Linear regression for sars-cov-2 coronavirus in wastewater and covid-19 dynamics infectious at baix llobregat, Spain

Nicolas Ayala-Aldana

nayalaaldana@gmail.com

https://orcid.org/0000-0002-3530-6734

University of Barcelona, Barcelona – Spain

Antonio Monleon-Getino

amonleong@ub.edu

https://orcid.org/0000-0001-8214-3205

University of Barcelona, Barcelona – Spain

Jaume Canela-Soler

jcanela@ub.edu

https://orcid.org/0000-0002-3280-6899

Hospital Clinic, University of Barcelona, Barcelona –Spain

ABSTRACT

Detection of SARS-CoV-2 RNA in wastewater is helpful to identify the presence of COVID-19 in the community. This method provides additional information, cheap and indicative of the COVID-19 contagion. The current research provides information about the Baix Llobregat case in Catalonia, Spain. Methods: This research used an open dataset from "Generalitat of Catalonia" for the Baix Llobregat. The time series of COVID-19 dynamics and COVID-19 genes in wastewater were analysed for 2020-2022. Simpler and multiple linear regression was performed for Genes N1 and N2 in wastewater and dynamics COVID-19 variables. Hypothesis analyses use a p-value<0.05 for statistics tests. Results: Linear regression between N1 and N2 COVID-19 genes shows a high correlation for 2020 and 2021. The best corresponding variable for the N1 gene was the cumulative incidence and the best associative variable for the N2 gene was %PCR-RAT positive. In multiple linear regression, the model acceptable results when considering RNA SARS-CoV-2 and the highest epidemiologic indicators with significant values (p<0.05). Discussion: COVID-19 in water waste could be useful to determine COVID-19 dynamics in the community. In this study, Cumulative Incidence and PCR-RAT% positive showed high performance in linear regression. The graphical results admit similar trends with COVID-19 genes in water waste and epidemiologic rates for time series.

Keywords: covid-19; pandemic dynamic; time series; environment epidemiology.

Regresión lineal para sars-cov-2 en aguas residuales y la dinámica infecciosa de covid-19 en el baix llobregat, España

RESUMEN

La detección del ARN del SARS-CoV-2 en las aguas residuales es útil para identificar la presencia del COVID-19 en la comunidad. Este método proporciona información adicional, barata e indicativa del contagio de COVID-19. La presente investigación estudia el caso del Baix Llobregat en Cataluña, España. Métodos: Esta investigación utilizó un conjunto de datos abiertos de la "Generalitat de Cataluña" para el Baix Llobregat. Se analizaron las series temporales de la dinámica de COVID-19 y de los genes de COVID-19 en las aguas residuales para 2020-2022. Se realizó una regresión lineal simple y múltiple para las variables Genes N1 y N2 en aguas residuales y los indicadores epidemiológicos de COVID-19. Se utilizó un valor p<0,05 para los análisis estadísticos. Resultados: La regresión lineal entre los genes N1 y N2 de COVID-19 muestra una alta correlación para 2020 y 2021. La variable con mejor correlación para el gen N1 fue la incidencia acumulada y la mejor variable asociativa para el gen N2 fue el %PCR-RAT positivo. En la regresión lineal múltiple, el modelo resulta aceptable al considerar el ARN SARS-CoV-2 y los indicadores epidemiológicos más altos con valores significativos (p<0,05). Discusión: La presencia de COVID-19 en aguas residuales podría ser útil para determinar la dinámica de COVID-19 en la comunidad. En este estudio, la incidencia acumulada y el PCR-RAT% positivo mostraron un alto rendimiento en la regresión lineal. Los resultados gráficos revelan tendencias similares con los genes de COVID-19 en los residuos del agua y las tasas epidemiológicas para las series temporales.

Palabras Claves: covid-19; dinámica de pandemia; series temporales; epidemiología ambiental.

INTRODUCTION

SARS-CoV-2 is a new type of coronavirus (a broad family of viruses that normally affect only animals) that can affect people and causes COVID-19. It was detected for the first time in December 2019 in Wuhan (China). Coronaviruses produce clinical conditions ranging from the common cold to more serious diseases (Hu et al., 2021; Zhou et al., 2020). SARS-CoV-2 is the virus of COVID-19 disease, and it is highly infectious. It can be transmitted through person-to-person contact and through direct contact with respiratory droplets generated when an infected person coughs. Due to high contagiousness, the disease has affected multiple countries on all continents.

According to the stool study, SARS-CoV-2 can replicate for 11 days in the gastrointestinal tract of patients, even after nasopharyngeal samples become negative. Regarding the presence and persistence of SARS-CoV-2 in wastewater, there is evidence that wastewater may contain RNA fragments and viable SARS-CoV-2 particles. Furthermore, viral RNA in stool is commonly observed in symptomatic and non-symptomatic patients (Chen et al., 2020) Currently, some studies have detected SARS-CoV-2 RNA in wastewater around the world and wastewater testing is an early warning tool to monitor the status and trend of COVID-19 infection (Gonzalez et al., 2020; McMahan et al., 2021). In addition, this tool is useful for evaluating public health response.

This study aims to determine predictive variables performance of the dynamics of COVID-19 using the concentration of viral genes of COVID-19 in waterwaste.

METODOLOGÍA

Area of Study

For this cross-sectional study, we studied the relation between SARS-CoV-2 in wastewater and COVID-19 contagious at Baix Llobregat in Spain. Baix Llobregat (figure 1) is a council of Catalonia, its population is 806,249 and its area is 241,8 (inhabitants/Km^2). Besides, Sant Feliu de Llobregat is the capital council of Baix Llobregat. This city has a population of 45,463 inhabitants and the wastewater plant treatment of this study is in the mentioned city. The SARS-CoV-2 N1 and N2 genes were measured in this wastewater plant once a week and the data were aggregated with the weekly COVID-19 indicators of Baix Llobregat.

In this context, we aimed to determine the relation between infectious epidemiological variables and N1 and N2 genes (CG/L) in waterwaste. The starting point is to extrapolate

the behaviour of the region Baix Llobregat from the measurement of the plant located in Sant Feliu de Llobregat. There is no other waterwaste plant treatment reported in this council.

COVID-19 time series was analysed from 06 July 2020 to 20 June 2022. In addition, the dataset included weekly data of COVID-19 contagious. Database was obtained from "Dades Covid" (Dades COVID, Generalitat de Catalunya.). Furthermore, genes concentrations were obtained as aggregated data from "Open Data" of "Generalitat de Catalunya". The variables of interest, in this database, are the concentration of N1 and N2 genes of COVID-19 in waterwaste (CG/L). The measure methodology is aimed at detecting the virus nucleocapsid by RT-qPCR once a week (Ellis et al., 2021).

Figure 1. Map of Baix Llobregat, Catalonia, Spain.

Note: The map illustrates the location of the Baix Llobregat in the Catalonia area, Spain. The yellow-filled area is Sant Feliu de Llobregat, the capital of the council.

Variables and Statistical Methods.

COVID-19 dynamics include the following epidemiological variables:

- Cumulative incidence: epidemiological variable computed as new cases over free-disease population for 14 days.
- Ro: Basic reproduction number (Ro).
- Confirmed cases: Confirmed cases for COVID-19 by PCR or rapid antigen test.
- RAT cases: COVID-19 new cases confirmed by rapid antigen test (RAT) in serological samples of patients.
- %PCR-RAT positives: number of positive cases by PCR test or RAT over the total people tested.
- Total Hospitalized: Hospitalized patients in care centres.

- Total Critical Healthcare: patients in intensive care within healthcare centres.

- Recovered Patients: People who do not have symptoms of COVID-19 and have overcome the disease.

The linear regression was used for the prediction of Genes N1 and N2 using independent variables of COVID-19 contagion in Baix Llobregat. Then, each highest association variable was analysed by a multiple linear regression model. The significance level considered was p-value <0.05 for both linear regressions. In addition, the SARS-CoV-2 load in wastewater was graphically compared with surveillance indicators of infection. Statistical data was performed in RStudio (v 4.0.2).

RESULTS AND DISCUSSION

Results.

Simpler linear regression was performed to study the relation of N1 and N2 COVID-19 genes in wastewater. Linear regression between N1 and N2 COVID-19 genes shows a high correlation for 2020 (R^2: 0.87) and 2021 (R^2: 0.95). Both variables reveal acceptance of the alternative hypothesis with a p-value <0.05. However, the linear regression for 2022 had no significant p-values. Perhaps, the high COVID-19 cases and the high weekly variance in 2022 affected the linearity between the N1 and N2 genes.

Figure 1. Simpler linear regression between N1 and N2 COVID-19 genes in water treatment plant for 2020-2022.
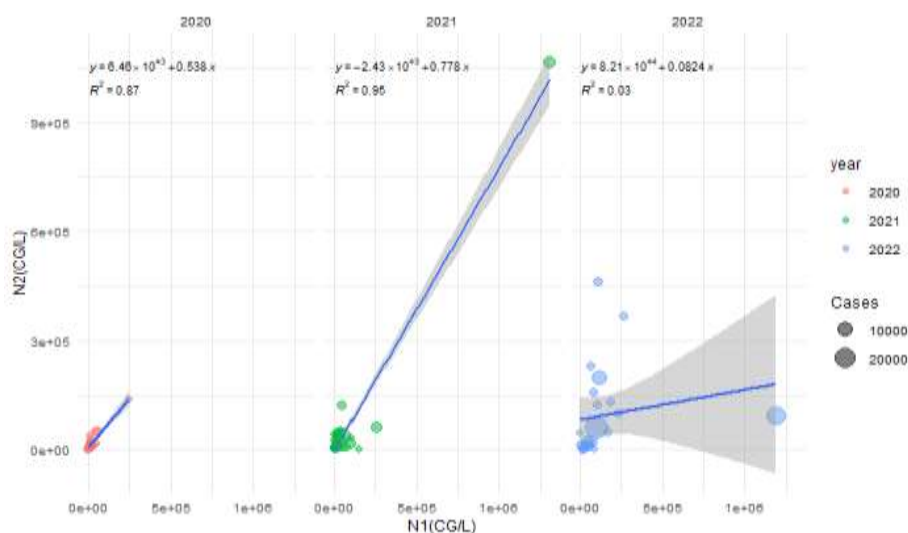

Fig1. Simpler linear regression was performed for N1 and N2 genes in wastewater (blue line). The analysis was segmented by year to provide insights into the relationship between the variables. The different colours show the time series for 2020 (red), 2021 (green) and 2022 (blue). The size of circles is the number of COVID-19 cases.

Table 1 shows the output of simpler linear regression between concentration of N1 Genes in water treatment plant and epidemic indicators of COVID-19 in Baix Llobregat. Simpler Linear Regression reveals statistical significance for 7 of 9 variables studied for β_1. Thus, cumulative incidence (R^2: 0.1804), confirmed cases (R^2:0.1387) and %PCR-RAT Positive (R^2:0.1393) have the most values for R-Squared.

Table 1. Prediction of N1 genes of COVID-19 in water treatment plant of Baix Llobregat in period 2020-2022.

Variable	R^2	β_0	β_1
cumulative incidence (14 days)	0.1804	25387.97 (p-value:0.19)	69.00 (p-value: 1.3e-05)
R_O	0.002011	91749 (p-value:0.104)	-19674(p-value: 0.661)
Confirmed cases	0.1387	30964.880 (p-value: 0.119272)	14.204 (p-value:0.000159)
PCR Cases	0.1188	-7716.72 (p-value:0.779194)	111.93 (p-value:0.000511)
TAR Cases	0.1249	40355.543 (p-value:0.035848)	14.352 (p-value:0.000357)
%PCR-TAR Positives	0.1393	-23326 (p-value:0.422898)	9274 (p-value:0.000154)
Total Hospitalized	0.1039	-19901.1 (p-value: 0.53307)	1099.0 (p-value:0.00121)
Total Critical Healthcare	0.01476	34392 (pp-value:0.313)	3673 (p-value:0.233)
Recovered patients	0.001818	60522.0 (p-value: 0.0258)	601.1 (p-value:0.6768)

Table 2 illustrate the output of simpler linear regression between concentration of N2 Genes in water treatment plant and epidemic indicators of COVID-19 at the Baix Llobregat. Simpler Linear Regression reveals statistical significance for 3 of 10 variables studied for N2 gen in wastewater. In this case, %PCR-RAT Positive (R^2:0.2553), PCR Cases (R^2:0.08098) and Total Hospitalized (R^2:0.05281) have the most values for R-Squared in simpler linear regression.

Table 2. Prediction of N2 genes of COVID-19 in water treatment plant of Baix Llobregat in period 2020-2022.

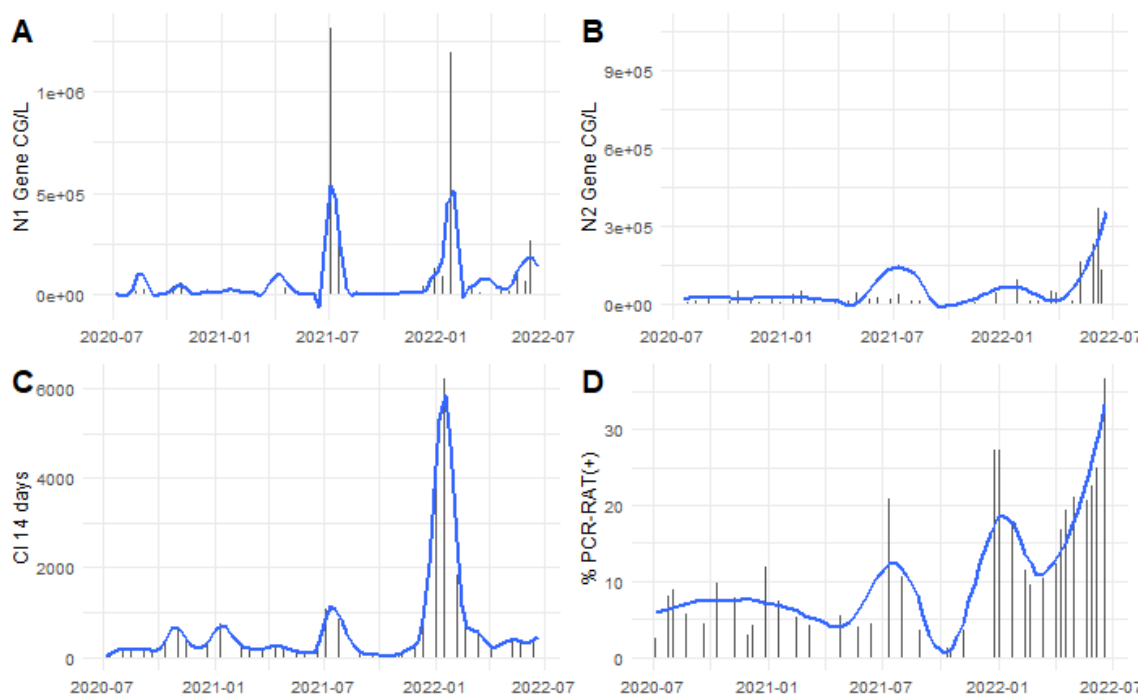| Variable	R^2	β_0	β_1
Cumulative incidence (14 days)	0.02429	37083.53 (p-value: 0.0162)	21.61 (p-value: 0.1358)
Ro	0.002991	27643 (p-value: 0.509)	17945(p-value: 0.603)
Confirmed cases	0.03133	35103.277 (p-value: 0.0234)	6.204 (p-value: 0.0896)
Confirmed case rate	0.03133	35103.28 (p-value: 0.0234	51.90 (p-value: 0.0896)
PCR Cases	0.08098	4175.88 (p-value: 0.8364)	67.17 (p-value: 0.0057)
TAR Cases	0.02068	40392.433 (p-value: 0.00619	5.379 (p-value: 0.1691)
%PCR-RAT Positives	0.2553	-40402 (p-value: 0.0423)	9311 (p-value: 2.39e-07)
Total Hospitalized	0.05281	3990.7 (p-value: 0.8659)	567.7 (p-value: 0.0267)
Total Critical Healthcare	0.001602	40440.1 (p-value: 0.103)	852.3 (p-value: 0.703)
Recovered patients	-0.008961	54118.9 (p-value: 0.00553)	-438.3 (p-value: 0.66987)

Multiple linear regression between the concentration of N1 Genes in the water treatment plant and epidemic indicators of COVID-19 was performed using the highest indicators performance for COVID-19 in the computed simpler linear regression (Table 2). The indicators selected in the model have the major performance for the prediction of N1 gene concentration. Approximately 64% of the independent variables could explain the trend of the dependent variable. For the case of N2 genes, approximately 58% of the independent variables could explain the trend of the dependent variable.

Table 3. Relation between N1 genes of COVID-19 in wastewater and epidemiological indicators by multiple linear regression.

Variables	β_1	Standard Error	P-value
N1 (CG/L) in waterwaste			
N2(CG/L)	9.574e-01	9.273e-02	< 2e-16
Cumulative incidence (14 days)	7.473e+01	1.286e+01	9.21e-08
Adjusted R-squared	0.6364		
N2 (CG/L) in waterwaste			
N1 (CG/L)	4.198e-01	5.044e-02	8.81e-13
% PCR-TAR Positives	5.179e+03	1.354e+03	0.000241
Adjusted R-squared	0.5792		

The time series of N1 and N2 genes concentration in wastewater are shown in figure 2. The Peak in cumulative cases and %PCR-RAT positive in July 2021 and January 2022 are related to an increase in N1 and N2 gene concentration.

Fig 2. Gene virus concentration and epidemiology indicators for COVID-19 in time series 2020-2022 at the Baix Llobregat.

Note: All graphs show the time series of COVID-19 dynamics and genes in water waste for the Baix Llobregat in 2020-2022. The epidemiological indicators computed with better performance in linear regression (table 1 and 2) were selected for time series . The data is represented as weekly values. **A:** Gene N1 (CG/L). **B:** Gene N2 (CG/L). **C:** Cumulate Incidence in 14 days. **D:** %PCR-RAT positive. Blue smooth line show LOESS trend in time series.

DISCUSSIONS.

The COVID-19 pandemic has caused a series of challenges worldwide to control its spread. A matter of interest has been the research for predictors or molecules in the environment that allow predicting increases in cases. Several studies have reported the detection of SARS-CoV-2 RNA in wastewater with increases in local incidence (Bivins et al., 2020; Larsen & Wigginton, 2020; Zhang et al., 2020). Regarding the models, some investigations concluded that linear correlation explains the relationship between the

concentration of genes in wastewater and the incidence of COVID-19 in determined cases. (Petala et al., 2022). Other studies have shown a similar or even higher correlation between new cases and concentrations of COVID-19 genes in wastewater both in Spain and other European cities (Koureas et al., 2021; Vallejo et al., 2022). The positive correlation between genes N1 and N2 in wastewater provides important information as a positive control mechanism. Furthermore, serious consideration needs to be given to biological viability because factors such as distance travelled, time in the sewer network, the composition of wastewater, collection methods and laboratory measurement affect RNA decomposition. Another factor to consider is that both symptomatic and non-symptomatic infected could excrete COVID-19 viral load. These factors could increase or decrease the concentration of viral RNA in wastewater. To sum up, RNA measurement in waterwaste offers a moderate or high relationship with incidence but it is difficult to generalize methods to understand the exact number of infected people from the concentration of RNA in waterwaste.

In our research, the Ro number, critical healthcare and recovered patients showed low relation with genes N1-N2 in wastewater. The proposed models are only applicable to the Baix Llobregat area, although they could be adapted to other regions. The wastewater data would be convenient to verify the reliability of epidemiological trends obtained from diagnosed cases. Perhaps, the measurement of COVID-19 in wastewater at institutions such as hospitals, schools or residences will provide interesting information on the local epidemiological situation.

Much more research is needed to understand how SARS-CoV-2 detected in wastewater is related to COVID-19 transmission. In our case, we cannot rule the concentration of genes in wastewater has a proportional ratio with the number of COVID-19 cases.

Limitations

The current research was limited to a single council using just one wastewater plan treatment. Expanding the study areas can provide valuable information and it would generalize the hypothesis of the study. In addition, this study does not consider asymptomatic cases that could increase the viral load in wastewater. Therefore, this would be a measurement bias of the epidemiological variables for instance cumulative incidence, PCR cases or critical healthcare patients.

External climatic conditions, such as humidity, temperature, and environmental pollution, could be studied to assess their impact on the concentration of the virus in wastewater. Regarding the characteristics of wastewater, other factors such as the length of the watercourse, salts in water or other biological forms affect the genetic availability of the virus.

CONCLUSIONS

Our results reveal that the measurement of capsid protein RNA material in wastewater and classical epidemiological indicators of COVID-19 are related in a linear regression model. The detection of the N1 gene is highly related to the N2 gene in wastewater. In combination, the two COVID-19 RNA metrics provided complementary insights into epidemiological dynamics.

Finally, in this study, the major predictors to explain the concentration of COVID-19 in wastewater were the cumulative incidence of fourteen days and the positivity of the PCR-RAT test. Other epidemiological variables such as the number of critical healthcare, Ro and recovered patients revealed a low correlation with COVID-19 RNA in wastewater.

REFERENCES.

Bivins, A., North, D., Ahmad, A., Ahmed, W., Alm, E., Been, F., Bhattacharya, P., Bijlsma, L., Boehm, A. B., Brown, J., Buttiglieri, G., Calabro, V., Carducci, A., Castiglioni, S., Cetecioglu Gurol, Z., Chakraborty, S., Costa, F., Curcio, S., De Los Reyes, F. L., … Bibby, K. (2020). Wastewater-Based Epidemiology: Global Collaborative to Maximize Contributions in the Fight against COVID-19. *Environmental Science and Technology*, *54*(13), 7754–7757. https://doi.org/10.1021/ACS.EST.0C02388/ASSET/IMAGES/LARGE/ES0C02388_0 001.JPEG

Chen, Y., Chen, L., Deng, Q., Zhang, G., Wu, K., Ni, L., Yang, Y., Liu, B., Wang, W., Wei, C., Yang, J., Ye, G., & Cheng, Z. (2020). The presence of SARS-CoV-2 RNA in the feces of COVID-19 patients. *Journal of Medical Virology*, *92*(7), 833–840. https://doi.org/10.1002/JMV.25825

Dades COVID. (n.d.). Retrieved July 18, 2022, from https://dadescovid.cat/

Ellis, P., Somogyvári, F., Virok, D. P., Noseda, M., & McLean, G. R. (2021). Decoding Covid-19 with the SARS-CoV-2 Genome. *Current Genetic Medicine Reports 2021 9:1*, *9*(1), 1–12. https://doi.org/10.1007/S40142-020-00197-5

Gonzalez, R., Curtis, K., Bivins, A., Bibby, K., Weir, M. H., Yetka, K., Thompson, H., Keeling, D., Mitchell, J., & Gonzalez, D. (2020). COVID-19 surveillance in Southeastern Virginia using wastewater-based epidemiology. *Water Research*, *186*, 116296. https://doi.org/10.1016/J.WATRES.2020.116296

Hu, B., Guo, H., Zhou, P., & Shi, Z.-L. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology*, *19*(3), 141–154. https://doi.org/10.1038/s41579-020-00459-7

Idescat. El municipio en cifras. Martorell (Baix Llobregat). (n.d.). Retrieved July 15, 2022, from https://www.idescat.cat/emex/?id=081141&lang=es

Idescat. The municipality in figures. Sant Feliu de Llobregat (Baix Llobregat). (n.d.). Retrieved July 15, 2022, from https://www.idescat.cat/emex/?id=082114&lang=en

Koureas, M., Amoutzias, G. D., Vontas, A., Kyritsi, M., Pinaka, O., Papakonstantinou, A., Dadouli, K., Hatzinikou, M., Koutsolioutsou, A., Mouchtouri, V. A., Speletas, M., Tsiodras, S., & Hadjichristodoulou, C. (2021). Wastewater monitoring as a supplementary surveillance tool for capturing SARS-COV-2 community spread. A case study in two Greek municipalities. *Environmental Research*, *200*, 111749. https://doi.org/10.1016/J.ENVRES.2021.111749

Larsen, D. A., & Wigginton, K. R. (2020). Tracking COVID-19 with wastewater. *Nature Biotechnology*, *38*(10), 1151. https://doi.org/10.1038/S41587-020-0690-1

McMahan, C. S., Self, S., Rennert, L., Kalbaugh, C., Kriebel, D., Graves, D., Colby, C., Deaver, J. A., Popat, S. C., Karanfil, T., & Freedman, D. L. (2021). COVID-19 wastewater epidemiology: a model to estimate infected populations. *The Lancet Planetary Health*, *5*(12), e874–e881. https://doi.org/10.1016/S2542-5196(21)00230-8/ATTACHMENT/D719E632-85E0-46C2-A1A4-10D3C1118991/MMC1.PDF

Petala, M., Kostoglou, M., Karapantsios, T., Dovas, C. I., Lytras, T., Paraskevis, D., Roilides, E., Koutsolioutsou-Benaki, A., Panagiotakopoulos, G., Sypsa, V., Metallidis, S., Papa, A., Stylianidis, E., Papadopoulos, A., Tsiodras, S., & Papaioannou, N. (2022). Relating SARS-CoV-2 shedding rate in wastewater to daily positive tests data: A consistent model based approach. *The Science of the Total Environment*, *807*(Pt 2). https://doi.org/10.1016/J.SCITOTENV.2021.150838

Vallejo, J. A., Trigo-Tasende, N., Rumbo-Feal, S., Conde-Pérez, K., López-Oriona, Á., Barbeito, I., Vaamonde, M., Tarrío-Saavedra, J., Reif, R., Ladra, S., Rodiño-Janeiro, B. K., Nasser-Ali, M., Cid, Á., Veiga, M., Acevedo, A., Lamora, C., Bou, G., Cao, R., & Poza, M. (2022). Modeling the number of people infected with SARS-COV-2 from wastewater viral load in Northwest Spain. *Science of The Total Environment*, *811*, 152334. https://doi.org/10.1016/J.SCITOTENV.2021.152334

Zhang, W., Du, R. H., Li, B., Zheng, X. S., Yang, X. Lou, Hu, B., Wang, Y. Y., Xiao, G. F., Yan, B., Shi, Z. L., & Zhou, P. (2020). Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. *Emerging Microbes & Infections*, *9*(1), 386. https://doi.org/10.1080/22221751.2020.1729071

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., … Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798), 270–273. https://doi.org/10.1038/s41586-020-2012-7