ORIGINAL ARTICLE

The impact of preventive measures on controlling Covid-19 pandemic: a statistical analysis study

Nabil Azouagh[1*] | Said El Melhaoui[2]

[1]Faculty of Sciences and Technics of Mohammedia, Hassan II University. 28806, BP 146 Mohammedia, Morocco.

[2]Faculty of Law, Economic and Social Sciences, Mohammed First University, 60 000 B.P. 724, Oujda, Morocco.

Abstract

Background: The purpose of this paper is to investigate the primary variables associated with the COVID-19 disease and to demonstrate how to evaluate the effect of the earlier consideration of the containment measure and the massive testing policy on controlling the spread of this pandemic. We introduced and analyzed, for the first time to our knowledge, a new variable referred to as the Gap, which was defined as the time between the appearance of the first case and the implementation of the containment measure.

Methods: A correlation, linear, and nonlinear regression-based statistical analysis was conducted to determine the impact of numerous variables and factors on the spread of this pandemic.

Results: 81.3% of the variability of total cases was explained by the variability of total tests, and 72.3% of the variability of total deaths was explained by the variability of total cases. In addition, we have constructed significant nonlinear models that explain 97.8% of the total cases' information and 89.4% of the total deaths' information as a function of the Gap variable. Furthermore, we have found no correlation between the total number of tests and the fatality rate.

Conclusion. Consideration of earlier containment is an effective measure that enables the prevention of a catastrophic disease spread scenario. In addition, the massive testing policy has no effect on the fatality rate. However, the performance of tests is highly effective at detecting new cases earlier, before they infect a large number of individuals, and is also an effective method for controlling the spread of this disease.

Keywords: COVID-19 disease, preventive measures, containment measure, massive policy of tests, linear and nonlinear models.

INTRODUCTION

The novel coronavirus pandemic (COVID-19) continues to claim thousands of lives every day (1). This disease is a severe acute respiratory syndrome with a basic reproductive number estimated from 1.4 to 6.49 (2, 3), it was reported on December 2019 in Wuhan (4), China, then quickly spread infecting millions of people around the whole world (1).

Facing the fact of the absence of any efficient cure combined to the exponential growth in infections (5), the world community has decided to take some preventives measures in order to control the propagation of this virus such as closing borders, schools, restaurants, the quarantine, the total containment and 'stay-at-home' policies (recommended or enforced), the contact tracking, the massive policy tests, social spacing and others measures (6, 7) which lead to minimize the contact between human beings and the earlier discovering of the new infected cases.

Two measures seems to summarize these efforts, the first one is the containment measure which here means a real 'stay-at-home' policy presenting in a restricting mobility except for the utmost necessity under a severe police watch. The second one is conducting a massive policy tests and contact tracking (8) to identify new infected cases.

Some countries hesitated to apply the containment measure, either for economics and social reason or simply because they believe that it was not necessary and not efficient. However, it seems that most countries who don't adopted an earlier containment, thus underestimated this option, suffered from a high propagation of the disease even by adopting other preventive measures as the United State of America, Turkey... (1). Now after quite four months of this crisis, the results of decisions and policies adopted by each country have been revealed in the total number of cases and deaths. Therefore, the problem of evaluating statistically the impact and the efficiency of the containment and the massive policy of tests is essential in this context.

In this work, we conducted a statistical study using COVID-19 data until April 9, 2020 (9) to reveal the relation between several factors of this disease, those of the total of cases and deaths, the fatality rate, the total number of test and a new proposed variable which is the time between the first case and the containment measure called here the Gap. We modeled the relationship between these variables which allowed us to identify the key factor and measure the real impact of these variables specially the containment and the massive policy tests. Therefore, we were able to assess the effect of delaying the preventive measures on the number of cases and deaths.

MATERIALS AND METHODS

2.1. Data base preparation

The initial data (107 countries most affected) was obtained from Worldomatter COVID- 19 coronavirus pandemic (9). For each country, we have selected three variables those of Total cases (TC), Total deaths (TD) and Total tests (TT). Then, based on CHAID classification method (10) we have create a representative sample of 33 countries showing almost all the kind of variability, see Table 1. We have calculated the fatality rate (FR) which is the ratio between confirmed deaths and confirmed cases. Then, basing on the official declaration of health services of these countries we have created the Gap variable which is the difference between the application of severe containment measure and the discovering of the first case. We stress that Gap variable of some countries which didn't consider the containment measure was imputed by the nearest countries Gap value regardless to the total cases.

2.2. Statistical analysis

The statistical analysis was carried out firstly by checking the correlation between variables. If this

Corresponding Author: *Nabil Azouagh,*
Faculté des Sciences et Techniques de Mohammedia.
Université Hassan II, BP 146 Mohammedia 28806,
Maroc
Email: azouaghnabil@gmail.com

correlation is strong we apply a linear regression where we seek the best model in term of fitness (coefficient of determination R-squared) and validity (non- autocorrelation of residuals,...) ([11]). The linear model has the following form:

$$Y = \alpha + \beta X + \varepsilon, \qquad (1)$$

where, α and β are the regression coefficients and ε the random part of the error. This model is characterized by :

$$\triangle Y = \beta \triangle X \qquad (2)$$

In case the correlation (linear dependence) appears to be weak, we use nonlinear models like exponential model, log-linear model and logarithm model ([12]).To perform this non-linear regression we need to apply a logarithm transformation to the adequate variable(s). For instance the log-linear model has the following form:

$$Y = kX^{\beta} + \varepsilon, \qquad (3)$$

where k and β are the coefficients of the model and ε represents the random variable which measure the error part. This model is characterized by:

$$\frac{\triangle Y}{Y} = \beta \frac{\triangle X}{X}, \qquad (4)$$

By applying the logarithm on both sides of (3) it results:

$$ln(Y) = ln(k) + \beta \, ln(X) + \epsilon, \qquad (5)$$

Where the error part is noted ϵ, then by considering $Y' = ln(Y)$, $k' = ln(k)$ and $X' = ln(X)$ (5) is reduced to a linear model:

$$Y' = k' + \beta X' + \epsilon. \quad (6)$$

We apply the traditional linear regression to estimate β and k'. Then, if the intercept is significant we deduce the coefficient k. If it's not significant we re-estimate the parameter β using the model without intercept.

RESULTS AND DISCUSSION

3.1. Relation between, total cases, total tests and fatality rate

Based on Table 2 there is a strong linear dependence between TC and TT which leads us to apply a linear regression. Thus, we obtained the following model:

$$TC = 0.16 \, TT + \varepsilon \qquad (7)$$

The R square is of 0.813 which entails that 81.3% of the variability of total cases is explained by the variability of total tests. Therefore, if the tests increased by 1000 we should expect to discover 160 new cases.

The results in Table 2 show the negative but insignificant correlation between total tests and fatality rate. Thus, there is no linear dependence between these two variables. We also checked that there is no nonlinear dependence of type exponential model, log- linear model and logarithm model. We believe that this inspected result is due to the fact that there is no efficient cure for this disease until now which means that finding the new cases earlier and taking care of them doesn't really affect the fatality rate. However, it will certainly decreases the rate of the disease spread since the patient infected will be quickly pull out from the population before they contaminate other people.

Relation between total cases, deaths and the Gap variable,

The results in Table 3 show a strong linear dependence between TC and TD variables which leads to consider a linear regression. However, a weak linear dependence is noted between the Gap variable and both TC and TD variables, in this case we will investigate some nonlinear dependence.

By regressing TD on TC we obtained:

$$TD = 0.051 \, TC + \varepsilon \qquad (8)$$

The R square is of 0.723 which entails that 72.3% of the variability of total deaths is explained by the variability of total cases. Therefore, if the infected cases increased by 1000 we should expect the rise of the deaths by 51.

Concerning the TC and Gap variables, the investigation of many nonlinear models allowed us to discover that the best model to describe the relation between the total cases and the Gap is the log-linear model. Thus, by applying the regression procedure described in subsection 2.2 we obtained the following model:

$$TC = Gap^{2.7} + \varepsilon, \qquad (9)$$

where we have succeeded to explain 97.8% of the TC information. The log-linear is characterized by

$$\frac{\triangle TC}{TC} = 2.7 \frac{\triangle Gap}{Gap} \qquad (10)$$

This results means that the growth rate of TC is 2.7 proportional to the growth rate of the Gap. So, a growth rate of 1% in the Gap will imply a 2.7% increase in the growth rate of the total cases. To exploit this result, let's take for instance the case of Morocco in the day of 9 April. Therefore, if they delayed almost a weak to apply the containment measure they would have double total cases. Indeed, by using (10) and knowing that the Gap is 18 days; instead of 1374 total cases on April 9, we would expect 2817.

Concerning the relationship between TD and the Gap variables, we deduce the same type of nonlinear model as for TC and the Gap which is the log-linear model. Thus, we applied the regression procedure to obtain:

$$TD = Gap^{1.73} + \varepsilon \qquad (11)$$

We have succeeded to explain 89.4% of the TD information. The log-linear is characterized by the following property:

$$\frac{\triangle TD}{TD} = 1.73 \frac{\triangle Gap}{Gap} , \qquad (12)$$

The growth rate of TD is 1.73 proportional to the growth rate of the Gap. Therefore, a growth rate of 1% in the Gap will entails a 1.73% increase in the growth rate of the total deaths. As before, let's take the case of Morocco in the day of 09 April. Thus, if they delayed almost 11 days to apply the containment measure they would have exceeded the double total deaths. Indeed, by using (12) and knowing that the Gap is 18 days, instead of 97 total deaths on April 9, we would expect almost 200.

CONCLUSION

Using Pearson's correlation, linear and nonlinear regression, we investigated the relationship between total cases, total tests, fatality rate, and for the first time, to our knowledge, the difference between the day of containment and the day of the first case, which we refer to as the Gap variable. Using a linear model, we have demonstrated the strong relationship between total cases and tests performed. Thus, we have demonstrated that an increase in the number of performed tests will reveal more new infected cases, thereby preventing further contamination caused by symptomless infected individuals. This result is, in our opinion, due to the absence of an effective treatment for this illness. In addition, a significant linear model has been developed to describe the variability of deaths as a function of new cases. Moreover, we demonstrated in a pragmatic manner that delaying the implementation of containment measures can double the rate of disease spread and increase the number of lives lost. In contrast to the vast majority (almost all) of those who initially hesitated to implement this preventative and effective measure, countries that considered earlier containment have not faced a catastrophic scenario. Therefore, it is strongly advised to implement stringent containment measures, especially when confronting a highly contagious new disease.

INFORMATION

TABLE 1: COVID-19 selected database at the date of April 9, 2020.

Countries	TC	TD	TT	Gap	FR
USA	468566	16691	2353096	51	0,0356
Spain	153222	15447	355000	43	0,1008
Italy	143626	18279	853369	39	0,1273
Germany	118235	2607	1317887	55	0,022
France	117749	12210	333807	53	0,1037
china	81865	3535	NA	57	0,0432
Iran	66220	4110	231393	54	0,0621
UK	65077	7978	298169	52	0,1226
Turkey	42282	908	276338	43	0,0215
Belgium	24983	2523	84248	43	0,101
Switzerland	24051	948	178500	20	0,0394
Netherlands	21762	2396	101534	43	0,1101
Canada	20765	509	370315	54	0,0245
Brazil	18145	954	62985	28	0,0526
Portugal	13956	409	140368	14	0,0293
Austria	13244	295	126287	17	0,0223
Russia	10131	76	1004719	59	0,0075
India	6725	227	177584	54	0,0338
Ireland	6574	263	53000	28	0,04
Denmark	5635	237	64002	13	0,0421
Japan	5347	99	61498	83	0,0185
Pakistan	4489	65	44896	30	0,0145
Colombia	2223	69	34910	18	0,031
Greece	1955	87	33634	26	0,0445
South Africa	1934	18	68874	26	0,0093
Argentina	1894	79	14850	16	0,0417
Algeria	1666	235	3359	27	0,1411
Iceland	1648	6	32663	25	0,0036
Morocco	1374	97	6116	18	0,0706
Moldova	1289	29	5108	10	0,0225
New Zealand	1239	1	51165	26	0,0008
Tunisia	643	25	9570	18	0,0389
Bulgaria	618	24	15899	9	0,0388

TABLE 2: Correlation between Total Cases (TC), Total Tests (TT) and Fatality Rate (FR) variables.

	TC	TT	FR
TC : Person's correlation	1	0.877	0.188
TT : Person's correlation	0.877	1	-0.017
FR : Person's correlation	0.188	-0.017	1

TABLE 3: Correlation between Total Cases (TC), Total Deaths (TD) and Gap variables.

	TC	TD	Gap
TC : Person's correlation	1	0.810	0.360
TD : Person's correlation	0.810	1	0.349
Gap : Person's correlation	0.360	0.349	1

REFERENCES

1. World Health Organization. Coronavirus disease 2019 (COVID-19). Available from: https: //cutt.ly/PynTDo5

2. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of Covid-19 is higher compared to SARS coronavirus. J Trav Med. 2020;27(2):taaa021. doi:10.1093/jtm/taaa021

3. Chinese Center for Disease Control and Prevention. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19). China CDC Weekly. 2020;(8):113–135.

4. Organization WH. World Health Organization. Coronavirus disease 2019 (COVID-19); 2020. Available from: https://cutt.ly/tynhdNX

5. Rapid Risk Assessment: Coronavirus disease 2019 (COVID-19) in the EU/EEA and the UK-ninth update; 2020. Available from: https://cutt.ly/9yNq40v

6. World Health Organization. Coronavirus disease 2019 (COVID-19). Available from: https: //cutt.ly/cynT9Nj [AccessedMay10,2020].

7. Coronavirus disease (COVID-19) technical guidance: Infection prevention and control;. Available from: https://cutt.ly/LynCTNS

8. World Health Organization. Coronavirus disease 2019 (COVID-19). Available from: https: //cutt.ly/QynxPth

9. WORLDOMATER. COVID-19 Coronavirus Pandemic; 2020. Available from: https://cutt.ly/mynYkVW.[AccessedApril10,2020].

10. Nisbet R, Elder J, Miner G. Academic Press; 2009.

11. Douglas C, Montgomery EA, Peck G, Vining GG. Introduction to linear regression analysis. John Wiley & Sons; 2012.

12. Christensen R. Log-linear models and logistic regression. Springer Science & Business Media; 2006.