



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



ELSEVIER

Contents lists available at ScienceDirect

## International Journal of Infectious Diseases

journal homepage: [www.elsevier.com/locate/ijid](http://www.elsevier.com/locate/ijid)

## Clinical and laboratory profiles of the SARS-CoV-2 Delta variant compared with pre-Delta variants<sup>☆</sup>



Shivang Bhakta<sup>1,\*</sup>, Devang K. Sanghavi<sup>1</sup>, Patrick W. Johnson<sup>2</sup>, Katie L. Kunze<sup>3</sup>, Matthew R. Neville<sup>4</sup>, Hani M. Wadei<sup>5</sup>, Wendelyn Bosch<sup>6</sup>, Rickey E. Carter<sup>2</sup>, Sadia Z. Shah<sup>5</sup>, Benjamin D. Pollock<sup>4</sup>, Sven P. Oman<sup>7</sup>, Leigh Speicher<sup>8</sup>, Jason Siegel<sup>1,9</sup>, Claudia R. Libertin<sup>6</sup>, Mark W. Matson<sup>10</sup>, Pablo Moreno Franco<sup>1,4,5</sup>, Jennifer B. Cowart<sup>7,\*\*</sup>

<sup>1</sup> Department of Critical Care Medicine, Mayo Clinic, Jacksonville, FL, USA

<sup>2</sup> Department of Quantitative Health Sciences, Mayo Clinic, Jacksonville, FL, USA

<sup>3</sup> Department of Quantitative Health Sciences, Mayo Clinic, Scottsdale, AZ, USA

<sup>4</sup> Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Jacksonville, FL, USA

<sup>5</sup> Department of Transplantation, Mayo Clinic, Jacksonville, FL, USA

<sup>6</sup> Division of Infectious Diseases, Mayo Clinic, Jacksonville, FL, USA

<sup>7</sup> Division of Hospital Internal Medicine, Mayo Clinic, Jacksonville, FL, USA

<sup>8</sup> Division of General Internal Medicine, Mayo Clinic, Jacksonville, FL, USA

<sup>9</sup> Department of Neurology, Mayo Clinic, Jacksonville, FL, USA

<sup>10</sup> Center for Digital Health – Data & Analytics, Mayo Clinic, Rochester, MN, USA

### ARTICLE INFO

#### Article history:

Received 3 March 2022

Revised 20 April 2022

Accepted 21 April 2022

#### Keywords:

COVID-19

Variants of concern

Delta variant

Genomics

Machine learning

Gradient boosting model

### ABSTRACT

**Objectives:** The emergence of SARS-CoV-2 variants of concern has led to significant phenotypical changes in transmissibility, virulence, and public health measures. Our study used clinical data to compare characteristics between a Delta variant wave and a pre-Delta variant wave of hospitalized patients.

**Methods:** This single-center retrospective study defined a wave as an increasing number of COVID-19 hospitalizations, which peaked and later decreased. Data from the United States Department of Health and Human Services were used to identify the waves' primary variant. Wave 1 (August 8, 2020–April 1, 2021) was characterized by heterogeneous variants, whereas Wave 2 (June 26, 2021–October 18, 2021) was predominantly the Delta variant. Descriptive statistics, regression techniques, and machine learning approaches supported the comparisons between waves.

**Results:** From the cohort (N = 1318), Wave 2 patients (n = 665) were more likely to be younger, have fewer comorbidities, require more care in the intensive care unit, and show an inflammatory profile with higher C-reactive protein, lactate dehydrogenase, ferritin, fibrinogen, prothrombin time, activated thromboplastin time, and international normalized ratio compared with Wave 1 patients (n = 653). The gradient boosting model showed an area under the receiver operating characteristic curve of 0.854 (sensitivity 86.4%; specificity 61.5%; positive predictive value 73.8%; negative predictive value 78.3%).

**Conclusion:** Clinical and laboratory characteristics can be used to estimate the COVID-19 variant regardless of genomic testing availability. This finding has implications for variant-driven treatment protocols and further research.

Published by Elsevier Ltd on behalf of International Society for Infectious Diseases.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

<sup>☆</sup> Summary: Laboratory and clinical characteristics differed between the Delta and pre-Delta variant in hospitalized cases. The gradient boosting model predicted patients with Delta-like characteristics with a high mean area under the receiver operating characteristic curve in a limited genomic sequencing setting.

\* Corresponding author: Department of Critical Care, Mayo Clinic, 4500 San Pablo Road, Jacksonville, FL 32224, USA.

\*\* Corresponding author: Division of Hospital Internal Medicine, Department of Internal Medicine, Mayo Clinic, 4500 San Pablo Road, Jacksonville, FL 32224, USA.

E-mail addresses: [bhakta.shivang@mayo.edu](mailto:bhakta.shivang@mayo.edu) (S. Bhakta), [cowart.jennifer@mayo.edu](mailto:cowart.jennifer@mayo.edu) (J.B. Cowart).

## Introduction

SARS-CoV-2, which causes COVID-19, has led to a significant global health crisis resulting in more than 5.8 million deaths as of February 14, 2022 (Coronavirus Resource Center, 2022). The high prevalence and transmissibility of SARS-CoV-2 in a population allows adaptive mutations in the viral genome, mostly mildly deleterious or neutral. A small number of these mutations may result in a significant phenotypical virus with an increase in transmissibility, increase in virulence, or decrease in the effectiveness of public/social health measures (Khateeb et al., 2021). The World Health Organization (World Health Organization, 2021) defines them as variants of concern (VOCs). Most countries have experienced many waves of this viral illness, generally coinciding with these new variant strains (Davies et al., 2021; Singh et al., 2021).

Owing to high transmissibility, the Delta variant (Pango lineage B.1.617.2) became the predominant variant in the United States in July 2021 (Centers for Disease Control and Prevention, 2021). Compared with the previous Alpha variant, the Delta variant was associated with a two-fold increased risk of hospitalization within 14 days of a positive test in England and Scotland (GOV.UK Coronavirus, 2021; Sheikh et al., 2021a). In a retrospective study from Singapore, Ong et al. (2021) found five-fold increased odds of disease severity with the Delta variant compared with the Alpha or Beta variant sequences. However, it is challenging to extrapolate data from other countries to the United States because of differences in healthcare resource use, patient characteristics, and socio-behavioral trends.

There are limited data regarding the clinical and biomarker characterization of Delta and other variants in the United States. Owing to limited US genomic surveillance (0.1–3.1% of positive SARS-CoV-2 tests) (Paul et al., 2021), the emergence of a new VOC wave may lag its identification occurring in a geographic locale. Consequently, surrogate identifying of the emergence of a new wave is needed. Laboratory markers may predict prognosis, including lymphocytopenia, inflammatory markers (e.g., C-reactive protein, ferritin), lactate dehydrogenase, high-sensitivity troponin I, abnormal coagulation parameters, and others that are commonly associated with poor outcomes (Poggiali et al., 2020; Henry et al., 2020; Sui et al., 2021). In this retrospective study, we aimed to compare the hospitalized patient characteristics of the Delta variant surge with the pre-Delta variant surge in a single-center hospital in Florida to clinically define and distinguish the variants.

## Methods

### Study Setting and Population

This study was conducted at Mayo Clinic in Jacksonville, Florida, and was deemed exempt from review by the institutional review board (IRB 21-002944). Hospitalized patients with a positive nasopharyngeal polymerase chain reaction test or antigen for SARS-CoV-2 on admission or during their hospital stay were reviewed. The vaccination status was assessed from our electronic medical record, which is updated from the Florida State Health Online Tracking System every 2 weeks for all patients >5 years of age residing in Florida. Vaccine breakthrough was defined as a positive polymerase chain reaction or antigen for SARS-CoV-2 obtained after 14 days from complete vaccination (after the second dose of a Pfizer, Moderna, or AstraZeneca vaccine or after the first dose of a Johnson & Johnson vaccine). The time of study predated the approval of mRNA vaccine third dose and viral vector vaccine second dose (or “boosters”) in the United States (US Food & Drug Administration, 2021).

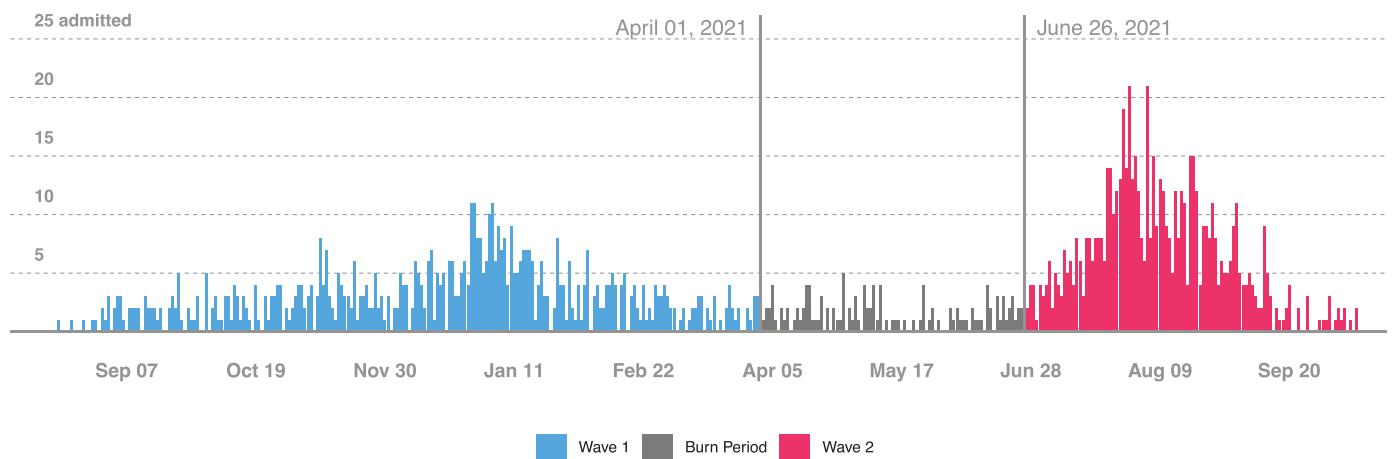
A COVID-19 wave was defined as the period characterized by a steady increase in hospitalizations that may stabilize and decrease

over time. A flattening of COVID-19 admissions for several weeks marked the end of a wave (<5 daily admits). For the Mayo Clinic, Florida, Wave 1 started on August 8, 2020, and ended on April 1, 2021. Wave 2 started on June 26, 2021, and ended on October 18, 2021. Data from the United States Department of Health and Human Services SARS-CoV-2 Interagency Group were used to identify the primary variant of each wave based on ecological data, as the capacity and resources to conduct viral genomic sequencing of specific hospitalized patients were not available. Wave 1 comprised a heterogeneous group of SARS-CoV-2 variants. Wave 2 was predominantly characterized by the Delta variant representing >50% of the sequenced genome (June 26, 2021) and subsequently >90% during peak hospital admissions. Therefore, a washout period of 12.2 weeks was established between Wave 1 and Wave 2 to minimize carryover effects from previous variants.

### Statistical Analysis

Data were analyzed using a mixture of standard descriptive statistics, regression techniques, and machine learning approaches to support comparing patient characteristics and hospital outcomes between the 2 waves of patients. First, data were analyzed for differences between waves using descriptive statistics. Absolute value standardized mean differences >10% were considered relevant differences between the waves. Kruskal-Wallis rank-sum tests were used to test for these differences more formally. Next, another goal was to determine whether more generalized, or clustered, combinations of variables could be associated with the waves. To address this, a gradient boosting machine (GBM) was estimated using baseline comorbidities, patient characteristics, and laboratory values performed closest to hospital admission. Rather than using the GBM to predict a clinical outcome such as 28-day mortality, the GBM was trained to predict the COVID-19 variant type predominant in that wave. In this way, the GBM was used to explore fundamental differences between patient cohorts with the hypothesis that the variants were unique while allowing for interactions and non-linear associations in the modeling form. Laboratory values had differing levels of missing rates in the 2 waves. To avoid missing data being used to predict wave likelihood, missing data were imputed using the MissForest algorithm before modeling. Data were split in a traditional train (80%) and test (20%) manner. The GBM was then constructed using a cartesian grid search to find optimal tuning parameters (learning rate, column sampling, row sampling, tree depth, and the number of trees) while using five-fold cross-validation to assess model performance. The model with the highest mean area under the curve across folds in the training data set was selected as the final model. In the final test data set, a receiver operating characteristic curve was generated, and traditional binary classification summaries such as sensitivity and specificity were used to assess model performance and misclassification in data not used during model development and selection.

To assess the role of each variable included in the model, two 2 graphical representations of the fitted model were used: a variable importance plot and a Shapley additive explanations (SHAP) plot. The variable importance plots provide a relative ranking of how much each variable improved the model fit; however, the figure does not capture the direction and magnitude of numeric changes in the variable with the classification outcome (i.e., wave). To summarize this latter impact in terms of both directions (e.g., the likelihood of the case being from Wave 2) and magnitude (e.g., how much this likelihood or estimated probability changes over the domain of the observed values), SHAP plots were added. SHAP plots show both global and local trends in model predictions: at the global level, values are ordered based on feature importance; at the local level, large SHAP contributions indicate measurements



**Figure 1.** Epidemiologic curve showing 2 COVID-19 disease admission waves in Mayo Clinic, Florida. The number of new COVID-19 cases (y-axis) is shown as the absolute number of admissions per day. The blue-colored first wave comprises a heterogeneous array of SARS-CoV-2 variants, named Wave 1. The red-colored second wave shows Wave 2. A buffer period of 12.2 weeks was established after the end of the first wave.

that had a high impact on individual predictions (far-left indicating greater likelihood of belonging to Wave 1 and far-right, Wave 2), and color represents the relative magnitude of value for each variable. Sensitivity and specificity are also reported after a threshold for 0.028 wave classification was selected to optimize F1 performance in training data. Given that the baseline laboratory markers of clotting, inflammation, and other assays were noted as important variables in the GBM fit, longitudinal analyses were conducted using censored mixed models with a random intercept. Data were winsorized at the 0 and 95th percentiles. For these models, the parameters of primary interest were the model wave indicator, which quantified how much difference there was in laboratory values at admission (time = 0), and the hospital day by wave interaction term, which quantified differences in the rate of change in the laboratory values between waves throughout the hospital stay.

Statistical analyses and graphical presentations were created using R version 4.0.3 (Vienna, Austria). When *P*-values were reported to be interpreted, a  $P < 0.05$  (2-sided) threshold was used to represent statistical significance.

## Results

The final sample ( $N = 1318$ ) included 653 cases in the pre-Delta variant Wave 1 and 665 cases in the Delta variant–dominant Wave 2. Figure 1 provides an overview of hospital admissions and identifies cases that were included in each wave, as well as a buffer period in which we observed an overlap between the pre-Delta and Delta variants. Descriptive statistics and tests for differences in baseline comorbidities and patient characteristics between waves are shown in Table 1. Several differences in baseline comorbidities and patient characteristics were observed, including age, race, ethnicity, and comorbidities such as hypertension, chronic kidney disease, chronic obstructive pulmonary disease, coronary artery disease, and congestive heart failure. Patients from Wave 2 were significantly younger, with fewer comorbidities and less immunosuppression, than those from Wave 1. Wave 2 patients were more likely to require intensive care but had lower unadjusted mortality than those from Wave 1 (Figure 2).

A comparison of laboratory assay values between the 2 waves is listed in Table 2. The Wave 2 group was characterized by higher values of inflammatory biomarkers, including C-reactive protein (CRP), ferritin, and lactate dehydrogenase (LDH) on admission. Other values found to be significantly different were serum creatinine, fibrinogen, international normalized ratio, activated partial thromboplastin time, prothrombin time, and segmented neu-

trophil count. An analysis of vaccine breakthrough cases between the waves was impossible because of a predominantly unvaccinated Wave 1 (94.9%) group.

The variable importance and SHAP plots are presented in Figure 3. An area under the curve of 0.854 (0.809, 0.899) (sensitivity: 86.4% [79.8%, 91.5%]; specificity 61.5% [52.1%, 70.4%]) in the test data set indicates a fundamental difference between cohorts (Figure 4). Notable variables with the highest predictive value in our model, according to SHAP values, were prothrombin time, international normalized ratio, LDH, fibrinogen, age of the patient, chronic lung disease, and CRP.

Censored mixed effects modeling of changes in laboratory data over time identified a statistically significant difference in random intercept estimates in 11 of 16 assays and random slope estimates (interaction of wave and days from admission to testing) in 14 of 16 assays (Supplementary Table 1).

## Discussion

Throughout the COVID-19 pandemic, much attention has been directed to the SARS-CoV-2 virus as if it were a single entity. With the emergence of the Delta and now Omicron variants, it has become clear that variants have the potential for differences in patient trajectories and characteristics that warrant further consideration and clarification. Genomic surveillance would be the most precise means of identifying the spread of new viral variants, but this technology has limitations. As of mid-2021, the United States ranked 33rd worldwide in genomic surveillance (Crawford and Williams, 2021), and even when genomic testing is performed, there are often significant time delays with results. Test results may also not be identifiable back to the individual patient level. One goal of this study was to better quantify how a constellation of factors shifted in hospitalized patients as the predominant variant changed in the community without access to individual patient-level genomic testing.

The ability to predict the phenotype of the predominant viral variant has implications for individual patient care as well as for hospitals and at the population level. The Delta variant–dominant Wave 2 was characterized by more inflammation and higher intensive care needs than the pre-Delta variant–dominant Wave 1. Knowing that a patient presents with Delta-like characteristics will allow better prognostication for that individual. If hospital cases increase with Delta-like characteristics, knowing this information will allow hospitals to plan internally and in networks to prepare for high intensive care use in equipment, space, and staffing. At the

**Table 1**  
Patient characteristics, comorbidities, and outcomes stratified by wave.

	Wave 1 (n=653)	Wave 2 (n=665)	Total (N=1318)	Standardized difference <sup>a</sup>
Age (years)	67 (20, 103)	60 (21, 101)	64 (20, 103)	<b>32.6%</b>
Sex (male)	392 (60.0%)	388 (58.3%)	780 (59.2%)	3.4%
Race				<b>16.2%</b>
American Indian/Alaskan Native	1 (0.2%)	2 (0.3%)	3 (0.2%)	
Asian	38 (5.8%)	27 (4.1%)	65 (4.9%)	
Black or African American	61 (9.3%)	83 (12.5%)	144 (10.9%)	
Pacific Islander	1 (0.2%)	0 (0.0%)	1 (0.1%)	
White	533 (81.6%)	524 (78.8%)	1057 (80.2%)	
Other/Unknown	19 (2.9%)	29 (4.4%)	48 (3.6%)	
Ethnicity				9.7%
Hispanic	27 (4.1%)	37 (5.6%)	64 (4.9%)	
Non-Hispanic	618 (94.6%)	614 (92.3%)	1232 (93.5%)	
Unknown	8 (1.2%)	14 (2.1%)	22 (1.7%)	
Chronic kidney disease	64 (9.8%)	44 (6.6%)	108 (8.2%)	<b>11.6%</b>
Chronic lung disease	391 (59.9%)	507 (76.2%)	898 (68.1%)	<b>35.6%</b>
Congenital heart disease	7 (1.1%)	6 (0.9%)	13 (1.0%)	1.7%
Congestive heart failure	104 (15.9%)	72 (10.8%)	176 (13.4%)	<b>15.0%</b>
Coronary artery disease	171 (26.2%)	119 (17.9%)	290 (22.0%)	<b>20.1%</b>
Diabetes mellitus	174 (26.6%)	162 (24.4%)	336 (25.5%)	5.2%
Hypertension	433 (66.3%)	342 (51.4%)	775 (58.8%)	<b>30.6%</b>
Immunosuppression <sup>b</sup>	125 (19.1%)	84 (12.6%)	209 (15.9%)	<b>17.9%</b>
Overall COVID-19 risk of complications score	4 (0, 10)	3 (0, 9)	4 (0, 10)	<b>30.6%</b>
End stage renal disease	50 (7.7%)	38 (5.7%)	88 (6.7%)	7.8%
Monoclonal antibodies	33 (5.1%)	32 (4.8%)	65 (4.9%)	1.1%
Dialysis	23 (3.5%)	13 (2.0%)	36 (2.7%)	9.6%
Transplant patient	92 (14.1%)	72 (10.8%)	164 (12.4%)	9.9%
Solid organ transplant	68 (10.4%)	42 (6.3%)	110 (8.3%)	<b>14.8%</b>
Solid organ transplant type				<b>20.2%</b>
Heart	7 (10.3%)	4 (9.5%)	11 (10.0%)	
Kidney	35 (51.5%)	24 (57.1%)	59 (53.6%)	
Liver	10 (14.7%)	6 (14.3%)	16 (14.5%)	
Lung	15 (22.1%)	8 (19.0%)	23 (20.9%)	
Pancreas	1 (1.5%)	0 (0.0%)	1 (0.9%)	
Vaccination status				<b>74.0%</b>
Unvaccinated	619 (94.8%)	474 (71.3%)	1093 (82.9%)	
Partially vaccinated	29 (4.4%)	40 (6.0%)	69 (5.2%)	
Breakthrough	5 (0.8%)	151 (22.7%)	156 (11.8%)	
Vaccine type at first immunization				<b>27.9%</b>
Johnson & Johnson	1 (2.9%)	13 (6.8%)	14 (6.2%)	
Moderna	14 (41.2%)	57 (29.8%)	71 (31.6%)	
Pfizer	19 (55.9%)	121 (63.4%)	140 (62.2%)	
Reason for testing				<b>60.6%</b>
N-Miss	468	164	632	
Asymptomatic	44 (23.8%)	19 (3.8%)	63 (9.2%)	
Symptomatic	141 (76.2%)	482 (96.2%)	623 (90.8%)	
Anti-spike antibody test				<b>36.4%</b>
N-Miss	534	43	577	
Negative (< 0.8 U/mL)	31 (26.1%)	268 (43.1%)	299 (40.4%)	
Positive (≥ 0.8 U/mL)	88 (73.9%)	354 (56.9%)	442 (59.6%)	
Positive anti-nucleocapsid antibody				<b>17.1%</b>
N-Miss	101	206	307	
Negative	358 (64.9%)	334 (72.8%)	692 (68.4%)	
Positive	194 (35.1%)	125 (27.2%)	319 (31.6%)	
Critical care services	177 (27.1%)	261 (39.2%)	438 (33.2%)	<b>26.0%</b>
Mechanical ventilation	52 (8.0%)	70 (10.5%)	122 (9.3%)	8.9%
Length of stay (days)				<b>10.4%</b>
N-Miss	0	2	2	
Median (range)	5 (1–193)	5 (1–155)	5 (1–193)	
Deceased	109 (16.7%)	85 (12.8%)	194 (14.7%)	<b>11.0%</b>

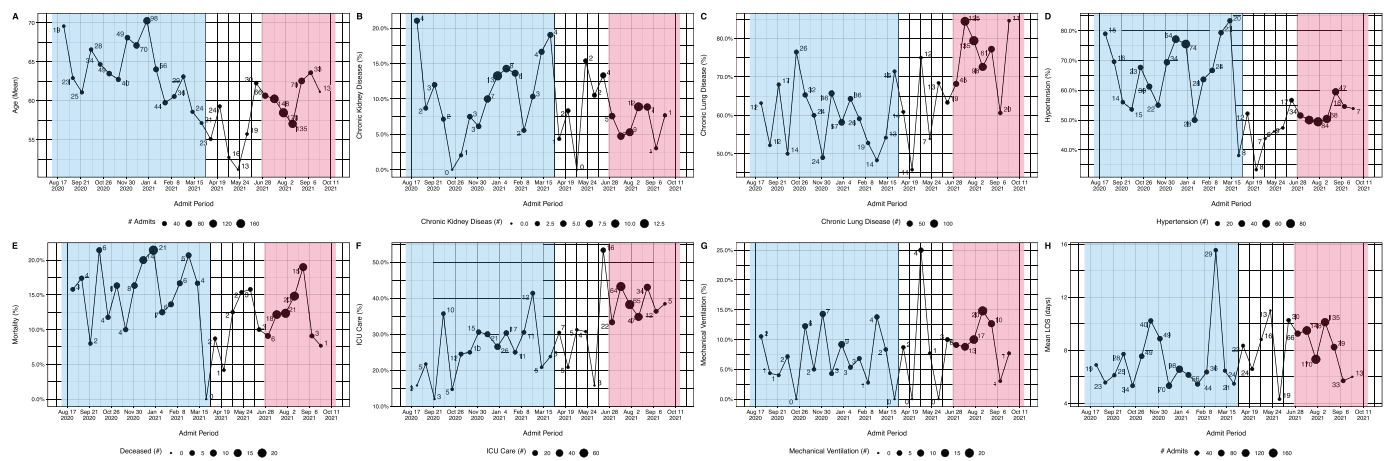
Categorical data are shown as count (percent). Numeric data are presented as median (range).

<sup>a</sup> Standardized difference = difference in proportions divided by standard error; imbalance defined as absolute value greater than 10% (text in bold formatting).

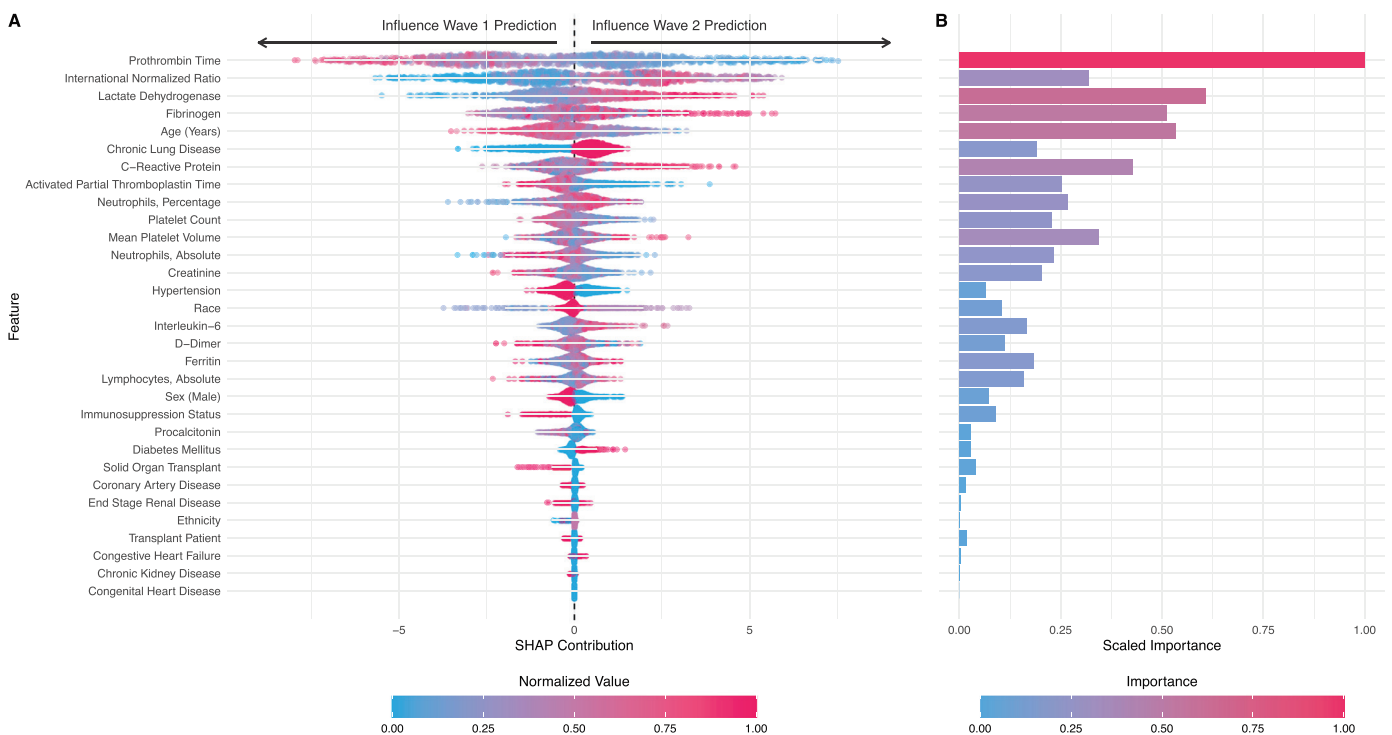
<sup>b</sup> Immunosuppression status was attributed to the following patients: diagnosed with human immunodeficiency virus infection, actively receiving chemotherapy, receiving immunosuppressive medications, or diagnosed with iatrogenic immunosuppression.

population level, a rise in Delta-like cases identified early would allow early adoption of public health measures to suppress the spread. In comparison, a non-Delta-like wave of cases in a highly vaccinated population would warrant a different public health response. As SARS-CoV-2 continues to mutate, VOCs will likely continue to spread. Public health strategies should be adaptable to not only rising case counts but also the possibility that a variant will cause more severe disease.

We used a machine learning technique, gradient boosting, to explore differences between waves. The GBM model identified multiple inflammatory and clotting factor variables that meaningfully shifted between the 2 waves. Certain markers were significantly higher in wave 2, such as coagulation studies, segmented neutrophils, fibrinogen, LDH, ferritin, and CRP (Table 2). CRP has been a surrogate marker for the degree of cytokines released in COVID-19, which is usually a higher level associated with a hy-



**Figure 2.** Patient characteristics and outcomes between Wave 1 and Wave 2. (a) Age; (b) chronic kidney disease; (c) chronic lung disease; (d) hypertension; (e) death; (f) intensive care unit; (g) mechanical ventilation; (h) length of hospital stay.



**Figure 3.** Shapley additive explanations (SHAP) plot for the gradient boosting model. (a) The figure plots every patient in the analysis as a point. The y-axis lists the input variables. The x-axis is a metric of the SHAP value associated with each variable and patient within the dataset (i.e., points plotted for each case based on the impact on prediction). The points plotted on the far-left have a greater impact on Wave 1 prediction and points plotted on the right have a greater impact on Wave 2 prediction. The normalized value of observation is color-based (red = higher values; blue = lower values). (b) The bar graph shows the input variables’ importance for wave prediction. The scaled importance is color-based (red = higher importance; blue = lower importance).

perinflammatory cytokine storm (Temesgen et al., 2022). Elevations in these specific studies support the hypothesis that the Delta variant is characterized by a relatively distinct, particularly hyperinflammatory profile. Other similar markers in both Wave 1 and Wave 2 included lymphopenia and elevations of procalcitonin, IL-6, D-dimer, and platelet count. Other studies have also indicated that absolute lymphopenia is not a predictor of outcomes by itself (Verma et al., 2022). These indicators remain elevated in severe COVID-19 but do not differ between the waves (Lopez-Castaneda et al., 2021; Malik et al., 2021).

Twhig et al. (2022) identified a higher risk of emergency care visits and hospital admission in unvaccinated patients infected with the Delta variant in the United Kingdom. In addition,

that study group reported that these patients were younger than those infected with the Alpha variant, similar to our findings. Ong et al. (2021), Luo et al. (2021), and Fisman and Tuite (2021) report a higher likelihood of hospitalization, intensive care unit admission, and/or death in patients infected with the Delta variant. Our study also demonstrates that Wave 2 patients, predominantly infected with the Delta variant, were more likely to require intensive care unit admissions. However, fewer deaths were seen in Wave 2 than in Wave 1. Several reasons may explain the lower number of deaths in Wave 2. A higher proportion of completely vaccinated individuals were admitted during this period. Most of our breakthrough patients were vaccinated with mRNA vaccines, which have approximately 90% effectiveness against hospitaliza-

**Table 2**  
Laboratory assays stratified by wave.

	Wave 1 (n=653)	Wave 2 (n=665)	Total (N=1318)	P-value <sup>a</sup>
Activated partial thromboplastin Time				<b>0.005</b>
N	334	469	803	
Median (range)	31.0 (17.0–225.0)	30.0 (17.0–300.0)	30.0 (17.0–300.0)	
C-reactive protein				<b>&lt;0.001</b>
N	617	639	1256	
Median (range)	53.8 (1.5–400.0)	71.2 (1.5–400.0)	63.2 (1.5–400.0)	
Creatinine				<b>0.003</b>
N	626	623	1249	
Median (range)	1.0 (0.3–12.5)	0.9 (0.2–20.3)	1.0 (0.2–20.3)	
D-dimer				0.79
N	615	636	1251	
Median (range)	825.0 (110.0–42000.0)	841.5 (110.0–42000.0)	831.0 (110.0–42000.0)	
Ferritin				<b>&lt;0.001</b>
N	604	611	1215	
Median (range)	365.5 (9.0–17569.0)	513.0 (5.0–30714.0)	433.0 (5.0–30714.0)	
Fibrinogen				<b>&lt;0.001</b>
N	391	514	905	
Median (range)	513.0 (108.0–1000.0)	567.0 (76.0–1000.0)	543.0 (76.0–1000.0)	
Interleukin-6				<b>0.002</b>
N	541	599	1140	
Median (range)	38.0 (1.0–3543.0)	44.0 (1.0–4500.0)	41.0 (1.0–4500.0)	
International normalized ratio				<b>0.010</b>
N	586	619	1205	
Median (range)	1.2 (0.8–5.4)	1.2 (0.9–5.2)	1.2 (0.8–5.4)	
Lactate dehydrogenase				<b>&lt;0.001</b>
N	597	621	1218	
Median (range)	268.0 (87.0–25000.0)	347.0 (65.0–3360.0)	299.0 (65.0–25000.0)	
Lymphocytes, absolute				0.18
N	587	619	1206	
Median (range)	0.9 (0.1–94.1)	0.9 (0.1–105.1)	0.9 (0.1–105.1)	
Mean platelet volume				0.27
N	605	643	1248	
Median (range)	10.3 (8.0–14.2)	10.2 (8.0–14.7)	10.2 (8.0–14.7)	
Neutrophils, percentage				<b>&lt;0.001</b>
N	587	619	1206	
Median (range)	74.8 (5.7–96.2)	78.3 (3.7–96.6)	76.6 (3.7–96.6)	
Neutrophils, absolute				<b>0.007</b>
N	587	619	1206	
Median (range)	4.5 (0.3–32.3)	5.2 (0.6–23.1)	4.8 (0.3–32.3)	
Platelet count				0.18
N	612	647	1259	
Median (range)	189.0 (2.0–1120.0)	195.0 (4.0–667.0)	193.0 (2.0–1120.0)	
Procalcitonin				0.42
N	590	631	1221	
Median (range)	0.1 (0.0–96.9)	0.1 (0.0–140.6)	0.1 (0.0–140.6)	
Prothrombin time				<b>&lt;0.001</b>
N	586	619	1205	
Median (range)	13.2 (9.5–62.4)	12.6 (9.7–58.0)	12.9 (9.5–62.4)	

Laboratory assays at the first test during a patient's admission. For values below the lower limit of detection, values were imputed to half the distance between 0 and the lower limit. For values above the upper limit of detection, values were winsorized at the upper limit.

<sup>a</sup> P-values arise from Kruskal-Wallis rank-sum tests. Values in bold formatting are statistically significant ( $p < 0.05$ ).

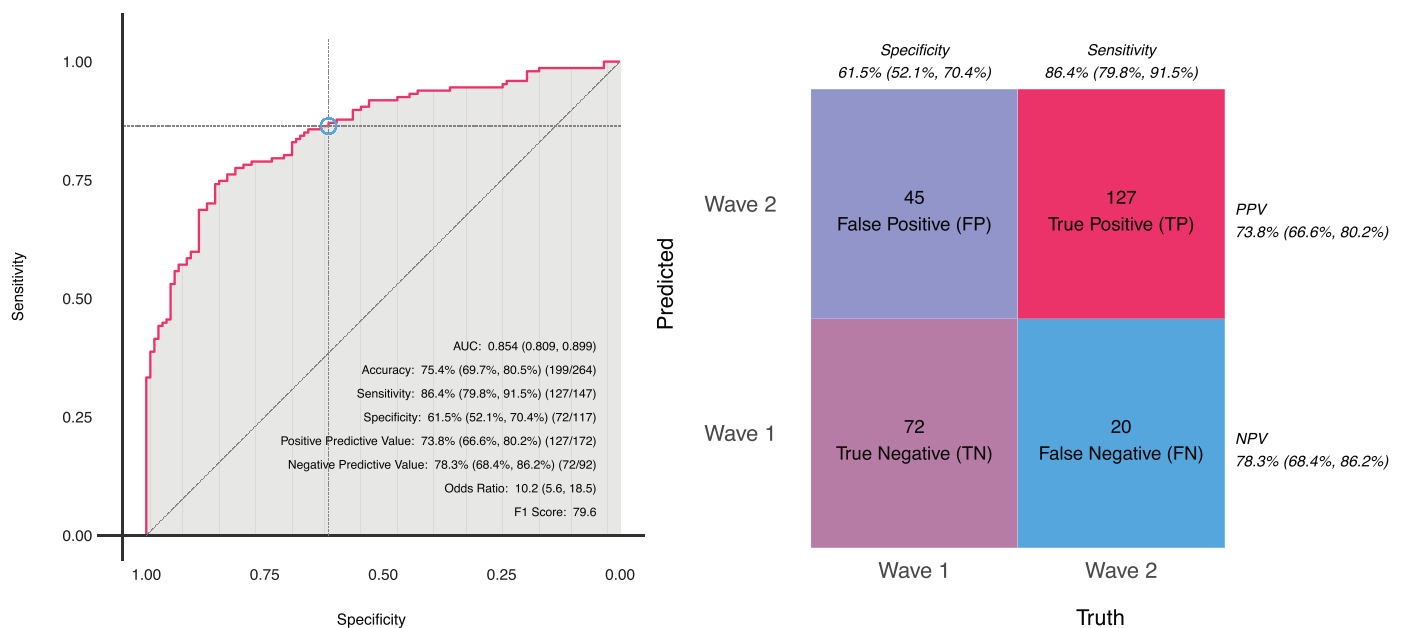
tion and death (Nasreen et al., 2022; Sheikh et al., 2021b). Besides a higher cumulative number of vaccinated individuals during the emergence of the Delta variant, our analysis did not account for changes in the treatment protocols during the pandemic.

Over the 2 years of the COVID-19 pandemic, the healthcare system has realized that it needs to adapt to the various SARS-CoV-2 variants based on the clinical characteristics of each strain, the population it involves, and the likelihood of mortality or morbidity. The vaccine's effectiveness also plays a huge role in the outcomes of these patients. Early on in the pandemic, healthcare's ability to handle patients' needs other than for COVID-19 was negatively impacted. With prediction modeling of future variants, decisions to maintain healthcare throughput can be better planned. The future of COVID-19 care, similar to the overall trend in healthcare, is personalization. The need of the hour is a targeted treatment to a particular phenotype of a patient based on the affecting genotype and the comorbidities of the patient. With genotyping of the COVID-19 variant not readily available, a model similar to what we demon-

strated can quickly help identify the variant affecting the patient. We hope to stratify the model further with data from the Omicron variant.

### Limitations

One of the central weaknesses of the study is the lack of a definitive classification of the variants that infected the hospitalized patients over the study period. This was in part due to the retrospective nature of the study and the evolving adoption of genomic sequencing for the variant. Thus, we have leveraged HHS.gov prevalence data to classify which of the SARS-CoV-2 variants predominated in given periods. Those data reflect variants sequenced from a given region comprising multiple states. Different percentages of variants can exist within a region but are presumed to be insignificant. Selection bias because of the study's single-center nature and misclassification bias of the variants in the wave could exist. We attempted to address these limitations by including all



**Figure 4.** Model performance and diagnostic summaries for the gradient boosting model.

The left panel shows a receiver operating characteristic curve along with associated diagnostic metrics. The blue circle represents the selected threshold, which was determined by the optimal F1 score in the training data. The right panel is a confusion matrix displaying false and true negatives/positives and associated metrics of specificity, sensitivity, negative predictive probability (NPV), and positive predictive probability (PPV). In all cases, metrics are associated with test data that were not used during model development or selection.

patients within the study period. Furthermore, because of the retrospective nature of this study, we relied on the records documented directly into the electronic health record. Although developed using data partitioned off for model testing (“validation”), the GBM was only trained on a single site without a separate prospective validation study. In addition, the model is currently limited in that it was trained only to discriminate the predominantly pre-Delta variants from a predominantly Delta strain. Finally, traditional diagnostic summary metrics do not account for misclassification bias; additional thresholds could be explored to minimize metrics such as false-negative rates. Further research will be required to determine whether the use of simple patient characteristics can readily identify changes in the predominant variant. Notably, shifts in predominant SARS-CoV-2 variants would likely be associated with model drift and a decrease in model performance, necessitating a classification model that is not strictly binary.

## Conclusion

The principal finding of our study is that a selection of readily obtainable laboratory studies and patient characteristics can be used to differentiate between cases of patients with SARS-CoV-2 infection hospitalized during a pre-Delta-predominant wave versus a Delta-predominant wave. The importance of this finding is that it may provide a future approach to developing simple statistical monitoring systems for future waves of infections. This may address the real-world challenges of sequencing the variant and adapting treatment accordingly. If a shift in patient characteristics is detected using readily obtainable data, genomic sequencing of variants could be expedited and prioritized. A significant genotypical shift resulting in VOCs may be readily apparent secondary to increased cases in the community and later seen as a wave in hospital admissions. As in-hospital COVID-19 cases decrease over time because of increased vaccine effectiveness or natural immunity, less clinically impactful genomic mutations, and improvements in outpatient treatments, a predictive model such as ours may com-

pare a patient’s phenotypical characteristics with known variants and treat them accordingly.

## CRediT authorship contribution statement

**Shivang Bhakta:** Conceptualization, Writing – review & editing, Formal analysis, Supervision, Project administration. **Devang K. Sanghavi:** Conceptualization, Writing – review & editing, Investigation, Supervision, Project administration. **Patrick W. Johnson:** Writing – original draft, Methodology, Writing – review & editing, Investigation, Data curation, Formal analysis. **Katie L. Kunze:** Writing – original draft, Methodology, Writing – review & editing, Investigation, Data curation, Formal analysis. **Matthew R. Neville:** Methodology, Writing – review & editing, Investigation, Data curation. **Hani M. Wadei:** Methodology, Writing – review & editing, Investigation, Data curation. **Wendelyn Bosch:** Conceptualization, Writing – review & editing, Methodology. **Rickey E. Carter:** Conceptualization, Methodology, Writing – review & editing, Data curation. **Sadia Z. Shah:** Conceptualization, Writing – review & editing, Supervision. **Benjamin D. Pollock:** Methodology, Writing – review & editing, Investigation, Data curation, Formal analysis. **Sven P. Oman:** Conceptualization, Writing – review & editing, Supervision. **Leigh Speicher:** Conceptualization, Writing – review & editing, Methodology. **Jason Siegel:** Conceptualization, Writing – review & editing, Methodology, Data curation. **Claudia R. Libertin:** Writing – review & editing, Investigation, Data curation, Supervision. **Mark W. Matson:** Conceptualization, Methodology, Writing – review & editing, Data curation. **Pablo Moreno Franco:** Conceptualization, Writing – review & editing, Visualization, Investigation, Supervision, Project administration. **Jennifer B. Cowart:** Conceptualization, Methodology, Writing – review & editing, Visualization, Investigation, Supervision, Project administration.

## Declarations of competing interest

The authors have no competing interests to declare.



## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Ethical approval

Approval was not required.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ijid.2022.04.050](https://doi.org/10.1016/j.ijid.2022.04.050).

## References

- Centers for Disease Control and Prevention, COVID data tracker. <https://covid.cdc.gov/covid-data-tracker/#datatracker-home>, 2021 (accessed 6 December 2021).
- Coronavirus Resource Center, Johns Hopkins University & Medicine. <https://coronavirus.jhu.edu/map.html>, 2022 (accessed 14 February 2022).
- Crawford DC, Williams SM. Global variation in sequencing impedes SARS-CoV-2 surveillance. *PLoS Genet* 2021;17.
- Davies NG, Abbott S, Barnard RC, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 2021;372:eabg3055.
- Fisman DN, Tuite AR. Evaluation of the relative virulence of novel SARS-CoV-2 variants: a retrospective cohort study in Ontario, Canada. *CMAJ* 2021;193:E1619–25.
- GOV.UK Coronavirus (COVID-19), SARS-CoV-2 variants of concern and variants under investigation in England: technical briefing. <https://www.gov.uk/government/publications/investigation-of-sars-cov-2-variants-technical-briefings>, 2021 (accessed 1 December 2021).
- Henry BM, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. *Clin Chem Lab Med* 2020;58:1021–8.
- Khateeb J, Li Y, Zhang H. Emerging SARS-CoV-2 variants of concern and potential intervention approaches. *Crit Care* 2021;25:244.
- Lopez-Castaneda S, Garcia-Larragoiti N, Cano-Mendez A, et al. Inflammatory and prothrombotic biomarkers associated with the severity of COVID-19 infection. *Clin Appl Thromb Hemost* 2021;27.
- Luo CH, Morris CP, Sachithanandham J, et al. Infection with the SARS-CoV-2 Delta variant is associated with higher recovery of infectious virus compared to the Alpha variant in both unvaccinated and vaccinated individuals. *Clin Infect Dis* 2021:ciab986.
- Malik P, Patel U, Mehta D, et al. Biomarkers and outcomes of COVID-19 hospitalisations: systematic review and meta-analysis. *BMJ Evid Based Med* 2021;26:107–8.
- Nasreen S, Chung H, He S, et al. Effectiveness of COVID-19 vaccines against symptomatic SARS-CoV-2 infection and severe outcomes with variants of concern in Ontario. *Nat Microbiol* 2022;7:379–85.
- Ong SWX, Chiew CJ, Ang LW, et al. Clinical and virological features of SARS-CoV-2 variants of concern: a retrospective cohort study comparing B.1.1.7 (Alpha), B.1.315 (Beta), and B.1.617.2 (Delta). *Clin Infect Dis* 2021:ciab721.
- Paul P, France AM, Aoki Y, et al. Genomic surveillance for SARS-CoV-2 variants circulating in the United States, December 2020–May 2021. *MMWR Morb Mortal Wkly Rep* 2021;70:846–50.
- Poggiali E, Zaino D, Immovilli P, et al. Lactate dehydrogenase and C-reactive protein as predictors of respiratory failure in COVID-19 patients. *Clin Chim Acta* 2020;509:135–8.
- Sheikh A, McMenemy J, Taylor B, Robertson C. Public Health Scotland and the EAVE II Collaborators. SARS-CoV-2 Delta VOC in Scotland: demographics, risk of hospital admission, and vaccine effectiveness. *Lancet* 2021;397:2461–2.
- Sheikh A, Robertson C, Taylor B. BNT162b2 and ChAdOx1 nCoV-19 vaccine effectiveness against death from the Delta variant. *N Engl J Med* 2021;385:2195–7.
- Singh J, Rahman SA, Ehtesham NZ, Hira S, Hasnain SE. SARS-CoV-2 variants of concern are emerging in India. *Nat Med* 2021;27:1131–3.
- Sui J, Noubouossie DF, Gandotra S, Cao L. Elevated plasma fibrinogen is associated with excessive inflammation and disease severity in COVID-19 patients. *Front Cell Infect Microbiol* 2021;11.
- Temesgen Z, Burger CD, Baker J, et al. Lenzilumab in hospitalised patients with COVID-19 pneumonia (LIVE-AIR): a phase 3, randomised, placebo-controlled trial. *Lancet Respir Med* 2022;10:237–46.
- Twohig KA, Nyberg T, Zaidi A, et al. Hospital admission and emergency care attendance risk for SARS-CoV-2 Delta (B.1.617.2) compared with Alpha (B.1.1.7) variants of concern: a cohort study. *Lancet Infect Dis* 2022;22:35–42.
- US Food & Drug Administration, Coronavirus (COVID-19) update: FDA takes additional actions on the use of a booster dose for COVID-19 vaccines. <https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-takes-additional-actions-use-booster-dose-covid-19-vaccines>, 2021 (accessed 8 December 2021).
- Verma JS, Libertin CR, Gupta Y, et al. Multi-cellular immunological interactions associated with COVID-19 infections. *Front Immunol* 2022;13.
- World Health Organization, Tracking SARS-CoV-2 variants. <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>, 2021 (accessed 10 February 2021).