

## Characterizing COVID-19 Clinical Phenotypes and Associated Comorbidities and Complication Profiles

Elizabeth R. Lusczek, PhD;<sup>1\*</sup> Nicholas E. Ingraham, MD;<sup>2\*</sup> Basil S. Karam, MD<sup>3</sup>; Jennifer Proper, BA;<sup>4</sup> Lianne Siegel, BA;<sup>4</sup> Erika S. Helgeson, PhD;<sup>4</sup> Sahar Lotfi-Emran, MD PhD;<sup>2</sup> Emily J. Zolfaghari, MS;<sup>5</sup> Emma Jones, MD;<sup>1</sup> Michael G. Usher, MD;<sup>6</sup> Jeffrey G. Chipman, MD;<sup>1</sup> R. Adams Dudley, MD;<sup>2,7</sup> Bradley Benson, MD;<sup>6</sup> Genevieve B. Melton, MD PhD;<sup>1,7</sup> Anthony Charles, MD MPH;<sup>8,9</sup> Monica I. Lupei, MD;<sup>10</sup> Christopher J. Tignanelli, MD.<sup>1,7,11</sup>

\*Both authors contributed equally to this manuscript

### Author Affiliations:

- <sup>1</sup> Department of Surgery, University of Minnesota, Minneapolis, MN
- <sup>2</sup> Department of Medicine, University of Minnesota, Division of Pulmonary and Critical Care, Minneapolis, MN
- <sup>3</sup> Department of Surgery, Medical College of Wisconsin, Milwaukee, WI
- <sup>4</sup> Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN
- <sup>5</sup> University of Minnesota Medical School, Minneapolis, MN
- <sup>6</sup> Department of Medicine, University of Minnesota, Division of General Internal Medicine, Minneapolis, MN
- <sup>7</sup> Institute for Health Informatics, University of Minnesota, Minneapolis, MN
- <sup>8</sup> Department of Surgery, University of North Carolina, Chapel Hill, NC
- <sup>9</sup> School of Public Health, University of North Carolina, Chapel Hill, NC
- <sup>10</sup> Department of Anesthesiology, University of Minnesota, Minneapolis, MN
- <sup>11</sup> Department of Surgery, North Memorial Health Hospital, Robbinsdale, MN

### Correspondence:

Nicholas Ingraham, MD	MMC 276 420 Delaware St SE Minneapolis, MN 55455	Office: (612) 626-1968 Fax: (612) 626-0439 Email: <a href="mailto:ingra107@umn.edu">ingra107@umn.edu</a>
-----------------------	--	--

NEI	<a href="mailto:ingra107@umn.edu">ingra107@umn.edu</a>	RR	<a href="mailto:rreilkof@umn.edu">rreilkof@umn.edu</a>	JR	<a href="mailto:rhei0005@umn.edu">rhei0005@umn.edu</a>
TS	<a href="mailto:schac008@umn.edu">schac008@umn.edu</a>	SE	<a href="mailto:saharle@umn.edu">saharle@umn.edu</a>	MGU	<a href="mailto:mgusher@umn.edu">mgusher@umn.edu</a>
JGC	<a href="mailto:chipm001@umn.edu">chipm001@umn.edu</a>	BT	<a href="mailto:thie0149@umn.edu">thie0149@umn.edu</a>	SI	<a href="mailto:ikram001@umn.edu">ikram001@umn.edu</a>
BB	<a href="mailto:benso040@umn.edu">benso040@umn.edu</a>	ERL	<a href="mailto:lusc0006@umn.edu">lusc0006@umn.edu</a>	MA	<a href="mailto:abass004@umn.edu">abass004@umn.edu</a>
CTJ	<a href="mailto:ctignane@umn.edu">ctignane@umn.edu</a>	JP	<a href="mailto:prope012@umn.edu">prope012@umn.edu</a>		
MIL	<a href="mailto:lupei001@umn.edu">lupei001@umn.edu</a>	AC	<a href="mailto:anthchar@med.unc.edu">anthchar@med.unc.edu</a>		
GMM	<a href="mailto:gmelton@umn.edu">gmelton@umn.edu</a>	LS	<a href="mailto:siege245@umn.edu">siege245@umn.edu</a>		
DW	<a href="mailto:wack0012@umn.edu">wack0012@umn.edu</a>	MHS	<a href="mailto:syed0021@umn.edu">syed0021@umn.edu</a>		

### Conflicts of Interest and Funding Source(s):

1. NIH NHLBI T32HL07741 (NEI)
2. This research was supported by the Agency for Healthcare Research and Quality (AHRQ) and Patient-Centered Outcomes Research Institute (PCORI), grant K12HS026379 (CJT) and the National Institutes of Health's National Center for Advancing Translational Sciences, grant KL2TR002492 and UL1TR002494.
3. NIH NHLBI T32HL129956 (JP, LS)

Reprints will not be available from the authors. All authors significantly contributed to developing, writing, and revising this manuscript.

## Abstract

**Background:** There is limited understanding of heterogeneity in outcomes across hospitalized patients with coronavirus disease 2019 (COVID-19). Identification of distinct clinical phenotypes may facilitate tailored therapy and improve outcomes.

**Objective:** Identify specific clinical phenotypes across COVID-19 patients and compare admission characteristics and outcomes.

**Design, Settings, and Participants:** Retrospective analysis of 1,022 COVID-19 patient admissions from 14 Midwest U.S. hospitals between March 7, 2020 and August 25, 2020.

**Methods:** Ensemble clustering was performed on a set of 33 vitals and labs variables collected within 72 hours of admission. K-means based consensus clustering was used to identify three clinical phenotypes. Principal component analysis was performed on the average covariance matrix of all imputed datasets to visualize clustering and variable relationships. Multinomial regression models were fit to further compare patient comorbidities across phenotype classification. Multivariable models were fit to estimate the association between phenotype and in-hospital complications and clinical outcomes.

**Main outcomes and measures:** Phenotype classification (I, II, III), patient characteristics associated with phenotype assignment, in-hospital complications, and clinical outcomes including ICU admission, need for mechanical ventilation, hospital length of stay, and mortality.

**Results:** The database included 1,022 patients requiring hospital admission with COVID-19 (median age, 62.1 [IQR: 45.9-75.8] years; 481 [48.6%] male, 412 [40.3%] required ICU admission, 437 [46.7%] were white). Three clinical phenotypes were identified (I, II, III); 236 [23.1%] patients had phenotype I, 613 [60%] patients had phenotype II, and 173 [16.9%] patients had phenotype III.

When grouping comorbidities by organ system, patients with respiratory comorbidities were most commonly characterized by phenotype III ( $p=0.002$ ), while patients with hematologic ( $p<0.001$ ), renal ( $p<0.001$ ), and cardiac ( $p<0.001$ ) comorbidities were most commonly characterized by phenotype I.

The adjusted odds of respiratory ( $p < 0.001$ ), renal ( $p < 0.001$ ), and metabolic ( $p < 0.001$ ) complications were highest for patients with phenotype I, followed by phenotype II. Patients with phenotype I had a far greater odds of hepatic ( $p < 0.001$ ) and hematological ( $p = 0.02$ ) complications than the other two phenotypes. Phenotypes I and II were associated with 7.30-fold (HR: 7.30, 95% CI: (3.11-17.17),  $p < 0.001$ ) and 2.57-fold (HR: 2.57, 95% CI: (1.10-6.00),  $p = 0.03$ ) increases in the hazard of death, respectively, when compared to phenotype III.

**Conclusion:** In this retrospective analysis of patients with COVID-19, three clinical phenotypes were identified. Future research is urgently needed to determine the utility of these phenotypes in clinical practice and trial design.

## Introduction

The coronavirus disease 2019 (COVID-19), a disease caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), has infected over 18 million and led to over 700,000 deaths since first appearing in late 2019.<sup>1</sup> Researchers are rapidly attempting to understand the natural history of and immune response to COVID-19.<sup>2</sup> Despite intense research since the arrival of this novel coronavirus<sup>3</sup>, only one pharmaco-therapeutic agent, dexamethasone, has been associated with reduced mortality in at-risk individuals.<sup>4</sup> COVID-19 results in a constellation of symptoms, laboratory derangement, immune dysregulation, and clinical complications.<sup>5</sup>

Emergency department presentation varies widely, suggesting various clinical phenotypes exist and, importantly, it is likely these various phenotypes respond differently to treatment. To illustrate, two early phenotypes of respiratory failure likely exist in COVID-19. A classic ARDS phenotype exists with poorly compliant lungs and poor gas exchange; however, a phenotype with normal lung compliance also exists in COVID-19 and is hypothesized to be driven by shunting secondary to pulmonary microthrombi.<sup>6,7</sup> An intricate, multidimensional view is required to adequately understand the disease and account for the variation in clinical outcomes. Furthermore, patients could benefit from phenotype-specific medical care, which may differ from established standards of care

Despite this need, few studies have characterized COVID-19 clinical phenotypes and evaluated their association with complications and clinical outcomes. The aim of this study was to characterize clinical phenotypes in COVID-19 according to disease-system factors using electronic health record (EHR) data pooled from 14 U.S. Midwest hospitals between March 7, 2020 and August 25, 2020.

## **Methods**

### Data Collection

The data source for this study included EHR reports from 14 U.S. Midwest hospitals and 60 primary care clinics. Patient and hospital-level data were available for 7,538 patients with PCR-confirmed COVID-19. Of these, 1,022 required hospital admission and were included in this analysis. The database included all comorbidities reported since March 29, 1997 for each patient and prior to their COVID-19 diagnosis. The database also included home medications, laboratory values, clinic visits, social history, and patient demographics (age, gender, race/ethnicity, language spoken, zip code, socioeconomic status indicators). For each COVID-19 hospitalization the database included all laboratory values, vitals, orders, medications, complications, length of stay, and hospital disposition. State death certificate data was linked with the database to enable capture of out-of-hospital death. Additionally, the database allowed linkage across the 14 hospitals, facilitating the tracking of transfers.

This study was approved by the University of Minnesota institutional review board (STUDY00001489).

## Participants

Patient-level data were obtained from the COVID-19 database from March 7, 2020 to August 25, 2020. The inclusion criterion was as follows: PCR-positive COVID-19 test requiring inpatient hospital admission to one of the 14 hospitals providing data. No hospitalized patients were excluded in this analysis to maximize generalizability. Follow-up data were available for a minimum of two weeks following admission for all patients.

## Clinical Variables for Phenotyping

We selected 33 variables for clustering based on their association with COVID-19 mortality, known COVID-19 pathophysiology, and presence in the database (no more than 50% missingness).<sup>8-11</sup> The following variables were included: age, body mass index (BMI), heart rate, respiratory rate, oxygen saturation, pulse pressure, systolic blood pressure, total protein, red cell distribution width, mean corpuscular volume, alkaline phosphatase, calcium, anion gap, bicarbonate, hematocrit, aspartate aminotransferase, glucose, absolute monocyte count, absolute neutrophil count, absolute lymphocyte count, white blood cell count, platelet, albumin, bilirubin, international normalized ratio (INR), lactate dehydrogenase, potassium, sodium, D-dimer, hemoglobin, C-reactive protein (CRP), creatinine, and gamma gap. For each variable we selected the first recorded value within the first 72 hours of the emergency department (ED) presentation that ultimately resulted in their hospitalization.

## Comorbidities

We selected 68 comorbidities documented for each patient from March 29, 1997 preceding their COVID-19 hospital admission in their electronic health record (Supplemental Table 3). All comorbidities were identified based on ICD-9, ICD-10, or problem list documentation within the electronic health record. An indicator variable was created for each comorbidity to denote the presence of the selected ICD-9, ICD-10, or problem list documentation at any time in the medical record. To facilitate analysis,

comorbidities were grouped by organ system into the following categories: cardiac, respiratory, hematologic, metabolic, renal, hepatic, autoimmune, cancer, and cerebrovascular disease.

### Complications and Clinical Outcomes

We selected 30 in-hospital complications measured during each patient's hospital stay for COVID-19 categorized into the following systems: cardiovascular, respiratory, hematologic, renal, hepatic, metabolic, and infectious (Supplemental Table 4). If applicable, complications could span multiple organ system variables. For example, ventilator associated pneumonia was included in both infectious and respiratory complications. Additional clinical outcomes included hospital length of stay (LOS), need for intensive care unit (ICU) admission, need for mechanical ventilation, and mortality. Mortality was defined as any in-hospital or out-of-hospital death based on death certificate data. All complications and outcomes were followed for a minimum of 2 weeks following hospital admission.

### Statistical Analysis

The overall rate of missingness of the 33 variables used for phenotyping, which included the first vitals and labs recorded for each inpatient within 72 hours of admission, was 19% (range 0% - 50%). We imputed missing values using multivariate imputations by chained equations implemented with the *mice* package (v.3.10.0).<sup>12,13</sup> Data were log-transformed before imputing missing values with predictive mean matching. A total of 40 imputed datasets were generated. The *diceR* package (v.1.0.0)<sup>14</sup> was used to perform k-means-based consensus clustering on each imputed dataset using 80% subsamples and 1,000 iterations. We considered grouping patients into 2-7 phenotypes and determined the optimal number was 3 by evaluating the consensus cumulative distribution function (CDF) plot, the delta area plot, and the consensus matrix heatmap. These figures were generated using the consensus clustering results for each imputed dataset, and all figures were qualitatively similar across datasets. For visualization purposes, these images are provided for a randomly selected dataset in **Supplemental Figures 1-4**. The final assignment of each patient into one of the three phenotypes was determined by majority voting across the 40 consensus clustering results. Principal component analysis (PCA) was performed on

the average covariance matrix to visualize the relationships among the three phenotypes and assess variable contributions.<sup>15</sup>

Continuous variables were summarized using the median and interquartile range (IQR) and compared across phenotypes using a Kruskal-Wallis test. Categorical characteristics and outcomes were summarized using counts and proportions and compared across phenotypes using a Pearson's chi-squared test or Fisher's exact test. Multinomial regression models were fit to further compare patient comorbidities across phenotype classification.

We next evaluated the relationship between phenotype and subsequent outcomes using both unadjusted and adjusted models. The adjusted models included sex,<sup>16,17</sup> race and ethnicity (white, Black, Asian, Hispanic, other, not reported),<sup>18</sup> and Elixhauser Comorbidity Index,<sup>19</sup> since these are known risk factors for the outcomes of interest and were not included in the clustering analysis. The associations between phenotype and complications, ICU admission and need for mechanical ventilation, were estimated using logistic regression models. Mortality was compared across phenotypes using Cox proportional hazard models and patients were censored at the last date of data collection, August 25, 2020. Hospital length of stay was compared across phenotypes using negative binomial regression models. The primary negative binomial model included individuals who died during hospitalization for whom length of stay was defined as the number of days until death. We performed a sensitivity analysis to assess the impact of mortality as a competing risk by refitting the length of stay model after removing the 127 patients who died. Two-sided p-values < 0.05 were considered statistically significant. P-values were not adjusted for multiple comparisons. Visualizations of comorbidities, complications, and outcomes by clinical phenotype were performed using the *circlize* package for R.<sup>20</sup> Comorbidities and complications were grouped into separate organ systems and the prevalence of each complication/comorbidity type was calculated as a percentage for each phenotype. All analyses were conducted using R version 3.6.3<sup>21</sup> and Stata version 16.1 (StataCorp).

## Results

The database included 1,022 patients requiring hospital admission with COVID-19. Among these patients, the median age was 62.1 [IQR: 45.9, 75.8] years; 481 [48.6%] male, 412 [40.3%] required ICU admission). Additionally, 437 [46.7%] were white, 188 [20.1%] were Black, 159 [17.0%] were Asian, 103 [11.0%] were Hispanic, 20 [2.1%] reported other race, and 28 [2.9%] did not report. Three clinical phenotypes were identified (I, II, III); 236 [23.1%] patients had phenotype I, 613 [60%] patients had phenotype II, and 173 [16.9%] patients had phenotype III.

### Variable Contributions to Clustering

The first two principal components (PCs) from PCA were used to visualize the relationship between phenotypes. PC1 and PC2 captured approximately 11% and 9% of the variance in the clustering variables, respectively. Thirteen components were needed to explain 70% of the variance (**Supplemental Figure 5**). While phenotypes II and III overlay substantially, phenotype I is more clearly defined in the right-hand side of the score plot of the first two principal components (**Figure 1**). Notably, this figure shows that distinctions between phenotypes are primarily driven by variation in PC1 as opposed to PC2. The variable contributions to PC1 (**Figure 2**) demonstrate that the largest contributors to the variation in PC1 are from LDH, absolute neutrophil count, and D-dimer. These variables therefore prominently contribute to separating the three phenotypes as shown in the biplot (**Figure 3**). Univariate tests showed that D-dimer and neutrophil count are highest in phenotype I and albumin is highest in phenotype III. Other variables influential to phenotype clustering are lactate dehydrogenase (highest in I), C-reactive protein (highest in I), white cell count (highest in I), red cell distribution width (highest in I), bilirubin (highest in I), AST (highest in I), hematocrit (highest in III), and hemoglobin (highest in III).

### Phenotype Characteristics

Differences across phenotypes with respect to patient demographics, admission vitals and labs, complications, comorbidities, and clinical outcomes are presented in **Table 1**. Patients with



phenotype I were older than patients in phenotypes II and III (67.2 [52.9, 79.0] years vs. 60.9 [45.9, 75.4] and 58.6 [34.8, 71.3] years respectively,  $p < 0.001$ ). Patients with phenotype III were more often female than patients with phenotype I or II (57.6% vs. 41.6% and 53.4%, respectively,  $p = 0.004$ ). Patients with phenotype I were more less likely white (38.8% vs. 45.6% vs. 60.7%, respectively,  $p = 0.002$ ) and more likely to be non-English speaking (47.9% vs. 39.2% vs. 23.7%, respectively,  $p < 0.001$ ). There were no statistically significant differences in BMI or socioeconomic status, as measured using the area deprivation index, between phenotypes (**Table 1**). Patients that presented with phenotype III had a more frequent history of smoking, alcohol abuse, neutropenia. Patients that presented with phenotype II had a less frequent history of hepatic disease than phenotypes I or III (**Table 1**).

When grouping comorbidities by organ system, cardiac ( $p < 0.001$ ), respiratory ( $p = 0.002$ ), hematologic ( $p < 0.001$ ), and renal ( $p < 0.001$ ) comorbidities were found to be significantly associated with phenotype. Cancer, hepatic, autoimmune, cerebrovascular, and metabolic comorbidities were not significantly associated with phenotype (**Table 1, Supplemental Figure 6**). Based on the estimated relative risk ratios, patients with renal (RRR 2.35; 95% CI 1.5-3.67;  $p < 0.001$ ), hematologic (RRR 2.64; 95% CI 1.75-3.98;  $p < 0.001$ ), and cardiac comorbidities (RRR 2.65; 95% CI: 1.68-4.17;  $p < 0.001$ ) were more likely to have phenotype I vs. III (**Figure 4**). Patients with respiratory comorbidities were 0.47 (95% CI: 0.31-0.72;  $p < 0.001$ ) times as likely to have phenotype I vs. III and 0.74 (95% CI: 0.52-1.04  $p = 0.09$ ) times as likely to have phenotype II vs. III (**Figure 4**).

#### Association between Phenotype and Clinical Outcomes

Clinical phenotypes I and II were associated with increased odds of respiratory (I: OR: 2.98, 95% CI 1.58 - 5.59; II: OR: 2.32, 95% CI: 1.29-4.17;  $p < 0.001$ ), renal (I: OR: 7.04, 95% CI 3.11-15.9; II: OR: 2.57, 95% CI: 1.15-5.74;  $p < 0.001$ ), and metabolic (I: OR: 4.85, 95% CI: 2.78-8.45; II: OR: 2.57, 95% CI: 1.52-4.34;  $p < 0.001$ ) complications, compared to phenotype III after adjusting for sex, race, and Elixhauser Comorbidity Index (**Supplemental Table 2**). There was a trend towards increased odds of hematologic complications among patients with phenotype I (I: OR: 2.11, 95% CI: 0.99-4.48,  $p = 0.05$ ) compared to III. Phenotype was associated with hepatic complications ( $p < 0.001$ ); however, while

phenotype I was associated with a 8.35-fold (OR: 8.35, 95% CI: 1.93-36.11,  $p < 0.001$ ) increase in the odds of hepatic complication, phenotype II did not differ significantly from phenotype III (OR: 0.56, 95% CI: 0.10-3.09,  $p = 0.51$ ). This is not surprising since only 4 individuals in phenotype II and 2 in phenotype III experienced hepatic complications during hospitalization (Table 1). Phenotype was also significantly associated with the rate of infectious complications ( $p < 0.001$ ) for phenotype 1 (OR 2.57, 95% CI 1.57 - 4.21;  $< 0.001$ ) but not did not reach statistical significance for phenotype 2 (OR 1.51, 95% CI 0.96 - 2.38;  $p = 0.07$ ) (**Supplemental Table 2 and Supplemental Figure 7**).

Clinical phenotypes differed in odds of ICU admission ( $p < 0.001$ ) and mechanical ventilation ( $p < 0.001$ ), hospital LOS ( $p < 0.001$ ), and risk of mortality ( $< 0.001$ ) on adjusted analysis which accounted for sex, race, and Elixhauser Comorbidity Index (**Table 2, Supplemental Figure 8**). Controlling for these risk factors and compared to phenotype III, phenotypes I and II were associated with 7.88-fold (OR: 7.88, 95% CI: 4.65-13.37) and 2.32-fold (OR: 2.32, 95% CI: 1.46-3.68) increases in the odds of ICU admission, respectively. Phenotypes I and II were associated with 25.59-fold (OR: 25.59, 95% CI: 7.69,- 85.17) and 7.45-fold (OR: 7.45, 95% CI: 2.27-24.43) increases in the odds of requiring mechanical ventilation. Phenotypes I and II were associated with 1.74-fold (IRR: 1.74, 95% CI: 1.45-2.10,  $p < 0.001$ ) and 1.22-fold (IRR: 1.22, 95% CI: 1.05-1.43,  $p = 0.01$ ) increases in hospital LOS. Phenotype I was associated with a 7.30-fold (HR: 7.30, 95% CI: 3.11-17.17,  $p < 0.001$ ) increase in risk of mortality, and Phenotype II had a 2.57-fold (HR: 2.57, 95% CI: 1.10-6.00,  $p = 0.03$ ) increase in the hazard of death compared to Phenotype 3. We performed a sensitivity analysis to assess the impact of mortality as a competing risk by fitting the LOS model before and after removing the 127 patients who died. The estimated effect sizes were similar between these two models (data not shown). Table 2 includes the LOS model with only survivors.

## Discussion

This is one of the first studies to report on clinical phenotypes associated with COVID-19. We identified three clinical phenotypes for patients with COVID-19 on hospital presentation. Most patients

presented with phenotype II, which is associated with a moderate course and an approximately 10% mortality. A subset of patients presented with the more severe phenotype I, which is associated with a staggering 27% mortality. Patients with cardiac, hematologic, and renal comorbidities were most likely to be characterized by phenotype I. Surprisingly, respiratory comorbidities appeared less related to phenotypes I or II and were most associated with phenotype III, which had the most indolent course. Despite this indolent course, patients with phenotype III had the highest rate of readmission which is likely in part due to the high survival rate. This also suggests patients with pre-existing respiratory comorbidities, while not at highest risk for mortality, may be at highest risk for long term sequelae following COVID-19. Patients that presented with phenotype I were most associated with the development of respiratory, hematologic, renal, metabolic, hepatic, and infectious complications. Surprisingly, cardiovascular complications did not significantly differ between phenotypes.

Elucidating patient risk factors and severe COVID-19 disease markers may allow early treatment implementation that may improve the patient's outcome. Multiple studies have documented COVID-19 risk factors; however, most have done so from a homogenous lens. For example, a prospective cohort study from New York City identified that the most considerable risks for hospital admission were age, male sex, heart failure, chronic kidney disease, and high BMI.<sup>22</sup> A large observational study conducted in the UK reported that increasing age, male gender, comorbidities such as cardiac disease, chronic lung disease, chronic kidney disease, and obesity were associated with higher mortality in COVID-19 positive patients admitted to the hospital.<sup>14</sup> A study from China found that increased odds of in-hospital death due to COVID-19 were associated with older age, higher SOFA score and D-dimers > 1.0 µg/mL on admission.<sup>23</sup> Another retrospective study reported that patients with severe COVID-19 disease and diabetes had increased leucocytes, neutrophils count, and increased C-reactive protein (CRP), D-dimers, fibrinogen levels.<sup>24</sup> A systematic review and meta-analysis found that the biomarkers associated with increased mortality include higher CRP, higher D-dimers, increased creatinine, and lower albumin levels.<sup>25</sup> However it is well known that patients do not have a singular natural history of disease. Multiple studies including this study found that only half of patients suffer a primarily respiratory disease.<sup>26,27</sup>

Patients suffer a constellation of cardiovascular, hematologic, renal, or hepatic progression of disease following COVID-19. It is likely patient baseline risk factors related to the virus,<sup>28</sup> home medications,<sup>16,29</sup> genetic predisposition<sup>30</sup>, race/ethnicity,<sup>18</sup> and other factors predispose patients to one of the various clinical manifestations and natural history of COVID-19.

Treatment of hospitalized patients should be tailored based on the clinical courses most likely for a patient given their *a priori* risk. For example, phenotypes with a higher risk of thrombotic events, may benefit from more aggressive anticoagulation. Phenotypes more prone to infectious complications, may benefit from more targeted immunomodulation instead of broad and systemic steroid therapy. A key first step to evaluate these treatment decisions is to characterize and describe clinical phenotypes requiring hospitalization. In this analysis we identified three clinical phenotypes for patients that required hospitalization for COVID-19. Few studies to date have attempted to elucidate clinical phenotypes. One study attempted to characterize clinical phenotypes at ICU admission using a dataset of 85 critically ill patients.<sup>31</sup> Similar to our analysis, they identified three distinct clinical phenotypes. Their low mortality cluster which they called cluster 1 was very similar to our phenotype III with a predominance of females, lower mortality rate, lower D-dimer and CRP levels. Similarly, their high mortality cluster was predominantly male, with elevated inflammation markers on ICU presentation. In this study, we not only characterized three clinical phenotypes, but extended findings outside of the ICU by characterizing the association of comorbidities with clinical phenotype and the association of clinical phenotypes with in-hospital complication and clinical outcomes.

Phenotype I can be termed the “Adverse phenotype” and was associated with the worst clinical outcomes. LDH, Absolute Neutrophil Count, D-dimer, AST, and CRP were most influential in phenotype I determination. The strong association of RDW with phenotype I was interesting. RDW was strongly associated with genetic age which is hypothesized to be a risk factor in Covid-19.<sup>30</sup> As people age, variability in red blood cell volumes increases. Similarly, Gamma Gap, a marker of immunoglobulin levels, was elevated in all three phenotypes (median >

3.5).<sup>32</sup> However, patients with clinical phenotype I were noted to have the largest increase in Gamma Gap. In this scenario elevated Gamma Gap was likely an indicator of systemic inflammation and has been associated in other inflammatory disease processes with prognosis. Other groups have previously reported on the importance of the Absolute Neutrophil to Absolute Lymphocyte count, here we noted that ANC/ALC was lowest for phenotype III and highest for phenotype I, in line with previous reports. Patients with cardiac, hematologic and renal comorbidities were most prone to develop phenotype I. Phenotype I was associated with numerous complications ( hematologic, hepatic, metabolic, renal, respiratory, and infectious) when compared to other phenotypes. It is interesting to note despite a higher rate of baseline cardiac comorbidities phenotype I was not associated with increased cardiac complications.

Phenotype III was associated with the best clinical outcomes and can be termed the “Favorable Phenotype”. Surprisingly, patients with phenotype III had a very high rate of respiratory comorbidities and the best clinical outcomes. What is most surprising is despite the lowest complication rate and mortality, this phenotype was associated with a greater than 10% rate of hospital readmission. It is possible that patients pre-existing respiratory comorbidities predisposed them to longer term sequelae which may have resulted in this readmission rate, although additional studies are needed to better elucidate these findings, specifically controlling for differences in survival. Patients with respiratory comorbidities such as asthma and COPD routinely use medications which may be protective in SARS-CoV-2 pathogenesis which may explain this protective effect. For example, our group has previously identified reduced mortality in COVID-19 for patients with asthma treated with beta2-agonists.<sup>16</sup> Patients with phenotype III were more likely to use inhaled steroids, nasal fluticasone, albuterol, and antihistamines.

Ultimately, a deeper investigation into clinical phenotypes and associated genomic, transcriptomic, and proteomic is needed. The ability to classify patients into clinical phenotypes can facilitate the linkage of exome data to better understand SARS-CoV-2 pathogenesis and natural history.

Understanding the COVID-19 severity, the biomarkers, and the risk factors is paramount during the COVID-19 pandemic.

## **Limitations**

Our study has several limitations, including that this is a retrospective study and therefore results may be biased or subject to residual confounding. Second, patients were followed for variable lengths of time. Patients that were admitted in March 2020 thus had approximately 5 months of follow-up whereas patients admitted in late August had limited time. We accounted for this by conducting a Cox proportional hazard analysis when analyzing in- and out- of hospital mortality. Additionally, when the data were pulled, only 54 patients (5%) remained hospitalized. While most patients developed complications within their first 2 weeks of hospital admission, it is possible that they may still develop clinical complications which is not reflected in this analysis.

## **Conclusion**

In this retrospective analysis of patients with COVID-19, three clinical phenotypes were identified. Future research is urgently needed to determine the utility of these phenotypes in clinical practice and trial design.

## **Conflicts of Interest and Source of Funding:**

1. NIH NHLBI T32HL07741 (NEI)
2. This research was supported by the Agency for Healthcare Research and Quality (AHRQ) and Patient-Centered Outcomes Research Institute (PCORI), grant K12HS026379 (CJT) and the National Institutes of Health's National Center for Advancing Translational Sciences, grant KL2TR002492 and UL1TR002494.
3. NIH NHLBI T32HL129956 (JP, LS)

## Acknowledgements

The authors thank Eric Murray and Fairview IT for collection of data

## Author Contribution

Concept and design: All authors

Acquisition, analysis, or interpretation of data: Luszczek, Proper, Siegel, Helgeson, Usher, Tignanelli

Drafting of the manuscript: All authors

Critical revision of the manuscript for important intellectual content: All authors

## References

1. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA*. 2020.
2. Ingraham NE, Lotfi-Emran S, Thielen BK, et al. Immunomodulation in COVID-19. *Lancet Respir Med*. 2020;8(6):544-546.
3. Ingraham NE, Tignanelli CJ. Fact Versus Science Fiction: Fighting Coronavirus Disease 2019 Requires the Wisdom to Know the Difference. *Crit Care Explor*. 2020;2(4):e0108.
4. Group RC, Horby P, Lim WS, et al. Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report. *The New England journal of medicine*. 2020.
5. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. *JAMA*. 2020.
6. Li X, Ma X. Acute respiratory failure in COVID-19: is it "typical" ARDS? *Crit Care*. 2020;24(1):198.
7. Diehl JL, Peron N, Chocron R, et al. Respiratory mechanics and gas exchanges in the early course of COVID-19 ARDS: a hypothesis-generating study. *Ann Intensive Care*. 2020;10(1):95.
8. Arentz M, Yim E, Klaff L, et al. Characteristics and outcomes of 21 critically ill patients with COVID-19 in Washington State. *Jama*. 2020;323(16):1612-1614.
9. Qin C, Zhou L, Hu Z, et al. Dysregulation of immune response in patients with COVID-19 in Wuhan, China. *Clinical Infectious Diseases*. 2020.
10. Wu C, Chen X, Cai Y, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA internal medicine*. 2020.
11. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*. 2020;395(10223):497-506.
12. Little RJA, Rubin DB. Statistical analysis with missing data. New York: Wiley; 1987. xiv, 278.
13. Stef van Buuren, Karin Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1-67. URL <https://www.jstatsoft.org/v45/i03/>.



14. Derek Chiu and Aline Talhouk (2020). diceR: Diverse Cluster Ensemble in R. R package version 1.0.0. <https://CRAN.R-project.org/package=diceR>.
15. Nassiri V, Lovik A, Molenberghs G, Verbeke G. On using multiple imputation for exploratory factor analysis of incomplete data. *Behav Res Methods*. 2018;50(2):501-517.
16. Bramante C, Ingraham N, Murray T, et al. Observational Study of Metformin and Risk of Mortality in Patients Hospitalized with Covid-19. *medRxiv*. 2020:2020.2006.2019.20135095.
17. Jin JM, Bai P, He W, et al. Gender Differences in Patients With COVID-19: Focus on Severity and Mortality. *Front Public Health*. 2020;8:152.
18. Ingraham NE, Purcell LN, Karam BS, et al. Racial/Ethnic Disparities in Hospital Admissions from COVID-19 and Determining the Impact of Neighborhood Deprivation and Primary Language. *medRxiv*. 2020:2020.2009.2002.20185983.
19. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Medical care*. 1998;8:27.
20. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circize implements and enhances circular visualization in R. *Bioinformatics*. 2014;30(19):2811-2812.
21. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
22. Petrilli CM, Jones SA, Yang J, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ*. 2020;369:m1966.
23. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020;395(10229):1054-1062.
24. Ye W, Chen G, Li X, et al. Dynamic changes of D-dimer and neutrophil-lymphocyte count ratio as prognostic biomarkers in COVID-19. *Respir Res*. 2020;21(1):169.
25. Tian W, Jiang W, Yao J, et al. Predictors of mortality in hospitalized COVID-19 patients: A systematic review and meta-analysis. *J Med Virol*. 2020.
26. Shi S, Qin M, Shen B, et al. Association of Cardiac Injury With Mortality in Hospitalized Patients With COVID-19 in Wuhan, China. *JAMA Cardiol*. 2020.
27. Yang X, Yu Y, Xu J, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med*. 2020;8(5):475-481.
28. Ingraham NE, Barakat AG, Reilkoff R, et al. Understanding the Renin-Angiotensin-Aldosterone-SARS-CoV-Axis: A Comprehensive Review. *Eur Respir J*. 2020:2000912.
29. Tignanelli CJ, Ingraham NE, Sparks MA, et al. Antihypertensive drugs and risk of COVID-19? *Lancet Respir Med*. 2020;8(5):e30-e31.
30. Kuo CP, L. Atkins JC, et al. . COVID-19 severity is predicted by earlier evidence of accelerated aging. *Medrxiv*. 2020.
31. Azoulay E, Zafrani L, Mirouse A, Lengline E, Darmon M, Chevret S. Clinical phenotypes of critically ill COVID-19 patients. *Intensive Care Med*. 2020;46(8):1651-1652.
32. Yang M, Xie L, Liu X, Hao Q, Jiang J, Dong B. The gamma gap predicts 4-year all-cause mortality among nonagenarians and centenarians. *Sci Rep*. 2018;8(1):1046.



## Figure Titles and Legends:

### Figure 1: Score Plot: PC2 vs. PC1

The principal component scores for PC1 and PC2 are plotted. Each point represents a patient in the dataset. Colors represent the cluster (phenotype) that the patient was assigned to by consensus clustering. Ellipses around each cluster/phenotype specify 95% confidence intervals, assuming a bivariate normal distribution.

Abbreviations: PC1 (principal component 1); PC2 (principal component 2)

### Figure 2: Contribution of Variables to PC1

The contributions of each of the 33 variables used in the clustering to principal component 1 are shown. The red line marks the expected average contribution of each variable if the contributions of the variables were uniform across the dataset. Variables contributing most to the observed pattern in PC1 are D-dimer and albumin.

Abbreviations: PC1 (principal component 1); Abs\_Nphil\_Ct (absolute neutrophil count); LDH (lactate dehydrogenase); CRP (C-reactive protein); WBC (white blood cell count); HCT (hematocrit); HGB (hemoglobin); Tbili (total bilirubin); RDW (red cell distribution width); AST (aspartate aminotransferase); Alk\_phos (alkaline phosphatase); RR (respiratory rate); CA (calcium); TP (total protein); INR (internal normalized ratio of prothrombin time); CO2 (carbon dioxide); K (potassium); O2SAT (oxygen saturation); BMI (body mass index); PLT (platelet); PP (pulse pressure); Na (sodium); SBP (systolic blood pressure); Abs\_mono\_ct (absolute monocyte count); MCV (mean corpuscular volume)

### Figure 3: PCA Biplot: PC2 vs. PC1

The scores (points) and loadings (arrows) of PC1 and PC2 are plotted for each patient and variable in the model. 95% confidence ellipses for the scores are shown. The biplot facilitates interpretation of the scores and loadings, assigning context to the variables which prominently contribute to the phenotypes.

Abbreviations: PC1 (principal component 1); PC2 (principal component 2); PCA (principal component analysis); Abs\_Nphil\_Ct (absolute neutrophil count); LDH (lactate dehydrogenase); CRP (C-reactive protein); WBC (white blood cell count); HCT (hematocrit); HGB (hemoglobin); Tbili (total bilirubin); RDW (red cell distribution width); AST (aspartate aminotransferase); Alk\_phos (alkaline phosphatase); RR (respiratory rate); CA (calcium); TP (total protein); INR (internal normalized ratio of prothrombin time); CO2 (carbon dioxide); K (potassium); O2SAT (oxygen saturation); BMI (body mass index); PLT (platelet); PP (pulse pressure); Na (sodium); SBP (systolic blood pressure); Abs\_mono\_ct (absolute monocyte count); MCV (mean corpuscular volume)

#### **Figure 4: Relative Risk Ratio of Comorbidities to Clinical Phenotypes**

Relative Risk ratios of comorbidities of phenotypes I and II compared to the reference group phenotype III.

#### **Supplemental Figure 1: Consensus Cumulative Distribution Functions**

Cumulative distribution functions (CDF) for a randomly selected imputed dataset are shown. A range of phenotypes (2-7) were considered, and the optimal choice of phenotypes is 3.

#### **Supplemental Figure 2: Delta Area**

The relative change in delta area under the cumulative distribution function is shown for the range of phenotypes ( $k=2-7$ ) for a randomly selected imputed dataset. The optimal choice of phenotypes is 3.

Abbreviations: CDF (cumulative distribution function)

### **Supplemental Figure 3: Consensus matrix with 3 clusters**

A consensus matrix heatmap is shown for a randomly selected imputed dataset clustered into 3 phenotypes. The heatmap allows visualization of consensus cluster assignments to evaluate cluster stability. Darker shades of green indicate higher stability.

### **Supplemental Figure 4: Consensus matrix with 4 clusters**

A consensus matrix heatmap is shown for a randomly selected imputed dataset clustered into 4 phenotypes. The heatmap allows visualization of consensus cluster assignments to evaluate cluster stability. Darker shades of green indicate higher stability. The choice of 4 clusters shows less stability than 3 clusters (see Supplemental Figure 3).

### **Supplemental Figure 5: Cumulative Proportion of Variance Explained**

The proportion of variance explained by each principal component is summed over all principal components. For example, PC1 and PC2 cumulatively explain 20% of the variation in the dataset.

Abbreviations: PC1 (principal component 1); PC2 (principal component 2)

### **Supplemental Figure 6: Comorbidities by Phenotype**

Chord diagram illustrates the prevalence of comorbidities (% observed) for the three clinical phenotypes.

### **Supplemental Figure 7: Complications by Phenotype**

Chord diagram illustrates the prevalence of complications (% observed) for the three clinical phenotypes.

### **Supplemental Figure 8: Clinical Outcomes by Phenotype**

Chord diagram illustrates the prevalence of clinical outcomes (% observed) for the three clinical phenotypes.

Abbreviations: ICU (intensive care unit); Vent (mechanical ventilation); Readmit (readmission to hospital or ICU); ECMO (extracorporeal membrane oxygenation).

### **Tables:**

#### **Table 1: Baseline demographics, comorbidities, and clinical outcomes of hospitalized COVID-19 patients with clinical phenotypes I, II, and III.**

\* Categorical variables presented as count (%), continuous variables presented as median (interquartile range) unless otherwise specified. Continuous variables were evaluated with Kruskal-Wallis tests. Categorical variables were evaluated with chi-square tests or Fisher's exact tests (counts < 5).

Abbreviations: ADI, area deprivation index; BMI, body mass index; ECMO, Extracorporeal membrane oxygenation; ICU, Intensive Care Unit

## **Table 2: Association of Clinical Phenotype with Clinical Outcomes**

Abbreviations: PH, proportional hazards; HR, hazard ratio; CI, confidence interval; OR, odds ratio; ICU, intensive care unit; IRR, incidence rate ratio; LOS, length of stay; LR, likelihood ratio

Legend: Reference group for all models is Phenotype III. All models adjusted for sex, race/ethnicity, and Elixhauser Comorbidity Index.

\* LOS model only included patients that survived.

## **Supplemental Table 1: Home medications and hospital day 5 laboratory values of hospitalized COVID-19 patients with clinical phenotypes I, II, and III.**

\* Categorical variables presented as count (%), continuous variables presented as median (interquartile range).

† Loop Diuretics include furosemide, torsemide, bumetanide.

Abbreviations: ALC, Absolute Lymphocyte Count; ANC, Absolute Neutrophil Count; ARB, Angiotensin receptor blocker; AST, Aspartate transaminase; CRP, C-Reactive Protein; ESR, Erythrocyte Sedimentation Rate; EKG, Electrocardiogram; IL, Interleukin; LDH, Lactate Dehydrogenase; Pro-BNP, proB-type Natriuretic Peptide; PLT, Platelets; PTT, Partial Thromboplastin Time; TNF-alpha, Tumor Necrosis Factor-alpha; WBC, White Blood Cell;

## **Supplemental Table 2: Association of Clinical Phenotype with In-Hospital Complications**

**Abbreviations:** OR, odds ratio; CI, confidence interval; LR, likelihood ratio

**Legend:** Reference group for all models is Phenotype III. All models adjusted for sex, race/ethnicity, and Elixhauser Comorbidity Index.

### **Supplemental Table 3: Comorbidities based on ICD-10 codes**

Supplementary Table 3: List of ICD 10 codes that were used to classify diagnosis.

Abbreviations: T1DM: Type 1 diabetes mellitus, T2DM Type 2 diabetes mellitus, HFpEF: heart failure with preserved ejection fraction, HF: heart failure, CAD: coronary artery disease, NAFLD: non-alcoholic fatty liver disease, NASH: non-alcoholic steatohepatitis, VTE: venous thromboembolism, HIT: heparin induced thrombocytopenia, DIC: disseminated intravascular coagulation, ITP: idiopathic thrombocytopenia, HLH: hemophagocytic lymphohistiocytosis, MI: myocardial infarction, COPD: chronic obstructive lung disease, ILD: interstitial lung disease, AICD: Automatic Implantable Cardioverter Defibrillator, VAD: ventricular assist device, CKD: chronic kidney disease, ESRD: end stage renal disease, Afib: atrial fibrillation, HIV: human immunodeficiency virus, Flu: influenza, IBD: inflammatory bowel disease.

### **Supplemental Table 4: Classification of complications by Group**

Supplementary Table 4: Complications were grouped into different organ/system-based classifications.

**Table 1:** Baseline demographics, comorbidities, and clinical outcomes of hospitalized COVID-19 patients with clinical phenotypes I, II, and III.

	Phenotype I	Phenotype II	Phenotype III	P-value
	N=236	N=613	N=173	
<b>Demographics*</b>				
<b>Age (years)</b>	67.2 (52.9-79.0)	60.9 (45.9-75.4)	58.6 (34.8-71.3)	<0.001
<b>Male</b>	132 (58.4%)	277 (46.6%)	72 (42.4%)	0.002
<b>Race / Ethnicity</b>				0.002
White	81 (38.8%)	257 (45.6%)	99 (60.7%)	
Black	53 (25.4%)	105 (18.7%)	30 (18.4%)	
Asian	39 (18.7%)	101 (17.9%)	19 (11.7%)	
Hispanic	26 (12.4%)	66 (11.7%)	11 (6.7%)	
Declined	3 (1.4%)	22 (3.9%)	3 (1.8%)	
Other	7 (3.3%)	12 (2.1%)	1 (0.6%)	
<b>Non-English Speaking</b>	113 (47.9%)	240 (39.2%)	41 (23.7%)	<0.001
<b>National ADI</b>	44.5 (25.0-56.0)	43.0 (25.0-56.0)	37.0 (26.0-62.0)	0.76
<b>BMI (kg/m<sup>2</sup>), mean (SD)</b>	29.5 (8.9)	30.8 (8.2)	30.4 (13.4)	0.21
<b>Smoker</b>	9 (3.8)	44 (7.2)	18 (10.4)	0.03
<b>Alcohol abuse</b>	14 (5.9)	47 (7.7)	28 (16.2)	<0.001
<b>Comorbidities*</b>				
<b>Elixhauser Comorbidity Index</b>	7.0 (4.0-10.0)	5.0 (3.0-9.0)	5.0 (2.0-8.0)	<0.001
<b>Cardiac</b>	194 (82.2%)	428 (69.8%)	110 (63.6%)	<0.001
<b>Respiratory</b>	55 (23.3%)	198 (32.3%)	68 (39.3%)	0.002
<b>Hematologic</b>	127 (53.8%)	220 (35.9%)	53 (30.6%)	<0.001
<b>Metabolic</b>	175 (74.2%)	477 (77.8%)	121 (69.9%)	0.08
<b>Renal</b>	92 (39.0%)	170 (27.7%)	37 (21.4%)	<0.001
<b>Hepatic</b>	46 (19.5%)	82 (13.4%)	25 (14.5%)	0.08
<b>Autoimmune</b>	40 (16.9%)	126 (20.6%)	23 (13.3%)	0.07
<b>Cancer</b>	29 (12.3%)	73 (11.9%)	16 (9.2%)	0.58
<b>Cerebrovascular disease</b>	52 (22.0%)	106 (17.3%)	33 (19.1%)	0.28
<b>Blood Type O</b>	72 (42.4%)	158 (39.0%)	39 (37.5%)	0.67
<b>In-hospital Complications*</b>				
<b>Cardiovascular</b>	16 (6.8%)	46 (7.5%)	13 (7.5%)	0.93
<b>Respiratory</b>	49 (20.8%)	104 (17.0%)	14 (8.1%)	0.002
<b>Hematologic</b>	27 (11.4%)	35 (5.7%)	10 (5.8%)	0.01
<b>Renal</b>	54 (22.9%)	60 (9.8%)	7 (4.0%)	<0.001
<b>Metabolic</b>	85 (36.0%)	141 (23.0%)	18 (10.4%)	<0.001
<b>Hepatic</b>	21 (8.9%)	4 (0.7%)	2 (1.2%)	<0.001
<b>Infectious</b>	76 (32.2%)	134 (21.9%)	27 (15.6%)	<0.001
<b>Clinical Outcomes*</b>				
<b>ICU Admission</b>	158 (66.9%)	220 (35.9%)	34 (19.7%)	<0.001
<b>Mechanical Ventilation</b>	98 (41.5%)	88 (14.4%)	4 (2.3%)	<0.001
<b>Hospital Readmission</b>	6 (2.5%)	29 (4.7%)	14 (8.1%)	0.03
<b>ECMO</b>	7 (3.0%)	1 (0.2%)	0 (0.0%)	<0.001

<b>In- or Out of hospital mortality</b>	63 (26.7%)	57 ( 9.3%)	7 ( 4.0%)	<0.001
<b>Admission Vitals and Labs*</b>	<b>Phenotype I</b>	<b>Phenotype II</b>	<b>Phenotype III</b>	<b>P value</b>
<b>Heart rate (mean (SD))</b>	96.17 (20.82)	93.93 (19.35)	90.16 (22.3)	0.01
<b>Respiratory rate</b>	22.0 (18.0-28.0)	20.0 (18.0-23.0)	18.0 (16.0-20.0)	<0.001
<b>Oxygen saturation</b>	94.0 (89.0-97.0)	95.0 (92.0-97.0)	97.0 (95.0-99.0)	<0.001
<b>Pulse pressure</b>	55.0 (43.5-70.5)	53.0 (43.0-68.0)	51.0 (40.0-62.0)	0.02
<b>SBP (mean (SD))</b>	133.29 (27.14)	132.46 (23.54)	134.10 (26.26)	0.72
<b>Total protein</b>	6.5 (5.9-7.0)	6.7 (6.20-7.2)	6.6 (6.2-7.1)	0.01
<b>Red cell distribution width</b>	14.1 (13.2-15.4)	13.5 (12.9-14.7)	13.5 (12.8-14.6)	<0.001
<b>Mean corpuscular volume</b>	90.0 (86.0-94.0)	89.0 (85.0-93.0)	92.0 (88.0-95.3)	<0.001
<b>Alkaline phosphatase</b>	88.0 (67.5-129.0)	71.0 (55.5-92.0)	72.0 (58.-88.0)	<0.001
<b>Calcium</b>	8.10 (7.6-8.5)	8.30 (8.0-8.7)	8.40 (8.1-8.9)	<0.001
<b>Anion gap</b>	9.0 (7.0-12.0)	8.0 (6.0-10.0)	7.0 (6.0-9.0)	<0.001
<b>CO2</b>	23.25 (21.0-26.0)	24.0 (22.0-27.0)	25.0 (23.0-27.8)	<0.001
<b>Hematocrit</b>	36.40 (32.3-40.2)	37.60 (33.6-41.1)	38.45 (35.7-41.5)	<0.001
<b>Aspartate aminotransferase</b>	55.0 (38.0-95.0)	35.0 (24.0-53.0)	29.0 (20.0-44.0)	<0.001
<b>Glucose</b>	122.0 (101.0-165.0)	112.0 (96.0-149.5)	104.0 (91.0-126.5)	<0.001
<b>Absolute monocyte count</b>	0.40 (0.3-0.8)	0.40 (0.3-0.6)	0.50 (0.3-0.7)	<0.001
<b>Platelets</b>	206.0 (160.0-290.0)	190.0 (149.0-243.0)	196.0 (142.5-247.5)	0.01
<b>Albumin</b>	2.40 (2.0-2.7)	2.80 (2.5-3.1)	3.10 (2.8-3.4)	<0.001
<b>Bilirubin</b>	0.70 (0.4-1.1)	0.40 (0.3-0.6)	0.40 (0.3-0.6)	<0.001
<b>INR</b>	1.11 (1.03-1.28)	1.06 (0.99-1.17)	1.08 (0.98-1.21)	0.001
<b>Lactate dehydrogenase</b>	460.5 (380.0-562.8)	308.0 (249.0-394.0)	231.0 (180.0-293.5)	<0.001
<b>Potassium</b>	4.0 (3.6-4.3)	3.80 (3.6-4.2)	3.80 (3.6-4.2)	0.101
<b>Sodium</b>	137.5 (134.0-141.0)	137.0 (135.0-139.0)	138.0 (136.0-140.0)	0.003
<b>D-dimer</b>	3.08 (1.71-5.57)	0.87 (0.59-1.27)	0.60 (0.36-1.05)	<0.001
<b>Hemoglobin</b>	11.90 (10.5-13.1)	12.20 (10.7-13.5)	12.40 (11.3-13.7)	0.01
<b>C-reactive protein</b>	157.0 (102.0-244.0)	89.0 (55.0-134.8)	12.0 (5.0-20.0)	<0.001
<b>Creatinine</b>	1.06 (0.77-1.62)	0.84 (0.69-1.13)	0.80 (0.68-1.03)	<0.001
<b>Absolute neutrophil count</b>	8.05 (5.75-11.42)	4.20 (3.0-6.0)	2.90 (1.8-4.3)	<0.001
<b>Absolute lymphocyte count</b>	0.90 (0.6-1.3)	0.90 (0.7-1.3)	1.30 (0.9-1.7)	<0.001
<b>WBC</b>	8.74 (5.68-15.42)	4.50 (3.0-6.71)	2.36 (1.31-3.77)	<0.001
<b>Gamma Gap</b>	9.80 (7.2-13.2)	5.90 (4.3-7.6)	4.90 (3.9-7.3)	<0.001

\* Categorical variables presented as count (%), continuous variables presented as median (interquartile range) unless otherwise specified.

Abbreviations: ADI, area deprivation index; BMI, body mass index; INR, internal normalized ratio of prothrombin time; ECMO, Extracorporeal membrane oxygenation; ICU, Intensive Care Unit



**Table 2:** Association of Clinical Phenotype with Clinical Outcomes

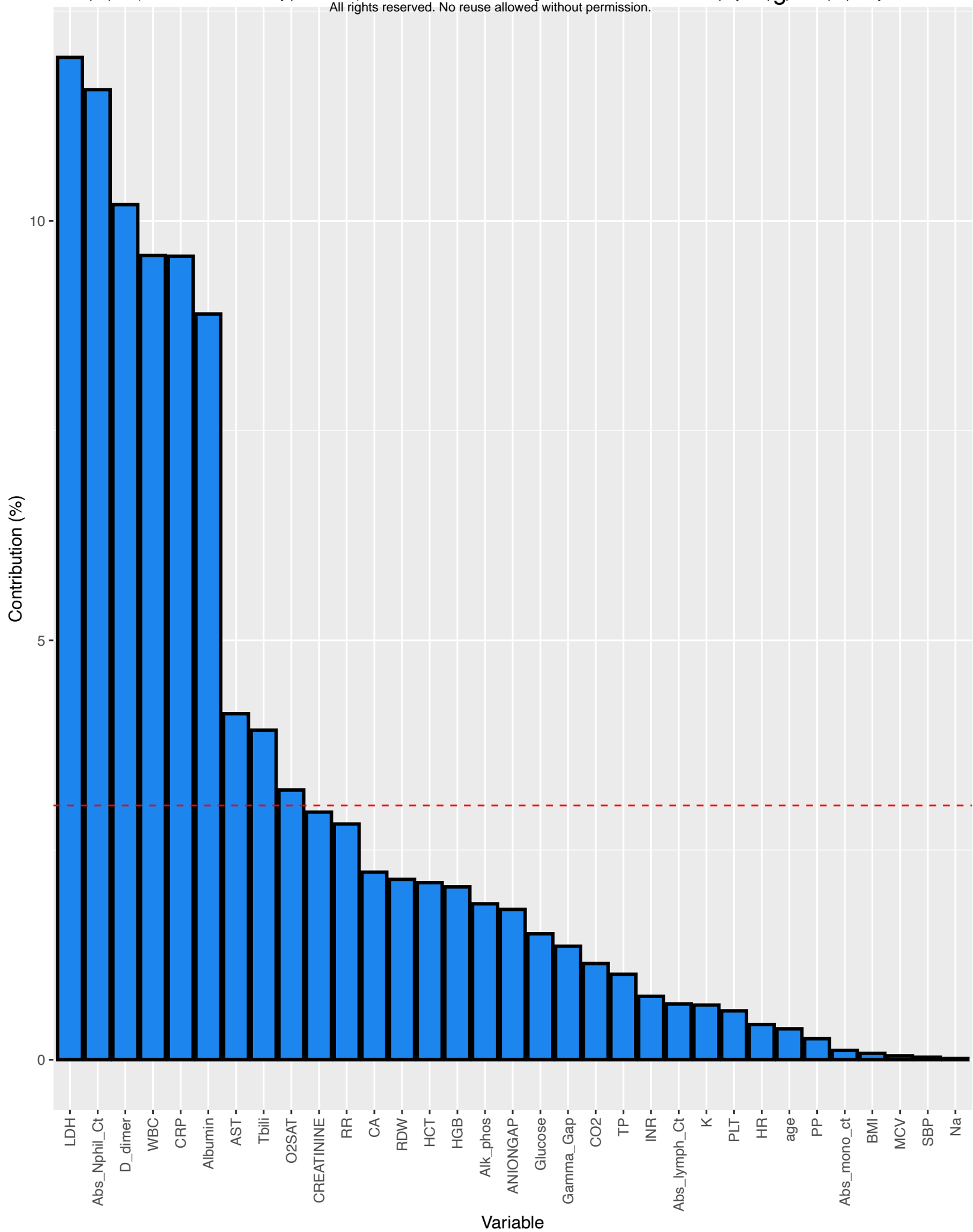
<b>In- and Out- of Hospital Mortality</b> ( <i>Cox PH</i> )	<b>HR</b>	<b>95% CI</b>	<b>P value</b>
<b>Mortality</b>			<0.001 (LR test)
Phenotype I	7.30	3.11 – 17.17	<0.001
Phenotype II	2.57	1.10 – 6.00	0.03
<b>Binary Outcomes</b> ( <i>Logistic Regression</i> )	<b>OR</b>	<b>95% CI</b>	<b>P value</b>
<b>ICU Admission</b>			<0.001 (LR test)
Phenotype I	7.88	4.65 – 13.37	<0.001
Phenotype II	2.32	1.46 – 3.68	<0.001
<b>Mechanical Ventilation</b>			<0.001 (LR test)
Phenotype I	25.59	7.69 – 85.17	<0.001
Phenotype II	7.45	2.27 – 24.43	<0.001
<b>Count Outcome</b> ( <i>Binomial Regression</i> )	<b>IRR</b>	<b>95% CI</b>	<b>P value</b>
<b>Hospital LOS*</b>			<0.001 (LR test)
Phenotype I	1.74	1.45 – 2.10	<0.001
Phenotype II	1.22	1.05 – 1.43	0.01

**Abbreviations:** PH, proportional hazards; HR, hazard ratio; CI, confidence interval; OR, odds ratio; ICU, intensive care unit; IRR, incidence rate ratio; LOS, length of stay; LR, likelihood ratio

**Legend:** Reference group for all models is Phenotype III. All models adjusted for sex, race/ethnicity, and Elixhauser Comorbidity Index.

\* LOS model only included patients that survived.





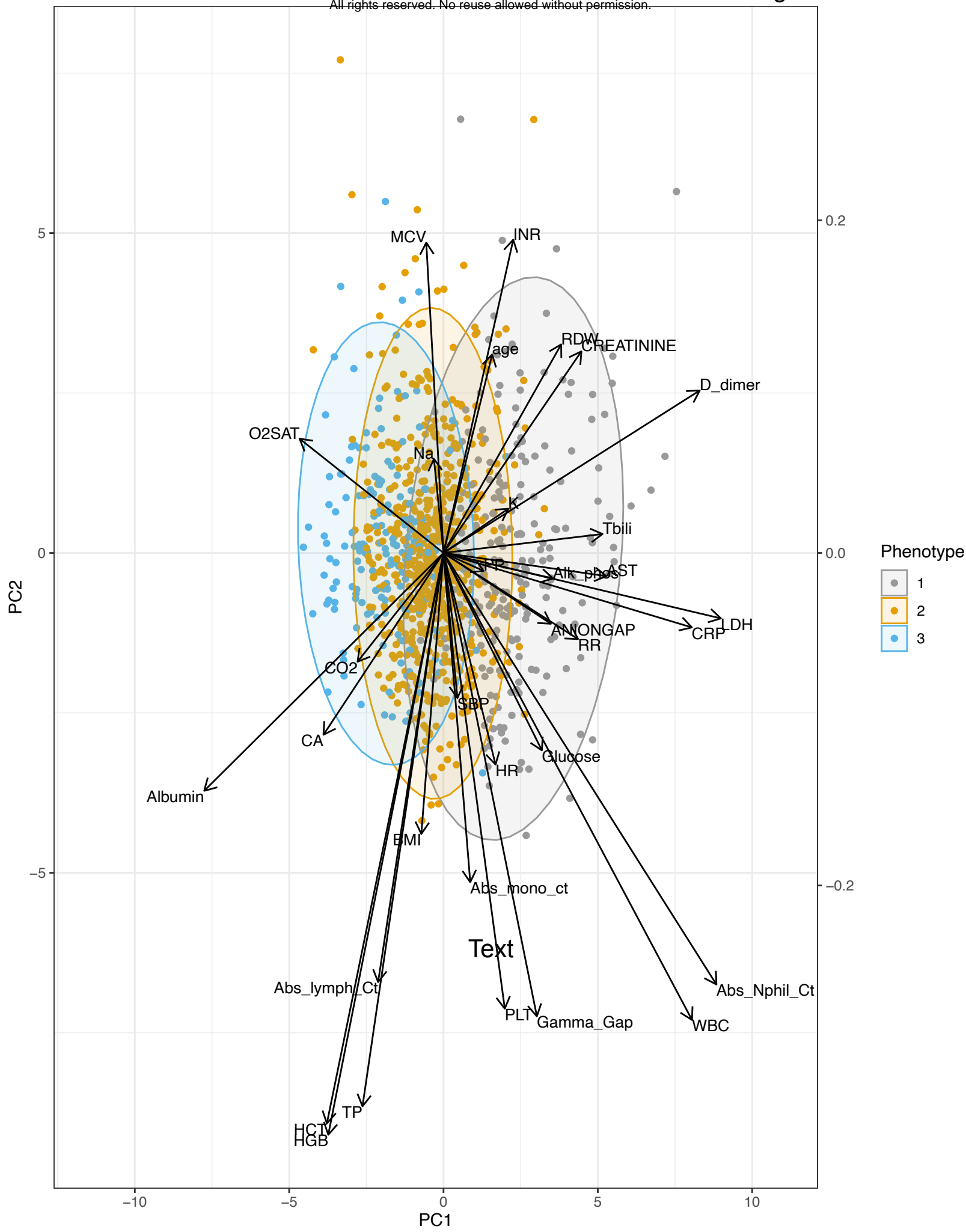


Figure 4

