# 10-Months Study Regarding the COVID-19 Spreading in Romanian Counties

Cristina Gena DASCALU[a,1], Magda Ecaterina ANTOHE[b,2] and Victor Lorin PURCAREA[c]

[a] *Medical Informatics and Biostatistics Dept., Faculty of Medicine, "Grigore T. Popa" University of Medicine and Pharmacy, Iasi, Romania*
[b] *"Grigore T. Popa" University of Medicine and Pharmacy, Iasi, Romania*
[c] *"Carol Davila" University of Medicine and Pharmacy, Bucharest, Romania*

**Abstract.** The paper analyzes the evolution of COVID-19 cases in Romanian counties over a period of 10 months, to highlight possible similarities that may contribute to a better understanding of the spreading pattern. The study uses the numbers of active cases for each county in Romania, as well as Bucharest and the whole country, reported daily by the Romanian Ministry of Health (https://datelazi.ro) between April 2nd, 2020 and January 25th, 2021. We compared the disease's evolution in Suceava county (the first outbreak of spread) with other counties in Romania in order to highlight the gaps between them. We calculated the cross-correlations between counties, interpreted as time series. The recorded lags varied between 1 – 15 days, the most counties having a lag of 6-7 days compared with Suceava. Therefore, on long term there are no important discrepancies between the regions in Romania regarding the evolution of the disease, which shows that the intervention efforts of the medical staff were uniform in efficiency. The existence of a lag of only one day between Suceava and the whole country shows that on long term, even in this county the situation is not very discrepant, belonging to the general evolution.

**Keywords.** COVID-19, time series, cross-correlation

## 1. Introduction

The COVID-19 / SARS-CoV-2 will be, probably (and hopefully), the biggest pandemic of the 21st century, similar in magnitude to the Spanish flu pandemic from the early twentieth century. Also characterized as "the most marketed pandemic in Earth history", it appeared at a time when the circulation of information in society reached unprecedented limits of speed, quantity and freedom. Thus, any news about the disease becomes public almost instantly, and anyone who is interested can immediately find the information they need. This is a great advantage in the fight against the pandemic, but it

---

[1] Corresponding Author, Cristina Gena Dascalu, Medical Informatics and Biostatistics Dept., Faculty of Medicine, "Grigore T. Popa" University of Medicine and Pharmacy, 16 Universitatii Str., Iasi, Romania; E-mail: cristina.dascalu@umfiasi.ro.

[2] Corresponding Author, Magda Ecaterina Antohe, Removable Prostheses, Oral Implantology and Dental Technique, Faculty of Dental Medicine, "Grigore T. Popa" University of Medicine and Pharmacy, 16 Universitatii Str., Iasi, Romania; E-mail: magda.antohe@umfiasi.ro.

can also have an adverse effect, because fake news and conspiracy theories are spreading just as quickly.

Our paper aims to analyze the evolution of illness cases in Romanian counties over a period of 10 months, in order to highlight possible similarities that may contribute to a better understanding of the spreading pattern (if any).

## 2. Method

The data used in our study are the official data reported daily by the Romanian Ministry of Health [1] between April 2nd, 2020 – the first day when the site was open and the Ministry started to publish such data, and January 25th, 2021. We recorded the daily number of active cases for each county in Romania, as well as Bucharest and the whole country and we calculated the daily number of new cases; in order to make comparisons between counties, we divided these values at the total number of inhabitants per counties, as it was reported by INS (Romanian National Institute of Statistics, [2]) and we expressed them as ratios at 100.000 habitants (in order to gain relevancy and consistency - otherwise, the raw values were very small in magnitude and difficult to interpret).

In Romania, at the beginning of the pandemic there was a major outbreak in Suceava county; as the disease set in and spread, this outbreak subsided, but other outbreaks appeared in all regions of the country (Brasov, Bucharest, Constanta, Iasi) - which first exploded and then attenuated in time. Now there is a significant decrease in cases, probably caused by the beginning of the vaccination campaign, but the future is still uncertain. In this context, our study aims to compare the evolution of the number of daily cases in Suceava county with other counties in the country and to highlight the gaps between them, because these gaps can have specific causes (geographical or socio-economic).

In order to identify such gaps, we used the cross-correlations analysis of time series. The cross-correlation is a measurement that tracks the movements of two sets of data relative to each other [3, 4, 5]. It is based on the concept of correlation [6], its range of values is -1 to 1 and can be used to measure information between two time series [7]; it measures the correlation coefficient between time series by shifting them for different lags and identifying the lag with the highest correlation, where the two series match the best [8] ; the lag characterizes the spread from one county to another, because it represents the elapsed time from the registration of a certain number of cases in one county to the registration of a similar number of cases in another county.

In the theory of signal processing, the cross-correlation between two signals $u(t)$ and $v(t)$ is defined as $w(t) = u(t) \otimes v(t) = \int_{-\infty}^{\infty} u^*(\tau - t)v(\tau)d\tau$ , where $u^*(\tau)$ is the complex conjugate of $u(\tau)$ (it makes no difference if $u(\tau)$ is real-valued [9]). In this regard, the concept of cross-correlation is closed to the concept of convolution between signals [10], taking in consideration that:

$$u(t) \otimes v(t) = \int_{-\infty}^{\infty} u^*(\tau)v(\tau + t)d\tau \text{ (cross-correlation) and}$$

$$u(t) * v(t) = \int_{-\infty}^{\infty} u(\tau)v(-\tau)d\tau \text{ (convolution)}.$$

The analysis was conducted in SPSS 27.0. We run the analysis without natural log transformation, as well as without assuming seasonally differences (the distribution of cases did not show the existence of any seasonal trend to be take into account); we calculated the matrix of cross-correlation and we represented the CCF diagram.

## 3. Experimental Results

The highest calculated correlation coefficients and the corresponding lags between Suceava and all the other counties in Romania, including also Bucharest (the capital) and the whole country are presented in Table 1. We also represented on a chart the distribution of recorded lags on the Romania's map (fig. 1), as well as the recorded correlation coefficients, in descending order (fig. 2).

**Table 1.** The identified lags and correlation coefficients between Suceava (757679 inhabitants) and the other counties of Romania

| County (no. of inhabitants - 2019) | Lag | Correlation coefficient | County (no. of inhabitants - 2019) | Lag | Correlation coefficient |
|---|---|---|---|---|---|
| Alba (376589) | 7 | 0.694 | Hunedoara (459671) | 7 | 0.631 |
| Arad (471155) | 7 | 0.664 | Ialomita (288379) | 1 | 0.516 |
| Arges (637880) | 5 | 0.425 | Iasi (944074) | 7 | 0.684 |
| Bacau (742053) | 6 | 0.672 | Ilfov (429946) | 6 | 0.578 |
| Bihor (617118) | 7 | 0.629 | Maramures (522154) | 1 | 0.680 |
| Bistrita-Nasaud (328286) | 7 | 0.685 | Mehedinti (280888) | 7 | 0.529 |
| Botosani (452328) | 7 | 0.688 | Mures (593024) | 6 | 0.683 |
| Brasov (634236) | 6 | 0.572 | Neamt (569851) | 14 | 0.663 |
| Braila (346773) | 6 | 0.485 | Olt (438318) | 7 | 0.550 |
| Buzau (468110) | 6 | 0.527 | Prahova (795931) | 6 | 0.521 |
| Caras-Severin (320124) | 7 | 0.705 | Satu Mare (387979) | 6 | 0.583 |
| Calarasi (311084) | 4 | 0.523 | Salaj (245088) | 6 | 0.673 |
| Cluj (730216) | 7 | 0.711 | Sibiu (466905) | 6 | 0.626 |
| Constanta (766315) | 1 | 0.552 | Teleorman (374887) | 7 | 0.642 |
| Covasna (226879) | 7 | 0.605 | Timis (752091) | 1 | 0.712 |
| Dambovita (522195) | 6 | 0.597 | Tulcea (238333) | 15 | 0.570 |
| Dolj (691276) | 6 | 0.669 | Vaslui (493234) | 1 | 0.564 |
| Galati (626201) | 7 | 0.501 | Valcea (397878) | 1 | 0.645 |
| Giurgiu (272768) | 7 | 0.503 | Vrancea (386223) | 8 | 0.529 |
| Gorj (359883) | 7 | 0.564 | BUCURESTI (2121794) | 1 | 0.607 |
| Harghita (331809) | 14 | 0.676 | ROMANIA (22177605) | 1 | 0.695 |

The recorded lags varied between 1 and 15 days (fig. 1). The 1-day lag was recorded between Suceava and Maramures, which is located in its immediate neighborhood, as well as between Suceava and Bucuresti, Timis and Constanta; Bucuresti is the capital of the country and the biggest urban agglomeration, and Timis and Constanta contain big cities (Constanta being also port at the Black Sea, and therefore an area with heavy human traffic). The other counties with 1-day lag were Ialomita, Vaslui and Valcea – which are generally rural areas, so there is no specific reason to facilitate the disease's spread. Among these counties, anyway, the highest cross-correlation coefficients were recorded for Timis (0.712), followed by Maramures (0.680), Valcea (0.645) and Bucuresti (0.607) – the only situation without obvious explanation being represented by Valcea county.

The biggest lags were identified in only 3 counties, for which we also can find good natural reasons: Harghita and Neamt (14 days), as well as Tulcea (15 days) – these are also counties without big cities, located in rural areas, rather isolated: Harghita and Neamt are located in the Carpathian Mountains, and Tulcea contains the Danube Delta. Therefore the people generally live here in small communities, the spread of an epidemic being hampered. Harghita and Neamt recorded high correlation coefficients (0.676 and, respectively, 0.663), which shows that, although later, they were still significantly affected by the spread of the disease, while Tulcea is less correlated (0.570) – fig. 2. The

figures actually show that it has remained so far one of the counties in the country least affected by the pandemic.



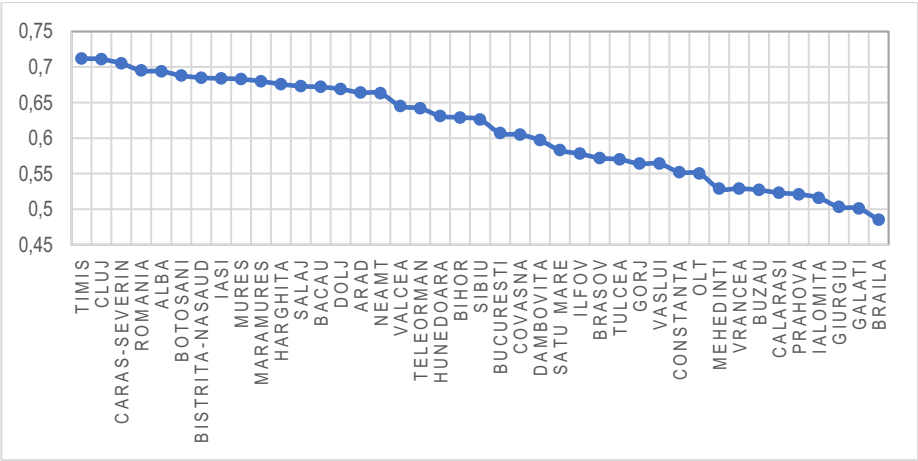**Figure 1.** The distribution of recorded lags in Romania's counties.



**Figure 2.** The recorded cross-correlation coefficients in descending order.

All the other counties of Romania recorded a lag compared to Suceava of 6 or 7 days, confirmed also by the raw figures. Indeed, the Suceava was the first important outbreak, but the pandemic, once in the country, spread rapidly to all regions. The most vulnerable were the big cities – the counties containing them recording high correlation coefficients (like Cluj – 0.711, Alba – 0.694, Iasi – 0.684, Mures – 0.683, Bacau – 0.672, Arad 0 .664) – fig. 2, as well as the Suceava's neighborhood (Botosani - 0.688 and Bistrita Nasaud – 0.685). The rural areas and the small cities, eventually located at a great distance from

Suceava or other important cities, were less correlated (Braila – 0.485, Giurgiu – 0.503, Prahova – 0.521, Calarasi – 0.523, Buzau – 0.527, Vrancea – 0. 529, Mehedinti – 0.529, Olt – 0.550) – fig. 2. A special situation is represented by Galati, which was also less correlated with Suceava – 0.501, even if it contains a big city with important human traffic (being port at the Danube river).

Romania as a whole recorded a lag of only 1 day compared with Suceava, being highly correlated with this county (0.695). Indeed, the disease has spread rapidly in all regions, and, although in March - May 2020 the number of registered cases was relatively low, except for those in Suceava, since June the cases have increased significantly, the outbreak in Suceava has attenuated, but new outbreaks have appeared (correlated with the holiday period, when people circulated more on the territory of the country, especially at sea and in the mountains).

## 4. Conclusions

The study reveals that Suceava county, which was the first outbreak of COVID-19 infection in Romania, is strongly correlated with the other counties in Romania and the entire country, at a lag which varies between 1 – 15 days, the most counties having a lag of 6-7 days compared with Suceava. Therefore, on long term there are no important discrepancies between the regions in Romania regarding the evolution of the disease, which shows that the intervention efforts of the medical staff were uniform in efficiency, not being influenced by the local socio-economic conditions. Also, the existence of a lag of only one day between Suceava and the whole country shows that on long term, even in this county the situation is not very discrepant, belonging to the general evolution.

## References

[1]   Code for Romania. COVID-19. Date la zi. Available at: https://datelazi.ro, Accessed 11 March 2021.
[2]   Institutul National de Statistica. Romania in cifre – breviar statistic. Available at https://insse.ro/cms/ro/tags/romania-cifre, Accessed 11 March 2021.
[3]   Holmes EE, Scheuerell MD, Ward EJ, Correlation within and among time series. In: Applied Time Series Analysis for Fisheries and Environmental Sciences. Available at: https://nwfsc-timeseries.github.io/atsa-labs/sec-tslab-correlation-within-and-among-time-series.html, Accessed 11 March 2021.
[4]   Zebende GF, da Silva PA, Machado FA. Study of cross-correlation in a self-affine time series of taxi accidents. Physica A. 2011;390:1677-83.
[5]   Machado FA, da Silva PA, Zebende GF. Autocorrelation and cross-correlation in time series of homicide and attempted homicide. Physica A. 2014; 400:12-19.
[6]   Riffenburgh RH, Gillen DL. Linear regression and correlation. In: Riffenburgh RH, Gillen DL. Statistics in Medicine, 4th Edition; London, UK: Elsevier Ltd., Academic Press; 2020; p. 357-90.
[7]   Shiavi R. Theory and application of cross correlation and coherence. In: Shiavi R. Introduction to Applied Statistical Signal Analysis, 3rd Edition; 2007; London, UK: Elsevier Ltd., Academic Press; p. 331-66.
[8]   Rehfeld K, Marwan N, Heitzig J, Kurths J. Comparison of correlation analysis techniques for irregularly sampled time series. Nonlin Processes Geophys. 2011;18:389–404.
[9]   Derrick TR, Thomas JM. Time Series Analysis: The Cross-Correlation Function. In: Stergiou N, editor. Innovative analyses of Human Movement, 1st Edition; 2004; Champaign, IL: Human Kinetics; p. 189-205.
[10]  Van Drongelen W. LTI Systems: Convolution, Correlation, Coherence, and the Hilbert Transform. In: Van Drongelen W. Signal Processing for Neuroscientists, 2nd Edition; 2018; London, UK: Elsevier Ltd., Academic Press; p. 251-88.