Metodología predictiva de la dinámica del número de casos de COVID-19: aplicación a China, Bélgica y Corea del Sur

Predictive methodology of the dynamics of the number of COVID-19 cases: application to China, Belgium, and South Korea

Javier Orswaldo Rodríguez Velásquez [a], Joao Cuesta Rivas [b], Sandra Catalina Correa Herrera [c].

a. MD. Insight Group. Hospital Universitario Nacional de Colombia. Bogotá, Colombia. ORCID: https://orcid.org/0000-0002-4585-3010.

b. Degree in science in Education. MSc. Microbiology. Education and Humanities Faculty, Universidad Militar Nueva Granada, Bogotá, D.C, Colombia. ORCID: https://orcid.org/0000-0002-0814-4186.

c. Psychologist. Insight Group. Hospital Universitario Nacional de Colombia. Bogotá, Colombia. ORCID: http://orcid.org/0000-0003-4998-1228.

DOI: 10.22517/25395203.24932

Abstract

Objectives: Multiple methodologies based on probability theory have been developed to establish predictions of dengue, malaria, HIV, obesity epidemics, among others. This research aimed to develop a new method for predicting the dynamics of the number of COVID-19 cases for China, Belgium, and South Korea based on the probability theory that allows the evaluation and comparison of their increment.

Material and methods: Probability ranges of the number of COVID-19 cases were established, which were assigned to each of the daily number of COVID-19 cases reported by China, Belgium, and South Korea that were evaluated during 74, 50, and 50 days respectively. The frequency and probability of each daily range for each country was calculated. Their total probability and the probability of the dynamics in intervals of 8 consecutive days were calculated, and the values between countries were compared to evaluate their differences.

Results: Probability values of 1.21E-30, 2.03E-22, and 3.15E-12 were established for China, Belgium, and South Korea, which allows the quantitative differentiation of the characteristics of their dynamics. The probability differences of the 8-day subspaces ranged from 0.003 to 1, allowing the temporal changes in the dynamics to be evaluated.

Conclusion: The ranges established for the evaluation of the number of COVID-19 cases allow to differentiate the behavior of epidemics between countries and to stratify the severity of expansion. Highlighting an underlying mathematical order for this phenomenon permitted quantitatively predict its spatiotemporal dynamic and indirectly, the efficacy of public health politics implemented for each country.

Keywords: Probability; Coronavirus; Pandemic; Public Health

Resumen

Objetivos: Se han desarrollado múltiples metodologías basadas en la teoría de la probabilidad para establecer predicciones de epidemias de dengue, malaria, VIH, obesidad, entre otras. Esta investigación tuvo como objetivo desarrollar un nuevo método de predicción de la dinámica del número de casos de COVID-19 para China, Bélgica y Corea del Sur, basado en la teoría de la probabilidad que permite evaluar y comparar su crecimiento.

Materiales y métodos: Se establecieron rangos de probabilidad del número de casos de COVID-19, los cuales fueron asignados a cada uno de los valores diarios del número de casos de COVID-19 reportados por China Bélgica y Corea del Sur evaluados durante 74, 50 y 50 días, respectivamente. Se calculó la frecuencia y probabilidad de cada rango diario para cada país. Se calculó su probabilidad total y la probabilidad de la dinámica en intervalos de 8 días consecutivos y se compararon los valores entre países para evaluar sus diferencias.

Resultados: se establecieron valores de probabilidad de 1.21E-30, 2.03E-22 y 3.15E-12 para China, Bélgica y Corea del Sur, lo que permite diferenciar cuantitativamente las características de su dinámica. Las diferencias de probabilidad de los subespacios de 8 días variaron de 0,003 a 1, lo que permitió evaluar los cambios temporales en la dinámica.

Conclusión: los rangos establecidos para la evaluación del número de casos de COVID-19 permiten diferenciar el comportamiento de las epidemias entre países y estratificar la severidad de la expansión, destacando un orden matemático subyacente para este fenómeno que permite predecir cuantitativamente su dinámica espacio-temporal y indirectamente, la eficacia de las políticas de salud pública implementadas para cada país.

Palabras clave: Probabilidad; Coronavirus; Pandemia; Salud Pública.

Introduction

Coronavirus-induced disease 2019 (COVID-19) was classified by the World Health Organization as a pandemic due to its alarming levels of spread and severity in almost all countries (1,2). Its serious impact forced the creation of therapeutic and preventive tools as well as public health policies capable of being adjusted in real time to the rapidly changing demands caused by its rapid expansion. Given that the appearance of COVID-19 in humans occurred very recently and its expansion has been very rapid, there is not yet a complete and systematic understanding of its transmission patterns, clinical characteristics, and risk factors for infection, which hinder the decision-making processes that minimize the impact of the disease (3). Therefore, it is essential to develop new methods that allow evaluating and predicting the dynamics of the pandemic over time.

Mortality is estimated to be around 3-4% of cases (4), with the highest risk groups being older adults with chronic diseases such as advanced diabetes, hypertension, heart disease, respiratory diseases, or cancer. (5). Three of the most affected countries by the pandemic are China, Belgium, and South Korea, which have reported 81,708, 20,814 and 10,284 cases respectively, as well as 3.331, 1.632 and 186 deaths as of April 6, 2020, according to Worldometer.info (6).

To solve the problems involved in epidemiological analysis and causal mathematical models, Rodríguez et al., have designed various mathematical and physical methodologies from an acausal perspective. This is conducted without considering the causal variables that can influence the variation of dynamics and focusing on its mathematical characteristics by making use of physical and mathematical theories such as probability, the random walk probabilistic model, second-order differential equations, the K-ratio, and recurring probability (7-10). Probability is defined as a dimensionless mathematical measure of the possibility of the occurrence of an event in the future and constitutes the foundation of most of the mentioned methodologies (11).

Among these methodologies, it is worth highlighting predictions achieved in ranges of weeks in municipalities, such as a weekly prediction method of malaria outbreaks in the municipalities of Colombia based on the probability and the K-ratio, which achieved a percentage of success of 99.86%. This constitutes a system that overcomes the problems of models based on endemic corridors, as well as the drawbacks of considering chance as the predominant factor in the dynamics (9). A method based on recurrent temporal probabilistic prediction was also developed, which allowed to obtain the number of malaria outbreaks and their duration in epidemiological weeks in different prototypical municipalities of Colombia. This permitted the dynamics of the epidemic to be evaluated in real time and to quantify when the onset of an outbreak is more probable with differences in order of magnitude between 10 and 103 (10). These methodologies are useful to develop surveillance and early warning systems for malaria outbreaks in Colombia.

The purpose of this study is to develop a predictive methodology based on the theory of probability of the dynamics of the number of COVID-19 cases applied in China, Belgium, and South Korea to evaluate the dynamics in time intervals of 8 days and make comparisons of their behaviors.

Table 1. Defined ranges for classifying the number of daily COVID-19 cases in China, Belgium, and South Korea.

Grade	Total number of cases	
	Minimal	Maximal
1	0	200
2	200	10 000
3	10 000	30 000
4	30 000	60 000
5	60 000	110 000
6	110 000	170 000
7	170 000	-

Consequently, the range to which each of the daily values reported for each country corresponded was determined. From these data, the frequency of appearance of each range in the evaluated days and the probability value associated with these were determined ranges for each country. The probability value associated with the range corresponding to each day was established below using equation 1.

$$P_i = \frac{Range\ i\ frequency\ of\ ocurrence}{Sum\ of\ frequencies\ of\ occurrence}\ Equation\ 1$$

Equation 2 was applied to determine a probability value for the entire individual dynamic of each country by multiplying each of the probabilities associated with each day.

$$P_{dynamic} = \prod_{i=1} P_{day_i}\ Equation\ 2$$

Subsequently, this procedure was repeated but taking time intervals of 8 days in such a way that days 1 to 8 correspond to the first dynamic, days 9 to 16 to the second dynamic, and so on. In this way, it was possible to establish a probability value associated with each of these time intervals for each country. These values were used to evaluate the characteristics of the pandemic expansion in the three evaluated countries and to compare their behavior.

Results

Without considering the frequencies and probabilities of 0, the values of the frequency of appearance of the ranges evaluated between 6 and 52 were presented for China, 7 and 23 for Belgium, and 2 and 42 for South Korea, as well as probability values between 0.081 and 0.703; 0.14 and 0.46; and 0.04 and 0.84 for these three countries respectively. The results show that the daily dynamics of the number of COVID-19 cases are not equiprobable, but that there are ranges that tend to be repeated. Based on this, it is possible to establish differences in the dynamics from the study of the distribution of frequencies and probabilities of the ranges evaluated. Thus, when observing the highest probability values for each country, it is observed that range 5 presents a probability of 0.703 for China while in Belgium the most probable range is 1, with a probability of 0.46. On the other hand, in South Korea the highest probability is 0.84, corresponding to rank 2 (Table 2).

Table 2. Frequency and probability of the established ranges for China, Belgium, and South Korea of the number of daily COVID-19 cases.

China				Belgium				South Korea		
R	**F**	**P**		**R**	**F**	**P**		**R**	**F**	**P**
7	0	0		7	0	0		7	0	0
6	0	0		6	0	0		6	0	0
5	52	0.703		5	0	0		5	0	0
4	7	0.095		4	0	0		4	0	0
3	6	0.081		3	7	0.14		3	2	0.04
2	9	0.122		2	20	0.4		2	42	0.84
1	0	0		1	23	0.46		1	6	0.12
TOTAL	**74**	**1**		**TOTAL**	**50**	**1**		**TOTAL**	**50**	**1**

*R: range. F: frequency of appearance; P: probability.

When establishing the total probability of the dynamics by multiplying the probabilities found throughout the dynamics of each country with equation 2, it was found that China presented a probability of 1.21E-30 while Belgium and South Korea had probabilities of 2.03E-22 and 3.15E-12 respectively. These results show clear differences between the evaluated dynamics of the number of COVID-19 cases over time, with notable differences of 8 and 18 orders of magnitude between China, Belgium, and South Korea accordingly, and 10 orders of magnitude between South Korea and Belgium.

Finally, 6 probability subspaces of 8 days were presented, the evaluations of the probability subspaces of the 8-days intervals are presented in Table 3. The results showed that from the three evaluated countries, China was the one with the strongest and fastest expansion of the number of COVID-19 cases. This is evident from the beginning of the timeline since the values of this variable were located in range 2 to later jump to range 4 and then remain stable for 3 intervals in range 5. In contrast, Belgium lasted two 8-day intervals in range 1, 3 intervals in range 2, and an interval in range 3, showing a much less drastic rise. Finally, South Korea maintained a stable behavior in range 2 during the 6 intervals evaluated, showing the most stable behavior of the three countries and therefore, a much more optimal management of the pandemic.

Table 3. Probabilities established for the 8-day probability subspaces for the dynamics of China, Belgium, and South Korea

China		Belgium		South Korea	
R	**P**	**R**	**P**	**R**	**P**
2	1	1	1	2	0.011
4	0.003	1	1	2	1
5	0.011	2	0.049	2	1
5	1	2	1	2	1
5	1	2	1	2	1
5	1	3	0.005	2	1

*R: Range. P: Probability

Discussion

This is the first work in which space-temporal predictions of the dynamics of the number of COVID-19 cases in China, South Korea, and Belgium are established using an innovative methodology. This methodology is based on probability theory and the establishment of ranges of specific interest in analogy to the energy levels at which electrons are found in the context of quantum physics. It allows the predictions of the number of COVID-19 cases from probabilities of intervals of 8 consecutive days, which also allows the quantification of the differences between the dynamics of the different countries.

This methodology is applicable and functional for epidemiological follow-up processes and constitutes a support tool for decision-making at the level of public health policies considering that it can be easily applied by the entities that carry out these follow-ups and that try to predict COVID 19 dynamics in real time. In view of the urgency of systematizing predictions over time, this methodology can be applied anywhere, anytime, and in any country in the world. Another advantage of its application is its independence from causal factors in the same way as the theories of modern physics, which eases the achievement of practical predictions in a simpler way than with multi-causal models. It is applicable to the evaluation of the global impact of the specific containment measures established by each country such as preventive isolations, quarantines, use of face masks and so forth.

Aiming at predicting the dynamics of the pandemic in each country, several models that seek to clarify this phenomenon have already been made (12-17). As an example, thanks to statistical models, it has been possible to predict that the demand for medical supplies will exceed the capacity of available resources in the coming weeks in the United States (13). A model that evaluates the behavior of the epidemic in China using artificial intelligence was also developed, it predicted a decline in late April and a possible outbreak if the quarantine was lifted (14). These and other mathematical processes pretend to establish relationships between some associated factors — such as comorbidities, age, sex, among others — and the dynamics of the number of cases of various diseases (16-17).

This approach, which seeks causal relationships, presents difficulties in establishing and quantifying the relationship between these factors and changes in disease dynamics. Unlike these models, the methodology proposed here is based on physical and mathematical laws that address this problem and establish predictive quantitative values from an acausal perspective. In the same way that in current theoretical physics, the causes that determine the behavior of phenomena are not conceived, but the mathematical orders that allow the understanding of their nature are established.

The approach that prevailed in this study was that the phenomena work in a "charged" probabilistic way in analogy with the energy levels of the electrons, which solves the problem of random fluctuations (18).

This acausal perspective previously allowed the development of innovative methodologies for the prediction of malaria outbreaks in Colombian municipalities (9) as well as the quantification of their quantity and duration (10) based on physical and mathematical theories. Similarly, this approach has led to the development of diagnoses and predictions in other areas of medicine, such as predictions of mortality (19), peptide binding (20), diagnoses of cell morphometry of preneoplastic and neoplastic cells of the cervix (21), neonatal and adult cardiac dynamics (22,23), and temporal predictions of the number of CD4+ T lymphocytes in patients with HIV/AIDS (24). These works demonstrate the utility of studying medical phenomena from physical and mathematical theories, revealing underlying mathematical orders to these phenomena

Funding

This work was supported by the Nueva Granada Military University [grant numbers: UMNG HUM 2357, Country]; the institution did not influence the research.

Acknowledgments

We thank the Universidad Militar Nueva Granada for their support to our research, especially to its Research Fund and the Vice-Rectory of Research. We also extend our thanks to Doctor Giancarlo Buitrago, director of the "Instituto de Investigaciones Clínicas" at Universidad Nacional de Colombia-Hospital Universitario Nacional de Colombia, and Doctor Rubén Caycedo, Head of the Surgery department of the Faculty of Medicine at Universidad Nacional de Colombia-Hospital Universitario Nacional de Colombia.

Conflict of interests None declared.

E-mail: grupoinsight2025@gmail.com

«Similarly, this approach has led to the development of diagnoses and predictions in other areas of medicine, such as predictions of mortality.»

References

1. Fehr AR, Perlman S. Coronaviruses: An Overview of Their Replication and Pathogenesis. Methods Mol Biol 2015;1282:1-23. doi: 10.1007/978-1-4939-2438-7_1

2. Wang Y, Sun J, Zhu A, Zhao J, Zhao J. Current understanding of middle east respiratory syndrome coronavirus infection in human and animal models. J Thorac Dis. 2018;10:S2260–S2271. doi: 10.21037/jtd.2018.03.80

3. Song Z, Xu Y, Bao L, Zhang L, Yu P, Qu Y, et al. From SARS to MERS, Thrusting Coronaviruses into the Spotlight. Viruses. 2019;11:59. doi: 10.3390/v11010059

4. de Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuane L, et al. Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. J Virol 2013;87:7790-2. doi:10.1128/JVI.01244-13

5. Zhong NS, Zheng BJ, Li YM, Poon LLM, Xie ZH, Chan KH, et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. Lancet 2003;362:1353-8.doi: 10.1016/S0140-6736(03)14630-2

6. Worldometer [Internet] COVID-19 Coronavirus Pandemic; 2020 [Citado 04 de abril de 2020]. Disponible en: https://www.worldometers.info/coronavirus/#countries

7. Rodríguez- Velásquez J, Vitery- Erazo S, Puerta G, Muñoz D, Rojas I, Pinilla- Bonilla L, Mora J, Salamanca D, Perdomo N. Dinámica probabilista temporal de la epidemia de dengue en Colombia. Revista Cubana de Higiene y Epidemiología. 2011; 49(1).Disponible en: http://bvs.sld.cu/revistas/hie/vol49_1_11/hiesu111.htm.

8. Rodríguez J, Jattin J, Soracipa Y. Probabilistic temporal prediction of the deaths caused by traffic in Colombia. Mortality caused by traffic prediction. Accident Analysis and Prevention. 2020;135:105332.

9. Rodríguez JO. Spatio-temporal probabilistic prediction of appearance and duration of malaria outbreak in municipalities of Colombia. Journal of Physics: Conf. Series. 2019;1160:012018.

10. Rodríguez J. Método para la predicción de la dinámica temporal de la malaria en los municipios de Colombia Rev. Panam. Salud Pública. 2010;27(3) 211-218

11. Feynman RP, Leighton RB, Sands M. Probabilidad. En: Feynman RP, Leighton RB, Sands M, (Eds.). Física. Vol. 1 (pp. 6-1, 6-16). Wilmington: Addison-Wesley Iberoamericana, S. A. 1964

12. Ayyoubzadeh SM1, Ayyoubzadeh SM2, Zahedi H3, Ahmadi M4, R Niakan Kalhori S1.Predicting COVID-19 incidence using Google Trends and data mining techniques: A pilot study in Iran. JMIR Public Health Surveill. 2020 . doi: 10.2196/18828.

13. IHME COVID-19. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. MedRxiv. 2020. doi: 10.1101/2020.03.27.20043752

14. Yang Z, Zeng Z, Wang K, Wong S, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis 2020;12(3):165-174

15. Jung S-M, Akhmetzhanov AR, Hayashi K, Linton NM, Yang Y, Yuan B, Kobayashi T, Kinoshita R, Nishiura H (2020) Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: inference using exported cases. J Clin Med 9:523

16. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KS, Lau EH, Wong JY (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. N Engl J Med 382:1199–1207

17. Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic?. Infect Dis Model. 2020;5:271–281. Published 2020 Mar 25. doi:10.1016/j.idm.2020.03.001

18. Feynman RP, Leighton RB, Sands M. Quantum Mechanics. In: The Feynman Lectures on Physics. Vol 3. California Institute of Technology. 1963

19. Rodríguez J. Dynamical systems applied to dynamic variables of patients from the Intensive Care Unit (ICU). Physical and mathematical Mortality predictions on ICU.J.Med.Med. Sci. 2015; 6(8): 102-108.

20. Rodríguez J, Bernal P, Prieto P, Correa C, Álvarez L, Pinilla L, et al. Predicción de unión de péptidos de Plasmodium falciparum al HLA clase II. Probabilidad, combinatoria y entropía aplicadas a las proteínas MSP-5 y MSP-6. Archivos de alergia e inmunología clínica. 2013; 44(1): 7-14.

21. Prieto SE, Rodríguez JO, Correa SC, Soracipa MY. Diagnosis of cervical cells based on fractal and Euclidian geometrical measurements: Intrinsic Geometric Cellular Organization. BMC Med Phys. 2014;14:2

22. Rodríguez J, Prieto S, Flórez M, Alarcón C, López R, Aguirre G, Morales L, Lima L, Méndez L. Physicalmathematical diagnosis of cardiac dynamic on neonatal sepsis: predictions of clinical application. J.Med.Med. Sci.2014; 5(5): 102-108.

23. Rodríguez J, Prieto S, Ramírez LJ. A novel heart rate attractor for the prediction of cardio-vascular disease. Informatics in Medicine Unlocked. 2019;15:100174

24. Rodríguez J, Prieto S, Pérez C, Correa C, Soracipa Y, Jattin J, et al. Predicción temporal de CD4+ en 80 pacientes con manejo antirretroviral a partir de valores de leucocitos. Infectio. 2020;24:103-107