Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases

B Oliveiros[1, 2, 3]; L Caramelo[4]; N C Ferreira[1, 3]; F Caramelo[1, 2, 3]

[1]Laboratory of Biostatistics and Medical Informatics, Faculty of Medicine, University of Coimbra, Portugal

[2]iCBR - Coimbra Institute for Clinical and Biomedical Research, Faculty of Medicine, University of Coimbra, Portugal

[3]CIBIT - Coimbra Institute for Biomedical Imaging and Translational Research, University of Coimbra, Portugal

[4]CITAB – Centre for the Research and Technology of Agro-Environmental and Biological Sciences, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal

Abstract

COVID-19 is having a great impact on public health, mortality and economy worldwide, in spite of the efforts to prevent its epidemy. The SARS-CoV-2 genome is different from that of MERS-CoV and SARS-CoV, although also expected to spread differently according to meteorological conditions. Our main goal is to investigate the role of some meteorological variables on the expansion of this outbreak.

In this study, an exponential model relating the number of accumulated confirmed cases and time was considered. The rate of COVID-19 spread, using as criterion the doubling time of the number of confirmed cases, was used as dependent variable in a linear model that took four independent meteorological variables: temperature, humidity, precipitation and wind speed. Only China cases were considered, to control both cultural aspects and containment policies. Confirmed cases and the 4 meteorological variables were gathered between January 23 and March 1 (39 days) for the 31 provinces of Mainland China. Several periods of time were sampled for each province, obtaining more than one value for the rate of disease progression. Two different periods of time were tested, of 12 and 15 days, along with 3 and 5 different starting points in time, randomly chosen. The median value for each meteorological variable was computed, using the same time period; models with $R^2_{adj} > 0.75$ were selected. The rate of progression and doubling time were computed and used to fit a linear regression model. Models were evaluated using $\alpha = 0.05$.

Results indicate that the doubling time correlates positively with temperature and inversely with humidity, suggesting that a decrease in the rate of progression of COVID-19 with the arrival of spring and summer in the north hemisphere. A 20ºC increase is expected to delay the doubling time in 1.8 days. Those variables explain 18% of the variation in disease doubling time; the remaining 82% may be related to containment measures, general health policies, population density, transportation or cultural aspects.

Keywords

COVID-19, doubling time, temperature, humidity, precipitation, wind speed, mortality

Introduction

The outbreak of pneumonia cases in Wuhan, China during last December led to great efforts to prevent a global epidemic. The alert from China CDC was rapidly transmitted to the World Health Organization,[1-3] excluding possible causes such as influenza, adenovirus, severe acute respiratory syndrome (SARS-CoV) and Middle-East respiratory syndrome (MERS-CoV).[1,3-4] The novel coronavirus, named SARS-CoV-2, and its genomic characterization was performed a few days after, permitting to devise a robust test method.[1-5] Although the genomic characterization revealed some relations both to SARS-CoV and MERS-CoV,[4-6] the new virus was found to be much more aggressive than those other coronavirus or the seasonal one[4-6]. When human-to-human transmission was proved, on the 20th of January, the onset of the disease (COVID-19) has changed.[2,7-10] According to the China CDC[11], the case fatality rate (CFR) was 0.2% at the end of January 2020 and 14.4% of the confirmed cases were considered severe or even critical. In the last week of February, 79441 cases were confirmed worldwide (97% in Mainland China) and the number of deaths was 2620 (95% in Mainland China).[12,13] The epidemiological curves of COVID-19 in China showed the progression of illness in the outbreak over time from December 8, 2019 up to February 11, 2020,[11] when there were a total of 72314 confirmed cases as the geo-temporal spread of COVID-19.[11] At that time, the majority of confirmed cases occurred in the northern hemisphere and until the last week of February 2020 no confirmed cases had been reported in South America or Africa, except for one

case in Egipt.[12] In fact, the first confirmed case in Brazil was reported on February 26, while Algeria and Nigeria reported the first cases respectively on the 25th and 27th of February. The discussion about the COVID-19 epidemic spread in the northern hemisphere, while low temperatures and high humidity are present, and the unknown, although expectable positive impact of spring and summer in sustaining the epidemy, as its spread into the southern hemisphere was not as epidemic, has aroused our question: how do meteorological variables, such as temperature and humidity, modulate COVID-19 duplication time?

Even though there is not yet strong evidence that meteorological conditions may have a role on COVID-19 outbreak or on human transmission, some studies have reported their role in guinea pigs influenza transmission[14] and enveloped virus survival[15] in droplets. Some evidence of a faster spreading of diseases in high humidity levels has been reported,[16] namely for the Legionella disease, although this infection is not caused by a virus. Few papers have been written since mid-February on this topic[17-20] even though the relationship is not perfectly established and more research is required.

We intend to add value to this discussion by evaluating the meteorological impact on COVID-19 duplication time.

Material and Methods

The statistical model developed was implemented in two steps: firstly, an exponential model relating the accumulated number of confirmed cases and time was considered. Secondly, the rate of spread was used as dependent variable in a linear model that took as independent variables temperature, humidity, precipitation and wind speed.

Only cases belonging to China were considered as an attempt to control both cultural aspects and policies adopted to contain the virus. Therefore, data from the 31 provinces of Mainland China were gathered from the 23rd January up to the 1st of March, completing 39 days. These data were completed with meteorological variables, comprising temperature, humidity, precipitation and wind speed, collected for the same period, using the Meteostat Application Programming Interface (API).[21] We searched for meteorological stations containing hourly measurements of these variables for the whole 39-day time period that were closest

to the latitude and longitude coordinates that were available in the files that contained the confirmed cases time-series.[13] These geo-localization coordinates correspond to the geometrical centerpoint of each Chinese province.[22] When meteorological data from a station was not possible to obtain around that position, another search point was chosen randomly from the set of closest nodes of an XY grid of nodes separated by 0.5 degrees in latitude and longitude and cantered in the originally desired geolocation.

To compute the rate of spread a simple exponential model was assumed, described by:

$$N = N_0 e^{\lambda t}, \tag{eq1}$$

where N_0 is the number of infected at instant zero, represents the rate of infection or the rate of spread and t is the time. A more natural way of interpreting is by transforming it into the doubling time, T_d, given by:

$$T_d = \frac{\ln(2)}{\lambda}. \tag{eq2}$$

The doubling time is the time needed to duplicate the number of infected subjects. Since the rate of progression changes over time and the exponential model does not hold any longer, we considered mainly the initial days of the time series, selecting several periods of time composed each one of a predetermined number of consecutive days, but with different starting points. The starting point was assumed by randomly choosing the first day of the period. For each province several periods were sampled, allowing to obtain more than one value of the rate of progression (Figure 1).

Two different periods of time of 12 and 15 days were tested, along with 3 and 5 different starting points.

For each rate of progression, a corresponding value of each meteorological variable was computed, taking into account the same period of time. We opted to use the median of the meteorological variables (temperature, humidity, precipitation and wind speed) as it is more robust than the mean and tends to better represent the central tendency of the variable. Only models attaining more than 0.75 for the adjusted R square value were selected.
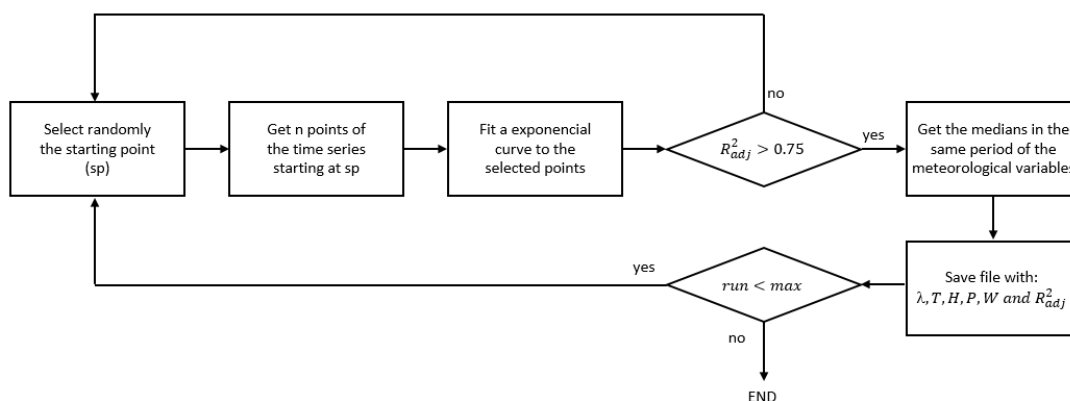
Figure 1 - Flowchart of the routine in R language used to compute the statistical models of COVID-19 spreading of new cases

The rate of progression was transformed into the doubling time, T_d, and recorded along with the median of temperature, humidity, precipitation and wind speed. These values were then used to fit a linear regression model aiming to assess how the meteorology is related to the doubling time.

The exponential models for the rate of progression were compute resorting to the R programming language[23], whereas the linear regression models were computed using IBM SPSS v25 with an adopted statistical significance level of 0.05.

Results

The number of confirmed infected cases of COVID-19 were initially analysed by plotting them against time. Figure 2 depicts the curves obtained for the provinces that at the first time point (23[td] of January) had only one or two cases.
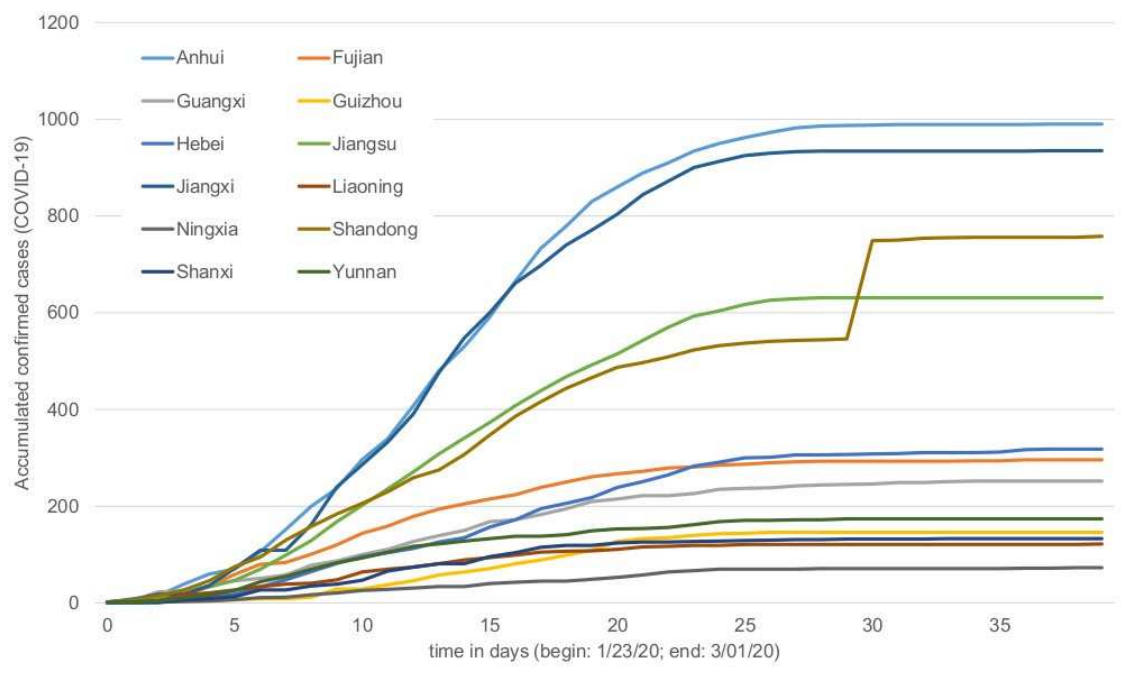
Figure 2 - Accumulated confirmed cases of COVID-19 in function of time. Only the provinces that had one or two cases at the beginning of the series are shown.

The analysis of Figure 2 shows that the number of accumulated cases is different between provinces and, in general, its rate decreases over time up to the point where it becomes null. Since the objective of the study was to analyse the rate of spread, we decided to use periods of 12 and 15 days to determine it. The initial 25 days were thought to be the most informative regarding the rate of spread. The fits of temporal evolution of confirmed cases of COVID-19 of the remaining provinces are not shown, but a similar profile can be obtained, leading to the same conclusion. Table 1 shows the statistics - mean, standard deviation, minimum and maximum - of the doubling time, T_d, for the two ranges of days (12 and 15 days) and the number of sampled periods used for each province.

Table 1- Statistics of the doubling time for two different time periods and two repetitions

	Period of 12 days	Period of 15 days
Number of sampled periods = 3	3.78 (1.34) 1.64/7.05	4.53 (1.68) 2.04/9.48
Number of sampled periods = 5	3.82 (1.27) 1.67/7.05	4.42 (1.53) 1.99/7.76

mean (sd); min/max

A two way ANOVA shows statistical differences $(F_{(1, 450)}=23.573; p<0.001)$ between the size, 12 or 15 days of the period employed to compute the doubling time, but no statistical differences $(F_{(1, 450)}=0.047; p=0.828)$ between the number of samples taken from each province. This result is in agreement with the hypothesis that the doubling time changes with time.

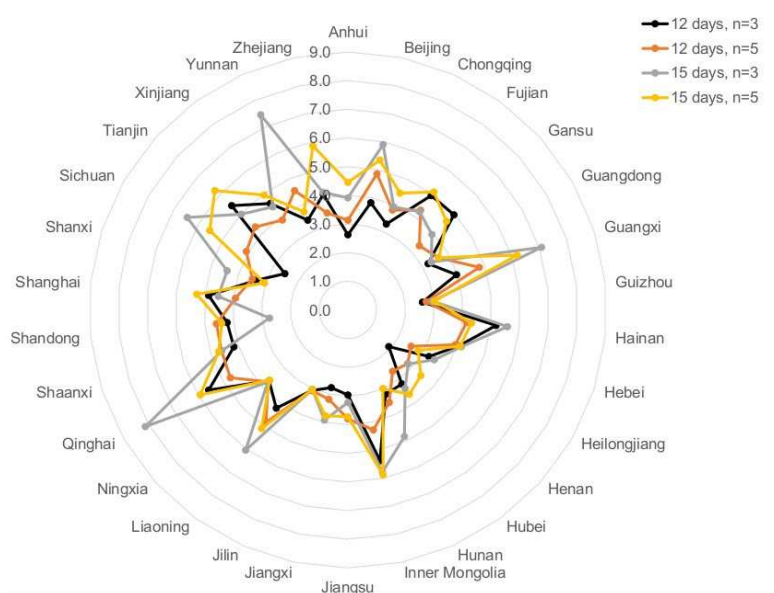The average of the doubling time duplication was determined for each province (Figure 3).

Figure 3 - Doubling time for each province considering the different forms of calculation.

Taking the values of doubling time T_d and the meteorological variables, a linear regression was performed. Table 2 shows, for the 4 conditions studied, the results achieved from the linear regressions, that assumed temperature and humidity as independent variables.

Table 2 - Results of the linear regressions between doubling time, temperature and humidity.

Condition	Model significance	R^2_{adj}	Coeficents of regression (B)
12 days n=3	$F(2, 81)=10.109$; $p<0.001$	0.180	T: B=+0.048 CI95%[0.005; 0.092]; p=0.028 H: B=-0.066 CI95%[-0.095; -0.036]; p<0.001
12 days n=5	$F(2, 140)=8.702$; $p<0.001$	0.098	T: B=+0.049 CI95%[0.018; 0.081]; p=0.002 H: B=-0.050 CI95%[-0.074; -0.026]; p<0.001
15 days n=3	$F(2, 83)=6.792$; $p=0.002$	0.120	T: B=+0.090 CI95%[0.033; 0.148]; p=0.002 H: B=-0.072 CI95%[-0.112; -0.031]; p=0.001
15 days n=5	$F(2, 138)=6.246$; $p=0.003$	0.070	T: B=+0.061 CI95%[0.020; 0.103]; p=0.004 H: B=-0.051 CI95%[-0.080; -0.021]; p=0.001

T= temperature; H=Humidity

Precipitation and wind speed did not reach statistical significance in any model (data not shown), thus Table 2 only refers results for temperature and humidity.

The results obtained for all models are statistically meaningful and despite their variation, it is possible to perceive that the coefficients of regression (B) are not statistically different as their confidence intervals overlap. On the other hand, the amount of variation explained, given by the adjusted R square value, differs between models. The model based on 12 days and 3 sampled periods is able to explain 18% of the variance in the doubling time, which means that temperature and humidity alone may describe 18% of the variation of confirmed COVID-19 infections. More importantly, the signal and value of the coefficients of regressors are of utterly importance to understand how the spread of COVID-19 is expected to be affected by temperature and humidity. According to all models, temperature increases the doubling time, which means that it delays the spread of COVID-19. Humidity, on the contrary, benefits it. The models differ, however, on the amount of contribution: for example, in the best scenario (model: 15 days, n=3) the doubling time is increased by 0.090 days for each Celsius degree increase, and is increased by 0.072 days for each unit decrease of the humidity value.

Discussion and conclusion

In this work, the way temperature and humidity affect the doubling time of COVID-19 spreading was determined. Results suggest that temperature correlates

positively with the doubling time and negatively with humidity. This means that, with spring and summer, the rate of progression of COVID-19 is expected to be slower. Still, these two variables contribute at maximum to 18% of the variation, being the remaining 82% related to other factors such as containment measures, general health policies, population density, transportation, cultural aspects, etc. Besides, the direct impact is also small: for example, if temperature raises 20ºC, it is expected that in the best-case scenario the doubling time increases on average 1.8 days.

These results are in agreement with other studies that suggested that the aerosol spread of the influenza virus is both dependent upon relative humidity and temperature, although performed in animal models[14], and that the virus survival in droplets is higher at high humidity levels with a significant decrease on its infectivity rate at mid-levels of humidity[15]. Additionally, other authors suggest that some diseases spread faster in high humidity levels[16], reporting an odds ratio for a community-acquired pneumonia case, diagnosed with leggionaire's disease[a], 3.1 times higher in high humidity level (above 80%) than when submitted to humidity levels below 50%, at temperatures of 16-27 ºC (60º-80ºF).

Wei Lo *et al*[17] recently reported a statistically significant association between absolute humidity and mean temperature on COVID-19 spread among China provinces. Furthermore, they have concluded that transmission and exponential growth of confirmed cases are occurring in China provinces in humidity conditions ranging from cold and dry (Jilin or Heilongjiang) to tropical (Guangxi or Singapore), suggesting that changes in weather, as expected by the arrival of spring and summer, will not necessarily lead to declines in outbreak unless extensive public health interventions are implemented, and that further studies on the effects of meteorology on COVID-19 transmission are needed.

On the other hand, Jin Bu *et al*[18] reached the conclusion that continuous warm and dry weather is conducive to the survival of the 2019-nCoV and speculate that conditions such as temperature ranging from 13 to 19°C and humidity between 50% and 80% are suitable for the survival and transmission of this new coronavirus. However, their predictions were performed using SARS data and meteorological

---

[a]Caused by bacteria, not virus, but symptoms are similar to flu, with pneumonia as an expected outcome.

conditions at that time and, as they report, 2019-nCoV has a high basic case reproduction number (R_0) lying between 2.2 to 6.7, causing much more infections than SARS.

Moreover, Mao Wang et al[19] have recently submitted a paper supporting that temperature could change the COVID-19 transmission and that there might be an optimal temperature for the viral transmission, suggesting that colder regions in the world should adopt strictest control measures. Yuwen Cai[20] did not find any correlation between the growth rate of the epidemy and daily mean temperature in either Wuhan or Hunan but found a weak correlation between the mortality in confirmed cases and daily mean temperature both in Wuhan ($r = -0.441$) and Hubei ($r = -0.440$), although not adjusted for the use of three makeshift hospitals, which proved to be effective.

The main focus of this work was to assess the relationship between the rate of spread of COVID-19 and some meteorological variables, which determines the type of model adopted. Although the reproduction number, R_0, is the parameter widely accepted to characterize the velocity of spreading, there are different forms of computing it, which tend to lead to different results[22]. On the other hand, the R_0 calculation is generally based on assumptions about the epidemic phenomenon such as serial interval distribution[25] or "the population is closed, that all cases are observed, and use daily case counts only"[26]. For the reasons mentioned, we opted for a simple/naive model that could assimilate the principal aspects of the variation of COVID-19 cases and translate it into a straightforward measurement (T_d) that could be easily comprehended. Obviously, this model has several drawbacks mostly regarding the optimal period where an exponential growth is verified. We studied two different periods sizes and a random starting point aiming to analyse the impact of this aspect and as an attempt to mitigate its consequences. The doubling time values vary with the period size 26% at maximum (Table 1), which is not neglectable. Even so, this difference only affects slightly the regression coefficients of temperature and humidity, since they do not show statistical differences.

Another point of possible bias is the COVID-19 data that do not cover all provinces from the beginning of the outbreak, making it difficult to study homogeneously in all provinces the time period that corresponds to the exponential growth.

Additionally, the meteorological variables in this study were obtained for locations near the centre of provinces, which typically do not correspond to the average location of the population. Measurements that better represent the central tendency of the meteorological variables felt by the general population of a particular region are currently being implemented.

A final remark about the short average doubling time values obtained (3.78 to 4.53 days - Table 1), which should be a motive of concern. For each doubling time, the number of infected doubles, so one month of sustained growth at a conservative pace of 5 days means an increase of the number of infected by a factor of 64.

Support

References

1. World Health Organization. Novel coronavirus – China. Geneva, Switzerland: World Health Organization 2020. https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/

2. Wang C, Hornby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet, Feb 2020. 395(10223): 470-473. https://doi.org/10.1016/S0140-6736(20)30185-9

3. Hui DS, Azhar EI, Madani TA, Ntoumi F, Kock R, Dar O, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - the latest 2019 novel coronavirus outbreak in Wuhan, China. Int J Infect Dis, Feb 2020. 91:264 – 6. http://dx.doi.org/10.1016/j.ijid.2020.01.009

4. Zhu N, Zhang D, Wang W, Li XW, Yang B, Song JD, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. N Engl J Med, Feb 2020. 382:727-733. http://dx.doi.org/10.1056/NEJMoa2001017

5. Chan JFW, Kok KH, Zhu Z, Chu H, To KKW, Yuan SF, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbs Infect, Jan 2020. 9(1): 221-236 http://dx.doi.org/10.1080/22221751.2020.1719902

6. Tan WJ, Zhao X, Ma XJ, Wang WL, Niu PH, Xu WB, et al. A novel coronavirus genome identified in a cluster of pneumonia cases—Wuhan, China 2019–2020. China CDC Weekly 2020;2(4): 61- 2. http://weekly.chinacdc.cn/en/article/ccdcw/2020/4/61

7. Paules CI, Marston HD, Fauci AS. Coronavirus infection—more than just the common cold. JAMA, Jan 2020. http://dx.doi.org/10.1001/jama.2020.0757

8. Munster VJ, Koopmans M, van Doremalen N, van Riel D, de Wit E. A novel coronavirus emerging in China – key questions for impact assessment. N Engl J Med, Feb 2020; 382:692-694. http://dx.doi.org/10.1056/NEJMp2000929

9. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet, Feb 2020. 395(10223): 497-506 http://dx.doi.org/10.1016/S0140-6736(20)30183-5

10. Chan JFW, Yuan S, Kok KH, To KKW, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet, Feb 2020. 395: 497-506. https://doi.org/10.1016/S0140-6736(20)30183-5

11. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19). China CDC Weekly 2020; 41(2):145-151. DOI: https://doi.org/10.3760/cma.j.issn.0254-6450.2020.02.003

12. Caramelo F, Ferreira N, Oliveiros B. Estimation of risk factors for COVID-19 mortality - preliminary results. Preprint at Med RXIV. 24th February 2020; doi: https://doi.org/10.1101/2020.02.24.20027268

13. https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd402994 23467b48e9ecf6 [accessed on 2020-02-24]

14. Lowen AC, Mubareka S, Steel J, Palese P. Influenza Virus Transmission is Dependent on Relative Humidity and Temperature. PLoS Pathol, 2007 Oct; 3(10): e151.doi: https://dx.doi.org/10.1371/journal.ppat.0030151

15. Prussin AJ, Schwake DO, Lin K, Gallagher DL, Buttling L, Marr LC. Survival of the Enveloped Virus Phi6 in Droplets as a Function of Relative Humidity, Absolute Humidity, and Temperature. Appl Environ Microbiol, 2018 May 31;84(12). pii: e00551-18. https://dx.doi.org/10.1128/AEM.00551-18

16. Simmering JE, Polgreen LA, Hornick DB, Sewell DK, Polgreen PM. Weather-Dependent Risk for Legionnaires' Disease, United States. Emerg Infect Dis. 2017;23(11):1843–1851. doi: https://dx.doi.org/10.3201/eid2311.170137

17. Luo W, Majumder MS, Liu D, Poirier C, Mandl KD, Lipsitch M, Santillana M. The role of absolute humidity on transmission rates of the COVID-19 outbreak. Preprint at Med RXIV. 17th February 2020; doi: https://doi.org/10.1101/2020.02.12.20022467

18. Bu J, Peng DD, Xiao H, Yue Q, Han Y, Lin Y, Hu G, Chen J. Analysis of meteorological conditions and prediction of epidemic trend of 2019-nCoV infection in 2020. Preprint at Med RXIV. 18th February 2020; doi: https://doi.org/10.1101/2020.02.13.20022715

19. Wang M, Jiang A, Gong L, Luo L, Guo W, Li C, Zheng J, Li C, Yang B, Zeng J, Chen Y, Zheng K, Li H. Temperature significant change COVID-19 Transmission in 429 cities. Preprint at Med RXIV. 25th February 2020; doi: https://doi.org/10.1101/2020.02.22.20025791

20. Cai Y, Huang Sr. T, Liu Sr. X, Xu Sr. G. The Effects of "Fangcang, Huoshenshan, and Leishenshan" Makeshift Hospitals and Temperature on the Mortality of COVID-19. Preprint at Med RXIV. 2th March 2020; doi: https://doi.org/10.1101/2020.02.26.20028472

21. https://api.meteostat.net/ [accessed on 2020-03-05]

22. https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series [accessed on 2020-03-05]

23. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

24. Obadia, T., Haneef, R. & Boëlle, P. The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. BMC Med Inform Decis Mak 12, 147 (2012). doi: https://doi.org/10.1186/1472-6947-12-147

25. Hens N, Van Ranst M, Aerts M, Robesyn E, Van Damme P, Beutels P. Estimating the effective reproduction number for pandemic influenza from notification data made publicly available in real time: a multi-country analysis for influenza A/H1N1v 2009. Vaccine. 2011 Jan 29;29(5):896-904. doi: https://doi.org/10.1016/j.vaccine.2010.05.010

26. White LF, Pagano M. Transmissibility of the influenza virus in the 1918 pandemic. PLoS One. 2008;3(1):e1498. Published 2008 Jan 30. doi: https://doi.org/10.1371/journal.pone.0001498