

Analysis of Data on Socio-Demographic and Clinical Factors of the COVID-19 Coronavirus Epidemic in Spain on Cases of Recovered and Death Cases

G. Sanglier Contreras¹, M. Robas Mora² & P. Jimenez Gómez²

¹Department of Architecture and Design. Engineering Area. Escuela Politécnica Superior, Universidad CEU San Pablo, Boadilla del Monte, Madrid, Spain.

²Microbiology Area, Pharmaceutical and Health Sciences Department, Faculty of Pharmacy, Universidad San Pablo CEU, Boadilla del Monte, Madrid, Spain.

Correspondence: Gastón Sanglier Contreras, Engineering Area, Escuela Politécnica Superior, Universidad CEU – San Pablo of Madrid, Spain. E-mail: sanglier.eps@ceu.es

Received: May 10, 2020

Accepted: June 30, 2020

Online Published: July 2, 2020

doi:10.5539/mas.v14n8p9

URL: <https://doi.org/10.5539/mas.v14n8p9>

Abstract

Carrying out a study of socio-demographic and clinical factors to determine which of these are more significant and have a greater influence on the speed of the spread of the virus, taking into account the behaviour of people who have died and been recovered in Spain. The objectives of this study have been to analyze the influence of socio-demographic and clinical factors on the speed of propagation of Covid-19, to determine the most relevant factors and to propose studies determining the prevalence of the disease.

The Chi-square model supported by the statistical program Statgraphics Centurion xvi has been used to determine the dependence or not of the different variables studied on the speed of propagation of the virus. In relation to the clinical variables, a cluster study has been carried out to see their dependence.

Very relevant conclusions have been obtained from the factor of age in the different analyzed bands, as well as from the little influence of the economic position of the people in the speed of propagation of the virus. The high population density and the areas studied are not always indicative of further spread of the disease

A linear function has been determined to link the clinical parameters studied that could be used in subsequent prevalence and seroprevalence studies. The fundamental variables in the study of the coronavirus have been indicated according to socio-demographic and clinical factors. We warn about environmental factors to be studied.

Keywords: infection, epidemiology, virus, data análisis, socio-demographics factor, clinical factor, Covid-19

1. Introduction

Although the epidemic has spread throughout the world, Spain has been one of the countries hardest hit by the pandemic. The first patient registered in Spain with coronavirus Covid-19 was known on January 31st on the island of La Gomera. Nine days later another case was detected on the island of La Palma. But it was not until Feb. 24 that the virus jumped to the mainland, with the first cases being detected in the communities of Madrid, Catalonia and Valencia (Sanglier, Robas & Jimenez, 2020).

Real-time analysis of the evolution of the coronavirus by Johns Hopkins University continues to add to its numbers. Today, the balance of infections in the world amounts to more than 3.3 million, with more than 239,000 deaths. Spain is now the fourth largest country in the world after France, the United Kingdom and Italy.

To date, Spain presents the following data: confirmed data of infected by PCR (Polymerase Chain Reaction) tests (216,582), deceased (25,100) and recovered (117,248). Their evolution from 31 January to 2 May is shown in the following graph.

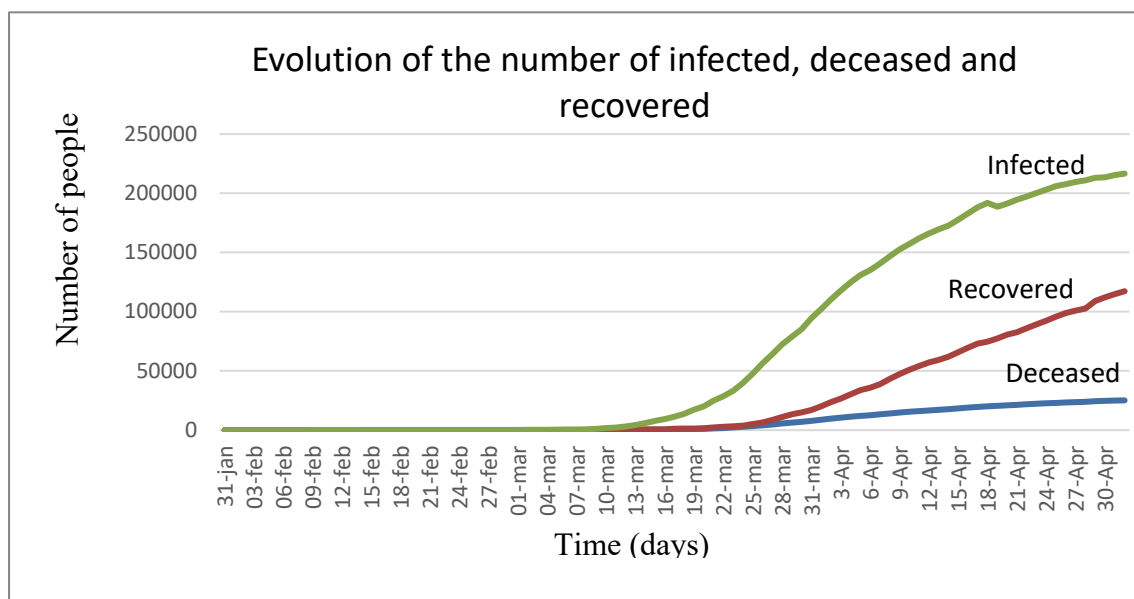


Figure 1. Evolution of the number of people infected, deceased and recovered in Spain during the Covid-19 pandemic

We look for the possible numerous reasons for the rapid spread of the coronavirus in the different communities of Spain. Researchers are analysing possible causes due to social-demographic factors (Chen et al., 2012; Clapham et al., 2016), environmental factors (Shaman & Kohl, 2009; van Regenmortel, 2000; Pica & Bouvier; 2011) and clinical factors (Yu et al., 2004; Zhu et al., 2019) among the main causes of the spread of the virus.

The aim of this research will focus on studying the effect of several variables contained in the factors mentioned above and seeing their involvement in the spread of the coronavirus CoVID-19 in Spain. The research will focus on studying the effect of variables such as sex, age, area (community), social level (income), etc. on the number of deaths and recoveries in Spain. The Chi-square methodology will be used to determine whether the variables analysed have an influence on the spread of the virus. Based on the variables analysed and the results obtained in this study on the influence on the spread of the virus, a prediction model could be determined that will help in the study of this type of pandemic (Canini & Perelson, 2014; Checkoway, Pearce & Crawford-Brown, 1989).

2. Materials and Method

The study used data provided by the Ministry of Health's Alert and Emergency Coordination Centre for cases of infection by Covid-19 disease from 23 March to 25 April, as these are the dates when the number of infected, dead and recovered cases first began to rise and then decline in a controlled manner.

Statistical analysis has been used as the Chi-square test to analyse the data collected and see its impact on the different variables to be taken into account in the study of the speed of the epidemic's spread.

The Chi-square or Pearson's test is a statistical test that allows us to recognize the association between two categorical variables whether they are dichotomous or polytomous. It is used to check the relationship between two variables in a contingency table that presents the frequency distribution (multivariate) of the variables. This test, also known as the independence test, is used to test the hypotheses of the categorical variables, and to check whether these variables are independent population variables or not (Dahari et al., 2005; De Wit, van Doremalen, Falzarano & Munster, 2016).

An analysis will be carried out using the Chi-square test of quantitative variables among them, such as sex, age, region, etc. against other quantitative variables such as the number of deaths or the number of people recovered according to the statistical data collected.

The Chi-square values calculated can be obtained by the formula:

$$X^2 = \sum_{1,j}^n \frac{(f_0 - f_e)^2}{f_e} \quad (1)$$

Being f_0 the observed frequency and f_e the expected frequency. To calculate the critical Chi-square, the values of the level of significance and the degree of freedom obtained will be taken into account.

The comparison between the expected Chi-square and the critical one will determine whether the null hypothesis or the alternative hypothesis raised in relation to the independence or not of the variables under study is fulfilled. The variables are significant if the probability of the association (P) does not exceed 1 in every 1,000 cases (0.0001); otherwise, it will be considered that it is not significant, therefore, the association between the variables will be rejected. We can also speak of a 95% confidence level, this means that if the P-value is less than 0.05 the data compared will be independent of each other, otherwise they will be independent.

Cross-tabulation methodology has been used to understand the relationship between the independent variables and the chaos of the deceased and recovered population. This cross tabulation method is used between a dependent variable and an independent variable, to understand how the dependent variable moves in relation to the independent variables (Dee & Shuler, 1997; Dickman & Heesterbeek, 2000; Montesinos-López & Hernández Suárez, 2007).

3. Results

3.1 Cases of Recovered and Deceased People

As of April 25, the number of confirmed cases of people infected with Covid-19 in Spain was 194,333, an increase of 1.5%. The total number of cases recovered was 78,430 and the number of deaths was 14,629.

For this study, carried out between 23rd February and 24th April for the reasons already mentioned, it has been determined that the number of people infected was 176,311, of which 99,005 were women and 77,306 men.

To study the effect of the sex-independent variable on the number of persons cured or recovered, the number of female cases recovered was 30,910 and the number of male cases recovered was 40,555.

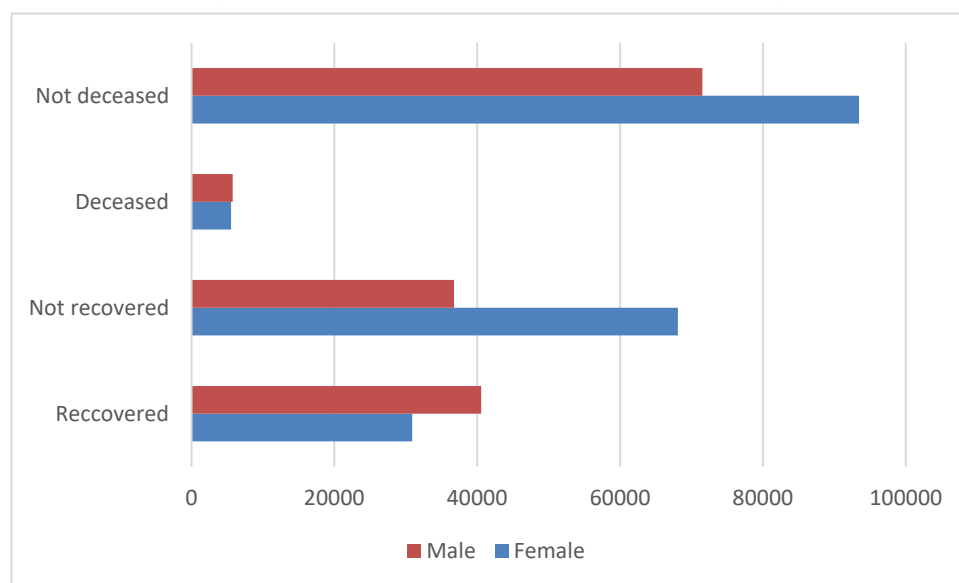


Figure 2. Distribution of deceased, undeceased, recovered and unrecovered cases base on sex.

3.2 Sex Influence

The Chi-square test will be used to determine the impact of the sex variable on the number of people recovered and on the number of deaths (Lee & Storch, 2014; Center for Disease Control and Prevention, 2009; Cember, 1985). The Table 1 of results obtained was as follows:

Table 1. Frequency table for the cases studied referring to the sex

	<i>Not deceased</i>	<i>Deceased</i>	<i>Not recovered</i>	<i>Recovered</i>	<i>Total per row</i>
Female	93462	5543	68095	30910	198010
	26.50%	1.57%	19.31%	8.77%	56.15%
Male	71532	5774	36751	40555	154612
	20.29%	1.64%	10.42%	11.50%	43.85%
Total per column	164994	11317	104846	71465	352622
	46.79%	3.21%	29.73%	20.27%	100.00%

Table 1 shows the counts in a 2x4 table. The first number in each cell in the table corresponds to the count or frequency. The second number shows the percentage of the whole table that that cell represents.

In the test of independence, a statistic of 8377.392 was obtained, with 3 degrees of freedom and a P-value of 0.0000. Since this value is less than 0.05, the hypothesis that the variables in the rows and columns are independent with a 95% confidence level can be rejected. Therefore, the variable observed in each row is related to its column. This means that the sex variable is statistically significant with the number of cases of recovered and deceased persons. The number of cases of deceased men is slightly higher than that of women, as is the case with those recovered, this is also observed in figure 2.

For the analysis of the age variable, with respect to the cases of infected persons, recovered persons and death, 9 age ranges have been taken into account for men and women (0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80 onwards). The data have been collected in the graph in figure 3.

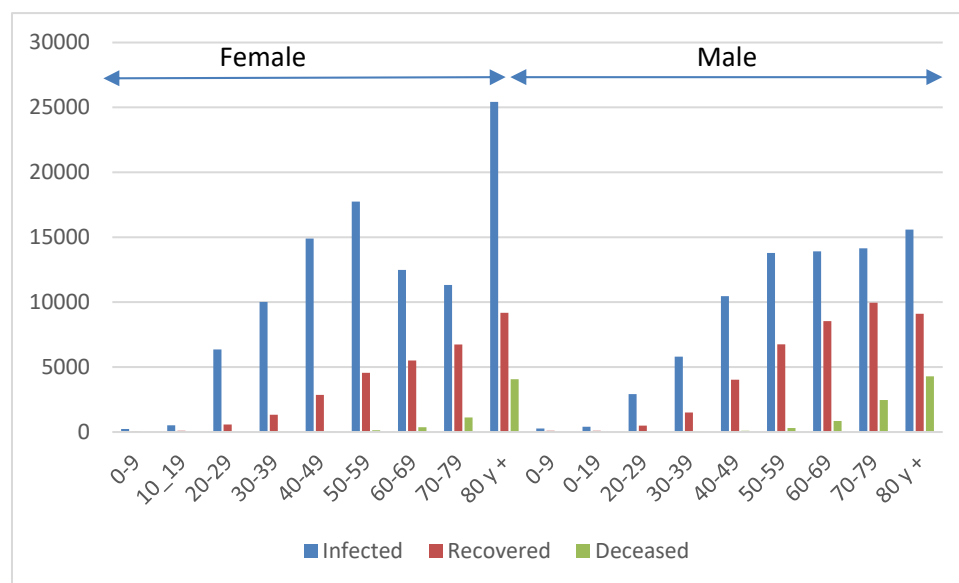


Figure 3. Distribution of infected, deceased and recovered cases base on age

3.3 Age Influence

As in the case of sex, the Chi-square test will be used to determine the impact of the age variable (in its different year bands) on the number of people recovered and on the number of deaths (Abrego & Del Río, 1998; Khiabani, Farrel, St George, & Rabadan, 2009; Lee & Storch, 2014; Konca et al., 2017).

The following table of frequencies shows the data entered for the respective age ranges for men and women, as well as all the weights obtained, which will give an idea of the influence of each variable with respect to the number of infected, recovered and dead people (Table 2).

Table 2. Frequency table for the cases studied referring to the age

<i>FEMALE</i>	<i>F(0-9)</i>	<i>F(10-19)</i>	<i>F(20-29)</i>	<i>F(30-39)</i>	<i>F(40-49)</i>	<i>F(50-59)</i>	<i>F(60-69)</i>	<i>F(70-79)</i>	<i>F(80+)</i>
Infected	235	522	6362	10003	14897	17749	12479	11323	25425
	0.09%	0.20%	2.43%	3.83%	5.69%	6.78%	4.77%	4,33%	9.71%
Recovered	77	101	573	1325	2859	4555	5504	6737	9179
	0.03%	0.04%	0.22%	0.51%	1.09%	1.4%	2.10%	2,57%	3.51%
Deceased	1	2	10	24	54	130	367	1125	4061
	0.01%	0.00%	0.00%	0.01%	0.02%	0.05%	0.14%	0,43%	1.55%
Total per column	313	625	6945	11362	17810	17810	18350	1985	38665
	0.12%	0.24%	2.65%	2.65%	6.80%	6.80%	7.01%	7,33%	14.77%

<i>MALE</i>	<i>F(0-9)</i>	<i>F(10-19)</i>	<i>F(20-29)</i>	<i>F(30-39)</i>	<i>F(40-49)</i>	<i>F(50-59)</i>	<i>F(60-69)</i>	<i>F(70-79)</i>	<i>F(80+)</i>	<i>Total per row</i>
Infected	265	401	2925	5809	10467	13916	13916	14150	15581	176435
	0.10%	0.15%	1.12%	2.22%	4.00%	5.32%	5.32%	5.41%	5.95%	67.41
Recovered	102	98	488	1499	4025	6754	8534	9955	9100	71465
	0.04%	0.04%	0.19%	0.57%	1.54%	2.58	3.26%	3.80%	3.48%	27.31
Deceased	1	1	14	27	92	305	855	2462	4293	13824
	0.00%	0.00%	0.01%	0.01%	0.04%	0.12%	0.33%	0.94%	1.64%	5.28
Total per column	368	500	3427	7335	14584	20975	23305	26567	28974	261724
	0.14%	0,19%	1,31%	2.80%	5.57%	5.57%	8.90%	10.15%	11.07%	100.00%

In the independence test, a statistic of 23917,882 was obtained, with 34 degrees of freedom and a P-value of 0.0000. Since this value is less than 0.05, the hypothesis that the variables in the rows and columns are independent with a 95% confidence level can be rejected. Therefore, the variable observed in each row is related to its column. This means that the age variable is statistically significant with the number of cases of infected persons, recovered persons and deaths.

The number of cases of infected women shows two significant increases in those between 40 and 49 years old with 5.69% and in women over 80 years old with 9.71% of cases. In men, the infection is more uniform between 50 and 80 years of age or more, with data ranging from 5.32 to 5.95%.

The number of recovered cases has an increasing trend in women between 1.09% and 3.51% between the ages of 40-49 and 80+ respectively. For men, the trend is upwards from 1.54% to 3.80% for the 40-79 age group, with a significant drop in the number of cases recovered in the over-80 age group to 3.48%, as shown in Figure 3. As for the number of deaths, the trend is upwards for women and men, with a slightly higher number of deaths among men, 1.64% compared to 1.55%.

It can be seen that for women, the age groups between 40-49, 50-59 and 80+ and for men between 50 and 80+ are the most important because they have a greater number of people infected in order to carry out studies on the influence of the virus and to determine critical factors in its spread.

3.4 Zones and Population Density Influence

Another factor to take into account is the analysis by communities and population density (Sadique et al., 2007; Lowen, Mubareka, Steel & Pelese, 2007) as possible factors of influence in the speed of the virus propagation, taking into account the chaos of the number of recovered and dead people. For this purpose, a table will be constructed collecting the data indicated for the different Autonomous Communities.

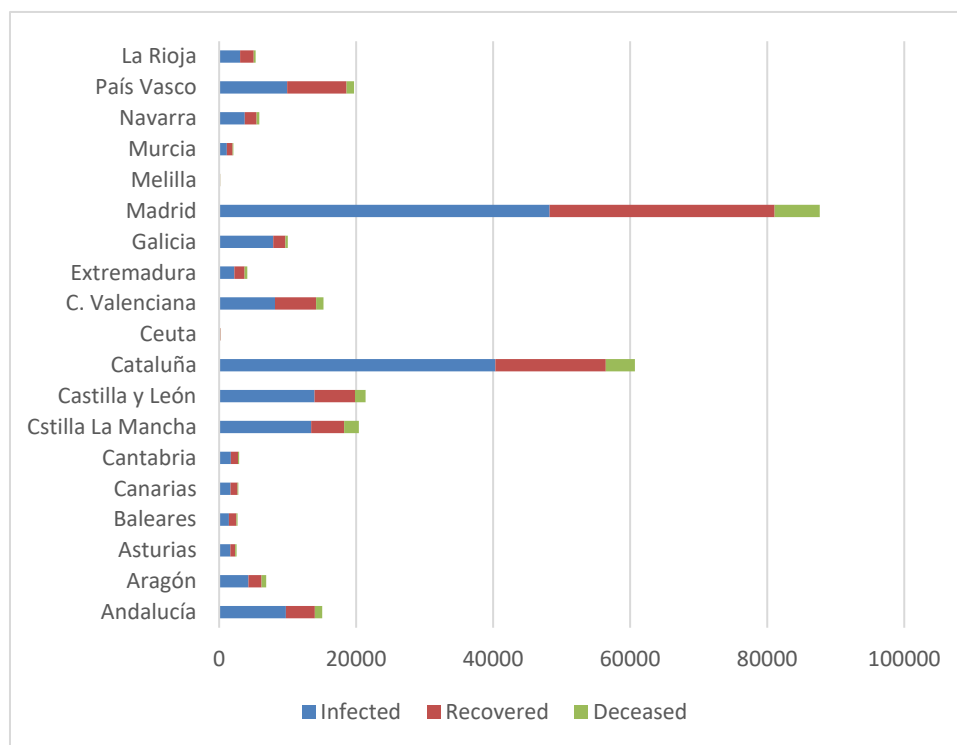


Figure 4. Distribution of infected, deceased and recovered cases base on Autonomous Communities

Figure 4 shows that the communities most affected by the number of coronavirus infections in Spain have been the communities of Madrid and Catalonia to a greater extent, followed by the communities of Castilla y León, the Basque Country, La Rioja and Andalusia. This also corresponds to a higher number of recovered cases and deaths.

If we now make a comparative study of the population density (number of inhabitants / surface area) against the number of infected persons, we obtain the graph in figure 5.

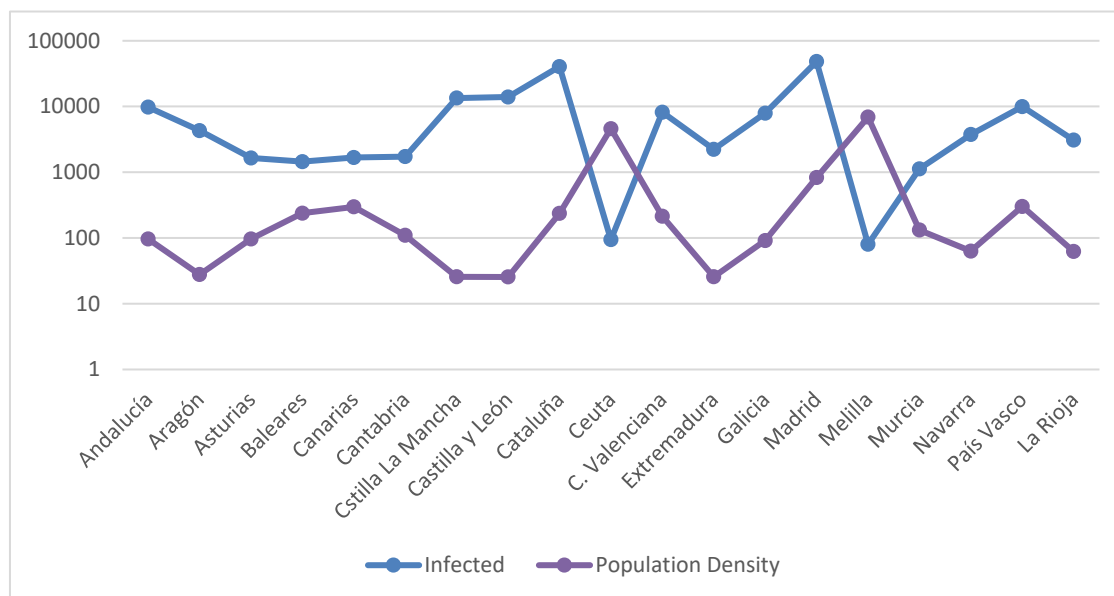


Figure 5. Distribution of infected versus population density in Autonomous Communities

As can be seen from figure 5, a higher population density, as presented by the communities of Ceuta, Melilla, Balearic and Canary Islands mainly, does not seem to correspond to a higher number of infected people. This could be due to the fact that the four Spanish communities are physically outside the peninsula. If we analyse other communities with an important population density such as Madrid and Catalonia, in this almost, if we

observe a higher number of infected people in both. Communities with a low population density such as Aragon, Castilla La Mancha, Castilla y León and Extremadura, do not present a small number of infected people as we would expect. It could be thought that the geographical location of the area (community) analysed, as well as the communications between them (bus, car, train, etc.) could be factors to consider for further studies. The table of frequencies obtained for the variables studied shows the following results:

Table 3. Frequency table for the cases studied referring to the Autonomous Communities and Population density

	<i>Andalucía</i>	<i>Aragón</i>	<i>Asturias</i>	<i>Baleares</i>	<i>Canarias</i>	<i>Cantabria</i>	<i>Castilla La Mancha</i>	<i>Castilla and León</i>
Infected	9742	4284	1644	1447	1674	1724	13431	13935
	3.24%	1.43%	0.55%	0.48%	0.56%	0.57%	4.47%	1.41%
Recovered	4245	1926	695	1084	1029	1035	4825	5916
	1.41%	0.64%	0.23%	0.36%	0.34%	0.34%	1.61%	1.97%
Deceased	1073	677	227	164	119	176	2147	1537
	0.36%	0.23%	0.08%	0.05%	0.04%	0.06%	0.72%	0.51%
Population Density	96	28	96	238	296	109	26	26
	0.03%	0.01%	0.03%	0.08%	0.10%	0.04%	0.01%	0.01%
Total per column	15156	6915	2662	2933	3118	3044	20429	21434
	5.05%	2.30%	0.89%	0.98%	1.04%	1.01%	6.80%	7.13%
	<i>Cataluña</i>	<i>Ceuta</i>	<i>C.Valenciana</i>	<i>Extremadura</i>	<i>Galicia</i>	<i>Madrid</i>	<i>Melilla</i>	<i>Murcia</i>
Infected	40336	94	8165	2225	7908	48244	80	1123
	13.44%	0.03%	2.72%	0.74%	2.63%	16.07%	0.03%	0.37%
Recovered	16109	98	5997	1504	1764	32839	77	841
	5.37%	0.03%	2.00%	0.50%	0.59%	10.94%	0.03%	0.28%
Deceased	4253	4	1078	398	370	6585	2	124
	1.42%	0.00%	0.36%	0.13%	0.12%	2.19%	0.00%	0.04%
Population Density	236	4586	214	26	91	827	6887	131
	0.08%	1.53%	0.07%	0.01%	0.03%	0.28%	2.29%	0.04%
Total per column	60934	4782	15454	4153	10133	88495	7046	2219
	20.30%	1.59%	5.15%	1.38%	3.38%	29.48%	2.35%	0.74%
	<i>Navarra</i>	<i>País Vasco</i>	<i>La Rioja</i>	<i>Total per row</i>	<i>Autonomous Communities</i>			
Infected	3741	9945	3074	172816				
	1.25%	3.31%	1.02%	57.56%				
Recovered	1730	8658	1981	92353				
	0.58%	2.88%	0.66%	30.76%				
Deceased	405	1092	289	20720				
	0.13%	0.36%	0.10%	6.90%				
Population Density	63	301	62	14339				
	0.02%	0.10%	0.02%	4.78%				
Total per column	5939	19996	5406	300228				
	1.98%	6.66%	1.80%	100.00%				

From the above data, it can be deduced that the number of people infected ranges from 0.03% to 16.07%, of people recovered from 0.03% to 10.94% and of people who died from 0.00% to 2.19%. The population density is between 0.01% and 2.29%.

The table of frequencies confirms the conclusions obtained above. The number of infected persons is higher for the autonomous communities of Madrid and Catalonia with 16.07% and 13.44%, also for the number of recovered persons with 10.94% and 5.37% respectively, and consequently for the number of deaths with 2.19% and 1.42%.

The communities of Melilla and Ceuta have the highest population densities with 2.29% and 1.42% of the total.

The graph in figure 6 has compared the number of confirmed infections in Covid-19 with the ratio of number infected divided by population density to see the influence on the first factor. It has been used for the ordinate axis logarithmic scale.

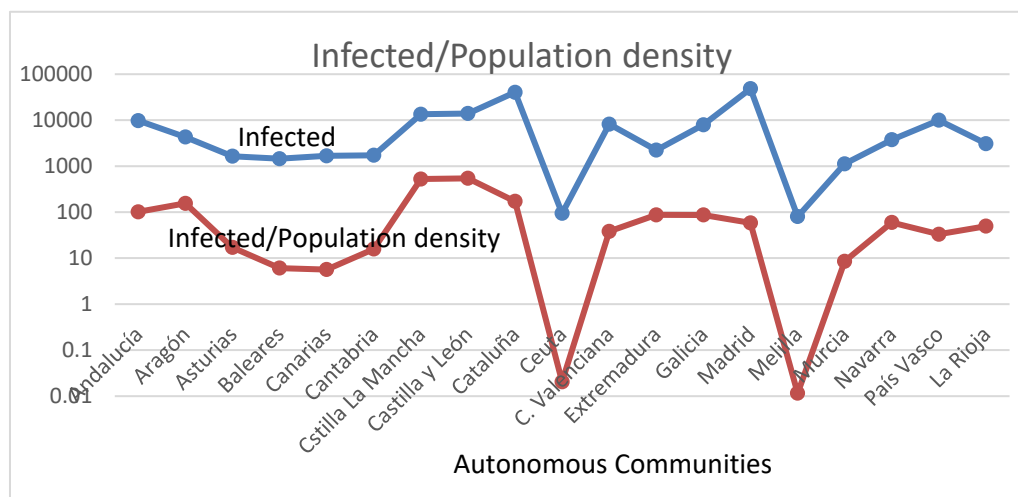


Figure 6. Number of infected versus number of infected divided by population density for the different Autonomous Communities

It is now perfectly clear that in communities with a lower population density, the number of infected people is smaller (Ceuta and Melilla). For other communities with high population densities, such as the Balearic Islands, the Canary Islands, Madrid, Catalonia and the Basque Country, it is diverse, and the same is true for low population densities (Aragon, Extremadura).

In the independence test, a statistic of 238616.112 was obtained with 54 degrees of freedom and a P-value of 0.0000. As in the cases studied previously, a value of less than 0.05 is obtained, therefore, the hypothesis that the variables in the rows and columns are independent with a 95% confidence level can be rejected. Therefore, the variable observed in each row is related to its column. This means that the variable of the zones (communities) is statistically significant with the number of cases of infected people, recovered people, dead people and with the established population density data.

3.5 Income Influence

Another socio-demographic factor to be analysed is the income of the population in Spain. How can this affect the economic level of people in the spread of the virus? For this parameter three economic ranges have been established, a low income ranging from 0-15,000 euros, the population that identifies with this status is 38.5%; an average income from 15,001-40,000 euros with 52.3 of population, and finally, a high income from 40,001 euros onwards with 9.2 of population.

If we take into account that the number of confirmed infected people was 176,311 (99,005 women and 77,306 men), we will have 67,880 people with low income, 92,210 people with medium income and 16,221 people with high income. If we do the Chi-square test for these data, the following table of numerical frequencies is obtained:

Table 4. Frequency table for the cases studied referring to low, average and high income

	<i>Low Income</i>	<i>Average Income</i>	<i>High Income</i>	<i>Total per row</i>
Total number of infected	67880	92210	16221	176311
	19.25%	26.15%	4.60%	50.00%
Female	38117	51780	9108	99005
	10.81%	14.68%	2.58%	28.08%
Male	29762	40431	7113	77306
	8.44%	11.47%	2.02%	21.92%
Total per column	135759	184421	32442	352622
	38.50%	52.30%	9.20%	100.00%

The table above shows the results obtained to determine whether or not the income of the Spanish population can influence or affect the number of people infected by coronavirus in the population of men and women. After performing the Chi-square analysis the value of the statistic performed with 4 degrees of libertad is 0.000, with a P-value of 1.0000, this indicates that since this value is greater than 0.05 (95% confidence level), it can be assured that the income data do not influence the speed of propagation of the virus.

3.6 Clinical Factors Influence

Apart from the socio-demographic data analysed above, there are other clinical data (Lourraine, 1998; Dowell, 2001; Elandt-Johnson, 1975) that could be of great importance to investigate in view of the possible spread of the pandemic, such as: hypertension, high cholesterol levels, allergies, mental health, tobacco and overweight that could affect the evolution of the disease.

The following table shows the data collected by the National Health System (SNS) in relation to the above parameters:

Table 5. Percentages of chronic health problems and other symptoms in infected women and men

<i>Percentage of population</i>	<i>Major chronic health problems</i>						
	<i>Hipertensión</i>	<i>Colesterol alto</i>	<i>Alergias</i>	<i>Salud Mental</i>	<i>Diabetes</i>	<i>Tabaco</i>	<i>Sobrepeso</i>
Female	20%	18%	15%	11%	8%	19%	17%
Male	20%	18%	15%	11%	8%	26%	18%

For the study of these independent variables, the respective data will be calculated using the previous weights, with respect to the infected, recovered and deceased persons between the time period of the study.

A multivariate statistical analysis by principal component analysis will be carried out with the aim of reducing the dimensionality of the study from the seven quantitative independent variables to new and few variables, and thus be able to make a better interpretation of the data (Murray & Morse, 2011; Lofgrnn, Fefferman, Naumov, Gorski & Naumova, 2007).

The weights corresponding to these variables will be determined in order to determine, by means of a single independent variable or main component, the influence on the number of infected people and to reduce as much as possible the number of independent variables of the study. For this purpose, the analysis program StatGraphics Centurion XVI v.2.0 will be used.

The data obtained are summarized in the following table where the data of percentages of variance and accumulated data appear, in addition to the eigenvalues of each of the variables. The technique consists of making a factorial decomposition, that is, taking the variance and covariance matrices of the variables and calculating the eigenvalues that will generate the associated eigenvectors and these are the weights or evaluations of the new variables called main components.

Table 6. Percentage of variance and cumulative for the main component

<i>Component</i>		<i>Percentage of</i>	<i>Percentage</i>
		<i>variance</i>	<i>cumulative</i>
<i>Number</i>	<i>Eigenvalor</i>		
1	5.9992	99.987	99.987
2	0.000698699	0.012	99.998
3	0.000103456	0.002	100.000
4	6.30581E-9	0.000	100.000
5	4.14768E-9	0.000	100.000
6	2.93775E-9	0.000	100.000

With a single main component (MC), 99.987% of the variance or accumulated variability can be captured, i.e. what was exposed with six variables is now explained with a single variable, which is a linear combination of the old variables.

The table of weights has been obtained to construct that new variable or main component whose values are shown in the following table:

Table 7. Weights for the new variable

<i>Component</i>	
<i>1</i>	
Hypertension (H)	0,408269
High cholesterol (HC)	0,408269
Allergies (A)	0,408269
Mental Health (MH)	0,408269
Tobacco (T)	0,408158
Overweight OW)	0,408257

This table gives the equation of the new main component that it has as an equation:

$$MP=0,408269*(H+HC+A+MH) + 0,408158*T + 0,408257*OW \quad (2)$$

Where the values in equation 2 have been standardized by subtracting their mean and dividing by their standard deviations. Next, a cluster análisis (Chan et al., 2020) will be performed to see if groupings can be made for some common characteristics of the factors analyzed. Similarities between the groups are sought. There are two criteria, one is the metric to be used, and the other is the clustering method, for the first a Euclidean square metric has been chosen, and for the second the nearest neighbour clustering method. The clustering carried out has been of observations with the standardized variables.

A cluster with 36 variables has been created. Clusters are groups of variables with similar characteristics. To form these clusters, the procedure starts with each variable in separate groups, then combines the two variables that were closest to each other to form a new group. After recalculating the distance between groups, the two closest groups are combined. This process is repeated until only one group remains. A graph called a dendrogram has been obtained to see the associations of variables. They have been grouped into three groups as shown in Figure 7 below.

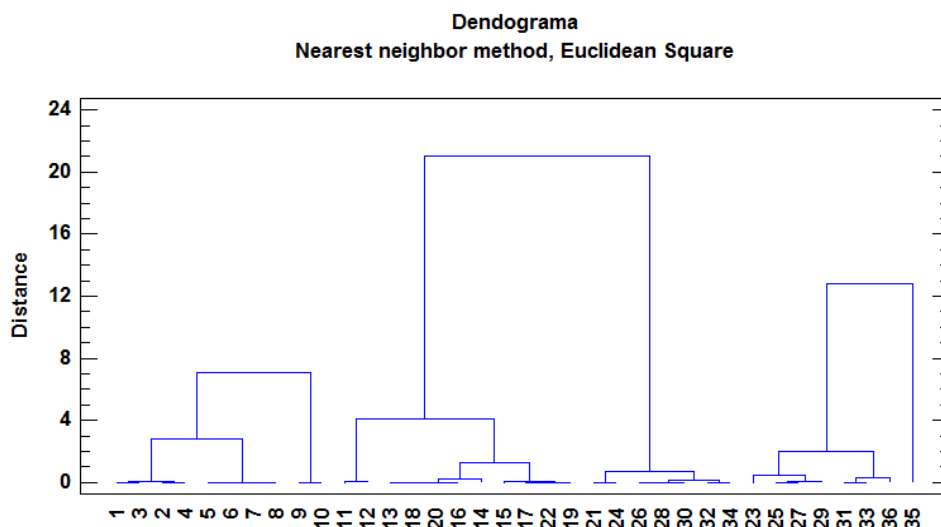


Figure 7. Dendrogram of the analysed variables

The cluster scatter diagram with the chosen metric and method parameters is as follows depending on the grouping made in the dendrogram

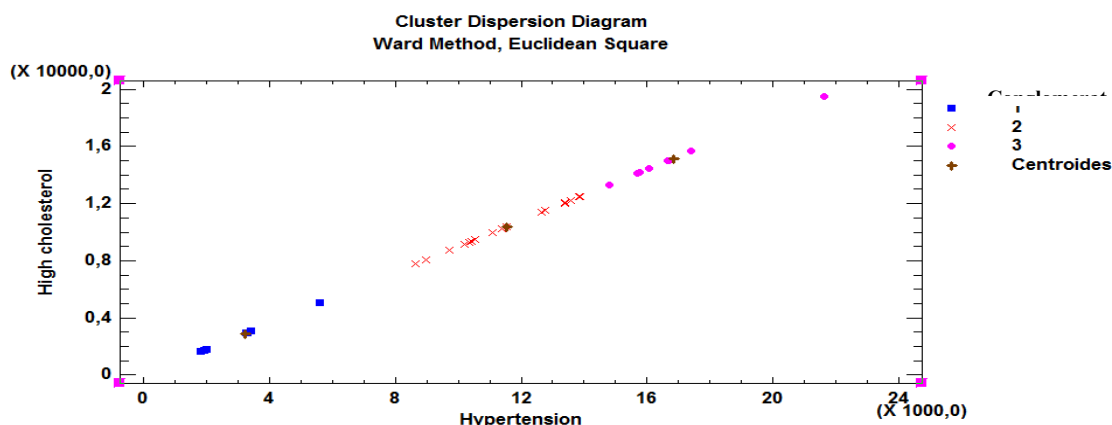


Figure 8. Cluster Dispersion Diagram

The three groupings made in the dendrogram are shown in the scatter plot. The first grouping appears in blue, the second in orange and the third in purple. The different centroids of the groupings made have also been represented in brown.

The graph in figure 9 shows a linear functional dependence of all the variables to which equation (1) obtained above responds for the main component that links the variables analysed

4. Discussion

The selection of the variables chosen in this study has been made very difficult. For the analysis of all the observable factors, both socio-demographic and clinical, the Chi-square method has been used. In the case of the clinical factors, a cluster or conglomerate type analysis has been carried out to see relationships between the variables. Variables have been left pending so as not to make the study unmanageable, and those considered most important have been chosen.

In the variable of age, many age ranges have been opened (9 ranges) because the authors have considered it very necessary given the high number of deaths that there were with the older people (70 years onwards). The results of the analysis have confirmed that men have a wider range of dangerous ages, from 50 to over 80 years of age. However, this is not the case for women, who are clearly more dangerous than women over 80 years of age, followed by a significant range but to a lesser extent for women between 40 and 59 years of age. This may be due to the fact that men, in general, present a more important clinical pathology picture than women.

A linear function of dependence between all the clinical variables has been found that could be included in a predictive model.

All the parameters studied in this work could be used in future work on prevalence and seroprevalence or in the last phase of the epidemic.

5. Conclusion

This study highlights the influence or not of the variables studied on the speed of spread of Covid-19 coronavirus in Spain. The variables studied have been divided into two groups, socio-demographic variables (sex, age, cohabitation area, population density and social level), and clinical variables (hypertension, cholesterol, allergies, mental health, tobacco and overweight). The environmental variables have been studied in another work resulting in very interesting conclusions in some of them such as temperature and air currents.

The application of the Chi-square technique indicates that of the socio-demographic variables studied all have importance in the propagation of the epidemic with the exception of people's income or economic level.

It is observed that for women, the age groups between 40-49, 50-59 and 80+ and for men between 50 and 80+ are the most important because they have a greater number of infected people in order to carry out studies on the influence of the virus and to determine critical factors in its propagation. These are the ages to take care of or to take into account in the event of a pandemic.

The different areas of coexistence or autonomous communities with high population density show preferences with a higher speed of infection of the virus, although this is not always true. It is important to pay attention to transport as a means of connection between and within the different communities. Transport is a very important element to study and monitor.

As for the clinical factors, they all show a linear functional dependence, so that they can be related through equation (1) for further studies.

All these results refer to deceased or recovered persons depending on the parameter and time period of study related to the speed of infection. These are very important data to take into account in studies of the evolution of pandemics worldwide.

Mitigation of the disease requires strict hygiene measures and containment processes, with de-escalation taking into account health parameters in particular. The study of the great amount of variables will help without a doubt in knowing better the disease to face better prepared in future times. Scientists in these cases must join efforts in one direction, and that direction can only be one, to save humanity.

5.1 Availability of Data

Our data are available upon request from the corresponding autor.

5.2 Conflict of Interest

None of the authors have conflicts of interest associated with this study to report.

5.3 Source of Funding

The authors received no specific funding for this work.

References

- Abrego L., Del Río C. & Hernández S. (1998). *La familia y el enfermo. En: Guía para la atención psicológica de personas que viven con el VIH/SIDA*. 2a. ed. México: Consejo Nacional para la Prevención y Control del SIDA, pp. 52-3.
- Canini L. & Perelson A.S. (2014). Viral kinetic modeling: state of the art. *J. Pharmacokinet. Pharmacodyn*, 41, 431-443. <https://doi.org/10.1007/s10928-014-9363-3>.
- Cember, H. (1985). *Introduction to Health Physics*. New York, Pergamon Press.
- Centers for Disease Control and Prevention (CDC) (2009). Surveillance for pediatric deaths associated with 2009pandemic influenza A (H1N1) virus infection. *United States*, 58(34), 941-947.
- Chan, J.F., Yuan, S., Kok, K.H., To, K.K., Chu, H. & Yang, J. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet*, 395(10223), 514-523.
- Checkoway H, Pearce N.E. & Crawford-Brown D.J. (1989). *Research methods in occupational epidemiology*. Nueva York: Oxford University Press, 117- 119.

- Chen Z., Xu P., Salyards G.W., Harvey S.B., Rada B., Fu F.Z. & He B. (2012). Evaluating a Parainfluenza Virus5-Based Vaccine in a Host with Pre-Existing Immunity against Parainfluenza Virus 5. *PLoS ONE*, 7(11), e50144. <https://doi.org/10.1371/journal.pone.0050144>.
- Clapham H. E., Quyen T.H., Kien D.T., Dorigatti I., Simmons C.P. & Ferguson N.M. (2016). Modelling virus and antibody dynamics during dengue virus infection suggests a role for antibody in virus clearance. *PLoS Comput. Biol.*, 12(e1004951). <https://doi.org/10.1371/journal.pcbi.1004951>.
- Dahari H., Major M., Zhang X., Mihalik K., Rice C.M., Perelson A.S., Feinstone, S.M. & Newmann A.U. (2005). Mathematical modeling of primary hepatitis C infection: noncytolytic clearance and early blockage of virion production. *Gastroenterology*, 128, <https://doi.org/10.1053/j.gastro.2005.01.049>.
- De Wit, E., van Doremalen, N., Falzarano, D. & Munster, V.J. (2016). SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol*, 14, 523–534.
- Dee K.U. & Shuler M.L. (1997). A mathematical model of the trafficking of acid-dependent enveloped viruses: application to the binding, uptake, and nuclear accumulation of baculovirus. *Biotechnol. Bioeng*, 54, 468–490. [https://doi.org/10.1002/\(SICI\)1097-0290\(19970605\)54:5<468::AID-BIT7>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1097-0290(19970605)54:5<468::AID-BIT7>3.0.CO;2-C).
- Diekmann O. & Heesterbeek J.A. (2000). *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Hoboken, New Jersey: John Wiley & Sons.
- Dowell S.F. (2001). Seasonal variation in host susceptibility and cycles of certain infectious diseases. *Emerg Infect Dis*, 7, 369–374.
- Elandt-Johnson R.C. (1975). Definition of rates: some remarks on their use and misuse. *Am J Epidemiol*, 102, 267–271.
- Jean, C., Quach, A. & Semret, M. (2013). Severity and outcome associated with human coronavirus OC43 infections among children. *Pediatr Infect Dis J*, 32, 325–329.
- Khiabani H, Farrell G.M., St George K. & Rabadan R. (2009). Differences in patient age distribution between influenza A subtypes. *F. journal*, 4(8), 6832.
- Konca, G., Korukluoglu, G., Tekin, M., Almis, H., Bucak, I.H. & Uygur, H. (2017). The first infant death associated with human coronavirus NL63 infection. *Pediatr Infect Dis J*. 2017, 36, 231–233.
- Lee, J. & Storch, G.A. (2014). Characterization of human coronavirus OC43 and human coronavirus NL63 infections among hospitalized children < 5 years of age. *Pediatr Infect Dis J*. 2014, 33, 814–820.
- Lofgren E., Fefferman N.H., Naumov Y.N., Gorski J. & Naumova E.N. (2007). Influenza seasonality: underlying causes and modeling theories. *J Virol*, 81, 5429–5436.
- Lourraine S. (1998). *Pérdida y el virus de la inmunodeficiencia humana. Agonía, muerte y duelo. 2a. reimpresión*. México: Ed. Manual Moderno, 181–202.
- Lowen A.C., Mubareka S., Steel J. & Palese P. (2007). Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog*, 3, e151.
- McKinney K.R., Gong Y.Y. & Lewis T.G. (2006). Environmental transmission of SARS at Amoy Gardens. *J Environ Health*, 68, 26–30.
- Montesinos-López O.A. & Hernández-Suarez C.M. (2007). Modelos matemáticos para enfermedades infecciosas. *Salud Publica Mex*, 49, 21826.
- Murray E.J. & Morse S.S. (2011). Seasonal Oscillation of Human Infection with Influenza A/H5N1 in Egypt and Indonesia. *PloS one*, 6, e24042.
- Pica, N. & Bouvier, N.N. (2011). *Environmental factors affecting the transmission of respiratory viruses*. <https://doi.org/10.1016/j.coviro.2011.12.003>.
- Sadique M.Z., Edmunds W.J., Smith R.D., Meerding W.J., de Zwart O., Brug J. & Philippe B. (2007). Precautionary behavior in response to perceived threat of pandemic influenza. *Emerg Infect Dis*, 13, 1307–1313.
- Sanglier, G., Robas, M. & Jiménez, P.A. (2020). Application of Forecasting Methods to a Mathematical Models for Adjusting the Determination of the number of people infected by Covid-19 in the Community of Madrid. *Contemporary Engineering Sciences*, 13(1), 89–101. HIKARI Ltd. <https://doi.org/10.12988/ces.2020.91453>.

- Sanglier, G., Robas, M. & Jiménez, P.A. (2020). Estimation of the Epidemiological Evolution Through a Modelling Analysis of the Covid-19 Outbreak. *Acta Scientific Microbiology*, 13(4), 152-158.
- Shaman J. & Kohn M. (2009). *Absolute humidity modulates influenza survival, transmission, and seasonality*. PNAS March 3, 2009 106 (9) 3243-3248; Edited by Burton H. Singer, Princeton University, Princeton, NJ., <https://doi.org/10.1073/pnas.0806852106>.
- Shen, K., Yang, Y; Wang, T., Zhao, D. & Jiang, R. (2020). *Diagnosis, treatment, and prevention of 2019 novel coronavirus infection in children: experts' consensus statement*.
- World J Pediatr. van Regenmortel MHV.(2000). *Virus taxonomy*. San Diego: Academic Press. 1162 p.
- Velasco Hernández JX. (2007). *Modelos matemáticos en epidemiología: enfoques y alcances*. Miscelánea Matemática, pp. 44-1127.
- Yu I.T., Li Y., Wong T.W., Tam W., Chan A.T., Lee J.H., Leung D.Y. & Ho T. (2004). Evidence of airborne transmission of the severe acute respiratory syndrome virus. *N Engl J Med*, 350,1731–1739.
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B. & Song. J. (2019). *Coronavirus Investigating, and Research Team A novel coronavirus from patients with pneumonia in China 2019*. N Engl J Med.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).