AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

Research and Applications

Using sequence clustering to identify clinically relevant subphenotypes in patients with COVID-19 admitted to the intensive care unit

Wonsuk Oh [1,2,3,4], Pushkala Jayaraman [2,3,4], Ashwin S. Sawant [5], Lili Chan[4,5,6,7],
Matthew A. Levin[2,3,8], Alexander W. Charney[2,9,10], Patricia Kovatch[2,11],
Benjamin S. Glicksberg [1,2,3], and Girish N. Nadkarni[1,2,3,4,5,6,7]

[1]Hasso Plattner Institute of Digital Health, Icahn School of Medicine at Mount Sinai, New York, New York, USA, [2]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA, [3]Mount Sinai Clinical Intelligence Center, Icahn School of Medicine at Mount Sinai, New York, New York, USA, [4]Division of Data Driven and Digital Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA, [5]Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA, [6]Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA, [7]Charles Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA, [8]Department of Anesthesiology, Perioperative and Pain Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA, [9]Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, New York, USA, [10]Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, New York, USA and [11]Department of Pharmacological Science, Icahn School of Medicine at Mount Sinai, New York, New York, USA

Corresponding Author: Girish N. Nadkarni, MD, MPH, Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1243, New York, NY 10029, USA (girish.nadkarni@mountsinai.org)

Received 19 July 2021; Revised 1 October 2021; Editorial Decision 31 October 2021; Accepted 2 November 2021

ABSTRACT

Objective: The novel coronavirus disease 2019 (COVID-19) has heterogenous clinical courses, indicating that there might be distinct subphenotypes in critically ill patients. Although prior research has identified these subphenotypes, the temporal pattern of multiple clinical features has not been considered in cluster models. We aimed to identify temporal subphenotypes in critically ill patients with COVID-19 using a novel sequence cluster analysis and associate them with clinically relevant outcomes.

Materials and Methods: We analyzed 1036 confirmed critically ill patients with laboratory-confirmed SARS-COV-2 infection admitted to the Mount Sinai Health System in New York city. The agglomerative hierarchical clustering method was used with Levenshtein distance and Ward's minimum variance linkage.

Results: We identified four subphenotypes. Subphenotype I ($N=233$ [22.5%]) included patients with rapid respirations and a rapid heartbeat but less need for invasive interventions within the first 24 hours, along with a relatively good prognosis. Subphenotype II ($N=418$ [40.3%]) represented patients with the least degree of ailments, relatively low mortality, and the highest probability of discharge from the hospital. Subphenotype III ($N=259$ [25.0%]) represented patients who experienced clinical deterioration during the first 24 hours of intensive care unit admission, leading to poor outcomes. Subphenotype IV ($N=126$ [12.2%]) represented an acute respiratory

distress syndrome trajectory with an almost universal need for mechanical ventilation.

Conclusion: We utilized the sequence cluster analysis to identify clinical subphenotypes in critically ill COVID-19 patients who had distinct temporal patterns and different clinical outcomes. This study points toward the utility of including temporal information in subphenotyping approaches.

INTRODUCTION

Coronavirus disease 2019 (COVID-19)[1] is a novel respiratory disease, leading to over 33 million confirmed cases with 0.6 million deaths in the United States by May 2021.[2] Efforts to reduce the burden of COVID-19 and its complications include diagnostic and prognostic models,[3] treatments,[4–9] and vaccines under emergency use authorization.[10–12] However, in-hospital mortality for the subset of hospitalized patients who need mechanical ventilation still exceeds 50%.[13] Among patients hospitalized with COVID-19, even those with similar baseline characteristics may follow different clinical trajectories and have different outcomes.[14–19]

Subphenotypes are subgroups of a disease with distinct biomarkers even if patients appear clinically similar at their early stages.[20–23] More research is needed to unveil novel COVID-19 subphenotypes, elucidate their pathophysiology, and investigate whether subphenotype-specific treatment approaches are needed. While many studies have focused on identifying novel subphenotypes using features available at the baseline[24–27] or a single temporal feature,[28,29] most of them overlook overall temporality.

Sequence cluster analysis[30,31] is a data mining technique to find groups in a sequential database such that each group contains similar sequences. This technique can be a rational approach to identify subphenotypes characterized by distinct patterns of disease progression over a period of time. Two approaches are widely applied: sequence similarity distance metrics and feature engineering. Sequence similarity distance metrics are a class of distance metrics that measure similarity by counting the number of operations required to transform one sequence to the other. Edit distance[32–34] and substitution matrix[34,35] are two well-known examples, primarily applied in DNA/RNA sequencing.[34] The resultant distance matrix can be used in clustering methods. Unlike the sequence similarity distance metrics, feature engineering in the context of sequence cluster analysis transforms spatiotemporal features into different (spatial) subspaces so that existing distance metrics can measure the similarity of the sequences. n-gram[36] and sequential pattern mining[37] are two examples. While n-gram is widely applied in natural language processing,[36] sequential pattern mining is used in learning subsequences to predict the next event.[38–40] The resultant features can be applied to the classical clustering, yet the feature space's sparseness can be challenging.

We aimed to derive temporal subphenotypes using sequence cluster analysis in critically ill patients with COVID-19. Figure 1 illustrates the workflow of this study. We employed the agglomerative hierarchical clustering method with Levenshtein distance, sequence similarity distance metrics, and Ward's minimum variance linkage on biomarkers and treatments during the first 24 hours of intensive care unit (ICU) stay to derive subphenotypes. We evaluated the association of the subphenotypes with two clinical outcomes: in-hospital mortality and hospital discharge at 30 days. We explored the temporal characteristics of the subphenotypes. Finally, we checked the robustness of the subphenotypes to sampling and the choice of the clustering method.

MATERIALS AND METHODS

Study design and participants

We retrospectively reviewed data from adult patients with laboratory-confirmed COVID-19 admitted to ICU at five Mount Sinai Health System (MSHS) hospitals in New York city between March and December 2020. A confirmed case of COVID-19 was defined by reverse transcription-polymerase chain reaction positivity for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from a nasopharyngeal swab sample, which is considered the gold standard for COVID-19 diagnosis.[41] We included patients at least 18 years of age with laboratory-confirmed SARS-COV-2 infection within 14 days prior or 24 hours posterior to ICU admission. We excluded patients with ICU admission dates outside the study period or ICU stay that was less than 24 hours. We also excluded those without recorded height and weight within the preceding 3 years, without a blood pressure recording during the first 24 hours of ICU stay, or without a recorded date of death or discharge from the hospital.

Data collection and measurements

We extracted electronic health record (EHR) data, specifically socio-demographic information (age, sex, race, and ethnicity). We also extracted discretized clinical/laboratory measurements and treatment, which are commonly assessed and intervened in the ICU setting[42] as listed in Table 1. Finally, we recorded in-hospital mortality and hospital discharge at 30 days as outcomes, to assess the usability of identified clusters.

Data preparation

We transformed biomarkers and treatment administrations during the first 24 hours of the ICU admission into a sequence consisting of 16 non-overlapping interval slots, each of which was 1.5 hours long and one of 10 distinct statuses. We did this in the following three steps.

First, we excluded biomarkers and treatments having less than 10% prevalence in the cohort. Table 1 lists the 10 biomarkers and 7 treatments, which constituted the set of 17 features.

Second, we created a series of 16 non-overlapping interval slots, each 1.5-hour long, to cover the first 24 hours of the ICU stay. For each feature, we marked 1 if the feature was present in a time-window, and 0 if it was not. We arrived at the 1.5-hour window size based on our subjective assessment of the duration of action of drugs, and treatment administration intervals available in our dataset.

Third, we applied dimensionality reduction techniques to the 17 features to obtain a 10-level variable for sequence cluster analysis.
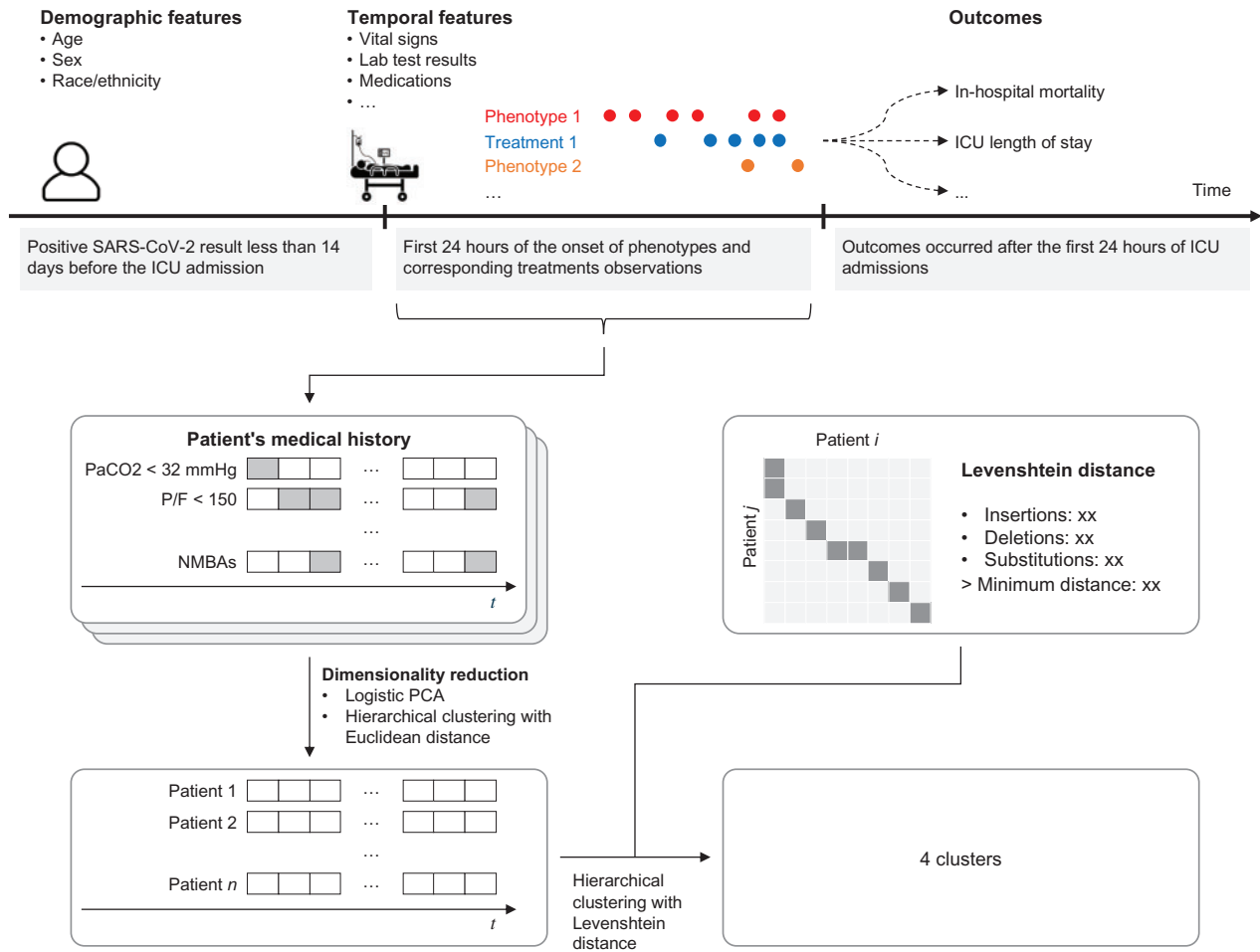
Figure 1. Study flow diagram. Both demographic features and those expected to evolve over time were used to identify subphenotypes. Outcomes were analyzed by subphenotype. Subphenotypes were derived using the agglomerative hierarchical clustering method with Levenshtein distance and Ward's minimum variance linkage.

This was done in two steps. We started by using Logistic principal component analysis[56] with 5-fold cross-validation to remove collinearity problems and to identify 10 principal components. We then used the agglomerative hierarchical clustering method with Euclidean distance and Ward's minimum variance linkage on the principal components to obtain a 10-level categorical variable. We used a consensus of 26 indices[57] to determine the optimal number of levels.

Subphenotype derivation

We used the agglomerative hierarchical clustering method with Levenshtein distance and Ward's minimum variance linkage because it needs only a distance matrix and not measurements themselves. We determined the number of resulting subphenotypes by average Silhouette width,[58] the slope changes in Gap statistic,[59] Clest,[60] elbow method on the total within the sum of squares,[61] and visual evaluation of the dendrogram.

Statistical analysis

We calculated frequencies and percentages for biomarkers and treatments over the first 24 hours of ICU stay and analyzed differences across the subphenotypes using the chi-square test. We used Kaplan–Meier survival analysis to compare time to mortality and

hospital discharge among subphenotypes. Survival time for 30-day in-hospital mortality was defined as 24 hours after the ICU admission to either date of in-hospital death or last known contact within 31 days after the ICU admission. Similarly, survival time for 30-day hospital discharge was defined as 24 hours after the ICU admission to either date of hospital discharge or last known contact within 31 days after the ICU admission or 31 days of those who experienced in-hospital death. We also explored temporal variation in the prevalence of the subphenotypes on a monthly basis from March to December 2020.

We evaluated the robustness of the described subphenotypes to sampling and choice of clustering algorithm in the following manner. First, we evaluated sensitivity to sampling by using the holdout method to randomly split patients into mutually exclusive training (80%) and test (20%) sets. The k-nearest neighbor classifier with Levenshtein distance was trained on the training set. The re-derived subphenotypes were assigned using the k-nearest neighbor classifier on the test set. We used heatmaps to visualize how subphenotypes and re-derived subphenotypes on the test coincided with each other. Second, we evaluated sensitivity to clustering methods by re-deriving subphenotypes using a k-medoid method with Levenshtein distance and using Heatmap to evaluate how they coincided with subphenotypes from the agglomerative hierarchical clustering method.

Table 1. Inclusion or exclusion of biomarkers and treatment

Category	Biomarker	Retained
Shock[43–45]	SBP <90 mmHg	✓
	HR >90/min	✓
	RR >20/min	✓
	Body temperature <96.8°F	✗
	$PaCO_2$ <32 mmHg	✓
	WBC >12 000/mm³	✓
	Band cell count >10%	✗
	Serum lactate >4 mmol/l	✗
HTN-C[46,47]	SBP >180 mmHg	✗
	DBP >120 mmHg	✗
ARDS[48,49]	P/F ratio <150	✓
DKA[50,51]	Glucose >250 mg/dl	✓
	Arterial pH <7.25	✓
	Serum bicarbonate <15 mmol/l	✗
	AG >12 mmol/l	✓
AKI[52,53]	Increase in Cr to two times the baseline	✓
	Cr >4.0 mg/dl	✗
ALF[54]	INR >2	✗
Category	Treatment	Retained
Shock[43–45]	Red blood cell transfusion	✓
	Vasoactive agents	✓
HTN-C[55]	Intravenous vasodilators	✗
	Loop diuretics	✓
ARDS[48,49]	Mechanical ventilation	✓
	Neuromuscular-blocking agents	✓
DKA[50,51]	Insulin	✓
AKI[52,53]	Hemodialysis	✓

Abbreviations: AG: anion gap; AKI: acute kidney injury; ALF: acute liver failure; ARDS: acute respiratory distress syndrome; Cr: serum creatinine; DKA: diabetic ketoacidosis; HR: heart rate; HTN-C: hypertensive crisis; INR: international normalized ratio; P/F: PaO2/FiO2 (fraction of inspired oxygen); PaCO2: partial pressure of carbon dioxide in arterial blood; RR: respiratory rate; SBP: systolic blood pressure; DBP: diastolic blood pressure; WBC: white blood cell count.

We performed all statistical analyses using R software version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria).[62] Source code is available at https://github.com/Nadkarni-Lab/ohw_jamia_2021.

RESULTS

Study population

We considered 1702 patients aged at least 18 and either admitted to an intensive care unit within 14 days of testing positive for SARS-COV-2 or having their first positive SARS-COV-2 test within 24 hours of ICU admission. We excluded patients with ICU admissions dates outside the study period (N = 30), or ICU stays of less than 24 hours (N = 122). We also excluded those without a blood pressure recording during the first 24 hours of ICU stay (N = 105), without recorded height and weight within the preceding 3 years (N = 69), or without a recorded date of death or discharge from the hospital (N = 2). The remaining 1036 patients formed our final cohort.

The mean age was 63 years (95% confidence interval (CI): 54, 74). Thirty-seven percentage were female, 20% were Black, and 27% were Hispanic. At 30 days, 44.8% of patients had died and 41.1% had been discharged alive from the hospital. We provide descriptive statistics for the cohort in Table 2, under the heading "Full cohort."

Characteristics of subphenotypes

We identified four distinct subphenotypes. Table 2 shows patient characteristics at the time of ICU admission and during the first 24 hours of ICU stay by subphenotype. Figure 2 shows a dendrogram of the resulting cluster hierarchy. Figure 3 shows the cumulative prevalence of abnormal biomarkers and treatments of each subphenotype. The x-axis denotes the time in hours after the ICU admission, and the y-axis represents the cumulative prevalence. The upper two rows depict the 10 biomarkers, and the lower two rows depict the 7 treatments. Figure 4A and B shows survival curves for in-hospital death and discharge from the hospital, respectively, for individuals with each subphenotype.

Subphenotype I (SP-I) included 233 patients in the cohort. These patients tended to be White (N = 72 [30.9%]) and non-Hispanic (N = 185 [79.4%]). They were characterized by being tachypneic (respiratory rate (RR) >20; N = 97 [41.6%]) while not yet being mechanically ventilated (N = 12 [5.2%]) at the time of ICU admission. At 24 hours, almost all these patients were tachypneic (N = 228 [97.9%]), even though the number needing mechanical ventilation remained low (N = 40 [17.2%]). In particular, 24 hours after the ICU admission, patients with SP-I experienced 5 times higher systolic blood pressure <90 mmHg (17.2% vs. 3.3%), 2 times higher HR >90 BPS (62.2% vs. 29.7%), 1.7 times higher RR >20 BPS (97.9% vs. 58.9%), 2 times higher partial pressure of carbon dioxide/fraction of inspired oxygen (P/F) < 150 (37.8% vs. 18.9%), 1.8 times higher vasoactive agents administration (18.5% vs. 10.5%), 1.5 times higher loop diuretics admiration (17.6% vs. 12.0%), and 3 times higher mechanical ventilation administration (17.2% vs. 5.7%) than with subphenotype II (SP-II). SP-I had the highest probability of survival (0.449 [CI 0.374, 0.538]) and the second-highest probability of discharge from the hospital (0.598 [CI 0.507, 0.673]) at 30 days, although the differences were not statistically significant.

SP-II included 418 (40.3%) patients in the cohort. These patients were younger (mean age 61.5 [IQR 53.0, 73.0]) and more likely to be Black (N = 96 [23.0%]) than the other three subphenotypes. They generally had the lowest prevalence of unfavorable physiological biomarkers during the ICU stay. They were also less likely to need vasoactive agents or mechanical ventilation. SP-II showed the highest probability of hospital discharge at 30 days (0.745 [CI 0.686, 0.794]).

Subphenotype III (SP-III) included 259 (25.0%) patients in the cohort. These patients tended to be older (mean age 65.3 [IQR 58.0, 73.0]) and were more likely to be males (N = 169 [65.3%]) than the other three subphenotypes. SP-III had the highest incidence of shock at ICU admission, and this trend continued over the first 24 hours of ICU stay. SP-III also had the highest requirement for support with vasoactive agents and over half of these patients needed mechanical ventilation, with the need for these interventions increasing over the first 24 hours. SP-III showed the second-lowest probability of survival (0.722 [0.649, 0.779]), although the difference was not statistically significant. The proportion discharged from the hospital (0.487 [CI 0.381, 0.574]) was comparable to SP-I and SP-IV.

Subphenotype IV (SP-IV) included 126 (12.2%) patients in the cohort and had a relatively larger proportion of patients of Hispanic or Latino (N = 42 [33.3%]) ethnicity. A distinct characteristic was

Table 2. Characteristics of cohort and subphenotypes

Feature	Full cohort	SP-I	SP-II	SP-III	SP-IV	P value
Patients, n (%)	1036 (100)	233 (22.5)	418 (40.3)	259 (25.0)	126 (12.2)	–
Demographics						
Age, years, n (IQR)	63.3 (54.0, 74.0)	64.8 (57.0, 76.0)	61.5 (53.0, 73.0)	65.3 (58.0, 73.0)	62.2 (53.0, 73.8)	0.014
Sex, male, n (%)	651 (62.8)	144 (61.8)	261 (62.4)	169 (65.3)	77 (61.1)	0.815
Race, n (%)						
White	273 (26.4)	72 (30.9)	100 (23.9)	77 (29.7)	24 (19.0)	0.075
Black or African American	207 (20.0)	37 (15.9)	96 (23.0)	55 (21.2)	19 (15.1)	
American Indian/Alaska native	2 (0.2)	1 (0.4)	0 (0.0)	1 (0.4)	0 (0.0)	
Asian	54 (5.2)	13 (5.6)	20 (4.8)	16 (6.2)	5 (4.0)	
Hawaiian native and Pacific islander	2 (0.2)	0 (0.0)	1 (0.2)	1 (0.4)	0 (0.0)	
Some others	498 (48.1)	110 (47.2)	201 (48.1)	109 (42.1)	78 (61.9)	
Ethnicity, n (%)						
Hispanic or Latino	282 (27.2)	48 (20.6)	123 (29.4)	69 (26.6)	42 (33.3)	0.035
Not Hispanic or Latino	754 (72.8)	185 (79.4)	295 (70.6)	190 (73.4)	84 (66.7)	
NYC boroughs						
Bronx	59 (5.7)	14 (6.0)	27 (6.5)	11 (4.2)	7 (5.6)	0.001
Brooklyn	267 (25.8)	67 (28.8)	115 (27.5)	64 (24.7)	21 (16.7)	
Manhattan	360 (34.7)	74 (31.8)	137 (32.8)	83 (32.0)	66 (52.4)	
Queens	257 (24.8)	48 (20.6)	109 (261)	76 (29.3)	24 (19.0)	
Staten Island	5 (0.5)	1 (0.4)	4 (0.10)	0 (0.0)	0 (0.0)	
Not applicable	88 (8.5)	29 (12.4)	27 (0.62)	25 (9.7)	8 (6.3)	
Survival probability						
10 days	0.71 (0.68, 0.74)	0.77 (0.71, 0.83)	0.81 (0.77, 0.85)	0.54 (0.49, 0.61)	0.64 (0.56, 0.73)	<0.001
20 days	0.49 (0.46, 0.53)	0.56 (0.49, 0.64)	0.56 (0.50, 0.63)	0.36 (0.30, 0.43)	0.42 (0.33, 0.52)	
30 days	0.39 (0.35, 0.43)	0.45 (0.37, 0.54)	0.43 (0.36, 0.51)	0.28 (0.22, 0.35)	0.38 (0.29, 0.48)	
Hospital discharge probability						
10 days	0.25 (0.22, 0.28)	0.22 (0.16, 0.28)	0.38 (0.33, 0.43)	0.12 (0.07, 0.17)	0.06 (0.01, 0.11)	<0.001
20 days	0.49 (0.45, 0.52)	0.46 (0.38, 0.53)	0.63 (0.57, 0.68)	0.37 (0.28, 0.45)	0.26 (0.14, 0.36)	
30 days	0.62 (0.58, 0.66)	0.60 (0.51, 0.67)	0.75 (0.69, 0.79)	0.49 (0.38, 0.57)	0.43 (0.29, 0.55)	
Abnormal biomarkers at the ICU admission						
SBP <90 mmHg, n (%)	54 (5.2)	6 (2.6)	3 (0.7)	35 (13.5)	10 (7.9)	<0.001
HR >90 bpm, n (%)	178 (17.2)	54 (23.2)	31 (7.4)	66 (25.5)	27 (21.4)	<0.001
RR >20 bpm, n (%)	273 (26.4)	97 (41.6)	69 (16.5)	73 (28.2)	34 (27.0)	<0.001
$PaCO_2$ <32 mmHg, n (%)	14 (1.4)	2 (0.9)	4 (1.0)	6 (2.3)	2 (1.6)	0.428
WBC >12 000, n (%)	49 (4.7)	7 (3.0)	16 (3.8)	22 (8.5)	4 (3.2)	0.011
P/F <150, n (%)	83 (8.0)	13 (5.6)	15 (3.6)	29 (11.2)	26 (20.6)	<0.001
Glucose >250 mg/dl, n (%)	77 (7.4)	14 (6.0)	26 (6.2)	25 (9.7)	12 (9.5)	0.237
pH <7.25, n (%)	41 (4.0)	2 (0.9)	10 (2.4)	21 (8.1)	8 (6.3)	<0.001
AG >12 mmol/l, n (%)	83 (8.0)	16 (6.9)	25 (6.0)	31 (12.0)	11 (8.7)	0.039
7 days Cr >2 times, n (%)	10 (1.0)	0 (0.0)	3 (0.7)	5 (1.9)	2 (1.6)	0.134
Treatments at the ICU admission						
Red blood transfusion, n (%)	9 (0.9)	1 (0.4)	3 (0.7)	4 (1.5)	1 (0.8)	0.569
Vasoactive agents, n (%)	136 (13.1)	9 (3.9)	8 (1.9)	88 (34.0)	31 (24.6)	<0.001
IV loop diuretics, n (%)	10 (1.0)	2 (0.9)	3 (0.7)	4 (1.5)	1 (0.8)	0.743
Mechanical ventilation, n (%)	159 (15.3)	12 (5.2)	4 (1.0)	70 (27.0)	73 (57.9)	<0.001
Neuromuscular blocker, n (%)	39 (3.8)	4 (1.7)	3 (0.7)	13 (5.0)	19 (15.1)	<0.001
IV insulin, n (%)	31 (3.0)	5 (2.1)	7 (1.7)	12 (4.6)	7 (5.6)	<0.001
Hemodialysis, n (%)	4 (0.4)	1 (0.4)	1 (0.2)	1 (0.4)	1 (0.8)	0.040
Abnormal biomarkers during the first 24 h of ICU admission						
SBP <90 mmHg, n (%)	219 (21.1)	40 (17.2)	14 (3.3)	118 (45.6)	47 (37.3)	0.003
HR >90 bpm, n (%)	519 (50.1)	145 (62.2)	124 (29.7)	175 (67.6)	75 (59.5)	<0.001
RR >20 bpm, n (%)	766 (73.9)	228 (97.9)	246 (58.9)	197 (76.1)	95 (75.4)	<0.001
$PaCO_2$ <32 mmHg, n (%)	125 (12.1)	37 (15.9)	40 (9.6)	39 (15.1)	9 (7.1)	0.236
WBC >12 000, n (%)	447 (43.1)	88 (37.8)	142 (34.0)	158 (61.0)	59 (46.8)	0.050
P/F <150, n (%)	372 (35.9)	88 (37.8)	79 (18.9)	137 (52.9)	68 (54.0)	<0.001
Glucose >250 mg/dl, n (%)	324 (31.3)	74 (31.8)	110 (26.3)	101 (39.0)	39 (31.0)	0.161
pH <7.25, n (%)	192 (18.5)	17 (7.3)	31 (7.4)	107 (41.3)	37 (29.4)	0.008
AG >12 mmol/l, n (%)	571 (55.1)	128 (54.9)	204 (48.8)	74 (67.2)	65 (51.6)	0.007
7 days Cr >2 times, n (%)	184 (17.8)	22 (9.4)	32 (7.7)	96 (37.1)	34 (27.0)	0.010
Treatments during the first 24 h of ICU admission						
Red blood transfusion, n (%)	46 (4.4)	8 (3.4)	13 (3.1)	20 (7.7)	5 (4.0)	0.681
Vasoactive agents, n (%)	406 (39.2)	43 (18.5)	44 (10.5)	230 (88.8)	89 (70.6)	<0.001

(continued)

Table 2. continued

Feature	Full cohort	SP-I	SP-II	SP-III	SP-IV	*P* value
IV loop diuretics, *n* (%)	154 (14.9)	41 (17.6)	50 (12.0)	44 (17.0)	19 (15.1)	0.798
Mechanical ventilation, *n* (%)	340 (32.8)	40 (17.2)	24 (5.7)	150 (57.9)	126 (100.0)	<0.001
Neuromuscular blocker (IV), *n* (%)	193 (18.6)	23 (9.9)	32 (7.7)	86 (33.2)	52 (41.3)	0.001
IV insulin, *n* (%)	162 (15.6)	16 (6.9)	41 (9.8)	75 (29.0)	30 (23.8)	0.039
Hemodialysis, *n* (%)	47 (4.5)	7 (3.0)	17 (4.1)	18 (6.9)	5 (4.0)	0.681

Abbreviations: NYC: New York City, AG: anion gap; Cr: serum creatinine; HR: heart rate; ICU: intensive care unit; P/F: PaO_2/FiO_2 (fraction of inspired oxygen); $PaCO_2$: partial pressure of carbon dioxide in arterial blood; RR: respiratory rate; SBP: systolic blood pressure; SP-I: subphenotype I, SP-II: subphenotype II, SP-III: subphenotype III, SP-IV: subphenotype IV; WBC: white blood cell count, IQR:interquartile range, BPM: beats per minute.
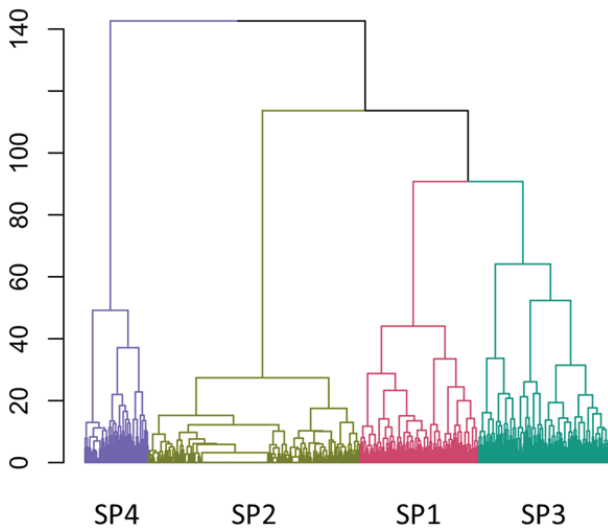

Figure 2. Dendrogram of the agglomerative hierarchical clustering method with Levenshtein distance and Ward's minimum variance linkage.

the high prevalence of respiratory failure. Over half of these patients were mechanically ventilated at the time of ICU admission, and all of them were mechanically ventilated at 24 hours. Associated with this were low P/F ratios and the need for neuromuscular blockade. The hospital discharge rate (0.430 [CI 0.286, 0.545]) at 30 days was similar to SP-I and SP-III.

Temporal (monthly) characteristics of subphenotypes

Figure 5 shows the temporal (monthly) characteristics of subphenotypes. The histogram on top shows the number of patients each month between March and December 2020, and the bar plot on the lower subgraph shows the corresponding percentage of each subphenotype. The number of COVID-19 patients surged during the two waves in March and November 2020, aligned with daily hospitalization trends in New York state.[56] However, the percentage of each subphenotype was not consistent across the months. In particular, the proportion of SP-I increased over the months (20.1% in March and 54.5% in November 2020), while that of SP-II and SP-III decreased (39.8% and 26.9% in March and 27.3% and 9.1% in November 2020, respectively). On the other hand, SP-IV showed a relatively stable prevalence (13.5% in March and 9.1% in November 2020), with small fluctuations.

Robustness of subphenotypes

Sensitivity to the choice of clustering algorithm

Figure 6A shows the heatmap of prevalence co-occurrence of subphenotypes from hierarchical clustering and k-medoid methods. The

x-axis represents re-derived subphenotypes from the k-medoid method, and the *y*-axis represents subphenotypes from the hierarchical clustering method. Overall, SP-I (67.8 %), SP-II (83.7 %), and SP-IV (75.3%) were robust to clustering methods, while SP-III (47.1 %) was not.

Sensitiveness of sampling

Figure 6B shows the heatmap of how the subphenotypes and the re-derived subphenotypes coincided with each other. We derived subphenotypes (*x*-axis) directly from the test set, while the re-derived subphenotypes (*y*-axis) were inferred using the k-nearest neighbor model learned on the training set and then applied to the test set. SP-I (88.2 %), SP-III (64.5 %), and SP-IV (69.0 %) appeared to be quite robust to sampling but SP-II (57.8 %) was less so.

DISCUSSION

We used sequence cluster analysis to identify subphenotypes of critically ill adult patients with COVID-19, based on biomarkers and treatments during the first 24 hours of ICU stay. While other investigators have looked at the derivation of subphenotypes using features available at baseline,[24–27] or time series with a single feature,[28,29] these may overlook temporal patterns that are apparent only when multiple features are considered over a period of time. We identified four subphenotypes with distinct temporal patterns during the first 24 hours and different clinical outcomes at 30 days.

Among those subphenotypes, SP-II consisted of patients with comparatively smaller physiological derangements and minimal need for invasive interventions, with relatively good outcomes. SP-I demonstrated significant physiological derangements but relatively low rates of invasive interventions and good outcomes comparable to SP-II. Thus, SP-I may represent patients with a higher level of physiological reserve at the time of ICU admission. SP-III and SP-IV showed a significant prevalence of hemodynamic instability and respiratory failure needing mechanical ventilation at the time of ICU admission. At the end of 24 hours of ICU care, shock needing vasoactive agents appeared to be more prevalent in the former, and respiratory failure requiring mechanical ventilation was universally needed for patients in the latter. This suggests that exploration of the temporal progressions of clinical features can help identify meaningful subphenotypes.

The relative change in the proportion of different subphenotypes over time may reflect changes in clinical practice over the course of the study period. For example, when the first wave of COVID-19 hit New York city in March 2020, hospital capacity was quickly overwhelmed due to the COVID-19's high infectivity[63] and infection fatality,[64] as well as lack of standard admission criteria and treatment guidelines. Accordingly, the SP-II subphenotype, which was associ-
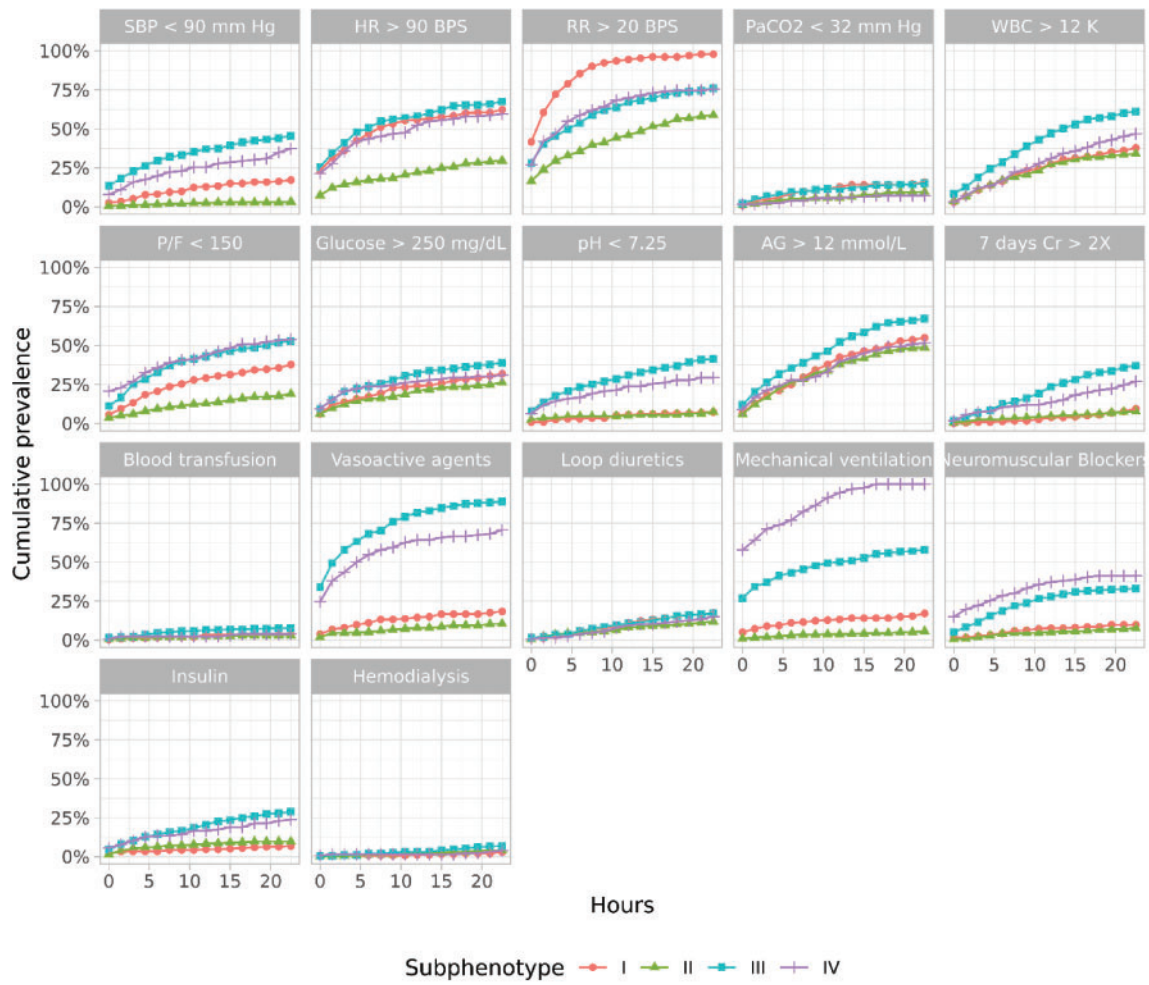
Figure 3. Cumulative incidence of abnormal biomarkers and treatments. AG: serum bicarbonate; Cr: serum creatinine; HR: heart rate; P/F: $PaCO_2/FiO_2$ (fraction of inspired oxygen); $PaCO_2$: partial pressure of carbon dioxide; RR: respiratory rate; SBP: systolic blood pressure; WBC: white blood cell count.
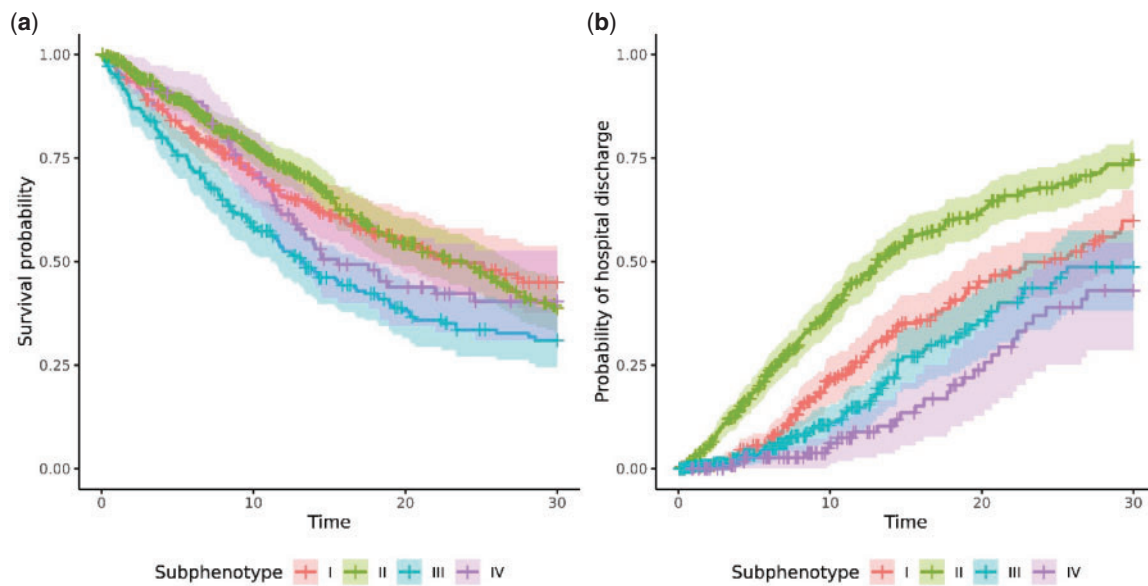
Figure 4. (A) Survival probability of 30-day in-hospital mortality. (B) Probability of 30-day hospital discharge.

ated with a better prognosis, comprised a significant number of ICU admissions during March and April 2020 (39.8% and 43.8%, respectively). However, with new interim guidelines,[65] these patients comprised a smaller proportion of ICU admission during the second wave in November and December 2020 (27.3% and 28.7%, respectively). The interim guidelines may also have worked on the limit of patients developing severe complications during the first 24 hours of the ICU admission. SP-III, which was associated with a high shock and mechanical ventilation rate, decreased from 26.9% and 32.5% in March and April 2020 to 9.1% and 12.6% in November and December 2020. This was accompanied by an increase in the percentage of SP-I, associated with less physiological derangements, less

Figure 5. Temporal characteristics of subphenotypes.

severe complications, and, potentially, better outcomes. Unfortunately, we did not see much change in the prevalence of SP-IV, which was associated with poor outcomes and an almost universal need for mechanical ventilation.

The derivation of these four subphenotypes was possible through sequence cluster analysis. Sequence cluster analysis lets multiple features be considered over a period of time, allowing us to separate subphenotypes even if some subphenotypes appear clinically similar at the admission or by the end of the observation period. We tested whether we could identify similar subphenotypes from nontemporal cluster analysis. We used flattened data with distance metrics for binary vectors and found that the resulting subphenotypes did not coincide with our original subphenotypes and appeared instead to separate patients primarily based on the intensity of the presence of biomarkers and treatment administration. We hypothesize that non-temporal clustering was less effective because biomarkers and treatment administrations during the first 24 hours of ICU stay follow multimodal distributions, causing valuable information to be lost when the data is collapsed into binary vectors. The details of this comparison can be found in the Supplementary Materials.

Model validation is important for machine learning-based studies.[66–68] The use of standardized terminologies decreases the burden of model transferability and facilitates model validation. We used standards like LOINC and CPT codes, and the RxNorm drug vocabulary to obtain data from the EHR. We have made our code freely available at https://github.com/Nadkarni-Lab/ohw_jamia_2021 to encourage reproducibility.

This study is not without limitations. First, we were not able to access some relevant information, like cardiac ejection fraction, electrocardiogram tracings, or radiology images. Second, our cohort of patients was limited to those receiving ICU-level care at five MSHS hospitals in New York city. As a result, our cohort was geographically limited, although it was demographically diverse. Third, we have not quantified the effect of resource availability on the case mix admitted to ICUs over the course of the pandemic. We also did
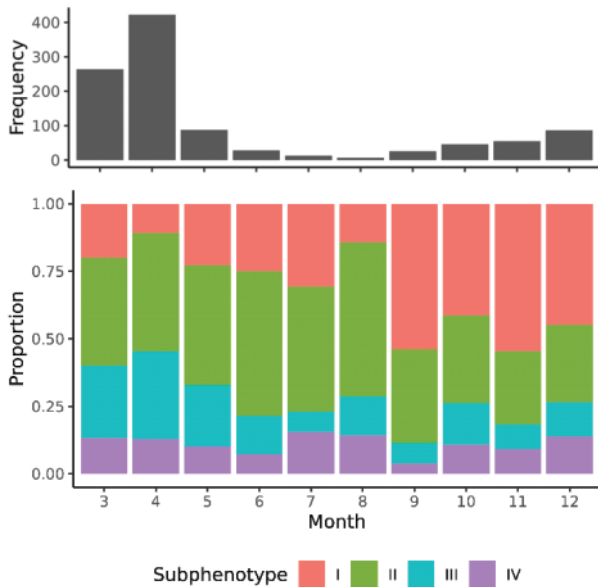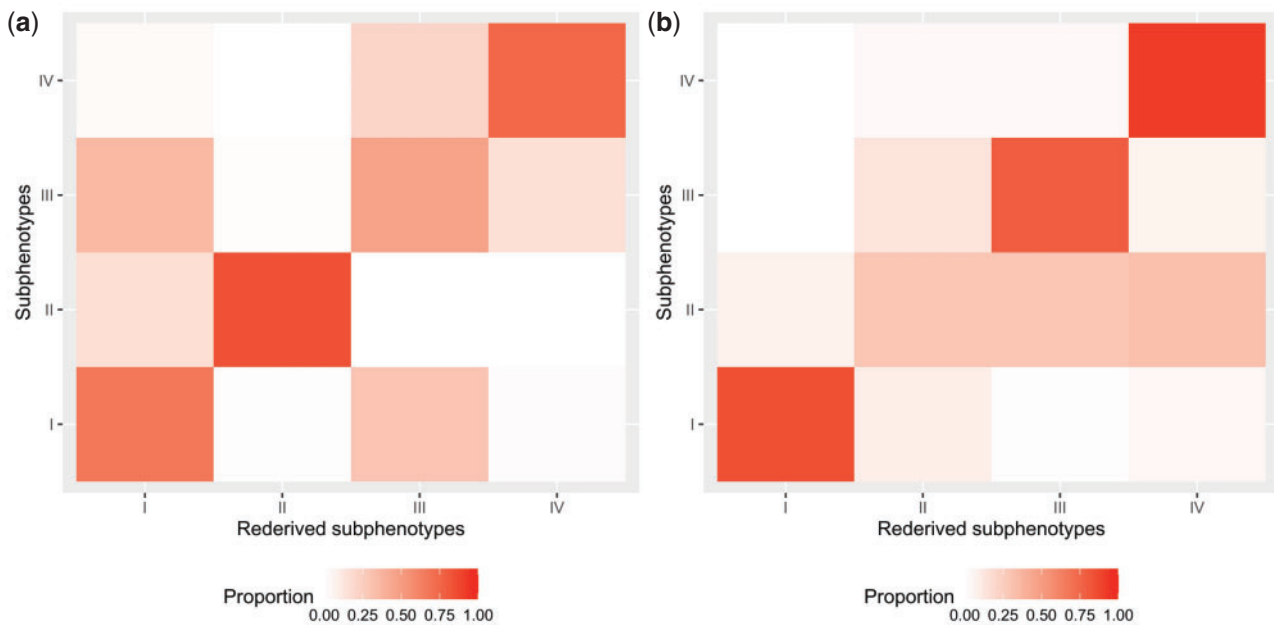

Figure 6. Robustness of subphenotypes: (A) sensitivity to clustering algorithm and (B) sensitivity to sampling.

not consider how the population at risk in New York city changed over the period of the study and the resulting impact on subphenotype prevalence. Fourth, we do not know how well our findings will describe patients after December 2020, as the standard of care for COVID-19 has evolved, different SARS-COV-2 variants have become prevalent, and a substantial proportion of the population has been vaccinated. Assessment of generalizability of the subphenotypes identified in this paper will require validation against cohorts from different time periods and geographical areas; the ability of our method to leverage temporal information for this remains. Fifth, we used data from only the first 24 hours of ICU stay to extract temporal patterns. This was because of our assessment of the clinical importance of the first 24 hours, and also a decrease in cohort size when we tried to extend the observation period. As a result, we may not capture some patterns that only appear later in the ICU course. We are currently conducting a follow-up study on National COVID Cohort Collaborative (N3C) data to study patterns that may emerge with a longer observation period.

CONCLUSION

The four subphenotypes we derived for critically ill patients with COVID-19 were associated with distinct temporal patterns and clinical outcomes and may form the basis of studies exploring subphenotype-specific treatment.

FUNDING

AUTHOR CONTRIBUTORS

WO and GNN designed the study. WO and PJ performed and analyzed the data. WO, ASS, LC, MAL, AWC, PK, BSG, and GNN wrote the paper with input from all authors.

ETHICS APPROVAL

This study has been approved by the Institutional Review Board at the Icahn School of Medicine at Mount Sinai as part of a protocol allowing for access to patient level data (approval no. 19-00951).

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST STATEMENT

GNN is a founder of Renalytix, Pensieve, and Verici and provides consultancy services to AstraZeneca, Reata, Renalytix, Siemens Healthineer, and Variant Bio, and serves a scientific advisory board member for Renalytix and Pensieve. He also has equity in Renalytix, Pensieve, and Verici. LC is a consultant for Vifor Pharma Inc and has received honorarium from Fresenius Medical Care. BSG provides consultancy services to University of California, San Francisco.

DATA AVAILABILITY STATEMENT

The data underlying this article will be shared on reasonable request to the corresponding author. Our institution has a data use committee and due processes requiring the transfer of data external to our institution.

REFERENCES

1. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19). *JAMA* 2020; 324 (8): 782–93.
2. Centers for Disease Control and Prevention. United States COVID-19 cases and deaths by state; 2020. https://covid.cdc.gov/covid-data-tracker/#cases_casesper100klast7days Accessed November 27, 2020.
3. Wynants L, Van Calster B, Collins GS, *et al*. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020; 369. doi:10.1136/bmj.m1328
4. Beigel JH, Tomashek KM, Dodd LE, *et al*. Remdesivir for the treatment of Covid-19—Final report. *N Engl J Med* 2020; 383 (19): 1813–26.
5. Wang Y, Zhang D, Du G, *et al*. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet* 2020; 395 (10236): 1569–78.
6. Jeronimo CMP, Farias MEL, Val FFA, Metcovid Team, *et al*. Methylprednisolone as adjunctive therapy for patients hospitalized with coronavirus disease 2019 (COVID-19; Metcovid): a randomized, double-blind, phase iib, placebo-controlled trial. *Clin Infect Dis* 2021; 72 (9): e373–e381.
7. Angus DC, Derde L, Al-Beidh F *et al*.; Writing Committee for the REMAP-CAP Investigators. Effect of hydrocortisone on mortality and organ support in patients with severe COVID-19: the REMAP-CAP COVID-19 corticosteroid domain randomized clinical trial. *JAMA* 2020; 324 (13): 1317–29.
8. Arabi YM, Mandourah Y, Al-Hameed F, *et al*. Corticosteroid therapy for critically ill patients with middle east respiratory syndrome. *Am J Respir Crit Care Med* 2018; 197 (6): 757–67.
9. Liu J, Zhang S, Dong X, *et al*. Corticosteroid treatment in severe COVID-19 patients with acute respiratory distress syndrome. *J Clin Invest* 2020; 130 (12): 6417–28.
10. Polack FP, Thomas SJ, Kitchin N *et al*.; C4591001 Clinical Trial Group. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med* 2020; 383 (27): 2603–15.
11. Baden LR, El Sahly HM, Essink B, *et al*. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N Engl J Med* 2021; 384 (5): 403–16.
12. Sadoff J, Gray G, Vandebosch A *et al*.; ENSEMBLE Study Group. Safety and efficacy of single-dose Ad26.COV2.S vaccine against Covid-19. *N Engl J Med* 2021; 384 (23): 2187–201.
13. Centers for Disease Control and Prevention. In-hospital mortality among hospital confirmed COVID-19 encounters by week from selected hospitals. Accessed June 14, 2021. https://www.cdc.gov/nchs/covid19/nhcs/hospital-mortality-by-week.htm.
14. Gattinoni L, Coppola S, Cressoni M, Busana M, Rossi S, Chiumello D. COVID-19 does not lead to a "typical" acute respiratory distress syndrome. *Am J Respir Crit Care Med* 2020; 201 (10): 1299–300.
15. Kumar A, Arora A, Sharma P, *et al*. Is diabetes mellitus associated with mortality and severity of COVID-19? A meta-analysis. *Diabetes Metab Syndr Clin Res Rev* 2020; 14 (4): 535–45.
16. Vaid A, Somani S, Russak AJ, *et al*. Machine learning to predict mortality and critical events in a cohort of patients with COVID-19 in New York city: model development and validation. *J Med Internet Res* 2020; 22 (11): e24018.
17. Zhou L, Romero N, Martínez-Miranda J, Conejero JA, García-Gómez JM, Sáez C. Heterogeneity in COVID-19 severity patterns among age-

gender groups: an analysis of 778 692 Mexican patients through a meta-clustering technique. *medRxiv* 2021.

18. Chan L, Chaudhary K, Saha A *et al.*; Mount Sinai COVID Informatics Center (MSCIC). AKI in hospitalized patients with COVID-19. *J Am Soc Nephrol* 2021; 32 (1): 151–60.

19. Vaid A, Chan L, Chaudhary K, *et al.* Predictive approaches for acute dialysis requirement and death in COVID-19. *Clin J Am Soc Nephrol* 2021; 16 (8). doi:10.2215/CJN.17311120.

20. Strehlow MC. Early identification of shock in critically ill patients. *Emerg Med Clin North Am* 2010; 28 (1): 57–66.

21. Cardoso LTQ, Grion CMC, Matsuo T, *et al.* Impact of delayed admission to intensive care units on mortality of critically ill patients: a cohort study. *Crit Care* 2011; 15 (1): R28.

22. Bhatraju PK, Mukherjee P, Robinson-Cohen C, *et al.* Acute kidney injury subphenotypes based on creatinine trajectory identifies patients at increased risk of death. *Crit Care* 2016; 20 (1): 372.

23. Oh W, Steinbach MS, Castro MR, *et al.* A computational method for learning disease trajectories from partially observable EHR data. *IEEE J Biomed Health Inform* 2021; 25 (7): 2476–86.

24. Wang X, Jehi L, Ji X, Mazzone PJ. Phenotypes and subphenotypes of patients with COVID-19. *Chest* 2021; 159 (6): 2191–204.

25. Vasquez CR, Gupta S, Miano TA, *et al.* Identification of distinct clinical subphenotypes in critically ill patients with COVID-19. *Chest* 2021; 160 (3): 929–43.

26. Bos LDJ, Paulus F, Vlaar APJ, Beenen LFM, Schultz MJ. Subphenotyping acute respiratory distress syndrome in patients with COVID-19: consequences for ventilator management. *Ann Am Thorac Soc* 2020; 17 (9): 1161–3.

27. Su C, Zhang Y, Flory JH, *et al.* Clinical subphenotypes in COVID-19: derivation, validation, prediction, temporal patterns, and interaction with social determinants of health. *npj Digit Med* 2021; 4 (1): 110.

28. Bhavani SV, Huang ES, Verhoef PA, Churpek MM. Novel temperature trajectory subphenotypes in COVID-19. *Chest* 2020; 158 (6): 2436–9.

29. Su C, Xu Z, Hoffman K, *et al.* Identifying organ dysfunction trajectory-based subphenotypes in critically ill patients with COVID-19. *Sci Rep* 2021; 11 (1): 15872.

30. Garcia-Garcia D, Hernandez EP, Diaz de Maria F. A new distance measure for model-based sequence clustering. *IEEE Trans Pattern Anal Mach Intell* 2009; 31 (7): 1325–31.

31. Zou Q, Lin G, Jiang X, Liu X, Zeng X. Sequence clustering in bioinformatics: an empirical study. *Brief Bioinform* 2018; 21 (1): 1–10.

32. Damerau FJ. A technique for computer detection and correction of spelling errors. *Commun ACM* 1964; 7 (3): 171–6.

33. Bergroth L, Hakonen H, Raita T. A survey of longest common subsequence algorithms. In: *Proceedings Seventh International Symposium on String Processing and Information Retrieval. SPIRE 2000*. IEEE Comput. Soc; 2000: 39–48. doi:10.1109/SPIRE.2000.878178.

34. Turk C, Turk S, Temirci ES, Malkan UY, Haznedaroglu İC. In vitro analysis of the renin–angiotensin system and inflammatory gene transcripts in human bronchial epithelial cells after infection with severe acute respiratory syndrome coronavirus. *J Renin Angiotensin Aldosterone Syst* 2020; 21 (2). doi:10.1177/1470320320928872.

35. Oscamou M, McDonald D, Yap VB, Huttley GA, Lladser ME, Knight R. Comparison of methods for estimating the nucleotide substitution matrix. *BMC Bioinformatics* 2008; 9 (1): 511.

36. Purushotham S, Meng C, Che Z, Liu Y. Benchmarking deep learning models on large healthcare datasets. *J Biomed Inform* 2018; 83 (October): 112–34.

37. Li T-R, Xu Y, Ruan D, Pan W. Sequential pattern mining. In: Ruan D, Chen G, Kerre EE, Wets G, eds. *Intelligent Data Mining*. Springer, Berlin Heidelberg; 2005: 103–122. doi:10.1007/11004011_5.

38. Low-Kam C, Raïssi C, Kaytoue M, Pei J. Mining statistically significant sequential patterns. In: *Proceedings of the 2013 IEEE International Conference on Data Mining (ICDM'13)*; 2013: 488–97. doi:10.1109/ICDM.2013.124.

39. Chen YL, Wu SY, Wang YC. Discovering multi-label temporal patterns in sequence databases. *Inf Sci* 2011; 181 (3): 398–418.

40. Wright AP, Wright AT, McCoy AB, Sittig DF. The use of sequential pattern mining to predict next prescribed medications. *J Biomed Inform* 2015; 53: 73–80.

41. Kevadiya BD, Machhi J, Herskovitz J, *et al.* Diagnostics for SARS-CoV-2 infections. *Nat Mater* 2021; 20 (5): 593–605.

42. Society of Critical Care Anesthesiologists. *ICU Resident's Guide*; 2017.

43. Rivers E, Nguyen B, Havstad S, *et al.* Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 2001; 345 (19): 1368–77.

44. Moranville MP, Mieure KD, Santayana EM. Evaluation and management of shock states: hypovolemic, distributive, and cardiogenic shock. *J Pharm Pract* 2011; 24 (1): 44–60.

45. van Diepen S, Katz JN, Albert NM *et al.*; American Heart Association Council on Clinical Cardiology; Council on Cardiovascular and Stroke Nursing; Council on Quality of Care and Outcomes Research; and Mission: Lifeline. Contemporary management of cardiogenic shock: a scientific statement from the American Heart Association. *Circulation* 2017; 136 (16): e232–68.

46. Chobanian AV, Bakris GL, Black HR *et al.*; the National High Blood Pressure Education Program Coordinating Committee. Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. *Hypertension* 2003; 42 (6): 1206–52.

47. Muntner P, Whelton PK, Woodward M, Carey RM. A comparison of the 2017 American College of Cardiology/American Heart Association Blood Pressure Guideline and the 2017 American Diabetes Association Diabetes and Hypertension Position Statement for U.S. adults with diabetes. *Diabetes Care* 2018; 41 (11): 2322–9.

48. Papazian L, Aubron C, Brochard L, *et al.* Formal guidelines: management of acute respiratory distress syndrome. *Ann Intensive Care* 2019; 9 (1): 69.

49. Torbic H, Duggal A. Neuromuscular blocking agents for acute respiratory distress syndrome. *J Crit Care* 2019; 49 (2019): 179–84.

50. Gosmanov AR, Gosmanova EO, Dillard-Cannon E. Management of adult diabetic ketoacidosis. *Diabetes Metab Syndr Obes* 2014; 7: 255–64.

51. Kitabchi AE, Umpierrez GE, Murphy MB, *et al.* Management of hyperglycemic crises in patients with diabetes. *Diabetes Care* 2001; 24 (1): 131–53.

52. Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. *Nephron Clin Pract* 2012; 120 (4): c179–84.

53. Palevsky PM, Liu KD, Brophy PD, *et al.* KDOQI US commentary on the 2012 KDIGO clinical practice guideline for acute kidney injury. *Am J Kidney Dis* 2013; 61 (5): 649–72.

54. Flamm SL, Yang Y-X, Singh S *et al.*; AGA Institute Clinical Guidelines Committee. American Gastroenterological Association Institute guidelines for the diagnosis and management of acute liver failure. *Gastroenterology* 2017; 152 (3): 644–7.

55. Whelton PK, Carey RM, Aronow WS, *et al.* 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Hypertension* 2018; 71 (6): E13–E115.

56. NYC Department of Health and Mental Hygiene. Long-Term Trends. https://www1.nyc.gov/site/doh/covid/covid-19-data-trends.page Accessed June 22, 2021.

57. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: an R package for determining the relevant number of clusters in a data set. *J Stat Softw* 2014; 61 (6): 1–36.

58. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 1987; 20 (C): 53–65.

59. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. *J R Stat Soc Ser B: Statist Methodol* 2001; 63 (2): 411–23.

60. Dudoit S, Fridlyand J. A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biol* 2002; 3 (7). doi:10.1186/gb-2002-3-7-research0036

61. Thorndike RL. Who belongs in the family? *Psychometrika* 1953; 18 (4): 267–76.

62. R Core Team. R: A Language and Environment for Statistical Computing; 2020.

63. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J Travel Med* 2020; 27 (2): 1–4.

64. Basu A. Estimating the infection fatality rate among symptomatic COVID-19 cases in the United States. *Health Aff (Millwood)* 2020; 39 (7): 1229–36.

65. National Institutes of Health. *COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines*. Vol 2019; 2020.

66. Yadaw AS, Li Y, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. *Lancet Digit Heal* 2020; 2 (10): e516–25.

67. Eng D, Chute C, Khandwala N, *et al.* Automated coronary calcium scoring using deep learning with multicenter external validation. *npj Digit Med* 2021; 4 (1): 88.

68. Wong A, Otles E, Donnelly JP, *et al.* External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med* 2021; 181 (8): 1065–70.