Prediction of Cases of Infection and Deaths Caused by COVID-19 in Mexico through the Construction of Probabilistic Models under Health Conditions in 2020

Juan Bacilio Guerrero Escamilla¹*, Sócrates López Pérez¹
and Yamile Rangel Martínez¹

¹Autonomous University of Hidalgo State In the Academic Area of Sociology and Demographics, Mexico.

Authors’ contributions

This work was carried out in collaboration among all authors. Author JBGE designed the study, performed the statistical analysis, wrote the protocol, and wrote the first draft of the manuscript. Authors SLP and YRM managed the analyses of the study, managed the literature searches. All authors read and approved the final manuscript.

Abstract

In the present research work, two probabilistic models are constructed, which are exponential regression and negative binomial regression. The first one refers to the number of positive cases of being infected by COVID-19. The second one refers to deaths. It was possible to estimate the dynamics of the phenomenon with both instruments, resulting in the presence of more than 106 thousand positive cases of COVID - 19, with an approximation of more than 9 thousand deaths, all of this, in approximately 4 months. In the first case, these were the results, which when updated with data issued by the federal government’s health sector in November, changed the contagion scenarios and the estimates of deaths from covid-19.

Keywords: COVID – 19; exponential regression; negative binomial regression.

*Corresponding author: E-mail: bacimate@gmail.com;
1 Introduction

1.1 Problem definition

In the first report from the World Health Organization (WHO), it was informed diverse cases of pneumonia with an unknown etiology and detected in the city of Wuhan, China's Hubei province on December 31\textsuperscript{st}, 2019. In the course of 3 days, the national authorities of China informed to WHO a total number of 44 patients, without identifying the causal agent; however, with an identification of a new type of coronavirus [1,2].

From January 2020, China shared the genetic sequences, which it allowed to distribute them with other countries and to develop specific diagnostic kits [3,4]. From this understanding, Thailand reported the first imported case, as well as Japan did around the same time. On January 20\textsuperscript{th}, 2020, Korea notified to WHO one of its first cases of coronavirus, and this information was confirmed by the laboratory of Wuhan.

1.2 Data set description

To the last reports obtained from the WHO during the month of April 2020, the transmission situation it has already been in world scene. This setting showed a total number of confirmed cases of 2, 397,216 from different countries, and a total number of new confirmed cases of 83,006, a total number of deaths of 162,956, and a total of new deaths of 5,109. These reports were taken into account from the countries, territories and areas with cases and deaths caused by COVID-19, and reported laboratory-confirmed cases, with data gathered from April 21\textsuperscript{st} 2020 emitted by WHO.

In the case of Americas with data obtained from Pan American Health Organization (PAHO), it shows a total number of confirmed cases of 925,291 in 54 affected countries and with 44,775 confirmed deaths. By this time, to the specific case of Mexico, there are 93,453 confirmed accumulated cases and 150,157 negative cases, 38,497 suspected cases and 10,167 deaths. (https://coronavirus.gob.mx/datos/, consulted from June 2nd, 2020).

The specialists considered that from 7,778 million of inhabitants in the world, 70% of them have the probability of being infected, provided that there is no intervention from the authorities, governments and the same population. From this 70% of infected ones moreover, 80% of these would be asymptomatic [5].

The other 20% would show severe symptoms, additionally of this population of probable infection the 14% would develop more serious symptoms with requirement for hospitalization, the other 6% would require to have intensive cares; although, the other 3% would die. In this last, fatality rate of virus SARS-CoV-2 might be. Liu et al, [6] In the case of fatality, each country determines the formula to calculate, in which Mexico has taken as the relation between the deaths and the infected.

1.3 Related work

Taking into consideration this panorama, we will have to make an analysis defining the settings of transmission, mortality rate and condition in which population might be affected. Nevertheless, this situation did not consider intervention and any action to influence on transmission behavior. As a consequence, the diverse analysis and modeling of the scientific community must think over to foresee the future of the spread of the virus and then, focus on the intervention of authorities, resource calculation, and sustained design of public policies [5].

This is a quick response of the scientific community to offer efficient answer with techno-scientific tools, and which define different settings with certainty in order to decide which models of intervention are adequate to follow, and which factors include as political option. Even though, in an initial time there is no
enough information to support these models, later we must trust on adjusting these deficiencies of data meanwhile important epidemiological knowledge are generated.

Making reference to the case of Mexico, there is an advantage about the previous behaviors of the virus showed in China and Europe, so this benefit helps with more certainty to the estimations of the mortality rate of Covid-19 and how quickly it spreads, how these estimations differ from diverse populations, and also how these probabilities of risk were quantified in Germany [7].

As a result, to the design of these models it can be added the own conditions of the countries. In the particular case of Mexico, in respect of its modeling of State, Political class, ways to construct the public policies, National legal system, National Health system, Public healthcare system, resources and quality of care, quality of citizenship, capacity of social organization, demographic composition, territorial distribution, sociodemographic characteristics and economic composition, distribution of wealth and levels of poverty, in addition to the mobility index and situation of morbidity and mortality [8].

In referring to the case of Mexico, whatever modeling of analysis of conditions of mobility of the population can be incorporated, thus the level of risk will change based on the composition of the comorbidity, in other words, more diseases in concert with the age. In a manner that, the increase of risk will be for people who have diabetes, so then, this group develops a viral infection, and it is more difficult to treat due to the fluctuations in the blood glucose levels, and possibly because of the presence of complications of diabetes (International Diabetes Federation).

From those who have obesity, they can get different complications in the health because of apnoea and decrease of oxygen in the cells, until difficulties to amount of fat to ensure attention of intubation, the respirator fit and their own mobilization in the hospital stay, among other cardiovascular diseases.

It is relevant to mention, these models among the model that is being proposed in the current research project, has as main objective to analyze infections curve to smooth it and to reduce the dynamic of infections and to give opportunity to the Hospital System of Mexico can attend to the cases of severe symptoms, and also to prepare and to improve health infrastructure or re-converting of diverse hospitals to intensive-care nursing. Consequently, mortality rate could be reduced, and the most severe cases of recovery could be treated. In this case, take all actions leading to the isolation of the population, thus the potential spread of the virus is from 1.4% to 5.5%, and it must be reduced between 0.9% and 0.5% [9].

Within the strategic objectives that WHO has in relation to Covid-19, the following are highlighted:

- To interrupt transmission from person to person, including the reduction of secondary infections among close contacts.
- To identify, to isolate patients and to treat early diagnosis patients, which includes providing optimized care to infected patients.
- To identify and to reduce transmission from the animal source.
- To address critical uncertainties with regard to clinical severity, the degree of transmission or infection and the treatment.
- To communicate critical information about risks and events to all communities and to counter misinformation.
- To minimize social and economic impact through multi-sector partnerships [3,4].
This can be achieved through the combination of public health measures; for instance, quick identification, diagnosis, case management, identification and monitoring of contacts, prevention and infection control in health-care settings, implementing health measures to travelers, awareness of the population and risk communication.

On the basis of the foregoing, the present research work has the aim of predicting the number of cases and deaths of COVID-19 in Mexican population through the construction of two mathematical models, taking as a reference the start dates of each phase of the pandemic.

To the construction of both models, the Operations Research Methodology is used, which is formed of four stages [10]:

1. Mathematical formulation: in this first stage, present lineal relations that exist among the cases and deaths regarding the time.
2. Construction of the model: in this second stage, adequate models are selected to predict the dynamic, therewith the calculation of parameters of both models is done.
3. Validating the model: starting from the estimated parameters, in this third stage, the calculation of the adjustment of both models, thus inasmuch as this approximates to 1, greater precision of estimation.
4. Analysis of the results: in this last stage the interpretation of results is done, once algorithms are obtained (mathematical equations) of both phenomena.

With these models, a structured view of reality is showed with the pandemic COVID-19. In this way, the purpose of both models is to provide elements that intervene in the behavior of the phenomenon in order to optimize the performance [11].

2 Theoretical Analysis

To predict the dynamic of cases and deaths of the pandemic COVID-19, consider the following:

- Positive cases: \( Y_c = f(T) \)
- Deaths: \( Y_d = f(T) \)

Such that:

- \( Y_c \) is the total of cases caused by COVID-19 per day.
- \( Y_d \) is the total of deaths caused by COVID-19 per day.
- \( T \) is the time.

It is based on the assumption that, both phenomena have a growing exponentially behavior [12]:

\[
Y_c \sim \exp \left( \frac{1}{\lambda} \cdot \frac{1}{\lambda^2} \right) \quad (1) \\
Y_d \sim \exp \left( \frac{1}{\lambda} \cdot \frac{1}{\lambda^2} \right) \quad (2)
\]

This is represented graphically (Magister, [12]).

Based on this context, the COVID-19 epidemic will not stop growing (in the presence of cases) until the population would be immune, resulting in a high fatality rate due to the collapse of the health system, this can be seen in the Fig. 3, within the function \( f(T) \).
From this setting, the authorities of Health System implemented the following actions (recommendation by World Health Organization) [3,4]:

- To interrupt transmission from person to person, including the reduction of secondary infections among close contacts.
- To identify, to isolate patients and to treat early diagnosis patients, which includes providing optimized care to infected patients.
- To identify and to reduce transmission from the animal source.
- To address critical uncertainties with regard to clinical severity, the degree of transmission or infection and the treatment.
- To communicate critical information about risks and events to all communities and to counter misinformation.
- To minimize social and economic impact through multi-sector partnerships.
These actions will have as effects lengthening of the epidemic in the time, and by doing so, it will decrease the fatality rate; as a result of this, the Health System will not collapse, thus enough infrastructure will be needed to attend a higher number of possible cases (this setting can be observed in the Fig. 3 of the function \( f_2(T) \)).

On the basis of the foregoing, the dynamic of the pandemic COVID -19 will approximate a normal distribution [13]:

\[
Y_c \sim N(\mu, \sigma^2)
\]  

(3)

This is graphically represented:

---

**Fig. 4. Phases of the pandemic COVID -19 per day**

*Source. Own authorship, Mexico 2020*

Taking as reference the Fig. 4, The World Health Organization identifies five stages of the pandemic COVID - 19 (WHO, 2020):

- First stage: the disease arrives to a country through a person or a small number of people who acquired the virus from abroad, then the number of cases is limited to a few dozens.
- Second stage: with the arrival of people from other countries with COVID – 19, they spread the virus to those who have had contact, and these in turn continue spreading the disease, so the number of cases starts to exceed the hundreds.
- Third stage: in this stage of the disease that has been presented all over the country, and there is a high number of community outbreaks, in which the number of sick people increases exponentially, having an effect to the saturation of the health centers.
- Fourth stage: it starts with the reduction of infections in local areas, with the possibility of the emergence of a second wave with a likely duration from three to nine months.
- Fifth stage: this is the culmination of the pandemic, which is declared by the World Health Organization.

Since the context of probability and statistics, the function of density of the pandemic COVID -19, it is expressed in the following way:
Cases

\[ E(Y_c) = \hat{\beta} e^{-\hat{\beta}_1 (T_i - \mu_T)^2} \]  

Deaths

\[ E(Y_d) = \hat{\lambda} e^{-\hat{\lambda}_1 (T_i - \mu_T)^2} \]

Where:

- The expected value of cases is higher or equal to the expected value of deaths \( E(Y_c) \geq E(Y_d) \).
- The time \( T_i \) is equal to the mean of time less deviation standard of time \( T_i = \mu_T - \sigma_T \).
- The mean of time is equal time \( i \) more the standard deviation of time \( \mu_T = T_i + \sigma_T \).
- The standard deviation of time is the root of the variance of time \( \sigma_T = \sqrt{\text{Var}(T)} \).

The presence of positive cases and deaths daily lead to the accumulation of these:

![Fig. 5. Positive cases and deaths of the pandemic COVID - 19 per day](Source. Own authorship, Mexico 2020)
Y*c & Y*d represent the maximum points of the pandemic, including the positive cases and deaths respectively. Added to this, the turning points represent (Y*ca y Y*cd) the positive cases and accumulated deaths, that means that are coordinates in which the functions of both events, the positive cases and the deaths, it changes its tendency; as a consequence, the application of the recommended actions by the World Health Organization.

3 Modeling

Taking into account the theoretical analysis, the mathematical formulation of the pandemic of COVID – 19 is expressed in the following way:

Positive Cases
\[
E(Y_c) = \beta_0 + \beta_1(T_i - \mu_T)^2 + u_i
\]

Deaths
\[
E(Y_d) = \alpha_0 + \alpha_1(T_i - \mu_T)^2 + u_i
\]

Where:
- \( \beta_1, 0 & 1 \) are the parameters to estimate to the positive cases.
- \( \alpha_1, 0 & 1 \) are the parameters to estimate to the deaths.
- \( u_i \) is the margin of error (which is the proportional part that can be explained for the lineal relation that exists between the time and the pandemic).

To the construction of both models, it is assumed that Yc & Yd are counts; therefore, both variables approximate to the Poisson distribution or negative binomial [14,5]:

Poisson
\[
P(Y) = \frac{\lambda^Y e^{-\lambda}}{Y!}; \quad E(Y) = \lambda \quad & \quad V(Y) = \lambda
\]

Negative Binomial
\[
P(Y) = \binom{Y-1}{r-1} \frac{1}{r} \left( \frac{1}{p} \right)^r \left( 1 - \frac{1}{p} \right)^{Y-r}; \quad E(Y) = \frac{r}{p} \quad & \quad V(Y) = \frac{rq}{p^2}
\]

Obtaining the expected value and the variance of both variables:

| Expected value | Effective Cases | Deaths |
|----------------|----------------|--------|
| E(Yc)          | 12,807.39      | 1,135.58|
| V(Yc)          | 300,717,848.50 | 2,484,323.29|

**Chart 1. Expected value and variance of the independent variables**

Source. Own Authorship

In the chart 1, it is observed that the prediction of the pandemic COVID-19 cannot be modeled through a Poisson regression, thus the expected value and the variance of both events do not approximate; therefore, the most effective way will be through a negative binomial regression [15]:

Positive cases
\[
[E(Y_c)] = \beta_0 + \beta_1(T_i - \mu_T)^2
\]

Deaths
\[
[E(Y_d)] = \alpha_0 + \alpha_1(T_i - \mu_T)^2
\]

Where the adjustment of both models is through the deviance, which informs about the variability of the data, and it is expressed in the following way:

\[
D^2 = \frac{ND - RD}{ND}, \quad \text{such that} \quad 0 \leq D^2 \leq 1
\]

ND is the Null Deviance and RD is the Residual Deviance
Such that:

- If $D^2 = 1$, tighter fit.
- If $D^2 = 0.5$, moderate adjustment.
- If $D^2 = 0$, minor adjustment.

To the estimation of parameters, it is known that since the beginning of the pandemic that was on March 18th, and the stage III was on October 25th, this represents the day 171.

Where:

- $V(T) = 4125.51$
- $\sigma_T = \sqrt{V(T)} = \sqrt{4125.51} = 64.23 \sim 64$

Therefore, the day 235 would represent the maximum number of cases per day, it would be the turning point which changes the tendency of the pandemic:

\[ 189 = \mu_T - 64 \rightarrow \mu_T = 171 + 64 = 235 \quad (13) \]

Therefore, the regression models would be the following:

Positive cases
\[ [E(Y_c)] = \hat{\beta}_0 + \hat{\beta}_1(T_i - 235)^2 \quad (14) \]

Deaths
\[ [E(Y_d)] = \hat{\alpha}_0 + \hat{\alpha}_1(T_i - 235)^2 \quad (15) \]

Applying the Maximum Likelihood Theorem \(^1\) has that:

It may be $X_1, \ldots, X_n$ a random sample of the population $X$ with probability function $P_0$ to each particular sample, the maximum likelihood function is defined as the function of joint probability of [16]:

\[ \gamma(\theta) = \gamma(X_1, \ldots, X_n; \theta) = P_0(X_1 = x_1, \ldots, X_n = x_n) \rightarrow \text{Discrete} \]

\[ \gamma(\theta) = f_0(X_1, \ldots, X_n; \theta) \rightarrow \text{Continuous} \]

Positive cases:
\[ \sqrt{E(Y_c)} = 88.561 - 0.00152(T_i - 235)^2; \quad D^2 = 0.8299 \quad (16) \]

Deaths:
\[ \sqrt{E(Y_d)} = 29.933 - 0.0005(T_i - 235)^2; \quad D^2 = 0.6152 \quad (17) \]

Clearing $E(Y_c)$ & $E(Y_d)$:

Positive cases
\[ E(Y_c) = [88.561 - 0.00152(T_i - 235)^2]^2 \quad (18) \]

Deaths
\[ E(Y_d) = [29.933 - 0.0005(T_i - 235)^2]^2 \quad (19) \]

Therefore, level adjustment of both models is with a level of confidence of 0.95 and with a level of significance of 0.05, the obtained algorithm (positive cases and deaths) keep 0.8299 & el 0.6152 of the variability of the data, in other words, both algorithms explain 82.99% & 61.52% of the dynamic of positive cases and deaths of the epidemic of COVID-19 [17].
4 Interpretation of Results and Discussion

On the basis of the algebraic expressions 18 & 19, the dynamic estimation of the cases and the deaths caused by the pandemic of COVID-19 can be interpreted in the following way:

Taking as a reference the Fig. 6 and based on the estimations of the equation 19, the stage III started on October 5th and it would end on January 7th of 2021. During this stage, from 7,000 to 8,000 cases may be presented on average per day, having as a highest point more than 8,030 cases on November 15th. This represents a turning point, in other words, from that date starts to decelerate the exponential growth of the pandemic (Fig. 7).
From January 9th to March 16th of 2021 will begin a decrease of cases per day, less than 6,000 cases, with the possibility of increasing activity of the virus. This fact will depend on the degree of mobility of the population. From March 17th to Jun 26th will be the culmination of the pandemic, having an estimation of less than 1000 cases per day, all this will cause a result of more than 2 million infections.

In relation to the deaths, in the Fig. 8 stage III would have a duration of 66 days from September 19th to January 20th of 2021. In this stage, there will be from 700 to 900 deaths per day, having its peak of more than 896 deaths on November 9th. In this date would happen the turning point of the pandemic, starting from the deceleration of the exponential growth of the pandemic (Fig. 8).

From August 12th will start the decrease of deaths (less than 200 per day). From September 23rd to December 4th there will be the estimated average of less than 50 deaths per day, all of this may result in 54 thousand deaths.

---

**Fig. 8. Prediction of deaths of the pandemic COVID - 19 per day**

*Source. Own Authorship, México 2020*

---

**Fig. 9. Prediction of accumulated deaths of the pandemic COVID - 19**

*Source. Own authorship, México 2020*
5 Conclusions

With the construction and the development of these probabilistic models, exponential regression and negative binomial regression, the necessary elements could be obtained to know about the dynamic of the virus spread SARS-CoV-2, the deaths caused by Covid-19, its regional behavior during the time. Furthermore, we find the behavior of Covid-19, in each of its five stages, meanwhile range of temporality is defined. Taking into account these elements as basis, from this model it can predict the behavior of the fatality and its dynamic of transmission. Based on that, from this ceteris paribus, the diverse settings of actions taken, intervention and the prospective of resources can be defined to contribute to the mitigation of transmission; therefore, the composition of deaths. The model allows that, meanwhile the ones in charge of making decisions and taking care of resources and health infrastructure, prioritize interventions to impact on the range of temporality of each stage.

These interventions are actions that influence on a phenomenon from the own design. Nonetheless, the main use of the model is to determine those elements impact on to slow down the infection, consequently take action in health infrastructure and administer appropriately the resources of health institutions. This in turn implies the simulation of the incidence of these actions in the effects of lengthening of the epidemic over time, thus, the fatality rate will decrease; in turn, the Health System will not collapse, because it will have enough infrastructure to attend the maximum number of possible cases (Fig. 3 of the function f_d(T)).

Based on the above scenario, our model shows that the dynamics of the COVID-19 pandemic will approximate a normal distribution. In addition, this same model ensures the prediction of the behavior of infections and deaths from the COVID-19 disease. That when making a comparative analysis with the prediction models of the group of scientists of the Council of Science and Technology (CONACyT) and the Secretary of Health and Assistance (SSA), of the federal government in Mexico, their projections are 102,274 deaths for the month June 2020 and 123,413 for November of the same year and with an increase of 858 people died per day. In our model, the results obtained on these same dates are 886 deaths per day, which when comparing both results, gives us a difference of 3.26%.

The results of this comparison, being so small, show us a robust and highly useful model, to be applied as a simulator at other levels of state and municipal government, and to make the appropriate decisions for the control and mitigation of the contagions. When relating these results to the territorial distribution of contagion and deaths, as well as the integration of other local elements such as the composition of economic, social, transport, employment, etc. Networks, we ensure with certainty the targeting of mitigation and control actions, contagion, maximizing resources and slowing down infections.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Wang, Dawei, Bo Hu, Chang Hu, Fangfang Zhu, Xing Liu. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) Outbreak in China Summary of a Report of 72314 Cases from the Chinese Center for Disease Control and Prevention. American Medical Association. Clinical Characteristics of Hospitalized Patients with Novel Coronavirus. 2019;138.

[2] Woo Chan, Jasper Fuk, Yuan Shuofeng, Hang Kok, Kin, Lancet. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. 2020;395:514–23 Published Online January 24, 2020. Available:https://doi.org/10.1016/S0140-6736(20)30154-9 See Comment pages 469 and 470

[3] WHO. World Health Organization. (published in 2020 May 25). “Questions and Answers about COVID – 19 disease” ([WHO]; 2020.)
Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

WHO. World Health Organization. Situation Report. 2020;1-92. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

Sridhar Devi, Majumder Maimuna S. Modelling the pandemic. Over-reliance on modelling leads to missteps and blind spots in our response The British Medical Journal BMJ. 2020;369. Publishers DOI: https://doi.org/10.1136/bmj.m1567 (published 2020 April 21).

Liu, Cynthia, Zhou, Qiongqiong, Li, Yingzhu, Garner, Linda V. Research and development on therapeutic agents and vaccines for COVID-19 and Related Human Coronavirus Diseases. ACS Cent. Sci. 2020;6:315–331.

Süddeutsche, Zeitung. Uhr Kampf gegen Pandemie. Innenministerium dringt auf massive Ausweitung von Corona-Tests.2020;6:01. Available: https://www.sueddeutsche.de/politik/coronavirus-tests-strategie-1.4858950

López Pérez, Guerrero Escamilla. Gravitational model, Megalopolis; Mexico; 2015.

Grill, Eva, Rothenbacher, Dietrich. Deutschen Gesellschaft für Epidemiologie; MPH. 2020;23:03.

Sánchez C. Investigación de Operaciones I°. [Operations Research I]. Mexico: Hispano-American University; 2004.

Peeri, Noah C, Shrestha, Nistha, Rahman, Md Siddikur. The SARS, MER Sand novel coronavirus (COVID-19) epidemics, the new stand biggest global health threats: what lessons have we learned? International Journal of Epidemiology. 2020;1–10. DOI: 10.1093/ije/dyaa033.369: m1567.

Magister M. Elementary functions for building mathematical models. Argentina: Ministry of Education; 2010.

Freund J, Miller I, Miller M. Mathematical statistics with applications. United States: Prentice Hall. Infected Pneumoniai Wuhan, China. JAMA. 2020;323:13_1.

Wackerly D, Mendenhall WY, Scheaffer R. Estadística matemática con aplicaciones. USA: CENGAGE Learning. Journal of Infectious Diseases. 2010;92(2020):214–217.

McCullagh P, Nelder J. Generalized linear models. United States: Chapman and Hall; 1983.

Mendenhall W. Beaver R, Beaver B. Introduction to probability and statistics. United States: CENGAGE Learning; 2010.

Zhaoa Shi, Lin, Qianyin, Rand, Jinjun. Preliminary estimation of the basic reproduction number of novel corona virus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. International; 2020.

© 2020 Escamilla et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.