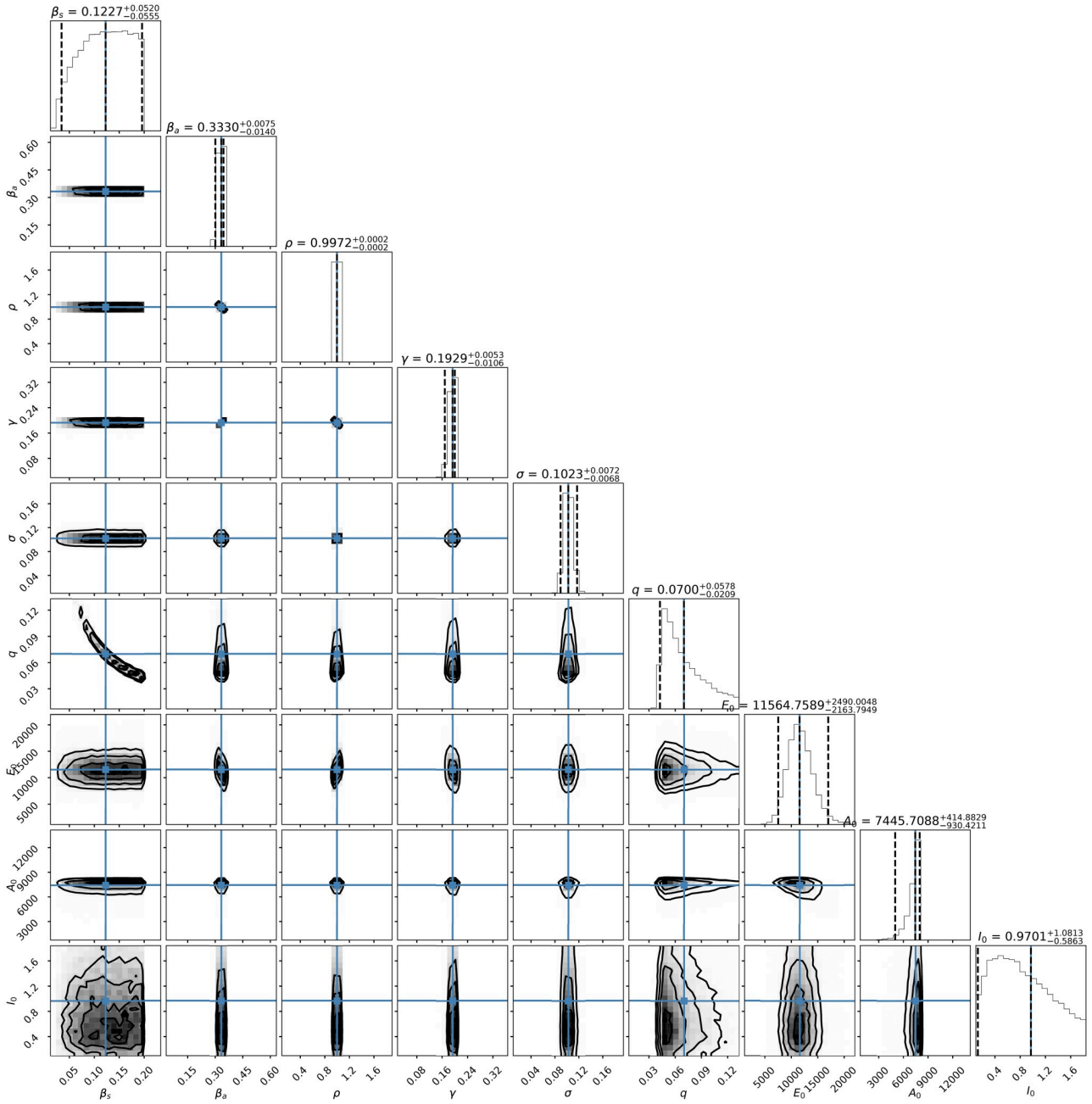

Fig 3. Joint probability density distributions of the estimated parameters. Joint probability density distributions of parameters of model (1) within 95% HPD using the Stan Package [63]. The blue lines represent the medians.

https://doi.org/10.1371/journal.pone.0259958.g003

simultaneously may lead to the nonidentifiability difficulty. In this work, I have assumed the value for the period of incubation of the disease given, equal to 5.1 days [25].

4 Clinical analysis with machine learning

In this section, I describe the methods to predict both hospital care and mortality using Machine Learning based on patients who have been diagnosed with morbidities such as

Table 2. Estimation of the parameters of the model (1).

	mean	2.5%	25%	50%	75%	97.5%
β_s	0.222620	0.044925	0.128393	0.211603	0.312559	0.433075
β_a	0.329556	0.299295	0.323674	0.332539	0.338223	0.344950
ρ	0.997225	0.996923	0.997114	0.997219	0.997329	0.997565
γ	0.190190	0.167331	0.186027	0.192736	0.196920	0.199693
σ	0.102499	0.089572	0.097630	0.102195	0.107137	0.116798
q	0.056694	0.019540	0.027548	0.040514	0.066996	0.193501
E_0	11780.966835	7526.224012	10093.482221	11627.748681	13273.949132	17010.362332
A_0	7182.017997	4917.603144	6872.829362	7445.705247	7765.232788	7977.776073
I_0	1.220856	0.130659	0.540569	0.970296	1.627982	3.748581
ϕ_0	4.048778	3.175233	3.712320	4.025514	4.363766	5.062059
ϕ_1	2.589235	2.060252	2.384828	2.576913	2.778927	3.192292
ϕ_2	2.389033	1.868520	2.191568	2.373326	2.573794	2.989858

Posterior estimation for the parameters of the contact tracing model (1) using the Stan Package [63]. First column correspond to the mean, then the respective percent of the Highest-Posterior Density.

https://doi.org/10.1371/journal.pone.0259958.t002

hypertension, obesity, diabetes and smoking. Thus, I describe the comorbidity associated with coronavirus in Mexico using the [33] dataset. I have performed Machine Learning techniques on it as follows. First, I implemented a predicted classifier for the kind of patient, a person already diagnosed with coronavirus and who has got one or more of the most relevant chronic diseases (i.e., hypertension or diabetes). I have used prediction methods in Machine Learning, such as Logistic Regression, Decision Tree, and K-Neighbors classifiers, the naive Bayes (Bernoulli), and even the powerful methods such as XGBoost and Random Forest through the Sci-Kit-learn package. Fig 5 shows the covariance matrix of the most relevant chronic diseases with respect to the two types of patient: outpatient or hospitalised patient. I can observe in Fig 5 that the most relevant chronic diseases with respect to the type of patient(outpatient or hospitalised) who has been diagnosed with coronavirus in Mexico are hypertension and diabetes. Table 3 shows the contingency table of these two chronic diseases with respect to the type of patient. Table 4 shows the contingency table of these two chronic diseases with respect to the patient's survival possibility. Fig 6A shows the relationship in percent between outpatients and hospitalised patients. Fig 6B shows the confusion matrix result using classical Machine Learning Methods. I could add more characteristics such as age(range) to obtain more true negative cases because the differences in proportion of outpatient and hospitalised decreases. Next, instead of considering the type of patient (outpatient and hospitalised), I consider if the patient survives or dies once diagnosed with coronavirus. Fig 7 shows the covariance matrix of the most relevant chronic diseases with respect to the two types of patient: survived or deceased. One can see Fig 7 that the most relevant chronic diseases with respect to the survival of a person who has been diagnosed with coronavirus in Mexico are hypertension and diabetes. Fig 8A shows the relationship in percent between outpatients and hospitalised patients. Fig 8B shows the confusion matrix result using Logistic Regression. We point out that similar results are obtained using other Machine Learning methods such as Decision Tree, and K-Neighbors, XGBoost and Random Forest. By adding more characteristics such as age (range), one obtains similar results to Fig 8B; that is, one obtains zero true negative predictions. I remind the reader that false negatives and false positives are the two type of errors of rejecting the hypothesis when it was actually true and accepting the hypothesis when it was actually false. Under different circumstances, one type of error may be more critical than the other. For example,

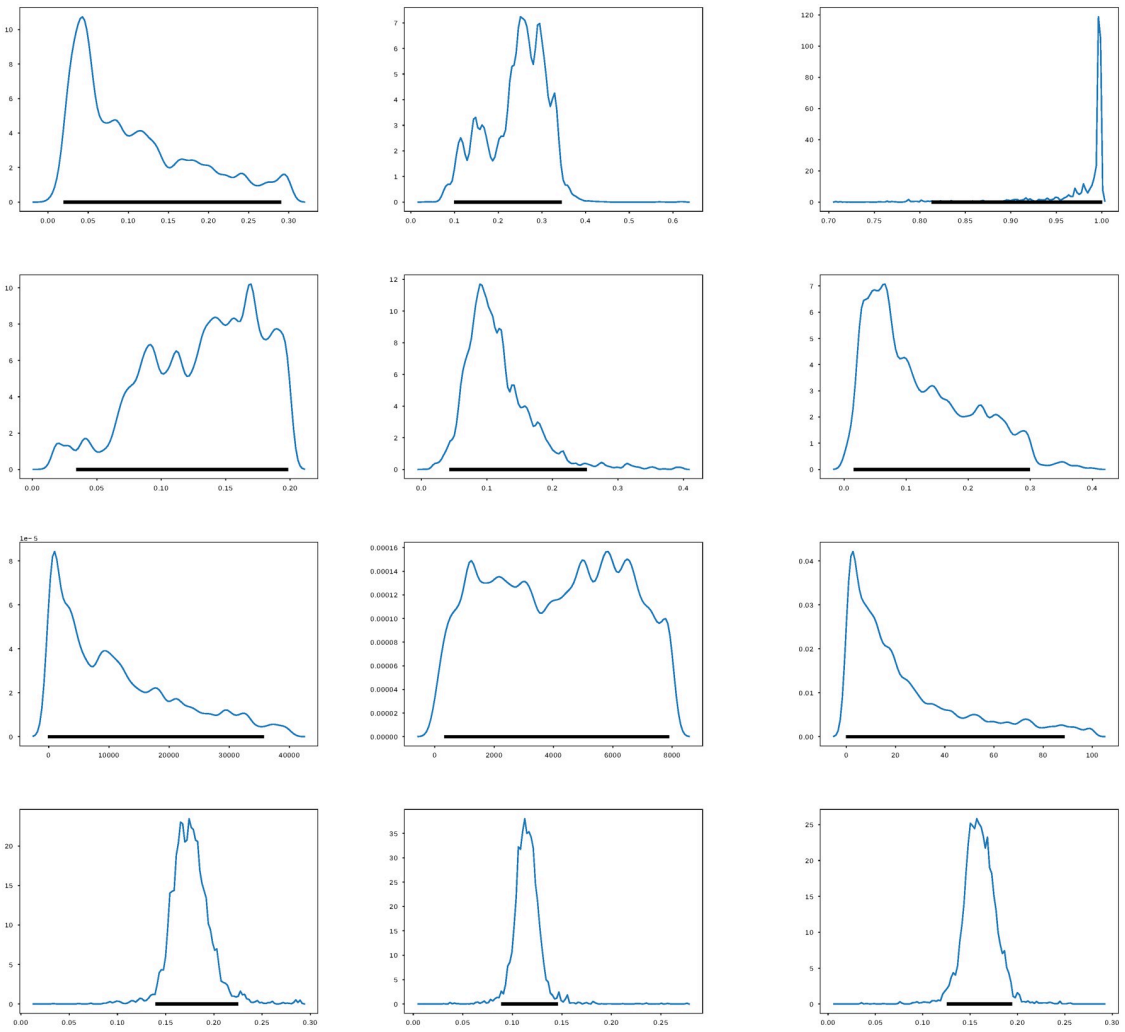

Fig 4. Credible intervals for the estimated parameters. Credible intervals for the estimated parameters of model (1) within %95 of HPD using the *t- walk* Package [64]. Top row from left to right, the parameters: β_s, β_a, ρ. Middle row from left to right: γ, σ, q. Middle row from left to right: E_0, A_0, I_0. Bottom row from left to right: ϕ_0, ϕ_1, ϕ_2.

https://doi.org/10.1371/journal.pone.0259958.g004

diagnosis of cancer would rather accept false positives than false negatives. The main difficulty in trying to predict if a person will survive assuming that they have either hypertension or diabetes is the rather unbalanced proportion between the two classes: survived and deceased. Unbalanced data is assumed with a category less than 20 percent. The lethality of coronavirus in the world is typically not greater than 15 percent.

As can be seen in Fig 8B, the true positives are very high but the prediction of true negatives is zero. I propose two options to deal with this difficulty. First, I have created a naive Bayes Multi-variate Bernoulli algorithm from scratch, as suggested in [68]. This algorithm was originally proposed as an anti-spam email filter. Analogous to their description of how to classify spam emails, a person with vector $x = \langle x_1, \ldots, x_m \rangle$; that is, with multiple features but each one is assumed to be a binary-valued variable. In the case of comorbidity, x represents the type of disease. The decision rule for Bernoulli naive Bayes is based on the probability that a vector x

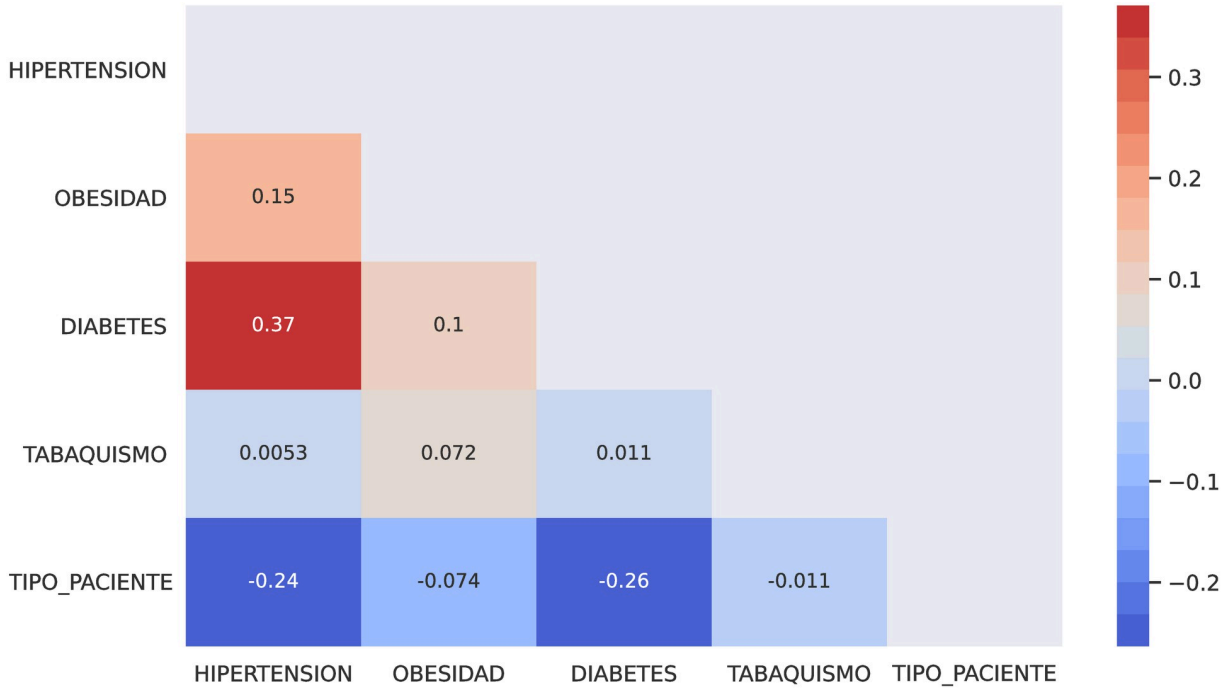

Fig 5. Covariance matrix of the most relevant chronic diseases with respect to the type of patient (outpatient/hospitalized) in Mexico.

https://doi.org/10.1371/journal.pone.0259958.g005

belongs in category c:

$$p(c|x) = \frac{p(c)p(x|c)}{p(x)}. \tag{8}$$

Given that the denominator does not depend on the category, NB classifies each "message" in the category that maximises the numerator in (8); that is, $p(c)p(x|c)$. In the case of a "spam filter", this is equivalent to classifying a message as spam whenever:

$$\frac{p(c_s)p(x|c_s)}{p(c_s)p(x|c_s) + p(c_h)p(x|c_h)} > \delta, \tag{9}$$

with $\delta = 0.5$, where c_h and c_s denote the ham and spam categories. The important part doing this algorithm from scratch is that I can vary δ to obtain more true negatives at the expense of true positives, or vice versa. In our case, I increased the true negatives, the number of true positives are very high using whatever classifier is mentioned. Consequently, I can tune the

Table 3. Contingency table of patient outpatient versus hospitalized.

Hipertension	Diabetes	Hospitalized	Outpatient
0	0	0.1871	0.8128
0	1	0.4757	0.5242
1	0	0.4069	0.5930
1	1	0.5846	0.4153

First two columns correspond the most relevant comorbidities with respect to COVID-19 in Mexico against the type of patient: outpatient and hospitalized using data [33].

https://doi.org/10.1371/journal.pone.0259958.t003

Table 4. Contingency table of the survival of a hospitalized patient.

Hipertension	Diabetes	Deceased	Survived
0	0	0.0626	0.9374
0	1	0.2068	0.7931
1	0	0.1936	0.8063
1	1	0.3035	0.6964

First two columns correspond the most relevant comorbidities with respect to COVID-19 in Mexico against the chance of survival of a hospitalized patient using data [33].

https://doi.org/10.1371/journal.pone.0259958.t004

threshold number of acceptance on the following formula 9. I selected $\delta = 0.45$ (instead of 0.5) and obtained the following confusion matrix. Fig 9A shows the confusion matrix result using the Naive Bayes method, the percent of true negatives has increased approximately to 2.6, and the false negative has decreased, although the false negative has also increased.

Second, I propose to use the Synthetic Minority Oversampling Technique (SMOTE) function to balance the minority class (people who passed away due to coronavirus). SMOTE briefly consists of synthesising elements for the minority class, based on those that already exist. This works randomly by picking a point from the minority class and then computing the k-nearest neighbors for this point. The synthetic points are added between the chosen point and its neighbors. Fig 9B shows the result using the SMOTE technique. Another filter to predict the survival/mortality of an individual apart from the age of the patient, could be if the patient is already admitted to the hospital, this could result in having not a fully unbalanced dataset.

5 Discussion and conclusions

In section 2, I formulate a contact tracing model for the transmission of the COVID-19 and forcast the number of coronavirus cases using Bayesian inference based on two independent software packages: the Stan package [63] and the *t- walk* package [64]. Future work should address the identifiability of the parameters of model (1), as suggested in [59, 66, 67], specifically the ρ parameter, because this parameter is multiplied by the period of incubation of the

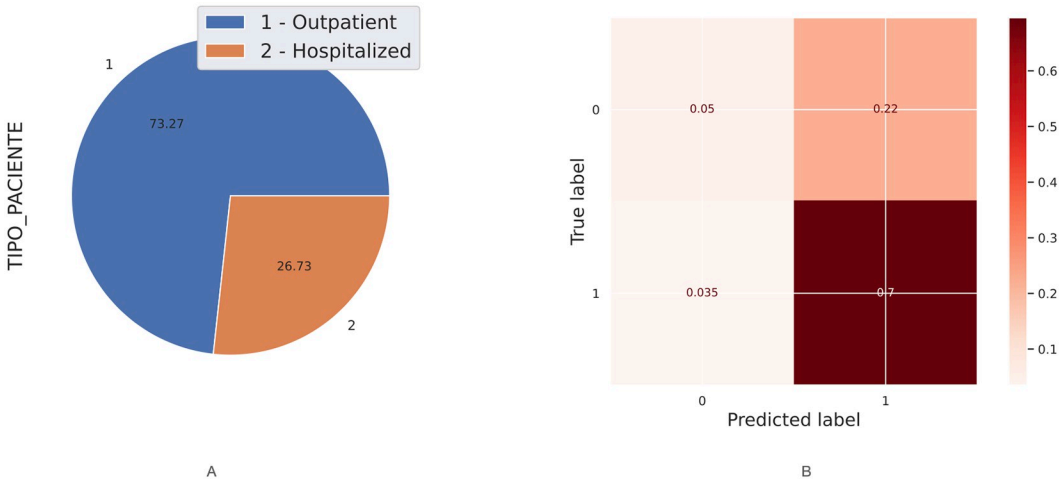

Fig 6. Analysis of outpatient versus hospitalized patients. (A): Percent relation between outpatient and hospitalized patients. (B): Confusion matrix using classical Machine Learning Methods.

https://doi.org/10.1371/journal.pone.0259958.g006

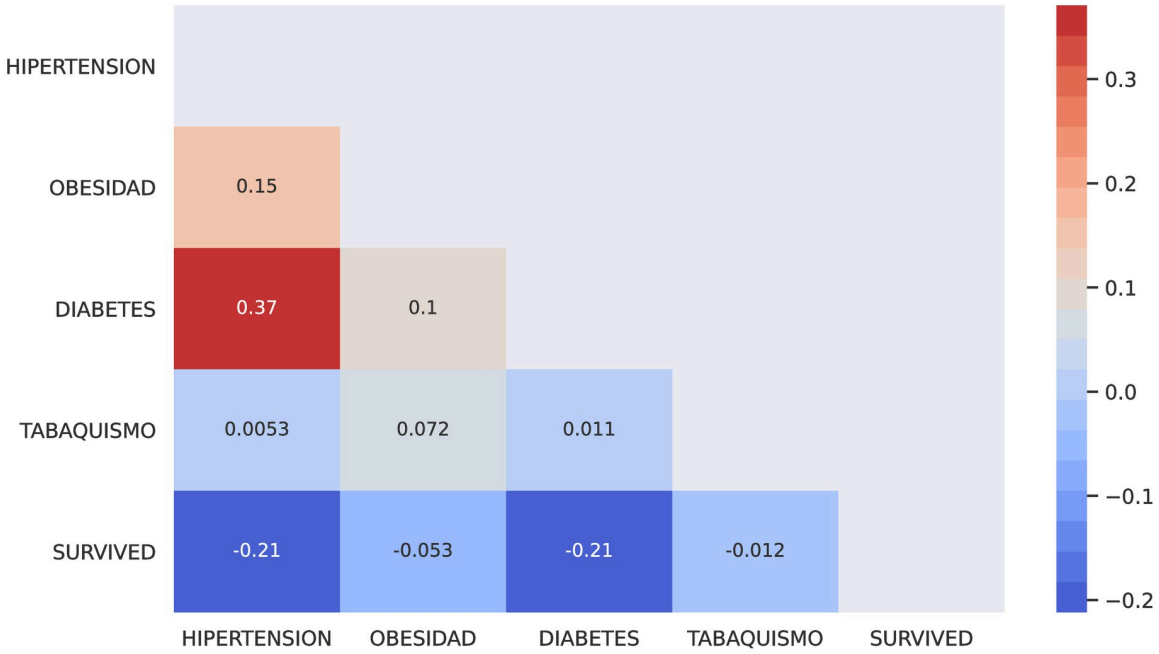

Fig 7. Covariance matrix of the most relevant chronic diseases with respect to the survival chance of a hospitalized patient in Mexico.

https://doi.org/10.1371/journal.pone.0259958.g007

disease, α. Thus, estimating both parameters simultaneously may lead to the nonidentifiability difficulty. In this work, I have assumed the value for the period of incubation of the disease given, equal to 5.1 days [25]. The value estimated for the parameter, ρ, which refers to the proportion of symptomatics and asymptomatics, was around.99, which indicates that a large percent are asymptomatic to this disease. This values is rather high compared with other results nowadays in the literature. This could due to the fact that it was assumed the value of incubation known, and this value could be incorrect for Mexico. I show trace plots, credible intervals, bands projections with medians and a MAP curve (for the *t- walk* case) and the joint crosstab

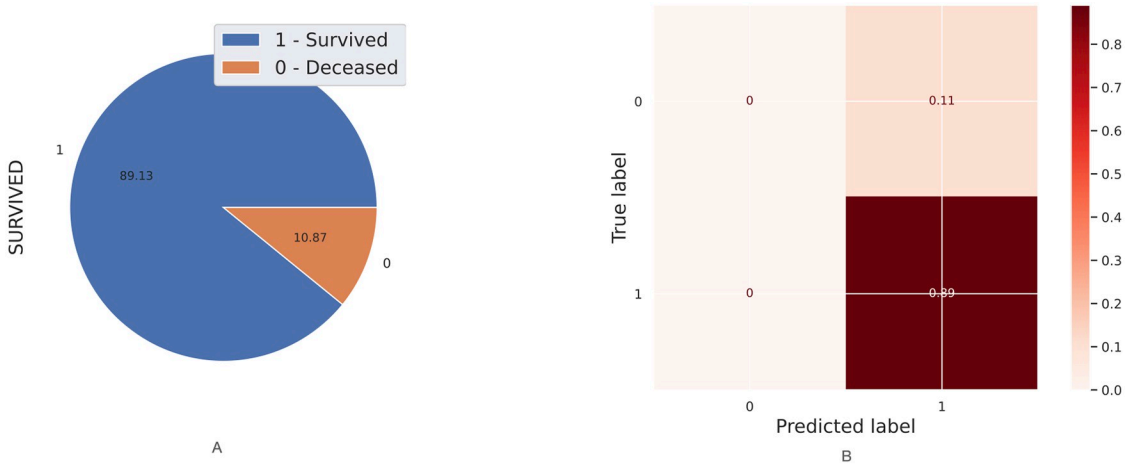

Fig 8. Analysis of survived versus deceased patients. (A): Percent relation between survived and deceased patients. (B): Confusion matrix using classical Machine Learning Methods.

https://doi.org/10.1371/journal.pone.0259958.g008

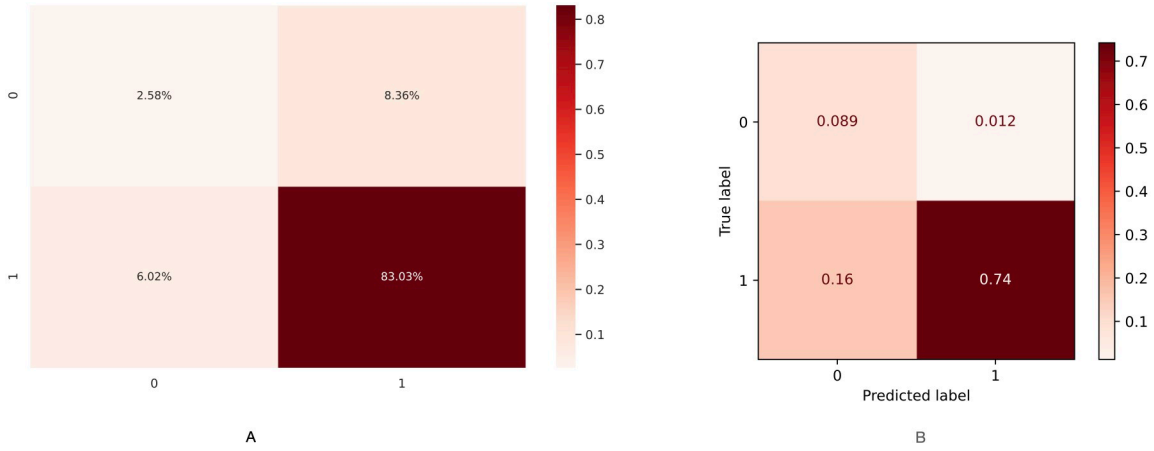

Fig 9. Improved prediction for the survival chance of a hospitalized patient. (A): Confusion matrix applying Naive Bayes with threshold $\delta = 0.45$. (B): Confusion matrix using the SMOTE.

https://doi.org/10.1371/journal.pone.0259958.g009

probability distributions given as a corner. From Fig 2, I can observe that the government of Mexico took some measures to control the transmission of the disease. The model has many implicit assumptions which may be incorrect, e.g., it assumes that the transmission rate is constant and homogeneous through the whole country, which is by far incorrect [34], that is, we can certainly say that every region state in Mexico has its own pandemic, and it is not true that mobility from the North to the South in Mexico is the same as in a specific state of Mexico. A better projection for Mexico City, which has a considerable percentage of coronavirus in the whole country can be found in [69]. Also, the model does not take into account the government interventions, which in each state were announced by a color of the traffic light, red meaning almost all the activities had to be suspended, yellow, some of the activities could reactivate, and green, a considerable percentage of activities could reactivate, depending of each state government. These interventions could be added in the transmission rate in model (1) as in [35]. Despite this, the contact tracing model proposed here could be useful for public health to have a big picture how the pandemic is developing in the country. Also, if a efficient surveillance system is implemented in a pandemic, i.e., where suspects are traced and counted with a small uncertainty, this model could be rather useful for Health systems to make appropriate interventions. Another asset of the current model proposed is that it is simple and computationally efficient.

In section 4, I explore methods using Machine Learning to predict the hospital care demand and mortality based on patients who have been diagnosed with comorbidities with COVID-19. Firstly, the most relevant comorbidities with COVID-19 associated with both hospital care demand and mortality are hypertension and diabetes. Observing the confussion matrix of the predictor for the hospital care demand or the type of patient of coronavirus, mostly true positives (outpatient) 70% are predicted well, but a small percentage 5% of true negatives (hospitalized) are predicted well, moreover, a considerable 22% of false positives is obtained and a small 3.5% of false negatives. Thus, from around a 26% of hospitalized patients (Fig 6), I can predict well only a 5% of the patients who need hospital care. Also, on the one hand, the error type II, i.e., the false positives, is rather big, meaning that using this binary classifier, I would send home people who indeed needed hospital care. On the other hand, the false negatives is small, 3.5%, meaning that I incorrectly send patients to the Hospital when they indeed do not need Hospital care, taking rest at home and following Doctor's advises would be enough. Under

different circumstances, one type of error (type I or II) may be more critical than the other. If the hospital occupancy is relatively high, e.g., equal or higher than 80%, having a high number of false negatives would be risky since the Hospital could collapse. Otherwise, having a high number of false positive would be preferable instead of having false negatives. This projection inaccuracy is due to the unbalanced on the data related with the outpatients versus hospitalized ones. Although, there is no fully unbalanced, this dataset present a considerable majority of outpatients with respect hospitalized people.

Something worse happens when trying to predict the mortality patients with COVID-19, only true positives (survived ones) 89% can be predicted well, and a 0% of true negatives (deceaced ones) can be predicted, and significant error type II, false positives one obtains, i.e., one would give to 11% of people, a survivable expectancy when in fact, they will decease. This projection inaccuracy again is due to the unbalanced on the data related with the outpatients versus hospitalized ones since the lethality of coronavirus in the world is typically not greater than 15 percent. Therefore, I present two methods to deal with unbalanced data because it is the first case of a coronavirus dataset in the world, especially for the case of survived/deceased: first, I propose to use the Naive Bayes method; and second, I propose to use the SMOTE technique. Using the Naive Bayes method leads to a decrease of true positives to 83% (before was 89%) but obtaining a nonzero true negatives percentage 2.58%, also the false positives decreased to the value of 8.36% (before was 11%) and the false negatives increased to a nonzero value of 6.02% (before was 0%). As it was mentioned above, if the hospital occupancy is equal or higher than 80%, having a high number of false negatives would be risky since the Hospital could collapse. Otherwise, having a high number of false positive would be preferable instead of having false negatives. In case of using the SMOTE technique leads to a decrease of true positives to 74% (before was 89%) but obtaining a nonzero true negatives percentage 8.9%, which is rather significant since the proportion of people who survived and deceased is 89.13% versus 10.87%. Also the false positives decreased to the value of 1.2%(before was 11%) and the false negatives increased to a nonzero value of 16.0% (before was 0%). Thus, the value of false negatives obtained using the SMOTE technique is 2.65 times greater than the false negatives value obtained using the Naive Bayes method. As it was explained, unless the hospital occupancy is higher than 80%, it is less risky to have a bigger number of false positive than false negatives.

Supporting information

S1 File.
(CSV)

S2 File.
(CSV)

S3 File.
(XLSX)

S4 File.
(XLSX)

Author Contributions

Conceptualization: Kernel Prieto.

Data curation: Kernel Prieto.

Formal analysis: Kernel Prieto.

Investigation: Kernel Prieto.

Methodology: Kernel Prieto.

Software: Kernel Prieto.

Validation: Kernel Prieto.

Visualization: Kernel Prieto.

Writing – original draft: Kernel Prieto.

Writing – review & editing: Kernel Prieto.

References

1. Kim S, Kwon HD, Lee J. Constrained optimal control applied to vaccination for influenza. Computers and Mathematics with Applications. 2016; 71:2313–2329. https://doi.org/10.1016/j.camwa.2015.12.044 PMID: 32288204
2. Heesterbeek H, Anderson R, Andreasen V, Bansal S, Daniela A. Modeling infectious disease dynamics in the complex landscape of global health. Science. 2020; 347 (6227).
3. Bolzoni L, Bonacini E, Soresina C, Groppi M. Time-optimal control strategies in SIR epidemic models. Mathematical Biosciences. 2017; 292:86–96. https://doi.org/10.1016/j.mbs.2017.07.011 PMID: 28801246
4. Kim S, Lee J, Jung E. Mathematical model of transmission dynamics and optimal control strategies for 2009 A/H1N1 influenza in the Republic of Korea. Journal of Theoretical Biology. 2017; 412:74–85. https://doi.org/10.1016/j.jtbi.2016.09.025 PMID: 27769686
5. Brown G, Porter A, Oleson J, Hinman J. Approximate Bayesian computation for spatial SEIR(S) epidemic models. Spatial and Spatio'temporal Epidemiology. 2018; 24(10):2685–2697. https://doi.org/10.1016/j.sste.2017.11.001 PMID: 29413712
6. Keeling M, Rohani P. Modeling Infectious Diseases in humans and animals. Princeton University Press; 2008.
7. Liu W, Levin S, Iwasa Y. Influence of non-linear incidence rates upon the behaviour of SIRS epidemiological models. J Math Biol. 1986; 23:187–204. https://doi.org/10.1007/BF00276956 PMID: 3958634
8. Anderson R, May R. Population biology of infectious diseases: Part I. Nature. 1977; 280:361–367. https://doi.org/10.1038/280361a0
9. Greenhalgh D. Hopf bifurcation in epidemic models with a latent period and nonpermanent immunity. Mathl Comput Modelling. 1997; 25(2):85–107. https://doi.org/10.1016/S0895-7177(97)00009-5
10. Waltman P. Deterministic threshold models in the theory of epidemics. vol. 1. Springer-Verlag, New York; 1974.
11. Liu W, Hethcote H, Levin S. Dynamical behaviour of epidemiological models with non-linar incidence rates. J Math Biol. 1987; 25:359–380. https://doi.org/10.1007/BF00277162 PMID: 3668394
12. Li M, Muldowney J. Global stability for the SEIR model in epidemiology. Mathematical Biosciences. 1995; 125(4):155–164. https://doi.org/10.1016/0025-5564(95)92756-5 PMID: 7881192
13. Li M, Wang L. Global stability in some SEIR epidemic models. In: Mathematical Approaches for Emerging and Reemerging Infectious Diseases: Models, Methods, and Theory. 2. Springer New York; 2002. p. 295–311.
14. Alonso-Quesada G, De la Sen M, Ibeas A. On the discretization and control of an SEIR epidemic model with a periodic impulsive vaccination. Commun Nonlinear Sci Numer Simulat. 2017; 42:247–274. https://doi.org/10.1016/j.cnsns.2016.05.027
15. Korobeinikov A. Lyapunov functions and global properties for SEIR and SEIS epidemic models. Mathematical Medicine and Biology. 2004; 21:75–83. https://doi.org/10.1093/imammb/21.2.75 PMID: 15228100
16. Kornienko I, Paiva L, de Pinho M. Introducing state constraints in optimal control for health problems. Procedia Technology. 2017; 17:415–422. https://doi.org/10.1016/j.protcy.2014.10.249
17. Adhikari R, Bolitho Aea. Inference, prediction and optimization of non-pharmaceutical interventions using compartment models: the PyRoss library. arXiv e-prints. 2020;.
18. Chatzilena A, Leeuwen E, Ratmann O, Baguelin M, Demiris N. Contemporary statistical inference for infectious disease models using Stan. Epidemics. 2019; 29. https://doi.org/10.1016/j.epidem.2019.100367 PMID: 31591003

19. Lin Q, Zhao Sea. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. International Journal of Infectious Diseases. 2020; 93:211–216. https://doi.org/10.1016/j.ijid.2020.02.058 PMID: 32145465

20. Flaxman S, Mishra S, Gandy Aea. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature. 2020; 584:257–261. https://doi.org/10.1038/s41586-020-2405-7 PMID: 32512579

21. Zhou C. Evaluating new evidence in the early dynamics of the novel coronavirus COVID-19 outbreak in Wuhan, China with real time domestic traffic and potential asymptomatic transmissions. medRxiv. 2020;.

22. Dehning J, Zierenberg J, Spitzner P, Wibral M, Neto J, Wilczek M, et al. Inferring change points in the COVID-19 spreading reveals the effectiveness of interventions. Science. 2020; 369(10). https://doi.org/10.1126/science.abb9789 PMID: 32414780

23. Song P, Wang L, Zhou Sea. An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China. medRxiv. 2020;.

24. McBryde E, Gibson G, Pettitt A, Zhang B, McElwain D. Bayesian modelling of an epidemic of severe acute respiratory syndrome. Bulletin of Mathematical Biology. 2006; 68:889–917. https://doi.org/10.1007/s11538-005-9005-4 PMID: 16802088

25. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Using posterior predictive distributions to analyse epidemic models: COVID-19 in Mexico City. Physical Biology. 2020; 17.

26. Dbouk T, Drikakis D. Fluid dynamics and epidemiology: Seasonality and transmission dynamics Fluid dynamics and epidemiology: Seasonality and transmission dynamics. Physics of Fluids. 2021; 33 (021901). https://doi.org/10.1063/5.0037640 PMID: 33746486

27. Chowell G, Fenimore P, Castillo-Garsow MA, Castillo-Chavez C. SARS outbreaks in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism. Journal of Theoretical Biology. 2003; 224:1–8. https://doi.org/10.1016/S0022-5193(03)00228-5 PMID: 12900200

28. Lipsitch M, Cohen Tea. Transmission dynamics and control of severe acute respiratory syndrome. Science. 2020; 300 (5627). https://doi.org/10.1126/science.1086616

29. Cao C, Chen W, Zheng S, Zhao J, Wang J, Cao W. Analysis of spatiotemporal characteristics of pandemic SARS spread in Mainland China. BioMed Research International. 2016; 2016:889–917. https://doi.org/10.1155/2016/7247983 PMID: 27597972

30. Bliznashki S. A Bayesian logistic growth model for the spread of COVID-19 in New York. medRxiv. 2020; 14(12).

31. Chandra V. Stochastic compartmental modelling of SARS-CoV-2 with approximate Bayesian computation. medRxiv. 2020;.

32. Roda W, Varughese M, Han D, Li M. Why is it difficult to accurately predict the COVID-19 epidemic? Infectious Disease Modelling. 2020; 5:271–281.

33. Covid-19 Mexico;. https://coronavirus.gob.mx/datos/.

34. Capistrán M, Capella A, Christen A. Forecasting hospital demand during COVID-19 pandemic outbreaks. arXiv e-prints. 2020;.

35. Acuña Zegarra M, Comas-García A, Hernández-Vargas E, Santana-Cibrian M, Velasco-Hernández J. The SARS-CoV-2 epidemic outbreak: a review of plausible scenarios of containment and mitigation for Mexico. medRxiv. 2020;. https://doi.org/10.1101/2020.06.30.20143560 PMID: 32637975

36. B T, Xia F, Tang S, Bragazzi N, Li Q, Sun X, et al. The effectiveness of quarantine and isolation determine the trend of the COVID-19 epidemic in the final phase of the current outbreak in China. International Journal of Infectious Diseases. 2020; 96:636–647. https://doi.org/10.1016/j.ijid.2020.05.113

37. A T, Konané F. Modeling the effects of contact tracing on COVID-19 transmission. Advances in Difference Equations. 2020; 1(509).

38. Z K, Van Bussel F, Hussain F. A predictive model for Covid-19 spread—with applidation to eight US states and how to end the pandemic. Epidemiology and Infection. 2020; 148(1-13).

39. Ngonghala C, Iboi E, Eikenberry S, Scotch M, MacIntyre C, Bonds M, et al. Mathematical assessment of the impact of non-pharmaceutical interventions on curtailing the 2019 novel Coronavirus. Mathematical Biosciences. 2020; 325:108364. https://doi.org/10.1016/j.mbs.2020.108364 PMID: 32360770

40. E A, Okyere E, Iddi S, Bonney J, Wattis A, Gomes R. Modelling COVID-19 Transmission Dynamics in Ghana. 2021;.

41. Girum T, Lentiro K, Geremew M, Migora B, Shewamare S. Global strategies and effectiveness for COVID-19 prevention through contact tracing, screening, quarantine, and isolation: a systematic review. Tropical Medicine and Health. 2020; 48(91):1–15. https://doi.org/10.1186/s41182-020-00285-w PMID: 33292755

42. Nussbaumer-Streit MVDACAPEKIWGSUCCZC B, Gartlehner G. Quarantine alone or in combination with other public health measures to control COVID-19: a rapid review. Cochrane Database of Systematic Reviews. 2020;(4).

43. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020; 368 (6491). https://doi.org/10.1126/science.abb6936 PMID: 32234805

44. Browne C, Gulbudak H, Webb G. Modeling contact tracing in outbreaks with application to Ebola. Journal of Theoretical Biology. 2015; 384:33–49. https://doi.org/10.1016/j.jtbi.2015.08.004 PMID: 26297316

45. Kwok K, Tang A, Wei V, Park W, Yeoh E, Riley S. Epidemic models of contact tracing: systematic review of transmission studies of Severe Acute Respiratory Syndrome and Middle East Respiratory Syndrome. Computational and Structural Biotechnology Journal. 2019; 17:186–194. https://doi.org/10.1016/j.csbj.2019.01.003 PMID: 30809323

46. Hethcote H, Zhien M, Shengbing L. Effects of quarantine in six endemic models for infectious diseases. Mathematical Biosciences. 2002; 180:141–160. https://doi.org/10.1016/S0025-5564(02)00111-6 PMID: 12387921

47. Hethcote H. The mathematics of infectious diseases. SIAM REVIEW. 2000; 42(4):599–653. https://doi.org/10.1137/S0036144500371907

48. Brauer F, van den Driessche P, Wu J. Mathematical Epidemiology. Springer; 2008.

49. Prieto K and Dorn O. Sparsity and level set regularization for diffuse optical tomography using a transport model in 2D. Inverse Problems. 2016; 33(1):014001. https://doi.org/10.1088/0266-5611/33/1/014001

50. Prieto K and Ibarguen-Mondragon E. Parameter estimation, sensitivity and control strategies analysis in the spread of influenza in Mexico. Journal of Physics: Conference Series. 2019; 1408(1):012020.

51. Smirnova A, DeCamp L, Liu H. In: Inverse problems and ebola virus disease using an age of infection model. Springer, Cham; 2016. p. 103–121.

52. Alavez-Ramirez J. Estimacion de parámetros en ecuaciones diferenciales ordinarias: identificabilidad y aplicaciones a medicina. Revista electrónica de contenido matemático. 2007; 21.

53. Capistrán M, Moreles M, Lara B. Parameter estimation of some epidemic models: the case of recurrent epidemics caused by respiratory syncytial virus. Bulletin of Mathematical Biology. 2009; 71(8):1890–1901. https://doi.org/10.1007/s11538-009-9429-3 PMID: 19568727

54. Stojanović O, Leugering J, Pipa G, Ghozzi S, Ullrich A. A Bayesian Monte Carlo approach for predicting the spread of infectious diseases. PLoS ONE. 2019; 14(12).

55. Luzyanina T, Bocharov G. Markov chain Monte Carlo parameter estimation of the ODE compartmental cell growth model. Mathematical Biology and Bioinformatics. 2018; 13:376–391. https://doi.org/10.17537/2018.13.376

56. Bettencourt L, Ribeiro R. Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. PlosOne. 2008; 3(5):e2185. https://doi.org/10.1371/journal.pone.0002185 PMID: 18478118

57. Boersch-Supan P, Ryan S, Johnson L. deBInfer:Bayesian inference for dynamical models of biological systems in R. Methods in Ecology and Evolution. 2017; 8:511–518. https://doi.org/10.1111/2041-210X.12679

58. Grinsztajn L, Semenova E, Margossian C, Riou J. Bayesian workflow for disease transmission modeling in Stan. arXiv e-prints. 2020;.

59. Chowell G. Fitting dynamic models to epidemic outbreak with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. Infectious Disease Modelling. 2017; 2:379–398. https://doi.org/10.1016/j.idm.2017.08.001 PMID: 29250607

60. Argüedas Y, Santana-Cibrian M, Velasco-Hernández J. Transmission dynamics of acute respiratory diseases in a population structured by age. Mathematical Biosciences and Engineering. 2019; 16(6):7477–7493. https://doi.org/10.3934/mbe.2019375 PMID: 31698624

61. Nayens T and Faes C and Molenberghs G. A generalized Poisson-gamma model for spatially oversdispersed data. Spatial and Spatio-temporal Epidemiology. 2012; 3:185–194. https://doi.org/10.1016/j.sste.2011.10.004

62. Coly S, Yao A, Abrial D, Garrido M. Disributions to model overdispersed count data. Journal de la Societe Francaise de Statistique. 2016; 157(2):39–63.

63. Carpenter B, Gelman A, Hoffman D, Goodrich B, Betancourt M, Brubaker M, et al. Stan: A probabilistic programming language. Journal of Statistical Software. 2017; 76(1):1–32. https://doi.org/10.18637/jss.v076.i01

64. Christen J, Fox C. A general purpose sampling algorithm for continuous distributions (the t-walk). Bayesian Anal. 2010; 5:263–282. https://doi.org/10.1214/10-BA603

65. House T, Ford A, Lan S, Bilson S, Buckingham-Jeffery E, Girolami M. Bayesian uncertainty quantification for transmissibility of influenza, norovirus and Ebola using information geometry. JRSoc Interface. 2016; 13. https://doi.org/10.1098/rsif.2016.0279 PMID: 27558850

66. Roosa K, Chowell G. Assesing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. Theoretical Biology and Medical Modelling. 2019; 16(1). https://doi.org/10.1186/s12976-018-0097-6 PMID: 30642334

67. Magal P, Webb G. The parameter identification problem for SIR epidemic models: identifying unreported cases. Journal of Mathematical Biology. 2018; 77:1629–1648. https://doi.org/10.1007/s00285-017-1203-9 PMID: 29330615

68. Metsis V, Androutsopoulos I, Paliouras G. Spam filtering with Naive Bayes-Which Naive Bayes? 3rd Conf on Email and Anti-Spam (CEAS). 2006;347.

69. Prieto K, Chavez-Hernandez MV, RomeroLeiton JP. On mobility trends analysis of COVID-19 dissemination in Mexico City. medRxiv. 2021; p. 1–27.