Forecasting the rate of cumulative cases of COVID-19 infection in Northeast Brazil: a Boltzmann function-based modeling study

Projeção da taxa de casos acumulados de COVID-19 no Nordeste brasileiro: um estudo de modelagem com base na função de Boltzmann

Previsión de la tasa de incidencia acumulada por infección de COVID-19 en la región nordeste de Brasil: un estudio de modelado basado en funciones de Boltzmann

Abstract

The COVID-19 death rate in Northeast Brazil is much higher when compared to the national average, demanding a study into the prognosis of the region for planning control measures and preventing the collapse of the health care system. We estimated the potential total cumulative cases of COVID-19 in the region for the next three months. Our study included all confirmed cases, from March 8 until April 28, 2020, collected from the official website that reports the situation of COVID-19 infections in Brazil. The Boltzmann function was applied to a data simulation for each set of data regarding different states. The model data were well fitted, with $R^2$ values close to 0.999. Up to April 28, 20,665 cases were confirmed in the region. The state of Ceará has the highest rate of accumulated cases per 100,000 inhabitants (75.75), followed by Pernambuco. We estimated that the states of Ceará, Sergipe and Paraíba will experience a dramatic increase in the rate of cumulative cases until July 31. Maranhão, Pernambuco, Rio Grande do Norte and Piauí showed a more discreet increase in the model. For Bahia and Alagoas, a 4.7 and 6.6-fold increase in the rate was estimated, respectively. We estimate a substantial increase in the rate of cumulative cases per 100,000 inhabitants in the region within three months, especially for Ceará, Sergipe and Paraíba. The Boltzmann function proved to be a simple tool for epidemiological forecasting that can help planning the measures to contain COVID-19.

COVID-19; Epidemiology; Mathematical Models; Pandemic
Introduction

In December 2019, a pneumonia outbreak related to COVID-19 was reported in the city of Wuhan, China, and soon spread to other regions. Three months later, the World Health Organization raised the state of contamination and declared it a pandemic due to the rapid geographical spread of the COVID-19 virus over a short-time scale.

Transmission of the etiological agent of COVID-19 occurs mainly through aspiration or contact with respiratory secretions of infected patients. Efforts to prevent it are being adopted worldwide, with an emphasis on social isolation of the population by restricting people circulation and grouping and blocking non-essential services, closing bars, hotels, shopping malls, restaurants, theaters, schools, universities, churches, elective clinical services, ports, airports and highways.

On April 28, 2020, COVID-19 had 3,098,391 confirmed cases worldwide, with 216,160 deaths in 185 regions. The first epicenter of the pandemic in China has substantially reduced the number of new cases since late February. Progressively, countries in Europe became the next global epicenter, followed by the United States. Current scenarios with the highest number of confirmed cases are the United States and Spain, with a total of 1,008,066 and 232,128 cases, respectively. The United States and Italy lead in number of registered deaths.

The epicenter of the coronavirus pandemic changes as the virus spreads to a new country. This depends on the speed of the response by health authorities to stop its transmission and provide adequate support to those who are sick and on the behavior of the population in respecting the recommended measures. However, cities in Latin America have shown low rates of social isolation, added to speeches that minimize their importance.

On April 28, 2020, no state in Brazil had reached the 63% social isolation rate. This factor, as well as regional inequalities in healthcare, low investment in public health and scientific research, a high poverty rate and government officials unfavorable to social isolation measures can be considered aggravating in the dissemination of COVID-19. Two months after the confirmation of the first COVID-19 case in the country, 71,866 cases and 5,017 deaths were confirmed, overtaking China in the same period of time and causing concern about the future of the pandemic in its regions.

Currently, the national incidence rate of COVID-19 is 145 cases per 1,000,000 inhabitants. The north of the country faces a public health calamity, with all intensive care beds occupied. The Southeast is the most affected region, considering the number of cases. Despite being outside the current epicenter of the coronavirus in Brazil, the Northeast Region has a death rate above the national average and is the second most affected region in absolute cases, which makes it important to study the prognosis of this region for planning control measures and preventing the collapse of the health system.

The potential of these situations can be estimated through complex mathematical models, such as those based on susceptible, infectious and recovered data (SIR), and simpler models in terms of understanding and application, such as the Boltzmann model, already used in studies in China. In our study, we estimated the potential total number of cases of COVID-19 in the Northeast Region of Brazil for the next three months by applying Boltzmann function-based regression analysis.

Methods

Design and study area

This epidemiological study used mathematical modeling and geoprocessing techniques. The spatial units of analysis were the nine states of Northeast Brazil (Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte and Sergipe) distributed over a total area of 1,561,177 km².

Data sources and measures

Our study included all confirmed cases of COVID-19 infection until April 28, 2020. COVID-19 infection was defined as a case with a positive result for viral nucleic acid testing in respiratory specimens.
or with a positive serological test. This data were collected from the official website \(^{13}\) that reports the situation of COVID-19 infections in Brazil. The data for model development were updated on April 29, 2020. The rates of cumulative cases of disease per 100,000 inhabitants were estimated considering the number of cases in each state divided by the population at risk based on the estimates for the states, obtained from the Brazilian Institute of Geography and Statistics (IBGE) \(^{17}\).

**Data analysis**

Data were organized in Microsoft Excel (https://products.office.com/) and incorporated into Microcal Origin software version 6.0 (https://microcal-origin.software.informer.com/6.0/). The Boltzmann function \(^{15,16,18}\) was applied to the data simulation for each set of data of the different geographical regions in Northeast Brazil. We obtained parameters of each function, in which the potential total number of confirmed cases is given by the parameter \(A_2\). The Boltzmann function for future simulation is expressed as follows:

\[
C(x) = A_2 + \frac{A_1 - A_2}{1 + e^{(x-x_0)/dx}} \quad (1),
\]

where \(C(x)\) is the cumulative number of confirmed cases after the first day \(x\); \(A_1, A_2, x_0\) and \(dx\) are constants. \(x_0\) corresponds to the inflection point and indicates the date on which the daily cases will reach their maximum. After that date, there will be a downward trend in total daily cases; \(dx\) is the adjustment coefficient, indicating the degree of increase in \(y\) (number of cases) as a function of the increase in \(x\) (days after the first case). In particular, \(A_2\) represents the estimated potential total number of confirmed cases. A *key date* (when number of daily new confirmed cases is lower than 0.1% in relation with total cases \(^{16}\)) was included in our study. Data from parameter \(A_2\) were used to estimate the rate of cumulative cases of COVID-19 per 100,000 inhabitants.

Maps were setup with the spread of the cumulative cases per 100,000 inhabitants of COVID-19 infection using actual and modelled Boltzmann data. Therefore, we used the cartographic base of Northeast Brazil available in the IBGE electronic database and reported data on COVID-19 infections \(^{17}\). Terra Datum model SIRGAS 2000 and the cartographic projection corresponding to the Mercator Transversal Universal system were used. The georeferenced data were incorporated into Quantum GIS version 3.10.5 (https://qgis.org/en/site/).

**Results**

The first cases of COVID-19 infection in Northeast Brazil were documented in the states of Alagoas and Bahia on March 8, 2020. Until April 28, 20,665 cases were confirmed across the region. The State of Ceará had the highest rate of accumulated cases per 100,000 inhabitants (75.75) during this period, followed by the states of Pernambuco (59.89), Maranhão (35.73), Rio Grande do Norte (24.44) and Alagoas (23.28). The lowest rates of cumulative cases were reported by the states of Sergipe (12.18), Piauí (12.46), Paraíba (15.75) and Bahia (17.08).

For Boltzmann data analysis, each region was well fitted and all \(R^2\) values were close to 0.999 (from March 8 to April 28, 2020) (Figure 1). The potential total number of confirmed cases for the next 150 days after the first case in each state of the Northeastern Region (Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte and Sergipe) were estimated as 5,110 ± 2,841, 11,836 ± 3,375, 100,627 ± 148,313, 3,462 ± 117, 10,854 ± 19,762, 10,364 ± 732, 1,071 ± 167, 1,662 ± 310, 3,537 ± 17,073, respectively (Figure 1). The *key date* was June 3, July 7, July 17, May 21, July 3, May 30, June 6, June 16 and June 20, for the states of Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte and Sergipe, respectively.

The potential number of cumulative cases per 100,000 inhabitants until May 27, 2020 was estimated for Ceará (542.46), Paraíba (170.73), Alagoas (150.3), Sergipe (124.04), Pernambuco (107.48), Bahia (63.29), Maranhão (48.86), Rio Grande do Norte (44.54) and Piauí (32.26). For July 31, 2020, the potential number of cumulative cases per 100,000 inhabitants was estimated as 1098.18, 270.05, 153.87, 153.13, 108.45, 79.53, 48.93, 47.39 and 32.73 for Ceará, Paraíba, Sergipe, Alagoas,
Figure 1

Fitting the cumulative number of COVID-19 cases to Boltzmann function in the states of the Northeast Region of Brazil.

1a) Alagoas State

1b) Bahia State

(continues)
Figure 1 (continued)

1d) Ceará State

1d) Maranhão State

(continues)
Figure 1 (continued)

1e) Paraíba State

1f) Pernambuco State

(continues)
Figure 1 (continued)

1g) Piauí State

1h) Rio Grande do Norte State
Pernambuco, Bahia, Maranhão, Rio Grande do Norte and Piauí, respectively. The mapping of the region shows the states with the potential rates of cumulative cases of COVID-19 over the period by color smoothing or intensification (Figure 2).

**Discussion**

By using data from March 8, 2020 to April 28, 2020 and the mathematical model incorporating these data, we provided an estimation of the rate of cumulative cases of COVID-19 infection per 100,000 inhabitants in Northeast Brazil for the next three months, specifically for May 27 and July 30, 2020. We estimated that the states of Ceará, Sergipe and Paraíba will see a dramatic increase in the rate of cumulative cases by up to 7.2, 10.4 and 10.8 times in a month and 14.5, 12.6 and 17.2-fold increases until July 31, respectively. Maranhão, Pernambuco, Rio Grande do Norte and Piauí were the states that showed a more discreet increase in the model, with the lowest potential of cumulative case rates until the end of the estimated period (1.4, 1.8, 1.9 and 2.6, respectively). For the states of Bahia and Alagoas, 4.7 and 6.6-fold increases, respectively, in the cumulative number of cases per 100,000 inhabitants were estimated for the period.

Knowing the number of infected by COVID-19 is essential to combat the spread of its etiological agent. However, we must note that our modeling approach did not consider many factors that may influence the number of cases recorded, along with the current situation of the pandemic. These factors refer to the number of tests made available by health services, the criteria for requesting tests, which are still restricted to certain cases, and the time spent on acquiring results, diagnoses and notification in the system. Thus, the actual number of cases in the period used for the model in our study may be even higher. In this context, one study estimated that Brazil could have eleven times more cases of COVID-19 than those officially registered 19.
Figure 2

Mapping of cumulative COVID-19 cases in Northeast Region of Brazil according to Boltzmann's function forecast.
Figure 2 (continued)
In the Northeast Region of Brazil, rapid tests for the serological diagnosis of COVID-19 infection only started after the second half of April 2020, being offered to a small percentage of the population with respiratory and flu symptoms, different from what is observed in other countries with a mass testing strategy. Furthermore, the results of exams made with real-time reverse transcriptase-polymerase chain reaction (RT-PCR) molecular testing still take many days to be released. These factors delay the registration of the actual number of confirmed cases and point to a possible reality of underreporting, since they omit patients with mild symptoms that do not seek health care, asymptomatic individuals or those in the incubation period of infection, who may be potential transmitters of the virus in the community.

Despite the tropical climate predominant in the region, a lower temperature is expected for the months considered in our study, which may be favorable to the spread of the severe acute respiratory syndrome coronavirus (SARS-CoV-2). This period is also favorable to the emergence of other common infections in Brazil, such as the common cold, influenza, dengue, Zika and chikungunya, which can lead to an even greater increase in demand for care and overburden health services.

Another challenge identified for the control of COVID-19 infection in these states are the urban structure of agglomeration of common communities in the peripheries and the threat of reduction of social isolation by the population, which has been showing signs of loosening. In the absence of a vaccine and specific treatment for cases of COVID-19 infection, it is strongly recommended and urgent that individuals of all ages collaborate to slow the estimated progression of the pandemic by avoiding overcrowding in hospital services, as well as the exposure of health professionals and more deaths. Up to April 30, 2020, at least 1,536 deaths from COVID-19 infection have been officially registered in the Northeast Region of Brazil. In addition to the epidemiological health numbers, the adoption of social isolation, the use of masks and the constant washing of hands by the population may also contribute to low more quickly the economic and emotional impacts caused by this pandemic.

Finally, our study showed that all data sets were well fitted to the Boltzmann function, which suggests that the model is suitable for analyzing cumulative cases of COVID-19. The validity of our results are based on model assumptions that states that the mechanisms and physical principles that govern the transmission of infectious diseases and human collective motion must be similar to those observed in the Boltzmann distribution probability. However, successful predictions depend on the accuracy of the data and it is understandable that the initial data do not reflect the true epidemic, since testing, identification, diagnostic and counting methods need time and effort to be properly established.

Another limitation of the methodology used corresponds to the non-linearity or sensitivity to conditions that may interfere with the future occurrence of COVID-19, since this estimate assumes that the overall conditions are not changing. Factors related to the host and its behavior, the pathogen’s ability to survive and the environmental influences can alter the estimate. Besides, the main advantage of the model used is that it only needs the cumulative number of confirmed cases; this represents a quick method for assisting central and local governments to deal with this emerging threat at the current critical stage.

**Conclusion**

Our results estimate a substantial increase in the rate of cumulative cases of COVID-19 infection per 100,000 inhabitants in Northeast Brazil over the next three months, with an emphasis on the states of Ceará, Sergipe and Paraíba. Maranhão, Pernambuco, Rio Grande do Norte and Piauí showed a more discreet increase in the modeling. For Bahia and Alagoas, 4.7 and 6.6-fold increases in the rate were estimated. All data sets were well fitted to the Boltzmann function, which was found to be a simple tool for epidemiological forecasting that could help planning measures to contain the COVID-19 pandemic. Social isolation measures may be the best strategy to slow the estimated progression of the infection in the studied period.
Contributors

All the authors contributed in the conception and design, acquisition of data, and analysis and interpretation of data; drafting the article or revising it critically for important intellectual content; final approval of the version to be published; and are responsible for all aspects of the work in ensuring the accuracy and integrity of any part of the work.

Additional informations

ORCID: Géssyca Cavalcante de Melo (0000-0002-6774-857X); Renato Américo de Araújo Neto (0000-0003-1337-7969); Karina Conceição Gomes Machado de Araújo (0000-0003-4433-5227).

References

Resumo

A Região Nordeste do Brasil tem uma taxa de letalidade muito mais elevada por COVID-19, comparado com a média nacional, o que exige uma investigação do prognóstico da doença para o planejamento de medidas de controle e para prevenir o colapso do sistema de saúde. Estimamos o total potencial de casos acumulados de COVID-19 na região nos próximos três meses. O estudo incluiu todos os casos confirmados de COVID-19, desde o primeiro caso confirmado, em 8 de março, até 28 de abril de 2020, coletados no site oficial que relata a situação das infeccões por COVID-19 no Brasil. A função de Boltzmann foi aplicada a uma simulação de dados para cada conjunto de dados dos diversos estados do Nordeste. Os dados do modelo mostraram bom ajuste, com valores de $R^2$ próximos a 0,999. Até 28 de abril, haviam sido confirmados 20.665 casos na Região Nordeste. O estado do Ceará apresenta a maior taxa de casos acumulados por 100.000 habitantes (75,75), seguido pelo estado de Pernambuco. Estimamos que Ceará, Sergipe e Paraíba apresentarão um aumento dramático na taxa de casos acumulados até 31 de julho. Maranhão, Pernambuco, Rio Grande do Norte e Piauí mostraram aumentos mais discretos de acordo com o modelo. Para Bahia e Alagoas, foram estimados aumentos de 4,7 e 6,6 vezes nas taxas, respectivamente. Estimamos um aumento sustancial na taxa de casos acumulados por 100.000 habitantes na Região Nordeste ao longo dos próximos três meses, especialmente no Ceará, Sergipe e Paraíba. A função de Boltzmann mostrou ser uma ferramenta simples para projeções epidemiológicas, podendo auxiliar no planejamento de medidas para conter a COVID-19.

COVID-19; Epidemiologia; Modelos Teóricos; Pandemias

Resumen

La región del nordeste brasileño cuenta con una tasa de mortalidad mucho más alta debido a la COVID-19, si se compara con la media nacional, por lo que es necesario un estudio en la prognosis de la región para planificar medidas de control y prevenir el colapso del sistema de salud. Estimamos el potencial total acumulativo de casos de COVID-19 en esta región durante los próximos tres meses. El estudio incluyó todos los casos confirmados de COVID-19, desde el primer caso, confirmado el 8 de marzo, hasta el 28 de abril de 2020, recogido del sitio web oficial que informa la situación de las infecciones por COVID-19 en Brasil. La función de Boltzmann se aplicó a la simulación de datos para cada conjunto de datos, referentes a diferentes estados. El modelo de datos estuvo bien ajustado, con valores $R^2$ cercanos a 0,999. Hasta el 28 de abril, se confirmaron 20.665 casos en la región. Ceará contó con la tasa más alta de incidencia acumulada por 100.000 habitantes (75,75), seguida de Pernambuco. Estimamos que Ceará, Sergipe y Paraíba sufrirán un dramático aumento en la tasa de incidencia acumulada de casos hasta el 31 de julio. Maranhão, Pernambuco, Rio Grande do Norte y Piauí mostraron un incremento más discreto en este modelo. En el caso de Bahía y Alagoas, se estimó un incremento de un 4,7 y 6,6, respectivamente. Estimamos un aumento sustancial en la tasa de incidencia acumulada de casos por 100.000 habitantes dentro de esta región, respecto a los tres próximos meses, especialmente en Ceará, Sergipe y Paraíba. La función de Boltzmann probó ser una herramienta simple para la previsión epidemiológica que puede ser de ayuda en la planificación de medidas para contener a la COVID-19.

COVID-19; Epidemiología; Modelos Teóricos; Pandemias