

RESEARCH ARTICLE

Trends of SARS-Cov-2 infection in 67 countries: Role of climate zone, temperature, humidity, and curve behavior of cumulative frequency on duplication time

Authors

Jaime Berumen, MD¹, Max J. Schmulson², Guadalupe Guerrero³, Elizabeth Barrera¹, Jorge Larriva-Sahd⁴, Gustavo Olaiiz⁵, Rebeca García-Leyva¹, Rosa María Wong-Chew⁶, Miguel Betancourt-Cravioto⁷, Hector Gallardo⁷, Germán Fajardo-Dolci⁸, Roberto Tapia-Conyer⁷

Affiliations

¹Unidad de Medicina Experimental, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico.

²Laboratorio de Hígado, Páncreas y Motilidad (HIPAM)-Unidad de Investigación en Medicina Experimental, Facultad de Medicina-Universidad Nacional Autónoma de México (UNAM).

³Hospital General de México, Dr. Eduardo Liceaga, Mexico City, Mexico.

⁴Instituto de Neurobiología, Universidad Nacional Autónoma de México (UNAM), Campus Juriquilla, Querétaro Mexico

⁵Centro de Investigación en Políticas, Población y Salud, Facultad de Medicina, Universidad Nacional Autónoma de México, Mexico City, Mexico.

⁶División de Investigación, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico

⁷Fundación Carlos Slim, Mexico City, Mexico

⁸Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico

***Correspondence**

Jaime Berumen, MD

E-mail: jaimeberumen47@gmail.com

Roberto Tapia-Conyer, MD

E-mail: tapiaconyer@gmail.com

Abstract

Aims. To analyze the role of temperature, humidity, date of first case diagnosed (DFC) and behavior of the growth-curve of cumulative frequency (CF) [number of days to rise (DCS) and reach the first 100 cases (D100), and the difference between them (ΔDD)] with the doubling time (Td) of COVID-19 cases in 67 countries grouped by climate zone.

Methods. Retrospective study based on the WHO registry of cumulative incidence of COVID-19 cases. 1,706,914 subjects diagnosed between 12-29-2019 and 4-15-2020 were analyzed based on exposure to SARS-CoV-2 virus, ambient humidity, temperature, and climate areas (temperate, tropical/subtropical). DCS, D100, ΔDD , DFC, humidity, temperature, Td for the first (Td10) and second (Td20) ten days of the CF growth-curve between countries and were compared according to climate zone, and identification of factors involved in Td, as well as predictors of CF using lineal regression models.

Results. Td10 and Td20 were ≥ 3 days longer in tropical/subtropical vs. temperate areas (2.8 ± 1.2 vs. 5.7 ± 3.4 ; $p=1.41E-05$ and 4.6 ± 1.8 vs. 8.6 ± 4.2 ; $p=9.7E-05$, respectively). The factors involved in Td10 (DFC and ΔDD) were different than those in Td20 (Td10 and climate areas). After D100, the fastest growth-curves during the first 10 days, were associated with $Td_{10} < 2$ and $Td_{10} < 3$ in temperate and tropical/subtropical countries, respectively. The fold change $Td_{20}/Td_{10} > 2$ was associated with earlier flattening of the growth-curve. In multivariate models, Td10, DFC and ambient temperature were negatively related with CF and explained 44.7% ($r^2 = 0.447$) of CF variability at day 20 of the growth-curve, while Td20 and DFC were negatively related with CF and explained 63.8% ($r^2 = 0.638$) of CF variability towards day 30 of the growth-curve.

Conclusions. Larger Td in tropical/subtropical countries is positively related to DFC and temperature. Td and environmental factors explain up to 64% of CF variability. However, pandemic containment measures may explain the remaining variability.

Key words: COVID-19; Temperate countries; Tropical and subtropical countries; Cumulative frequency; Pandemic; Doubling time; Speed of infection spread; Containment measures

1 Introduction

In December 2019, a new epidemic of severe acute respiratory syndrome (SARS) virus (SARS-CoV-2) broke out in China. The World Health Organization (WHO) denominated this novel viral as coronavirus disease-2019 (COVID-19) and declared it a pandemic by March 11th-2020.¹⁻³ The SARS-CoV-2 spread from

China to other Asian countries including Thailand, South Korea and Japan and then Australia. Subsequently, new cases appeared in Europe, the United States, Canada, and Iran, and by the end of February 2020 there were already cases in Brazil, Mexico, Greece, and Norway, among others. On March 13, there was an increase in the number of cases in western

European countries and the illness also emerged in African countries such as Sudan, Ethiopia, Kenya, and Guinea. On March 22, there were new cases in countries from all continents with increasing numbers in the United States and South America.⁴ By April 11th, 2020, a total of 1,254,464 cases had been confirmed worldwide, with 68,184 (5.4%) deaths.⁵ According to the above reports, the infection rate among the total population was much higher in Europe than in the Americas. In Europe for example, as of April 5th, 2020, the infection rate varied from 0.06% in Sweden to 0.27% in Spain.⁵ In contrast, in the Americas, up to the same date in April, the United States reported the highest number of cases with a total of 327,920. However, this only comprised 0.01% of its total population, and Canada had a 0.04% infectious rate. The data from these North American countries, were low, but still 10 times more frequent than in the rest of the countries of Central and South America, with rates ranging from 0.04% in Panama, 0.001% in Mexico, 0.002% in Colombia and 0.0003% in Guatemala.⁵ The above described infectious rates, decreasing from Europe to North America and South America, is in agreement with the time frame of the spreading path that this pandemic has followed.⁴

However, the much lower infection rates in South America may also reflect an underestimation of the illness, although other variables including climatologic differences deserve to be studied. Cold and dry environment may enhance the susceptibility to COVID-19 infection.^{6, 7} Thus, it is plausible that the higher temperatures and humidity in Latin

America countries, explain the much lower infectious rates of COVID-19 when compared to the Northern hemisphere.⁸ Therefore we aimed to analyze the role of environmental factors (temperature and humidity), the date of first case diagnosed (DFC) and the behavior of the growth curve of cumulative frequency (CF) [number of days to rise (DCS) and reach the first 100 cases (D100), and the difference between them (ΔDD)] with the doubling time (Td) of COVID-19 cases in 67 countries grouped by climate zone, as well as the identification of factors involved in Td and predictors of CF using lineal regression models. We hypothesized that the elevation of the curve, the time to reach the first 100 cases, as well as the time for duplicating the number of cases, should be faster in number of days in countries within temperate areas when compared to those in tropical or subtropical areas.

2 Methods

This is a retrospective incident case study that analyzed exposure to SARS-CoV-2 virus, ambient humidity, temperature, and climate areas (temperate, tropical/subtropical). The study populations were the daily confirmed CF of diagnosed cases of COVID-19 from 67 countries, officially reported by the WHO, from December 29, 2019 to April 15, 2020 (1,706,914 subjects).³ Eighteen countries are located in temperate or cold areas, and 49 in tropical or subtropical regions.³ The CF and the DFC in each country were obtained from this reports.

No informed consent was obtained as this was not an interventional study nor a direct survey of the study subjects. This

study was based solely on the analysis of a public international registry of subjects that have been tested for SARS-CoV-2 RT-PCR over the world. In addition, no personal identifications are present in this dataset therefore there is no bridge in the privacy of the study subjects. However, the Research and Ethics Committee of the Faculty of Medicine of the Universidad Nacional Autónoma de México was consulted, and they replied that ethical approval was not required.

2.1 Main outcome measures

Comparison of DCS, D100, Δ DD, DFC, humidity, temperature, Td for the first (Td10) and second (Td20) ten days of the CF growth curve between countries according to climate zone, and identification of factors involved in Td, as well as predictors of CF using lineal regression models.

2.1.1 Average, minimum, and maximum temperatures

The average temperature and relative humidity were collected from Time and Date AS database during the months of January, February and March for each of the 67 countries analyzed.⁹

2.1.2 Calculation of doubling time and the parameters of the growth curve of CF of COVID-19 cases

The CF of COVID-19 cases of each country was plotted in Excel and the exponential equation was obtained. DCS, D100 and Δ DD were graphically identified with the WHO data. The Td was obtained from the slope (λ) of the exponential graph ($N=N_0e^{\lambda t}$) as follows: $Td=\ln(2)/\lambda$.¹⁰ It was

calculated for the first (Td10), second (Td20) and third (Td30) 10-day periods of the CF curve, as well as for the entire 30-day period (TdT), starting from day D100. Td30 was not calculated for tropical/subtropical countries as the CF curve have not reached that period yet at the time of writing this article.

2.2 Statistical analysis

Numerical variables were described with medians and interquartile range (IQR) or means and standard deviations. The significance of differences between the groups (temperate vs. tropical/subtropical zones) was assessed with the Mann-Whitney U-test or the t-test. The Pearson's correlation was performed for some numerical variables. The association of significant variables with Td was explored using univariate (ULR) and multivariate (MLR) linear regression models. Finally, we built models to predict the CF of COVID-19 cases expected to happen on day 20 and 30 of the pandemic growth curve with the Td and the rest of variables using ULR and MLR models. To enter the DFC and the climate zone in the linear regression models the nominal data were transformed into numeric values, as follows: 1) the DFCs in the 67 countries were sorted in ascending order, and numerical values in ascending order, starting with the number 1, were assigned to each date, 2) for the climate zones, the values of 1 and 2 were assigned to temperate and tropical/subtropical zones, respectively. The association was expressed as the β coefficient and 95% confidence interval (CI), and the contribution to the variability of Td or CF

was expressed as adjusted r^2 . Variables with $p < 0.20$ in ULR analysis were considered for entry in MLR models. Confounders were defined as those variables for which the percentage difference of β coefficient between the adjusted and non-adjusted variables in the MLR model were higher than 10% ($p > 0.1$). The factors and interactions were included in the model in one block. A post hoc power analysis was performed for each linear regression model using the software G * Power 3.1.9.2, considering the sample size, the β and an $\alpha = 0.05$. All statistical tests were two-sided. The statistical analyses were conducted using SPSS version 20 software (SPSS Inc., Chicago, IL, USA).

3 Results

3.1 Analysis of the initial phase of the growth curve

The CF grew from the day the first case of COVID-19 was diagnosed in each of the 67 countries is depicted in Fig 1. The day the curve started to ascend (DCS) and the slope of the curves were observed. There is a wide variation between countries, but similar patterns are generally observed across continents or climatic regions. Unexpectedly, in countries located in temperate areas, such as most of the European countries, USA, Canada, Japan and Korea, the rise of the curve began late, 20 days after the first case was diagnosed. Conversely, in most countries located in tropical or subtropical areas such as those in the Middle East, Africa, Mexico, Central and South America, Asia and Oceania, the rise of the curves began much earlier. However, in this group there are notable

exceptions such as New Zealand, Australia, Egypt, and India, where the curve started belatedly. In fact, the median number of days it took for the curves to start to ascend was much higher in temperate countries than in tropical/subtropical ones [median (IQR) 29 (8-32) vs. 12 (8-16), $p=0.015$; Man-Whitney U-test]. On the other hand, the slope of the curves was much higher in temperate than in tropical/subtropical countries (Fig 1). An indirect way to evaluate the slope, is to measure the number of days it takes for the curve to reach 100 cases, once the ascend begins (ΔDD). While the D100 is much higher [median (IQR) 32 (12-36) vs. 20 (14-25); $p=0.108$, Man-Whitney U-test], the ΔDD is much lower in temperate countries [median (IQR) 3 (2-5) vs 7 (6-9); $p=0.002$, Man-Whitney U-test] than in tropical or subtropical areas. This suggests that the rate for doubling the number of positive cases is higher in temperate countries.

Interestingly, we found that the DFC had a linear negative correlation with the DCS ($r = -0.75$; $p = 1.3E-29$, Pearson correlation; Fig 2 A) and the D100 ($r = -0.71$; $p = 4.4E-19$, Pearson correlation; Fig 2 B); that is, the later the first case was diagnosed, the faster the DCS and D100 days were reached, and vice versa. These correlations could be related to temperature and relative humidity, since the ascending dates from January to March correlated with those two variables ($r = 0.658$, $p = 2.7E-26$, and $r = 0.233$, $p = 0.001$, Pearson correlation). However, in the analysis of the data, only the DCS day correlated negatively with the temperature ($r = -0.345$; $p = 5.4E-7$, Pearson correlation).

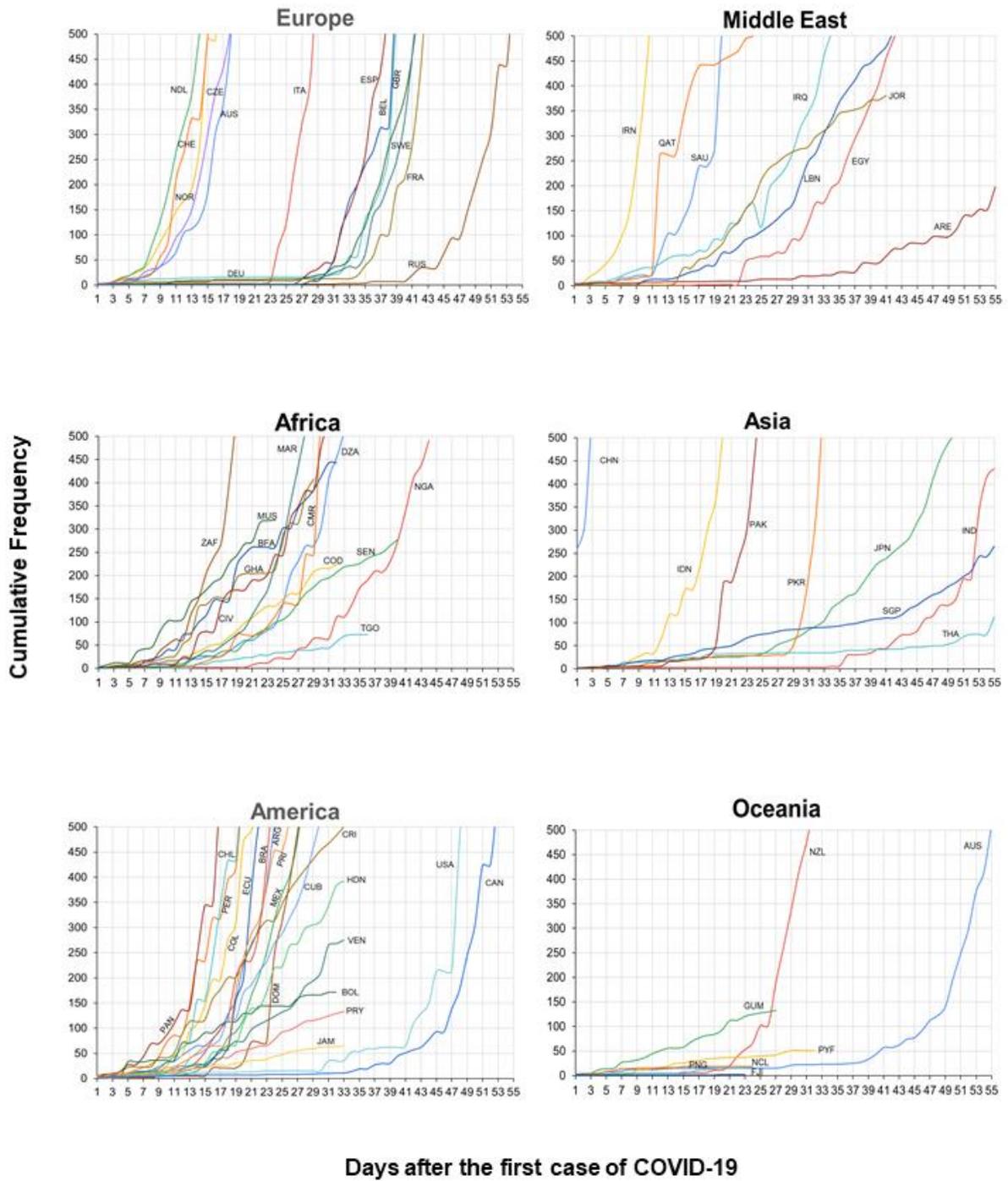


Figure 1

Fig 1. Days when the curves started to climb (DCS) and reach the first 100 (D100) COVID-19 cases. The curves show the cumulative frequency (CF) starting on the day when the first case of COVID-19 was diagnosed in each country. Based on these graphs we selected the days when the curves begin to climb and reach the first 100 cases. Abbreviations of each country: CHN: China, THA: Thailand, JPN: Japan, PKR: Republic of Korea, USA: USA, SGP: Singapore, AUS: Australia, FRA: France, CAN: Canada, DEU: Germany, ARE: United Arab Emirates, IND: India, ITA: Italy, RUS: Russian Federation, SWE: Sweden, GBR: United Kingdom, ESP: Spain, SAU: Saudi Arabia, BEL: Belgium, EGY: Egypt, IRN: Iran, LBN: Lebanon, IRQ: Iraq, CHE: Switzerland, AUT: Austria, BRA: Brazil, NOR: Norway, DZA: Algeria, PAK: Pakistan, NDL: Netherlands, NGA: Nigeria, NZL: New Zealand, MEX: Mexico, ECU: Ecuador, QAT: Qatar, CZE: Czech Republic, DOM: Dominican Republic, IDN: Indonesia, JOR: Jordan, MAR: Morocco, SEN: Senegal, CHL: Chile, ARG: Argentina, CMR: Cameroon, ZAF: South Africa, PER: Peru, COL: Colombia, CRI: Costa Rica, PAN: Panama, BOL: Bolivia, BFA: Burkina Faso, COD: Republic of Congo, HDN: Honduras, CIV: Ivory Coast, CUB: Cuba, PRI: Puerto Rico, VEN: Venezuela, GHA: Ghana, MUS: Mauritius, TGO: Togo, PRY: Paraguay, JAM: Jamaica, PNG: Papua New Guinea, NCL: New Caledonian, GUM: Guam, PYF: French Polynesia, FJI: Fiji.

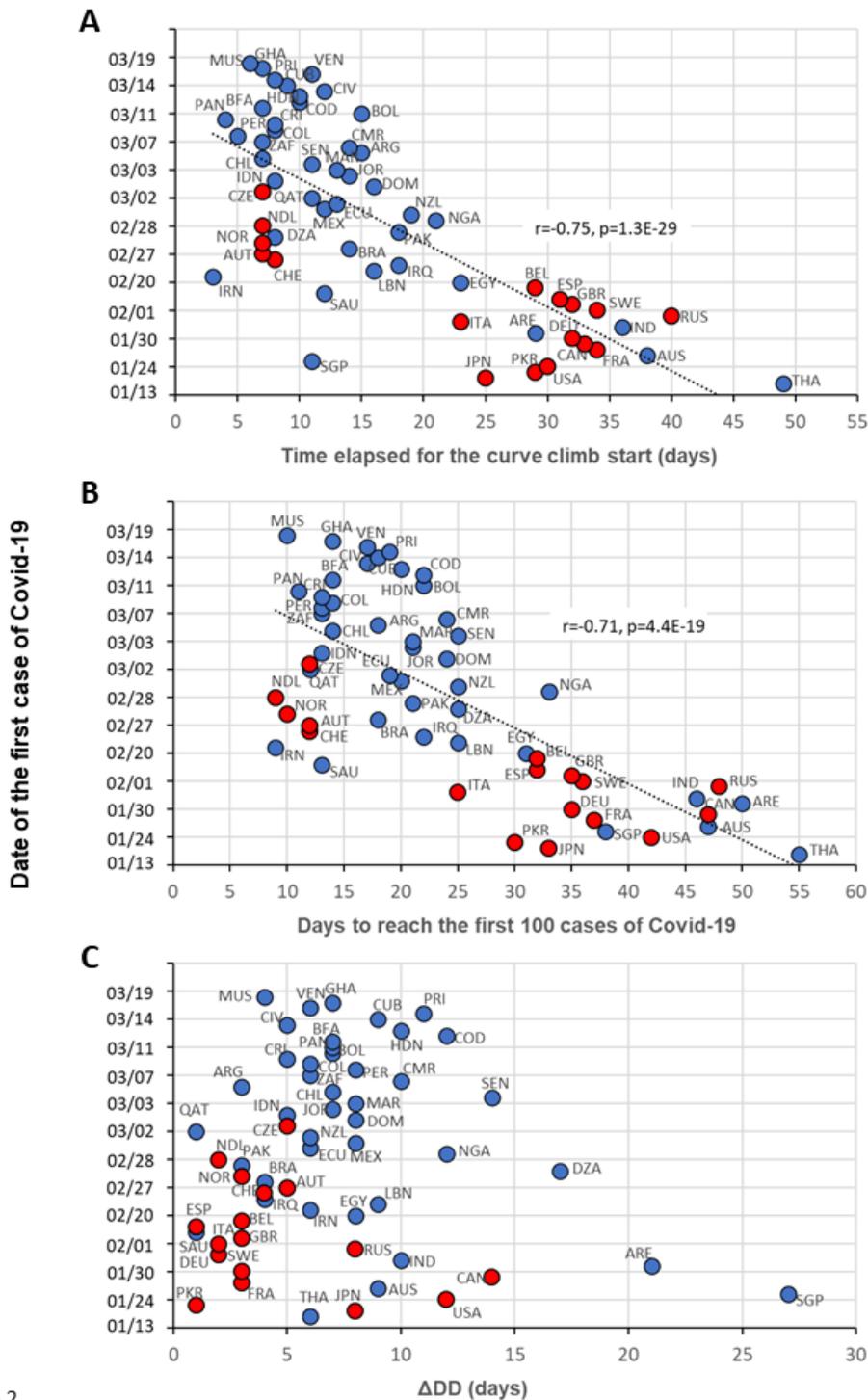


Figure 2

Fig 2. Correlation between the date of the first COVID-19 case diagnosed and DCS and D100. The blue circles indicate the countries located in tropical/subtropical zones and red circles the countries located in temperate zones. Correlation coefficients and p-values were calculated with the Pearson’s correlation test. Abbreviations of each country are described in Fig 1-legend.

3.2 Analysis of the growth curve in the first 30 days after the D100 day: calculation and analysis of the doubling time

The cumulative frequency, converted to \log_{10} , was plotted against the number of days of evolution of the epidemic in each country from D100 (Fig 3). In addition, doubling times of 1 to 4 days were calculated and included in the graph to locate the growth curve of each country between these intervals (black dotted lines in Fig 3). Changes in the trends of the accumulated frequencies over time are represented. For example, Fig 3 shows how the epidemic quickly grew in China, Korea and Iran during the first 10 days

closer to $T_d=1$ day (dotted line, Fig 3), then the curves start to lie down after the 10th and 20th days of the evolution of the epidemic. Something similar was seen in the European countries. To compare the growth curves in more detail, the doubling time was calculated for the first (T_{d10}), second (T_{d20}), and third (T_{d30}) 10 days, and for the entire 30-day period (T_{dT}) in each country. The T_{dT} was lower in temperate areas when compared with tropical/subtropical areas [4.2 ± 1.5 vs. 6.5 ± 3.1 ; $p=3.08E-04$]; this difference became greater when comparing the T_{d10} (2.8 ± 1.2 vs. 5.7 ± 3.4 ; $p=1.41E-05$) and the T_{d20} (4.6 ± 1.8 vs. 8.6 ± 4.2 ; $p=9.7E-05$) (all with t-test).

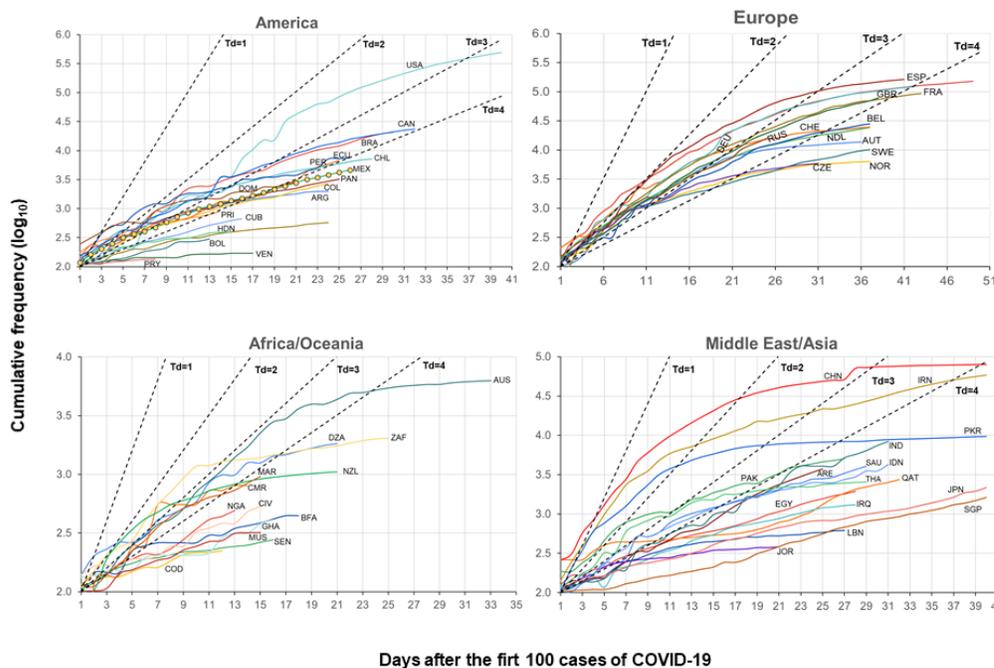


Figure 3

Fig 3. Trends of accumulative frequency in \log_{10} starting the day when 100 COVID-19 cases were reached. The accumulated frequency (CF) is plotted against the number of days when 100 cases of COVID-19 were diagnosed in each country. The dashed black lines indicated the doubling times (T_d) in days, that is the number needed to duplicate the number of infected people from one ($T_d=1$) to 4 ($T_d=4$) days. Abbreviations of each country are described in Figure 1-legend.

In addition, Fig 4 A-B depict the individual Td10, Td20 and Td30 from several temperate and tropical or subtropical countries. Countries that had a very rapid increase in the growth curve such as Italy, Spain, China, Korea, and Iran (Fig 3), have a Td10 very close to or below 2. It is important to point out that Td20 was higher than Td10 in all countries except the USA (Fig 4 A) and Singapore (Fig 4 B), which indicates that the rate for doubling the number of infected people was lower in most countries during days 11 to 20 of the curve. In the case of the USA, the infection duplication rate grew during that time-period. In fact, the Td20 falls below 2, and as seen in Fig 3, the slope of the curve increases from day 7 and the curve ascends almost as a straight line between Td= 2 and Td= 3. Something similar happened with Singapore (Fig 3), however, the curve remained below the Td= 4 line, which indicates that the Td > 4 days. In temperate countries, where the value of Td20 is two times that of Td10 [Fold Change

($FC_{Td20/Td10}$)], such as Korea ($FC_{Td20/Td10} = 4.5$), Sweden ($FC_{Td20/Td10} = 2.5$), Norway ($FC_{Td20/Td10} = 2.6$), or that have a close value as in the case of the Czech Republic ($FC_{Td20/Td10} = 1.9$; Fig 4), coincides with the early flattening of the curves, growing towards the area of Td= 4 or to a larger Td (Fig 3). It seems clear that the increase of the Td20, more than Td30, is essential for the early flattening of the curve. Examples of this include Austria ($FC_{Td20/Td10} = 1.6$) and Switzerland ($FC_{Td20/Td10} = 1.5$) that have a high Td30 but did not have a $FC_{Td20/Td10} > 2$, and the curves flattened later. This is even more difficult for Spain, France and Italy that started with a very low Td10 and $FC_{Td20/Td10} = 1.7$. In tropical and subtropical areas, in addition to Iran that had a Td10= 1.8, countries that had a Td10 ≤ 3 , such as Brazil, Chile, Ecuador and South Africa, had fast-growing curves during the first 10 days. However, most of them, except for Chile, had a $FC_{Td20/Td10} \geq 2$.

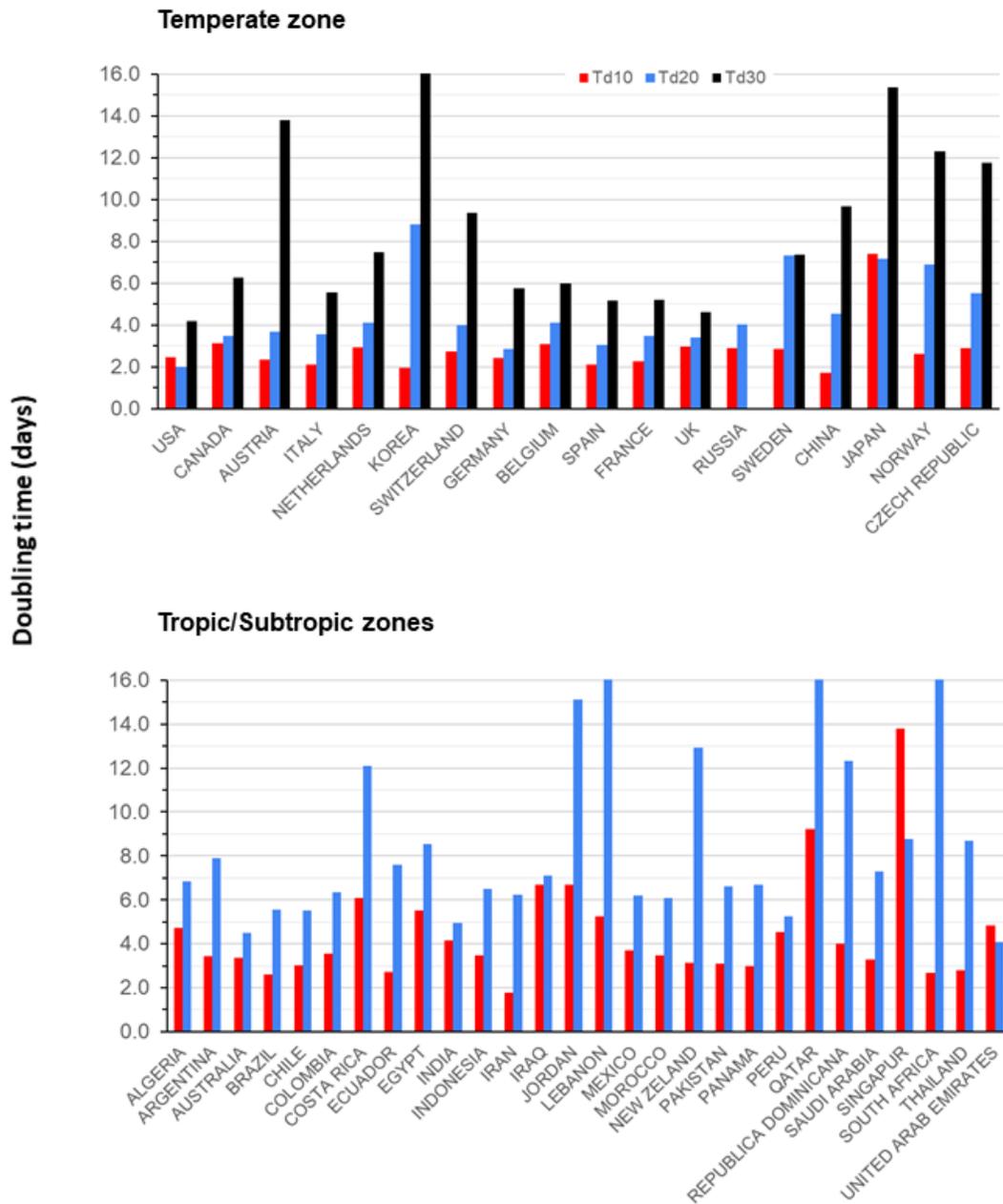


Figure 4

Fig 4. Doubling times (Td) in countries within temperate and tropical/subtropical zones. The plots show the Td from the first (blue), second (red) and third (black) 10 days of the cumulative frequency curves. Panel A shows countries from temperate zones and panel B countries from tropical/subtropical zones.

3.3 Identification of variables that could explain the doubling time using linear regression models

The relation of each variable (DCS, D100, Δ DD, DFC, average relative humidity, average temperature, climate zone) with Td10 and Td20 was investigated in ULR models. In addition, Td10 was included in the Td20 model (Table 1). Five of these variables were individually associated with the Td10, four showed a positive association (Δ DD, DFC, average temperature, and climate zone), while DCS had a negative association. The average relative humidity and the D100 day were not associated with the Td10. The average temperature ($\beta= 0.17$ 95%CI: 0.092 to 0.247, $p= 4.7E-05$) and the DFC ($\beta=0.079$ 95%CI: 0.037 to 0.121, $p= 4E-04$) were the

variables that most influenced the value of Td10; in fact, they also showed the highest values of r^2 (0.237 and 0.182, respectively). However, in the MLR analysis, the only variables that remained in the model were the DFC ($\beta= 0.077$ 95%CI: 0.037 to 0.116, $p= 2.6E-04$) and the Δ DD ($\beta=0.242$ 95%CI: 0.097 to 0.388, $p= 1.5E-03$). These two variables explain 29.7% of the variability of the Td10 ($r^2= 0.297$) (Table 1). It is not rare that temperature stands out of the model, as there is a very high correlation between temperature and ascending dates in the calendar from January to March. However, the exclusion of the temperature in the model and the permanence of the DFC may have an additional explanation.

Table 1. Linear regression models to evaluate explanatory variables of doubling time the first and second 10 days of the growth curve of COVID-19 in 69 countries.

Variables	Lineal regression models					
	Univariate			Multivariate		
	β (95% CI)	p-value	r^2	β (95% CI)	p-value	r^2
Td10^a (n=59)						
DCS ^c	-0.072 (-0.146-0.002)	5.8E-02	.044	0.007 (-0.083-0.097)	8.8E-01	0.297
D100 ^d	-0.019 (-0.089-0.05)	5.8E-01	0			
Δ DD ^e	0.243 (0.081-0.406)	4.1E-03	.121	0.242 (0.097-0.388)	1.5E-03	
DFC ^f	0.079 (0.037-0.121)	4.0E-04	.182	0.077 (0.037-0.116)	2.6E-04	
Average Relative humidity	-0.012 (-0.079-0.056)	7.3E-01	0			
Average temperature	0.17 (0.092-0.247)	4.7E-05	.237	0.064 (-0.044-0.173)	2.4E-01	
Climate zone	2.86 (1.191-4.529)	1.1E-03	.154	-0.663 (-3.359-2.034)	6.2E-01	

Td20^b (n=45)						
Td10 ^a	0.793 (0.291-1.295)	2.7E-03	.172	0.999 (0.468-1.53)	4.9E-04	0.426
DCS ^c	-0.087 (-0.184-0.011)	8.2E-02	.047	0.322 (0.075-0.57)	1.2E-02	
D100 ^d	-0.071 (-0.16-0.017)	1.1E-01	0	-0.332 (-0.549--0.115)	3.7E-03	
ΔDD ^e	-0.009 (-0.243-0.226)	9.4E-01	0			
DFC ^f	0.103 (0.024-0.181)	1.1E-02	.120	0.077 (-0.028-0.182)	1.5E-01	
Average Relative humidity	-0.027 (-0.123-0.069)	5.7E-01	0			
Average temperature	0.151 (0.018-0.284)	2.7E-02	.088	-0.161 (-0.345-0.023)	8.4E-02	
Climate zone	4.022 (1.877-6.167)	4.8E-04	.232	5.704 (2.396-9.012)	1.2E-03	

^aTd10 and ^bTd20 = doubling time the first 10 and 20 days of the curve, respectively, ^cDCS= day start the curve climb, ^dD100= day when the 100th was diagnosed, ^eΔDD=D100-DCS, ^fDFC= date of the first COVID-19 case diagnosed.

In the ULR models for Td20, all variables are significant, except relative humidity and ΔDD. But only four variables remained in the MLR model, including Td10 ($\beta=0.999$ 95%CI: 0.468 to 1.53, $p=4.91E-04$) and the climate zone ($\beta=5.7$ 95%CI: 2.396 to 9.012, $p=1.22E-03$), which appeared to be the most important factors influencing the value of Td20 (Table 1). For Td20, the temperature no longer seems to be an important factor, while important components of the initial curve behavior remained, such as DCS and D100 (Table 1). All these factors explain 42.6% ($r^2=0.426$) the variability of Td20. When the model is stratified by zone, the Td10 variable remains in both area models as the most important variable (data not shown).

3.4 Prediction of total cases of COVID-19 at days 20 and 30 of the growth curve: MLR models

We explored whether the Td10, Td20, and TdT, with the other variables, could predict the COVID-19 CF that would be reached 20 and 30 days after the day D100 in the growth curve using ULR and MLR models. A clear relationship of Td10 with CF at day 20 is observed in the ULR model ($r^2=0.189$), as discussed above in relation to Figures 3 and 4, however, the relationship of TdT to the CF is much higher ($r^2=0.268$), therefore this variable of the doubling time was the one introduced in the MLR model (Table 2). Also, ΔDD, DFC, the average relative humidity, the average temperature, and climate zone passed the cut-off value in the ULR models (Table 2). However, in the MLR analysis only TdT, DFC and the

average temperature remained in the model. These three variables had a negative relationship with the CF, that is, as their value increases the CF decreases,

and vice versa, when their value decreases, the CF increases. The explanatory variables explain 44.7% of the variability of CF for day 20 of the curve.

Table 2. Prediction of the accumulative frequency of COVID-19 at day 20 and 30 of the curve with the doubling time and other variables using linear regression models.

Variables	Univariate regression models			Multivariate regression models		
	β (95% CI)	P-value	r^2	β (95% CI)	p-value	r^2
Prediction of total cases of COVID-19 at day 20 of the curve*						
Td10 ^a	-1590.9 (-2729.3 - -452.6)	7.5E-03	.182			0.449
Td20 ^b	-978.6 (-1646.5 - -310.7)	5.3E-03	.197			
TdT ^c	-2381.4 (-3713 - -1049.8)	8.8E-04	.268	-1912.9 (-3161.6--664.2)	3.7E-03	
DCS ^d	27.9 (-199.9-255.8)	8.1E-01	0			
D100 ^e	-30.8 (-233.1-171.5)	7.6E-01	0			
Δ DD ^f	-374.4 (-898.5-149.7)	1.6E-01	0	-219.6 (-700-260.8)	3.6E-01	
DFC ^g	-252.3 (-449.4--55.3)	1.4E-02	.158	-164 (-342.4-14.3)	7.0E-02	
Average Relative humidity	93.5 (-160.1-347.1)	4.6E-01	0			
Average temperature	-477.9 (-764.2 - -191.7)	1.7E-03	.242	-259.9 (-545-25.2)	7.3E-02	
Climate zone	-8612.2 (-13559 - -3665.5)	1.2E-03	.257	-1395.8 (-9320-6528.3)	7.2E-01	
Prediction of total cases of COVID-19 at day 30 of the curve**						
Td10 ^a	-5003.8 (-13121-3113.5)	2.1E-01	0			0.638
Td20 ^b	-14255.8 (-23467.2--5044.4)	4.7E-03	.365	-17576.1 (-27101.2--8051)	1.7E-03	
Td30	-1008.6 (-2691.2-674.1)	2.2E-01	0			
TdT ^c	-15583.5 (-26521.1--4645.9)	8.1E-03	.323			
DCS ^d	682.9 (-1207-2572.7)	4.5E-01	0			
D100 ^e	541.6 (-1225.4-2308.4)	5.3E-01	0			

ΔDDf	-239.6 (-4176.2-3697)	9.0E-01	0		
DFC ^g	-1437.3 (-4068.5-1193.9)	2.6E-01	0	-2537.5 (-4556.9--518)	1.8E-02
Average Relative humidity	-634.2 (-3049.7-1781.2)	5.9E-01	0		
Average temperature	-1844.2 (-4980.7-1292.3)	2.3E-01	0	-3299.4 (-8736.1-2137.2)	2.1E-01

^aTd10 and ^bTd20 = doubling time the first 10 and 20 days of the curve, respectively, ^cTdT= entire 30-day period, ^dDCS= day start the curve climb, ^eD100=day when the 100th case was diagnosed, ^f $\Delta DD=D100-DCS$, ^gDFC= date of the first COVID-19 case diagnosed.

*The parameters of the multivariate model: $b=23,067$, $r=0.67$, Durbin-Watson=1.65, ANOVA: $p<0.0001$

** The parameters of the multivariate model: $b=161,049$, $r=0.8$, Durbin-Watson=1.9, ANOVA: $p=0.002$

For the prediction of the CF at the 30th day of growth curve the model has a better performance (Table 2), but was explored only for temperate areas, since in tropical/subtropical areas very few countries of those analyzed in the current study had reached 30 days of evolution in the curve at the time of writing this article. The Td20 performed better than the TdT, so it was this variable that was introduced in the MLR model. In fact, Td20 alone explains 36.5% of the CF's variability at the 30th day of the growth curve. This indicates that the speed of the growth curve during the previous 10 days (days 11 to 20), is essential for controlling the CF during the following 10-day period of the outbreak evolution, that is, from 21 to 30 days after the day D100. In the MLR analysis, only Td20 and the DFC remained in the model, which explained 63.8% ($r^2=0.638$) of the variability of the CF towards the 30th day of the curve. Similarly, both factors have a negative influence on the CF. The rest of the variability (36%) must be related to other factors, such as the containment

measures used by each country to control the epidemic.

4 Discussion

4.1 Key results

In this study we disclosed that the behavior of the growth curve (DCS, D100 and ΔDD) during the period between the diagnosis of the initial and the first 100 cases, were related to temperature, the DFC and the doubling time (Td). In addition, these values were substantially different between countries located in temperate and tropical/subtropical areas, especially the Td during the first ten days (Td10), after D100, was on average 3 days longer in tropical/subtropical than in temperate countries. We also identified that the factors involved in Td the first ten (Td10) days (DFC and ΔDD) are different than those involved the second ten (Td20) days (mainly Td10 and climate zone) of the growth curve. The fastest growth curves during the first 10 days, after D100 day, were associated with $Td10 < 2$ and $Td10 < 3$ in temperate and tropical/subtropical

countries, respectively. And the fold change $Td_{20}/Td_{10} > 2$ was associated with earlier flattening of the growth curve. In the MLR predictive models, the Td_{10} , the DFC and the ambient temperature, were negatively related to the CF for the 20th day of the growth curve, while only Td_{20} and DFC were negatively related to the CF reached on day 30 of the growth curve.

4.2 Strengths and limitations

This study has several strengths including the large sample of incident cases of COVID-19 collected from the WHO database, to analyze the behavior of the epidemic curves for more than 90 days. Another strength is the fact that epidemic curves and environmental variables of 67 countries distributed in different climatic zones and six continents, were studied. These two issues allowed the comparison of several variables in various segments of the epidemic growth curve and to establish the differences between the two climatic zones.

Although most tropical countries have weaker sanitary and medical systems and their population often cannot stay at home without working, its epidemic curves grew on average slower than in temperate countries. The above issue strongly supports the results of this paper. However, the study has some limitations, such as the inability to incorporate in the analysis the number of tests carried out or the containment measures implemented in each country, during the different periods of the epidemic curves, thus, the findings of this work should be carefully interpreted. Although it is possible that these issues

most likely do not affect or have very little impact on the analysis during the first 10 days of the epidemic curve, it is likely that they substantially affect the analysis in the second ten and third ten days of the epidemic curves. Despite this, the fold change ratio between Td_{20} and Td_{10} was no different between the two compared climate zones.

4.3 Comparison with previous studies and possible explanations of our findings

Notably, the findings from the analysis of the first stage of the growth curve, from the day on which the first case was diagnosed until the first 100 cases of COVID-19 were diagnosed, as well as the DFC and probably the value of Td_{10} , would have a modest influence from the number of tests or containment measures that were carried out in the different countries, since most of them were similar. The DFC, the average DCS and ΔDD were significantly different among temperate and tropical/subtropical countries. Although the DFC is directly related to temperature, this variable remains independent and heavier in the MLR models associated to Td_{10} and CF. The above suggest that the DFC is not only related to ambient temperature but perhaps also related to the way the infection was disseminated. The fact that the epidemic reached the tropical/subtropical countries at a later time could have been an explanation as a delay in the shockwave from China, first to other temperate countries in the northern hemisphere and from there to tropical/subtropical countries.⁴ For example, the infection from European countries and the USA, took an

estimate, between the number of days for the rise of the curve and an important establishment of the infection, of about 1 month to be exported to Latin America and other countries of tropical/subtropical areas. The average time delay, relative to December 29th, 2019, when the infection spread from China, was 58.5 ± 15.2 days. The question is why in these countries the cumulated positive cases curve rose rapidly but with a smaller slope than in countries of temperate zones, and why in the later ones, once the rise of the curve began, the slope was steeper. In discussing the last issue first, an important possibility could be that in European and other temperate countries, asymptomatic individuals were the first to accumulate and these subjects spread the infection.¹¹ This seems very feasible as up to 80% of infected individuals are asymptomatic.¹¹⁻¹⁴ In addition, common use of mass transportation and small housing spaces in European cities or those such as New York with a high population density, could have facilitated infection dissemination. Further, the spread could have then affected the older population which in temperate countries make up a much higher percentage of the population pyramid than in countries of tropical/subtropical areas. For example, in countries like Spain and Italy the 65 years and older age-group, constitutes 19% and 23% of their populations, respectively. Conversely, the same age-group represents only 7% of the Mexican and Ecuadorean populations.¹⁵ The older population may have greater susceptibility to infection in cold climates, mainly because a decreased immune response associated with aging, and diminished airway function.¹⁶⁻¹⁹ This

would explain the late rise with a steeper slope in the epidemic curve in those countries. In contrast, in tropical/subtropical countries the early ascent, with a slower slope of the curve, could be related to the initial presence of imported cases, mainly from European countries and the US, and the dissemination to their contacts. These imported cases, mainly of middle or higher socioeconomic groups who have the possibility of traveling, could have later spread the infection on a smaller scale, mainly because they do not use mass public transportation. On the other hand, in the population pyramid in those countries there is a large proportion of young people, which could have contributed to a lower frequency rate, due to a silent outbreak or an infection with less severe clinical manifestations. For example, the median age of confirmed cases with COVID-19 is much lower in Mexico, according to the daily data reported by the Secretary of Health of this country (45 years old),²⁰ than that in China (56 years old)²¹ and Italy (64 years old).²²

In addition, the expectation in these tropical/subtropical countries, is that there is a protective effect of ambient temperature, at least on the airway function.⁶ However, temperature variation as a function of climatic status also has a profound influence in virtually all stages of the host virus interaction,²³⁻²⁵ therefore, the relationship of ambient temperature and doubling time depicted here becomes highly relevant. There is a direct effect of temperature on the enzyme-mediated reactions resulting in cell homeostasis and immune responses.²³⁻²⁶ More explicitly, the

frequency of upper respiratory tract viral infections relates inversely to every degree Celsius that the temperature drops.^{23, 25, 26} Although available information strongly suggests a climatic influence on COVID-19 infection and spreading,^{10, 24} the novelty of this illness, warrants further assessment of this issue. However, systematic studies are required to rule out possible ethnic and/or genetic influences upon the spreading of and susceptibility to the disease.²⁷

In the case of Td20, there is a possibility that this calculation is biased or at least severely related to the number of tests and containment measures that each country has undertaken and may have contributed, at least in part, to the differences between temperate and tropical/subtropical countries. The fact that the Td20s difference (2.6 times Td10) was much greater than the Td10s differences between the zones (should be only 1.17 times Td10 as expected by the line regression data), suggests that the differences in Td20 between countries in the different climate areas, may be related to factors such as the number of tests performed and the implemented measures to mitigate the epidemic. However, the relationship or change folds between Td20 and Td10 must remain valid, on the assumption, at least in principle that, the policies on the number of tests and implemented containment measures, did not change too much in each country. In addition, no statistically significant difference was found between $FC_{Td20/Td10}$ in countries within temperate or tropical/subtropical regions. This implies that the value of $FC_{Td20/Td10} \geq 2$ would be a good parameter for both regions to

positively assess the evolution of the epidemic during the 21-30 day-periods of the curves.

4.4 Implications

Having a different pattern of infection spread between temperate and tropical/subtropical countries could only slow the speed in which the virus is being transmitted, and although this is good news for health services utilization, it does not necessarily imply that the proportion of the population infected will be smaller in the tropical/subtropical countries.⁸ The latter will depend on the timing, related to the CF growth curve, in which the containment measures are established, and their magnitude to reduce the spread of the epidemic in each country.^{8, 28-30} On the other hand, the value of Td10 allows the evaluation of the growth of infections during the second 10 days of the epidemic curve, while the ratio Td20/Td10 allows the study to evaluate the growth of the curve during the third 10 days and with it, may help in evaluating how effective are the implemented containment measures.

5 Conclusions

This study showed that the behavior of the growth curve (DCS, D100 and ΔDD) in the first stage of the epidemic is related to the date that the first COVID-19 case was diagnosed, the ambient temperature and doubling time of infection cases, which are different between countries located in temperate and tropical/subtropical areas. The Td10 is on average 3 days longer in

tropical/subtropical countries than in those located in temperate zones and can predict the growth of the curve for the following 10 days of the evolution of the epidemic after D100. These differences appear to be related to ambient temperature and the date of the first case that was identified and how the infection spread in both climatic zones.

In addition, the Td10 and Td20 values helped predict the cumulative frequency of COVID-19 at the 20th and 30th days of the epidemic after D100 day, while the Td20/Td10 ratio helps to evaluate the growth curve behavior during the next third 10 days of the epidemic evolution.

6 References

1. Guan WJ, Ni ZY, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *The New England journal of medicine*. Feb 28 2020;doi:10.1056/NEJMoa2002032
2. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *The New England journal of medicine*. Feb 20 2020;382(8):727-733. doi:10.1056/NEJMoa2001017
3. World Health Organization WHO. Coronavirus disease 2019 (COVID-19) Situation report 76. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200405-sitrep-76-covid-19.pdf?sfvrsn=6ecf0977_2
4. Gutierrez P. Coronavirus world map which countries have the most cases and deaths. Coronavirus updates. April 5-2020, 11.25 UTC. 2020. [theguardian.com/world/2020/apr/05/coronavirus-world-map-which-countries-have-the-most-cases-and-deaths](https://www.theguardian.com/world/2020/apr/05/coronavirus-world-map-which-countries-have-the-most-cases-and-deaths)
5. Worldometer. COVID-19 Coronavirus Pandemic. Accessed April 11th-2020, <https://www.worldometers.info/coronavirus/>
6. Kudo E, Song E, Yockey LJ, et al. Low ambient humidity impairs barrier function and innate resistance against influenza infection. *Proceedings of the National Academy of Sciences of the United States of America*. May 28 2019;116(22):10905-10910. doi:10.1073/pnas.1902840116
7. Sun Z, Thilakavathy K, Kumar SS, He G, Liu SV. Potential Factors Influencing Repeated SARS Outbreaks in China. *International journal of environmental research and public health*. Mar 3 2020;17(5)doi:10.3390/ijerph17051633
8. de Angel Sola DE, Wang L, Vazquez M, Mendez Lazaro PA. Weathering the pandemic: How the Caribbean Basin can use viral and environmental patterns to predict, prepare and respond to COVID-19. *Journal of medical virology*. Apr 10 2020;doi:10.1002/jmv.25864
9. Timeanddate.com. World Temperatures — Weather Around The World. <https://www.timeanddate.com/information/copyright.html>. 2020.
10. Oliveiros B, Caramelo L, Ferreira NC, Caramelo F. Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases. *medRxiv*. 2020; preprint doi: <https://doi.org/10.1101/2020.03.05.20031872>.
11. Day M. Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. *Bmj*. Mar 23 2020;368:m1165. doi:10.1136/bmj.m1165
12. World Health Organization WHO. Q&A: Similarities and differences – COVID-19 and influenza. Accessed April 5th, 2020, <https://www.who.int/news-room/q-a-detail/q-a-similarities-and-differences-covid-19-and-influenza>

13. Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (COVID-19) (2020).
14. Serra M. Coronavirus, Castiglione d'Adda è un caso di studio: "Il 70% dei donatori di sangue è positivo". *La Stampa*
<https://www.lastampait/topnews/primo-piano/2020/04/02/news/coronavirus-castiglione-d-adda-e-un-caso-di-studio-il-70-dei-donatori-di-sangue-e-positivo-138666481>. April 2-2020.
<https://www.lastampa.it/topnews/primo-piano/2020/04/02/news/coronavirus-castiglione-d-adda-e-un-caso-di-studio-il-70-dei-donatori-di-sangue-e-positivo-1.38666481>
15. OECD. Elderly population (indicator). <https://data.oecd.org/pop/elderly-population.htm>. Accessed 4-22-2020, 2020.
16. Haq K, McElhaney JE. Ageing and respiratory infections: the airway of ageing. *Immunology letters*. Nov 2014;162(1 Pt B):323-8. doi:10.1016/j.imlet.2014.06.009
17. Gruver AL, Hudson LL, Sempowski GD. Immunosenescence of ageing. *The Journal of pathology*. Jan 2007;211(2):144-56. doi:10.1002/path.2104
18. Meyer KC. Lung infections and aging. *Ageing research reviews*. Jan 2004;3(1):55-67. doi:10.1016/j.arr.2003.07.002
19. Solana R, Tarazona R, Gayoso I, Lesur O, Dupuis G, Fulop T. Innate immunosenescence: effect of aging on cells and receptors of the innate immune system in humans. *Seminars in immunology*. Oct 2012;24(5):331-41. doi:10.1016/j.smim.2012.04.008
20. Reporte diario COVID-19. Secretaría de Salud. Gobierno de México. Abril 5 de 2020. <https://coronavirus.gob.mx/>
21. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. Mar 28 2020;395(10229):1054-1062. doi:10.1016/S0140-6736(20)30566-3
22. Livingston E, Bucher K. Coronavirus Disease 2019 (COVID-19) in Italy. *Jama*. Mar 17 2020;doi:10.1001/jama.2020.4344
23. Martens WJ. Climate change, thermal stress and mortality changes. *Social science & medicine*. Feb 1998;46(3):331-44. doi:10.1016/s0277-9536(97)00162-7
24. Sajadi MM, Habibzadeh P, Vintzileos A, Shokouhi S, Miralles-Wilhelm F, Amoroso A. Temperature, humidity, and latitude analysis to predict potential spread and seasonality for COVID-19 (March 5, 2020). *SSRN*. 2020;
25. Eccles R. An explanation for the seasonality of acute upper respiratory tract viral infections. *Acta Otolaryngol*. 2002;122(2):183-91.
26. Alderson MRHT. Season and mortality. *Health Trends*. 1985;17:87-96.
27. Grasselli G, Pesenti A, Cecconi M. Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy: Early Experience and Forecast During an Emergency Response. *Jama*. Mar 13 2020;doi:10.1001/jama.2020.4031
28. World Health Organization WHO. Statement on the second meeting of the

- International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). 2020:
29. Li C, Romagnani P, von Brunn A, Anders HJ. SARS-CoV-2 and Europe: timing of containment measures for outbreak control. *Infection*. Apr 9 2020;doi:10.1007/s15010-020-01420-9
30. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *The New England journal of medicine*. Mar 26 2020;382(13):1199-1207. doi:10.1056/NEJMoa2001316