Available online at http://scik.org
J. Math. Comput. Sci. 2024, 14:4
https://doi.org/10.28919/jmcs/8301
ISSN: 1927-5307

AN INVESTIGATION ON THE APPLICATION OF DIAGNOSTIC ANALYTICS TO ASCERTAIN THE CAUSE OF THE PANDEMIC′S INFECTION SPIKE

C. J. SANTOSH*, ANURAG SHAKYA

Institute of Business Management and Commerce, Mangalayatan University, Extended NCR, 33rd Milestone, Mathura-Aligarh Highway, Aligarh, 202145, Uttar Pradesh, India

Abstract. Tracking all of the COVID-19 subvariants proved difficult due to the virus's rapid dissemination throughout the pandemic. Initially, many million individuals worldwide perished as a result of the World Health Organization's inability to monitor various COVID-19 instances. Finding the underlying reason for the pandemic's spike in infections is the aim of this investigation, which will include diagnostic analysis. The outcomes of applying diagnostic analytics are also emphasized. While the results are important, the focus is primarily on the event's deficiencies and how diagnostic analytics may help decision-makers identify the shortcomings of their previous decisions so they can make better ones moving forward. This study uses basic statistical techniques to examine the connection between variables as well as observational judgements drawn from charts. The results of this investigation confirm the value of diagnostic analytics in combating the COVID-19 pandemic and its function in aiding decision-makers in managing the infection wave.

Keywords: Data analytics; diagnostic analytics; pandemic; COVID-19.

2020 AMS Subject Classification: statistics; diagnostics; linear inference and regression.

1. Introduction

A contagious viral infection was initially discovered in Wuhan, China, in January 2020 [1]. With a very high transmission rate, the COVID-19 infection quickly spread to every country in

*Corresponding author
E-mail address: cjsantosh@hotmail.com
Received October 27, 2023

the world [2]. Numerous nations promptly enforced travel restrictions and lockdowns. Aside from social separation, face masks, seclusion, and hand cleanliness, these preventative measures became standard practices [3]. There have been reports of worldwide turmoil, an unusual incidence of hospitalization, and fatalities. It was stated that hospital beds, face masks, and hand sanitizers were in limited supply. One of the biggest challenges today is distributing resources globally in the most efficient ways. Overcoming these obstacles required the application of data analytics. Government officials were able to identify high-risk zones and put control measures in place by employing analytical approaches to predict patterns and potential trends. This study aims to apply diagnostic analytics, pinpoint the underlying reason for the infection spike, and determine the contributing elements, both positive and negative. We may assist in flattening the curve by making judgements based on the identification of positive and negative aspects.

2. Background

The four phases of data analytics are descriptive, diagnostic, predictive, and prescriptive [4]. The initial phase of data analytics is called descriptive analytics. To determine what transpired during an event, historical data is examined. The elements may be favorable or unfavorable. An event's success is influenced by positive variables, whilst negative aspects lead to failures. The components that contribute to an occurrence may all be found via descriptive analytics. The second phase of data analytics is called diagnostic analytics. It makes use of past data to pinpoint the main reason for an event's success or failure. There is an enormous quantity of data created when an event occurs. Multiple factors will correlate with one another and contribute to the success or failure of an event. The link and strength of these factors may be established quantitatively. The third step of data analytics is predictive analytics. It aids in forecasting future trends. Predictive analytics forecasts the future using previous data. Prescriptive analytics is the final level of data analytics, in which the next course of action is determined.

3. Methods

For analytics, we used data from Our World in Data, which can be found at the URL: hppts://ourworldindata.org/covidcases [5]. The information on daily confirmed cases, new fatalities, freshly provided vaccination, newly done COVID-19 tests, positive rate, reproduction rate, and stringency index was downloaded in CSV format. We also took into account information from the European Centre for Disease Prevention and Control, which can be found at https://www.ecdc.europe.eu/en/publications-data/data-virus-variants-covid-19-eueea. The CSV-formatted data provides information on COVID-19 sequencing data, COVID-19 variations causing infection, and the fraction of variants causing the rise [6]. This study takes into account the Omicron sub-variants. Microsoft Excel was used to assemble the data from both websites. To execute diagnostic analytics, we must first detect the patterns with descriptive analytics. Using diagnostic analytics, we must be able to determine the root cause based on the patterns found. To examine the link between variables, the Pearson correlation coefficient analysis will be used.

4. THEORY / CALCULATIONS

4.1. Descriptive Analytics.

The initial level of data analytics is descriptive analytics. The historical data is evaluated at this step to detect patterns. Figure [1] shows that there was an increase in infection in the European country of Belgium between Week 41 of 2021 and Week 20 of 2022. We discovered the trend in numerous European nations after doing comprehensive research. Figure [2] depicts the nations that saw a similar type of surge pattern during this time period.

4.2. Diagnostic Analytics.

The second step of data analytics is diagnostic analytics. A root cause analysis was undertaken to discover the source of the infection rise in the European country of Belgium between Week 41 of 2021 and Week 20 of 2022. The analysis took into account factors such as immunization count, new test count, stringency index, reproduction rate, and positive rate. The vaccination count is the number of persons who have been vaccinated in a certain day [7]. The total number of tests performed on a day is represented by the new test count [8]. The stringency index assesses the government's responsiveness to COVID-19 infection control [9]. The virus's transmissibility is measured by its reproduction rate [10]. From all the samples collected in a

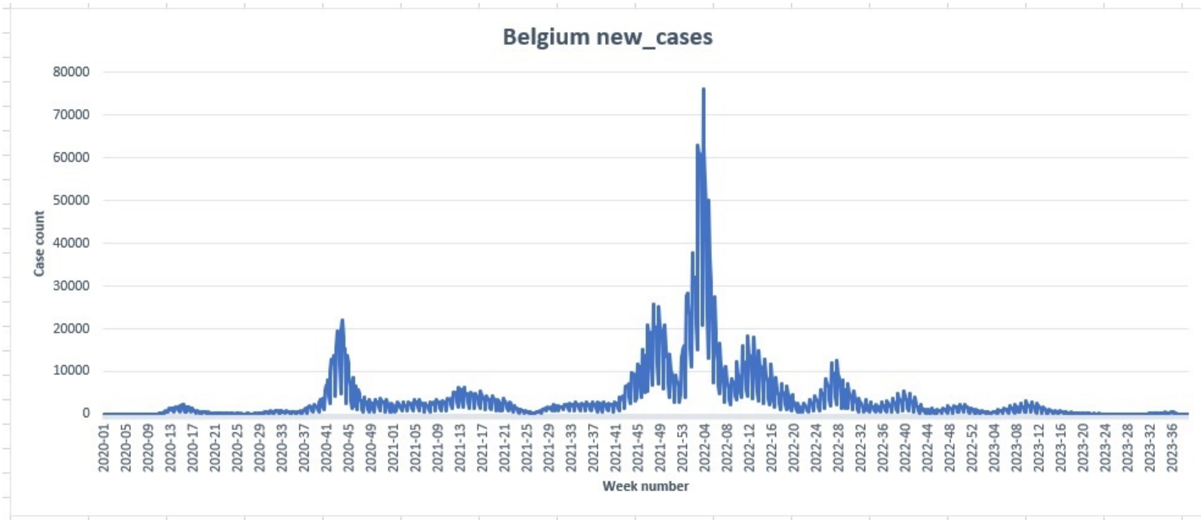

FIGURE 1. New cases - Belgium (2021 - 2022)

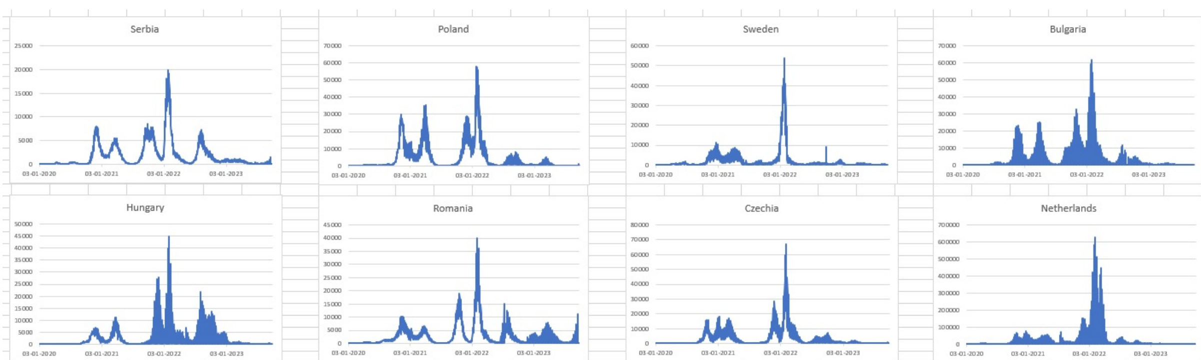

FIGURE 2. Surge Pattern - European Countries

single day, the number of COVID-19 positives discovered is known as the positive rate [11]. The positive rate depends on the number of tests conducted.

4.3. Positive factor that helped flatten the curve.

Our comprehensive study has shown that the curve has flattened due to a significant rise in new tests and immunizations. The cause for the decline in infection rate between week 04 and week 20 of 2022 is illustrated in the line chart in Figure [3].

4.4. Negative factor that led to the surge in infections.

The positive rate has adversely increased, as seen in Figure [4]. Additionally, we may deduce that following week number 41 of 2021, the stringency index significantly decreased. This indicates that between week number 41 of 2021 and week number 20 of 2022, there was an increase in the positive rate and a fall in the stringency index, which caused the infection spike. The Pearson correlation coefficient approach will be utilized to examine the link between the positive rate and new cases.

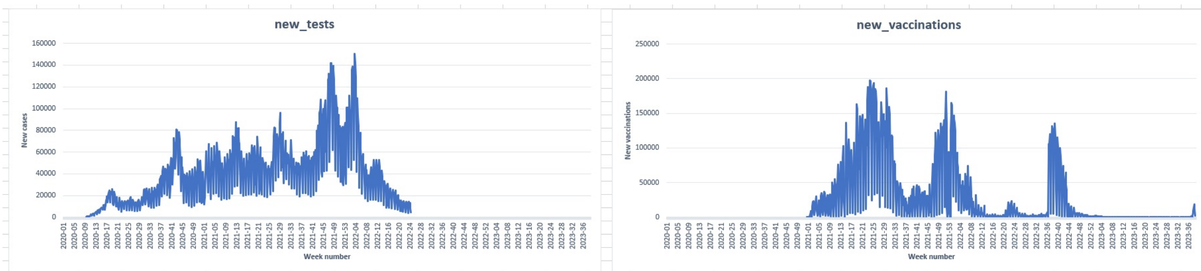

FIGURE 3. Factors with a positive contribution

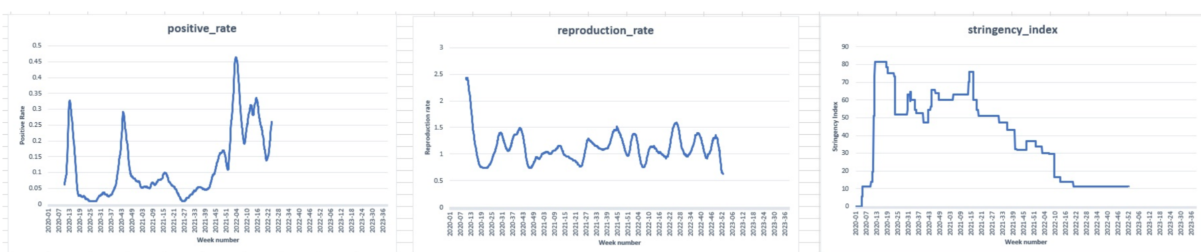

FIGURE 4. Factors with a negative contribution

4.5. Relationship between vairables.

The link between the two quantitative variables will be evaluated using the Pearson correlation coefficient (r). To find the line of best fit, we will visualize the Pearson correlation coefficient and assess the strength of the link. The significance of the correlation coefficient will be examined using a t-test.

The formula used to calculate the Pearson correlation coefficient is mentioned below

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{1}$$

Where,

r = correlation coefficient

x_i = variable x values

$\bar{x}$ = mean of x

y_i = variable y values

$\bar{y}$ = mean of y

$\bar{x}$ = 12256.50

$\bar{y}$ = 0.22

$$r = \frac{186364.87}{316769.06}$$

$$r = 0.588330413$$

4.6. t-Test.

Null Hypothesis H_0:No significant linear relationship between x and y

Alternative Hypothesis H_1: Significant linear relationship between x and y

$$t = \frac{r.\sqrt{n-2}}{\sqrt{1-r^2}} \tag{2}$$

$$t = \frac{0.588.\sqrt{238-2}}{\sqrt{1-0.588^2}}$$

$$t = \frac{0.588.(15.3622)}{\sqrt{0.654256}}$$

$$t = \frac{9.0329}{0.80886}$$

$$t = 11.16$$

Using a table of Critical values to make a decision

r = 0.588, and the sample size n is 238, df = n - 2 = 236. The critical value is 1.6513. Test Statistic >= Critical value: The null hypothesis is rejected, and the correlation is statistically significant. A scatter plot, Figure [5] shows both variables are quantitative, normally distributed, and have a linear relationship with a strong positive correlation.

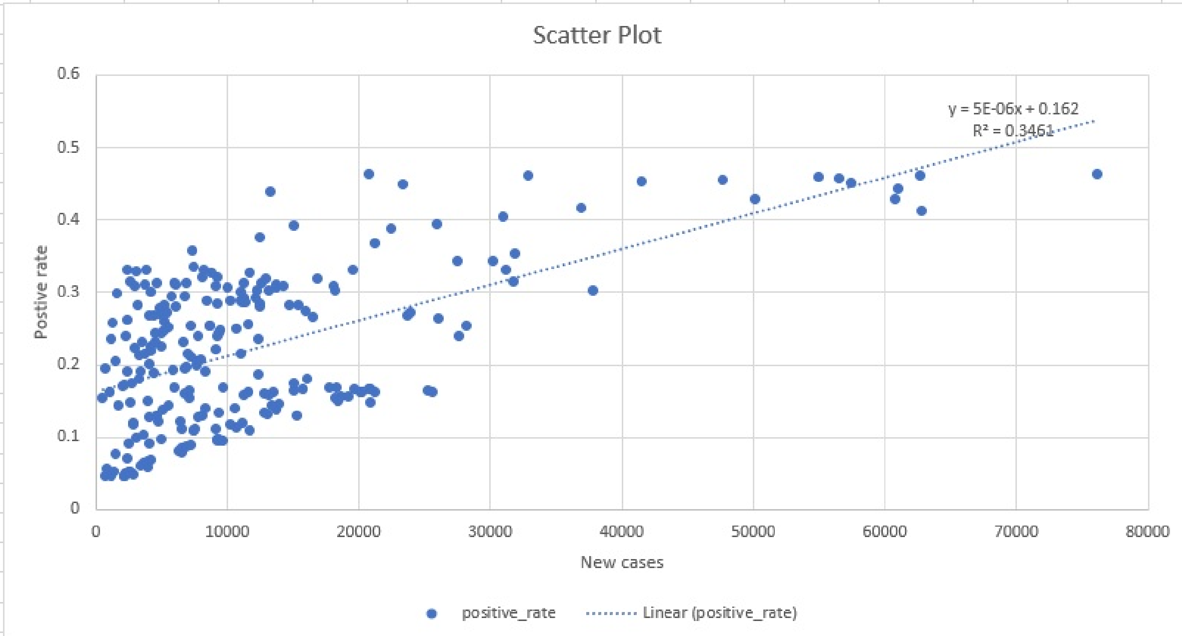

FIGURE 5. Scatter Plot - Linear relationship

4.7. Dominant variant analysis.

We identified four COVID-19 sub-variants—BA.1, BA.2, B.1.617.2, and UNK—as the cause of the spike in infections during this time by examining the GISAID data. A table including variations that were prominent throughout this time period is presented in Figure [6].

5. RESULTS

Based on the research above, we may conclude that an extreme drop in the stringency index and a significant increase in the positive rate will have a negative effect, leading to a spike in infection. According to this study, diagnostic analytics may be a game-changer during a pandemic and plays a significant role in determining the core cause, while descriptive analytics is crucial in detecting the pattern.

6. DISCUSSION

Many nations used stringent control measures including lockdowns, travel restrictions, social separation, and isolation during the pandemic. Few nations' stringency indices hovered around 81.48, which assisted them in flattening the curve. However, an elevated stringency index has an

country	country_code	year_week	source	percent_variant BA.1	percent_variant BA.2	percent_variant B.1.617.2	percent_variant UNK
Belgium	BE	2021-40	GISAID	0	0	100	0
Belgium	BE	2021-41	GISAID	0.1	0	99.8	0.1
Belgium	BE	2021-42	GISAID	0	0	99.9	0.1
Belgium	BE	2021-43	GISAID	0	0	100	0.1
Belgium	BE	2021-44	GISAID	0	0	100	0.1
Belgium	BE	2021-45	GISAID	0	0	99.9	0
Belgium	BE	2021-46	GISAID	0.2	0	99.7	0
Belgium	BE	2021-47	GISAID	0.1	0	99.9	0
Belgium	BE	2021-48	GISAID	2.4	0	97.6	0
Belgium	BE	2021-49	GISAID	11.7	0	88.3	0.1
Belgium	BE	2021-50	GISAID	37.5	0	61.9	0
Belgium	BE	2021-51	GISAID	53.8	0.1	45.7	0.1
Belgium	BE	2021-52	GISAID	79.4	0.2	20.3	0.2
Belgium	BE	2022-01	GISAID	90.7	0.7	7.2	0.2
Belgium	BE	2022-02	GISAID	94.1	3.2	2.6	0.9
Belgium	BE	2022-03	GISAID	95.1	4.1	0.7	0.2
Belgium	BE	2022-04	GISAID	81	17.9	0.9	0.4
Belgium	BE	2022-05	GISAID	77	22.5	0.4	0.1
Belgium	BE	2022-06	GISAID	64.9	34.4	0.2	0.4
Belgium	BE	2022-07	GISAID	58.6	40.6	0.1	0.1
Belgium	BE	2022-08	GISAID	49.2	50.2	0.1	0
Belgium	BE	2022-09	GISAID	41.9	56.9	0.2	0
Belgium	BE	2022-10	GISAID	22.7	77	0.2	0
Belgium	BE	2022-11	GISAID	13.4	86.4	0.1	0.1
Belgium	BE	2022-12	GISAID	9	91	0	0
Belgium	BE	2022-13	GISAID	5.4	94.4	0	0.1
Belgium	BE	2022-14	GISAID	3.8	95.6	0.1	0
Belgium	BE	2022-15	GISAID	2.9	96.7	0	0
Belgium	BE	2022-16	GISAID	1.1	98.3	0	0.2
Belgium	BE	2022-17	GISAID	1.1	96.9	0	0.1
Belgium	BE	2022-18	GISAID	0.5	95.6	0	0
Belgium	BE	2022-19	GISAID	0.5	92.1	0	0
Belgium	BE	2022-20	GISAID	0.7	88.1	0	0
Average				27.24	37.66	34.12	0.11

FIGURE 6. Dominant variants

adverse effect on a nation's economy. Nationwide demonstrations against limitations sparked social instability and helped the illness spread. The best way to stop the virus from spreading is to increase the number of tests and vaccinations administered daily together with non-economic control measures like face masks, hand cleanliness, and social distancing.

7. CONCLUSION

The transmission rate was greater for the COVID-19 sub-variants BA.1, BA.2, and BQ.1. Worldwide, the hospitalization rate was high despite the low death rate. As a result, there was a lack of resources everywhere on the planet. It is critical to keep an eye on the infection spike using validated analytical techniques. It is much simpler for government officials to comprehend what went wrong and what has to be done going forward when they can visualize the components that contributed to the infection spike. The infection spike can be efficiently controlled with the use of a concrete diagnostic method-based remedy.

CONFLICT OF INTERESTS

The author declares that there is no conflict of interests.

REFERENCES

[1] Y.C. Wu, C.S. Chen, Y.J. Chan, Overview of the 2019 novel coronavirus (2019-nCoV): the pathogen of severe specific contagious pneumonia (SSCP), J. Chin. Med. Assoc. 83 (2020), 217–220.

[2] G.D. Barmparis, G.P. Tsironis, Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach, Chaos Solitons Fractals. 135 (2020), 109842. https://doi.org/10.1016/j.chaos.2020.109842.

[3] B. Odugbemi, O. Odusanya, T. Odugbemi, et al. COVID-19: A review of the effectiveness of non-pharmacological interventions, Nigerian Postgrad. Med. J. 27 (2020), 261–267. https://doi.org/10.4103/npmj.npmj_208_20.

[4] V. Muneeswaran, P. Nagaraj, U. Dhannushree, et al. A framework for data analytics-based healthcare systems, in: J.S. Raj, A.M. Iliyasu, R. Bestak, Z.A. Baig (Eds.), Innovative Data Communication Technologies and Application, Springer Singapore, Singapore, 2021: pp. 83–96. https://doi.org/10.1007/978-981-15-9651-3_7.

[5] E. Mathieu, H. Ritchie, L. Rodés-Guirao, et al. Coronavirus Pandemic (COVID-19), Our World in Data, (2020), https://ourworldindata.org/coronavirus.

[6] S. Elbe, G. Buckland-Merrett, Data, disease and diplomacy: GISAID's innovative contribution to global health, Glob. Chall. 1 (2017), 33–46. https://doi.org/10.1002/gch2.1018.

[7] E. Mathieu, H. Ritchie, E. Ortiz-Ospina, et al. A global database of COVID-19 vaccinations, Nat. Hum. Behav. 5 (2021), 947–953. https://doi.org/10.1038/s41562-021-01122-8.

[8] J. Hasell, E. Mathieu, D. Beltekian, et al. A cross-country database of COVID-19 testing, Sci. Data. 7 (2020), 345. https://doi.org/10.1038/s41597-020-00688-8.

[9] S. Chen, L. Guo, T. Alghaith, et al. Effective COVID-19 control: A comparative analysis of the stringency and timeliness of government responses in Asia, Int. J. Environ. Res. Public Health. 18 (2021), 8686. https://doi.org/10.3390/ijerph18168686.

[10] N.H.L. Leung, Transmissibility and transmission of respiratory viruses, Nat. Rev. Microbiol. 19 (2021), 528–545. https://doi.org/10.1038/s41579-021-00535-6.

[11] T.R. Mercer, M. Salit, Testing at scale during the COVID-19 pandemic, Nat. Rev. Genet. 22 (2021), 415–426. https://doi.org/10.1038/s41576-021-00360-w.