ARTICLE
https://doi.org/10.22239/2317-269x.01774

# Predictions of the maximum number of confirmed cases and deaths of COVID-19 in Brazil

## Previsões de máximo de casos confirmados e óbitos de COVID-19 no Brasil

**Carlos Maximiliano Dutra***

Universidade Federal do Pampa (Unipampa), Uruguaiana, RS, Brasil

* **E-mail:** carlosdutra@unipampa.edu.br

Received: Aug 31, 2020
Approved: Nov 9, 2020

## ABSTRACT

**Introduction:** COVID-19 started in December 2019 in China and in February 2020 in Brazil. As of August 29, 2020, Brazil had 3,717,156 confirmed cases (60.6% of the total in America and 15.0% of the total World) and 117,665 deaths (59.4% of the total in America and 14.0% of the World total), the second place in the world. **Objective:** To estimate maximum limits for the quantitative number of confirmed cases and deaths of COVID-19 in Brazil. **Method:** The maximum estimates of confirmed cases and deaths were estimated by applying nonlinear regression with adjustment of the Gompertz function to the COVID-19 data recorded until the end of August 2020. **Results:** As results were obtained a maximum of 7,189,300 confirmed cases, with 95.0% of the maximum reached in February 2021 and 164,810 deaths, with 95.0% of the maximum reached in December 2020. **Conclusions:** These forecasts express the trend of temporal evolution of COVID-19 reported data by Our World in Data. They do not consider underreporting in the number of contaminated people and deaths, taking into account the limited testing capacity of the population and the existence of unaccounted asymptomatic contaminated people. These factors added to the uncontrolled relaxation of individual preventive measures and social distancing may have a significant effect on the predictions of the mathematical model used in this work and on the understanding of the transmission of the virus in Brazil.

**KEYWORDS:** Prediction; Coronavirus; Covid-19; Gompertz

## RESUMO

**Introdução:** A COVID-19 teve início em dezembro de 2019 na China e em fevereiro de 2020 no Brasil. Até 29 de agosto de 2020, o Brasil possuía 3.717.156 casos confirmados (60,6% do total da América e 15,0% do total mundial) e 117.665 óbitos (59,4% do total da América e 14,0% do total mundial), ocupando o segundo lugar no mundo. **Objetivo:** Estimar limites máximos para os quantitativos de casos confirmados e óbitos por COVID-19 no Brasil. **Método:** A estimativa dos máximos de casos confirmados e óbitos foi realizada através da aplicação de regressão não linear com ajuste da função Gompertz aos dados de COVID-19 registrados até final de agosto de 2020. **Resultados:** Como resultados foram obtidos máximo de 7.189.300 casos confirmados, com 95,0% do máximo atingidos em fevereiro 2021 e 164.810 óbitos, com 95,0% do máximo atingidos em dezembro 2020. **Conclusões:** Essas previsões expressam a tendência da evolução temporal dos dados notificados de COVID-19 pelo *Our World in Data*. Eles não consideram a subnotificação no número de contaminados e de óbitos, tendo em conta a limitada capacidade de testagem da população e a existência de contaminados assintomáticos não contabilizados. Estes fatores somados ao relaxamento descontrolado das medidas preventivas individuais e do distanciamento social poderão ter um efeito significativos nas previsões do modelo matemático usado neste trabalho e no entendimento da transmissão do vírus no Brasil.

**PALAVRAS-CHAVE:** Previsão; Coronavírus; COVID-19; Gompertz

## INTRODUCTION

The first cases of infection by SARS-CoV2 occurred in China, in December 2019, according to the World Health Organization (WHO).[1] On March 11, 2020, WHO declared COVID-19 a pandemic. There were then 118,319 confirmed cases and 4,292 deaths in 113 countries.[2] Until that date, South America as a whole had 109 confirmed cases and one death. Brazil, in turn, had only 34 confirmed cases,[2] and the first death was reported on March 12, 2020. About 5.5 months later, on August 29, 2020 (10:30 am GMT time), based on the COVID-19 database of *Our World in Data* (available on https://ourworldindata.org/coronavirus-source-data), the world had had 24,760,129 confirmed cases and 837,579 deaths; while South America reached approximately 6.1 million confirmed cases and 198,100 deaths. Also on August 29, 2020, Brazil had 3,717,156 confirmed cases (60.6% of the total in South America and 15.0% of the world total) and 117,665 deaths (59.4% of the total in South America and 14.0% of the world total). The country then ranked second on the list of countries with the most confirmed cases of COVID-19.

News about this rapid growth in the number of confirmed cases and deaths was followed by people all over the world in an unprecedented way, with daily updates on websites like *COVID-19 Dashboard – John Hopkins University* (https://coronavirus.jhu.edu/map.html); *Bing COVID* (https://www.bing.com/covid); and *Worldometer Coronavirus* (https://www.worldometers.info/coronavirus/), as well as in data repositories like *Our World in Data*. With the availability of data on the increase in the number of cases and deaths over time, several studies have been conducted to make predictions about the impact of COVID-19 in several countries. Perez et al.[3] used the nonlinear regression method of the Gompertz growth function together with the cumulative data of confirmed cases and deaths up to May 11, 2020 for Italy, Spain, China, and South Korea. They obtained predictions with values that were compatible with those effectively reported in the aforementioned databases. Medina-Mendleta et al.[4] performed nonlinear regression adjustment with the Logistical and Gompertz models with data on confirmed cases/deaths from COVID-19 in Spain and Italy. They correctly determined the peak dates in the distribution of confirmed cases/deaths, predicted the limit maximum for the cumulative values of confirmed cases/deaths, and still made predictions for Cuba. These predictions corresponded to the first "wave" of COVID-19 infection. In November 2020, European countries like Spain (https://www.worldometers.info/coronavirus/country/spain/) and Italy (https://www.worldometers.info/coronavirus/country/italy/) faced a second "wave", with significantly greater numbers of contagion than the first "wave". This new increase may be modeled again through nonlinear regression for new estimates in these countries.

The present study used the Gompertz nonlinear regression method to estimate the total number of confirmed cases and deaths in Brazil.

## METHOD

According to Domingues,[5] the Gompertz function was developed by Benjamin Gompertz in 1938 to estimate the limit for the growth of a population of solid tumors. Over time, it appears that this cumulative growth has the following phases: Lag – initial slow-growth phase; Exponential – with accelerated growth and then deceleration in the so-called inflection point; Plateau – growth converging to an upper limit of the population.

In equation (1) we have the Gompertz model defined according to the parameters used to characterize the cumulative growth of deaths due to COVID-19.

$$N(t) = A.e^{e^{-B.(t-C)}} \quad (1)$$

Where N (t) is the function of the number of individuals accumulated in the population of interest as a function of time "t"; A is the maximum or asymptotic limit value for the number of accumulated deaths; B is the relative growth at the inflection point (approximately the number of new cases divided by the number of cases accumulated at the time of the inflection point); and C corresponds to the time t at which the inflection point occurs.

Dutra et al.[6] applied the nonlinear regression method with the Gompertz model to the cumulative data of confirmed cases from eight countries that were in the pandemic stability phase. They found that when the method is applied after a country has reached the peak in the distribution of new cases, there is a convergence that makes it possible to reproduce the data already recorded in the pandemic stability phase in these countries. This demonstrated that predictions using the Gompertz nonlinear regression method under these conditions tend to come about.

In order to determine the maximum limits of confirmed cases and deaths for Brazil, the first step was to identify the distribution of data on new daily cases and deaths over time, and then whether they had already left the phase of accelerated exponential growth and peak distribution had been achieved. If so, nonlinear least squares regression of the Gompertz model was performed on data from confirmed cases and accumulated deaths per day. Through this statistical method, the values of the parameters of the Gompertz function were determined, especially parameter "A", which results in the maximum limit for the number of confirmed cases or accumulated deaths for Brazil.

From Equation (1), the elapsed time of the pandemic to determine a certain number of confirmed cases or accumulated deaths can be determined by isolating the time variable "t", as shown in Equation (2)

$$t = \frac{1}{B} \cdot \ln\left(\ln \frac{A}{N(t)}\right) + C \quad (2)$$

Where parameters A, B and C are obtained by nonlinear regression and describe function N (t). Equation (2) is used to make

temporal predictions about the pandemic, when the Gompertz N (t) function will reach a certain percentage of the maximum value A: 80.0%; 90.0% and 95.0%.

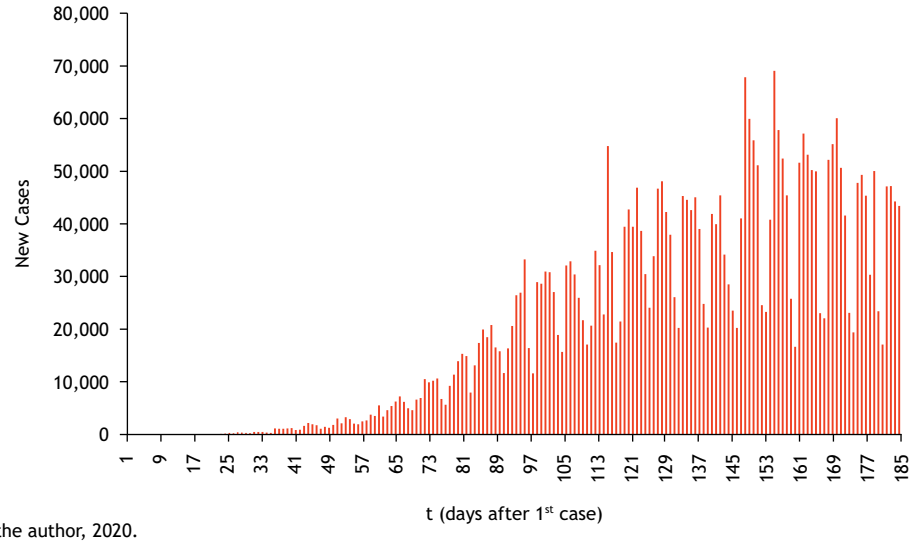
## RESULTS AND DISCUSSION

To make the estimates determined by the method, we used data on confirmed cases and deaths from COVID-19 in Brazil, updated until August 29, 2020 (10:30 am GMT), available in the *Our World in Date* COVID-19 data repository.

Figure 1 shows the evolution of daily confirmed cases over the days after the first case reported in Brazil. We notice that the accelerated growth has ceased and the formation of a peak has already occurred. This configuration indicates that the variation of confirmed cases accumulated over time can be modeled by the Gompertz function.

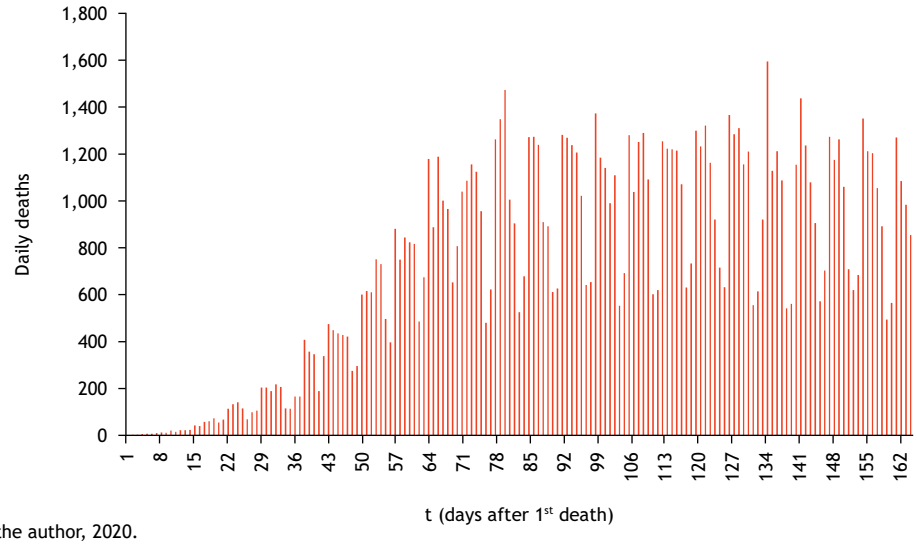
Figure 2 shows the evolution of daily deaths over the days after the first death reported in Brazil. It appears that the temporal distribution of deaths has also left its exponential and accelerated growth phase, with a steady and stable growth and a flattened peak.

From the analysis of Figures 1 and 2, it is possible to use Gompertz nonlinear regression together with the distributions of confirmed cases and deaths accumulated over time to estimate the maximum confirmed cases and deaths in Brazil. Using the free statistical PAST software[7] (available on https://folk.uio.no/ohammer/past/), nonlinear regression was done to determine the parameters of the Gompertz function as shown in Table 1.




Source: Prepared by the author, 2020.

**Figure 1.** Distribution of new cases of COVID-19 over the days after the first case recorded in Brazil.




Source: Prepared by the author, 2020.

**Figure 2.** Distribution of daily deaths due to COVID-19 as a function of time from the 1st death recorded in Brazil.

**Chart 1.** Summary of the results of the adjusted parameters of the Gompertz function to the data on confirmed cases and accumulated deaths from COVID-19 in Brazil.

| Accumulated | A | B | C |
|---|---|---|---|
| Confirmed cases | 7.1893e06 | -0.0160 | 157.86 |
| 95% confidence interval | (7.024e06; 7.348e06) | (-0.0163; -0.0157) | (156.41; 159.24) |
| Deaths | 164810 | -0.0185 | 106.72 |
| 95% confidence interval | (161200;169100) | (-0.0189; -0.0182) | (105.50; 107.25) |

Source: Prepared by the author, 2020.

Based on the parameters determined in Table 1, the maximum accumulated confirmed cases of 7,189,300 and the maximum cumulative deaths of 164,810 for Brazil can be estimated. These results reflect the analysis of nonlinear regression of the temporal evolution of confirmed cases and deaths of COVID-19 duly reported to the health systems and compiled by the Ministry of Health. It should be noted that these numbers of confirmed cases do not represent the actual number of people infected by SARS-CoV2. The underreporting of cases has been variable throughout the evolution of the COVID-19 pandemic in Brazil due to limited testing and the political guidance for doing rapid tests and RT-PCR adopted in Brazilian states. Until April 20, 2020, the reporting rate of confirmed cases of COVID-19 was estimated at 9.2%, which indicates underreporting and that the actual number of infected people would be 8-11 times greater than the number of reported cases.[8]

Figure 3 shows the graphical result of the nonlinear regression adjustment, where the red line represents the adjusted Gompertz function and the blue points represent the accumulated case data.

Figure 4 shows the graphical result of the nonlinear regression adjustment, where the red line represents the adjusted Gompertz function and the blue points represent the accumulated death data.

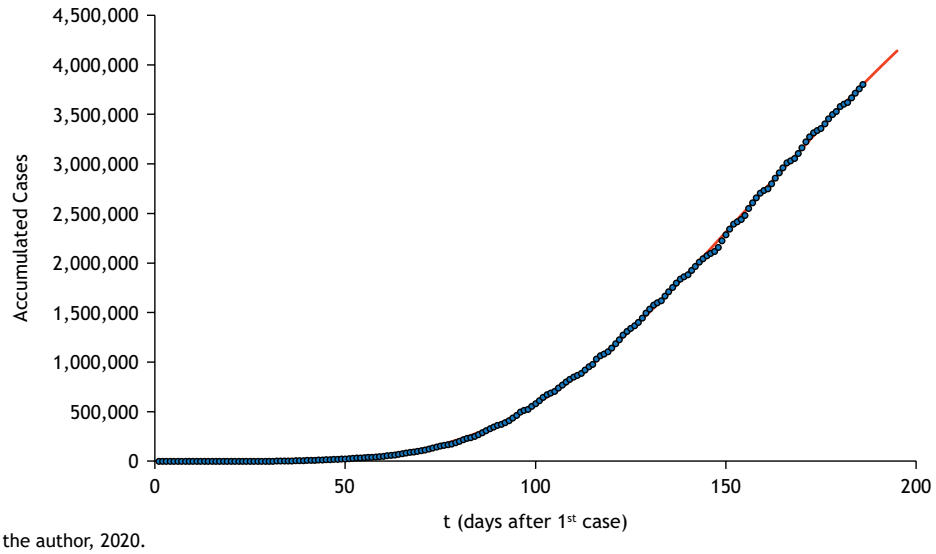
From the maximum accumulated confirmed cases of 7,189,300 and the maximum accumulated deaths of 164,810, it is possible to estimate the lethality rate (defined as the number of deaths from a given disease and the total number of patients multiplied by 100) which, in this case, will be 2.29% for COVID-19 in Brazil.

Using Equation 2 and the parameters determined in Chart 1, it is possible to make the following predictions for the COVID-19 pandemic to reach 80.0%, 90.0% and 95.0% of the maximums predicted by Gompertz nonlinear regression for confirmed cases and deaths in Brazil according to Chart 2.

Table 2 shows that 95.0% of the maximum accumulated confirmed cases will occur in February 2021 and 95.0% of the maximum accumulated deaths will occur in December 2020.

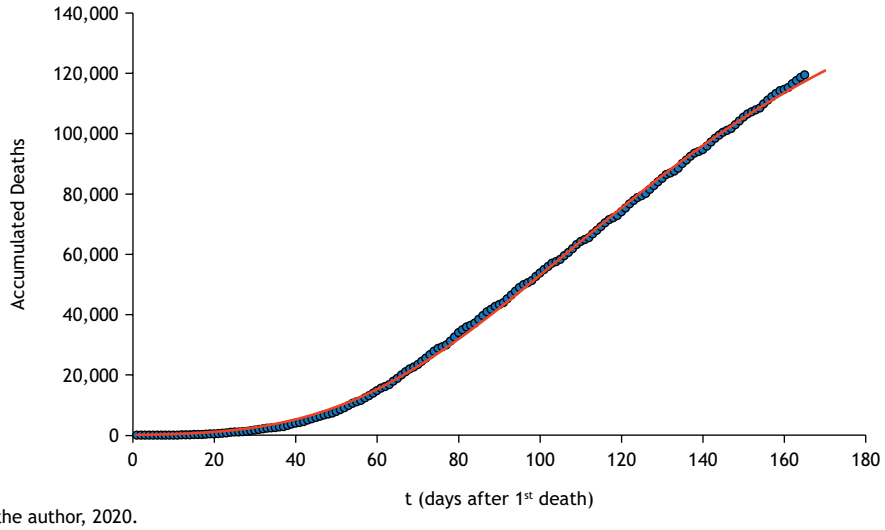
## CONCLUSIONS

This study estimated the maximum confirmed cases and deaths due to COVID-19 for Brazil using the Gompertz nonlinear regression method along with the temporal evolution of the reported data on COVID-19 until August 29 2020, according to data compiled by the Ministry of Health. The same methodology has been used successfully in other international studies on the evolution of COVID-19. It was possible to estimate a maximum of 7,189,300 confirmed cases and 164,810 deaths with a lethality of 2.29%,




Source: Prepared by the author, 2020.

**Figure 3.** Graphical result of the Gompertz nonlinear regression to the data on accumulated cases of COVID-19 in Brazil.




Source: Prepared by the author, 2020.

**Figure 4.** Graphical result of the Gompertz nonlinear regression to the data on accumulated deaths from COVID-19 in Brazil.

**Chart 2.** Summary of predictions about data on confirmed cases and accumulated deaths from COVID-19 in Brazil.

| Confirmed cases | 80.0% of the maximum | 90.0% of the maximum | 95.0% of the maximum |
|---|---|---|---|
| | Nov/11/2020 | Dec/20/2020 | Feb/03/2021 |
| Deaths | 80.0% of the maximum | 90.0% of the maximum | 95.0% of the maximum |
| | Sep/22/2020 | Nov/01/2020 | Dec/10/2020 |

Source: Prepared by the author, 2020.

with 95.0% of the maximum confirmed cases reached in February 2021 and 95.0% of the maximum deaths reached in December 2020. The predictions of this study should be seen as projections of the evolution of the number of contaminated people and deaths recorded by the Ministry of Health. These are approximate values to the real numbers that we will have at the end of the pandemic due to the limited number of tests done in the country and the lack of knowledge about the number of asymptomatic contaminated people throughout the entire pandemic process.

Determining the numbers and the exact moment when the pandemic will end in Brazil is very difficult, especially if we consider that the virus transmission varies according to the efficiency of social distancing measures and the testing and service capacity of the health system. This uncertainty about the evolution of the COVID-19 pandemic makes it even more important to improve testing and monitor the development of the disease so that we can measure its impact on society more accurately.

## REFERENCES

1. World Health Organization - WHO. Coronavirus disease 2019 (COVID-19) situation report 1. Geneva: World Health Organization; 2020[acesso 17 ago 2020]. Disponível em: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf
2. World Health Organization - WHO. Coronavirus disease 2019 (COVID-19) situation report 51. Geneva: World Health Organization; 2020[acesso 17 ago 2020]. Disponível em: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf
3. Perez FJD, Chinarro D, Otin RP, Mouhaffel AG. Growth forecast of the COVID-19 with the gompertz function, case study: Italy, Spain, Hubei (China) and South Korea. Int J Adv Eng Res Sci. 2020;7(7):67-77.
4. Medina-Mendieta JF, Cortés M, Iglesias M. COVID-19 forecasts for Cuba using logistic regression and gompertz curves. Medicc Rev. 2020;22(3):32-9. https://doi.org/10.37757/MR2020.V22.N3.8
5. Domingues JS. Análise do modelo de Gompertz no crescimento de tumores sólidos e inserção de um fator de tratamento. Biomat. 2011;21:103-12.
6. Dutra CM, Farias FM, Melo CAR. New approach of non-linear fitting to estimate the temporal trajectory of the COVID-19 cases. Braz J Health Rev. 2020;3(3):6341-56. https://doi.org/10.34119/bjhrv3n3-186
7. Hammer O, Harper DAT, Ryan PD. PAST: paleontological statistic software package for education and data analysis. Palaent Eletr. 2001;4(1):1-9.

8. Prado MF, Antunes BBP, Bastos LSS, Peres IT, Silva AAB, Dantas LF et al. Analysis of COVID-19 under-reporting in Brazil. Rev Bras Ter Intensiva. 2020;32(2):224-8. https://doi.org/10.5935/0103-507X.20200030

## Authors' Contributions

Dutra CM – Conception, planning (study design), data acquisition, analysis and interpretation, and writing of the manuscript. The author approved the final draft of the manuscript.

## Conflict of Interest

Authors have no potential conflict of interest to declare, related to this study's political or financial peers and institutions.