medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Identifying organ dysfunction trajectory-based subphenotypes in critically ill patients with COVID-19

Authors: Chang Su, PhD[1], Zhenxing Xu, PhD[1], Katherine Hoffman, MS[1], Parag Goyal, MD[2,3], Monika M Safford, MD[2,3], Jerry Lee, BA[4], Sergio Alvarez-Mulett MD[5,3], Luis Gomez-Escobar, MD[5,3], David R Price, MD[5,3], John S Harrington, MD[5,3], Lisa K Torres, MD[5,3], Fernando J Martinez, MD[5,3], Thomas R. Campion, Jr., PhD[1], Rainu Kaushal, MD[1,3], Augustine M.K. Choi, MD[5,3], Fei Wang*, PhD[1], Edward J. Schenck*, MD[5,3]

[1]Department of Population Health Sciences, Weill Cornell Medicine, New York, NY;

[2]Division of General Internal Medicine, Joan and Sanford I. Weill Department of Medicine, Weill Cornell Medicine, New York, NY;

[3]NewYork-Presbyterian Hospital, Weill Cornell Medicine, New York, NY;

[4]Weill Cornell Medical College, Weill Cornell Medicine, New York, NY;

[5]Division of Pulmonary & Critical Care Medicine, Joan and Sanford I. Weill Department of Medicine, Weill Cornell Medicine, New York, NY;

*Corresponding authors

Author Contributions:

ES and FW for conceptualization, investigation, writing, reviewing and editing of the manuscript. CS for data analysis, drafting, editing and reviewing manuscript. ZX, KH for data analysis. TC for providing data support. PG, MS, SAM, LGE, DP, JS, LT, FM, RK, AC for discussion, commenting and editing the manuscript.

Funding: This study received support from NewYork-Presbyterian Hospital (NYPH) and Weill Cornell Medical College (WCMC), including the Clinical and Translational Science

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Center (CTSC) (UL1 TR000457) and Joint Clinical Trials Office (JCTO). FW and CS are supported by NSF IIS 2027970, 1750326, ONR N00014-18-1-2585.

At a Glance Commentary

Scientific Knowledge on the Subject: COVID-19 associated respiratory failure leads to a significant risk of morbidity and mortality. It is clear that there is heterogeneity in the viral-induced host response leading to differential outcomes, even amongst those treated with mechanical ventilation. There are many studies of COVID-19 disease which use intubation status as an outcome or an inclusion criterion. However, there is less understanding of the post intubation course in COVID-19.

What This Study Adds to the Field: We have developed and validated a novel subphenotyping model based on post-intubation organ dysfunction trajectory in COVID-19 patients. Specifically, we identified clear worsening and recovering organ dysfunction trajectory subphenotypes, which are more predictive of outcomes than illness severity at baseline. Dynamic inflammatory markers and ventilator variables rather than baseline severity of illness, demographics and comorbidities differentiate the worsening and recovering subphenotypes. Trajectory subphenotypes offer a potential road map for understanding the evolution of critical illness in COVID-19.

Word Count: 415

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract

Rationale. COVID-19-associated respiratory failure offers the unprecedented opportunity to evaluate the differential host response to a uniform pathogenic insult. Prior studies of Acute Respiratory Distress Syndrome (ARDS) have identified subphenotypes with differential outcomes. Understanding whether there are distinct subphenotypes of severe COVID-19 may offer insight into its pathophysiology.

Objectives. To identify and characterize distinct subphenotypes of COVID-19 critical illness defined by the post-intubation trajectory of Sequential Organ Failure Assessment (SOFA) score.

Methods. Intubated COVID-19 patients at two hospitals in New York city were leveraged as development and validation cohorts. Patients were grouped into mild, intermediate, and severe strata by their baseline post-intubation SOFA. Hierarchical agglomerative clustering was performed within each stratum to detect subphenotypes based on similarities amongst SOFA score trajectories evaluated by Dynamic Time Warping. Statistical tests defined trajectory subphenotype predictive markers.

Measurements and Main Results. Distinct worsening and recovering subphenotypes were identified within each stratum, which had distinct 7-day post-intubation SOFA progression trends. Patients in the worsening suphenotypes had a higher mortality than those in the recovering subphenotypes within each stratum (mild stratum, 29.7% vs. 10.3%, p=0.033; intermediate stratum, 29.3% vs. 8.0%, p=0.002; severe stratum, 53.7% vs. 22.2%, p<0.001). Worsening and recovering subphenotypes were replicated in the validation cohort. Routine laboratory tests, vital signs, and respiratory variables rather

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

than demographics and comorbidities were predictive of the worsening and recovering subphenotypes.

Conclusions. There are clear worsening and recovering subphenotypes of COVID-19 respiratory failure after intubation, which are more predictive of outcomes than baseline severity of illness. Organ dysfunction trajectory may be well suited as a surrogate for research in COVID-19 respiratory failure.

Abstract Word Count: 268

Keywords (mesh 3-5): Coronavirus disease 2019; Coronavirus Infections; Respiratory Failure; Organ Dysfunction Scores; Intubation

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Main text

Introduction

The COVID-19 pandemic has created an unprecedented opportunity to explore a large cohort of patients infected with a single pathogen thus providing a window to examine patient variability in response to a uniform insult. Despite this opportunity, distinct subphenotypes of severe-COVID-19 associated respiratory failure remain largely unexplored(1-3). SARS-CoV-2 infection often leads to hypoxemic respiratory failure requiring treatment with mechanical ventilation which meets clinical and pathologic criteria for Acute Respiratory Distress Syndrome (ARDS)(4-6). In COVID-19 respiratory failure, like other forms of ARDS, there is significant risk of morbidity and mortality. However, there is clear heterogeneity in outcomes, even in those treated with mechanical ventilation(4, 5, 7-9). The baseline clinical characteristics and predictors of mortality of those requiring mechanical ventilation have been described(4, 7, 8, 10). These studies offer some insight into a differential host response but are limited to characterizing patients at baseline.

In prior studies of ARDS(11, 12), unique subphenotypes have been described, which identify hyperinflammatory and hypoinflammatory populations with differential demographics, clinical characteristics, inflammatory markers and outcomes. These subphenotypes are primarily characterized by host response inflammatory markers and patterns of organ injury, but are agnostic of the type of insult or infection. In COVID-19, baseline risk stratification may be insufficient to characterize subphenotypes that accurately reflect the complexity of the disease arc(13). Serial, temporally ordered,

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Sequential Organ Failure Assessment (SOFA)(14-17) and comprehensive Electronic Health Records (EHR) data are well suited to develop data-driven subphenotypes(18), where the goal is to identify coherent patient groups with similar clinical courses. Dynamic time warping (DTW)(19) is a well-established technique for evaluating the similarities among temporal sequences(20, 21). DTW is particularly well suited to evaluate longitudinal changes in organ dysfunction in COVID19. Characterizing a more complete representation of the disease course in COVID19 may offer insight into its pathophysiology.

We used DTW to conduct a two staged post-intubation trajectory analysis of SOFA-based organ dysfunction in patients with COVID19 to identify unique subphenotypes. In order to understand the differential disease course, we then explored clinical and biologic features including demographics, comorbidities, clinical characteristics, inflammatory markers, and treatments predictive of these trajectories.

Methods

This was a retrospective two staged modeling analysis on two cohorts of intubated COVID-19 patients. The overall workflow of our study is illustrated in Figure 1.

Study design and cohort description

We used individual patient data from two New York Presbyterian (NYP) system hospitals located in New York city: the New York Presbyterian Hospital-Weill Cornell

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Medical Center (NYP-WCMC), an 862-bed quaternary care hospital, and the New York Presbyterian-Lower Manhattan Hospital (NYP-LMH), a 180-bed non-teaching academic affiliated hospital. Patients were admitted from Mar 3, 2020 to May 12, 2020. SARS-CoV2 diagnosis was made through reverse-transcriptase–PCR assays performed on nasopharyngeal swabs. The critical care response to the pandemic has been previously described(22). The NYP-WCMC cohort was used as the development cohort to derive subphenotypes, and the NYP-LMH cohort was used for validation. The focus of this study was critically ill patients with COVID-19 who were treated with intubation (Supplemental Appendix 1).

Data collection

We collected all data from either the Weill Cornell-Critical carE Database for Advanced Research (WC-CEDAR), Weill Cornell Medicine COVID Institutional Data Repository (COVID-IDR), or via manual chart abstraction (REDCap). WC-CEDAR aggregates and transforms data from institutional electronic health records for all patients treated in ICUs in NYP-WCMC and NYP-LMH(23). The COVID-IDR contains additional aggregate EHR data on all patients who were tested for SARS-CoV-2 at NYP-WCMC or NYP-LMH. The REDCap database contains high-quality manually abstracted data on all patients who tested positive for COVID-19 at NYP-WCMC or NYP-LMH(24). In our analysis, the patient information incorporated included demographics, laboratory tests, vital signs, and respiratory variables obtained from WC-CEDAR, comorbidity information obtained

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

from the REDCap database, and medication data obtained from the COVID-IDR. Data analyzed were detailed in Supplemental Appendix 2.

SOFA calculation

The SOFA score is the sum of six organ dysfunction subscores, including cardiovascular, central nervous system (CNS), coagulation, liver, renal, and respiration(14, 17). In this study, the CNS, coagulation, liver, and renal subscores were derived according to the standard SOFA scoring system(14). The respiration subscore was calculated using a combination of the traditional and modified scoring method(25). The cardiovascular SOFA subscore was calculated with additional vasopressors according to a norepinephrine equivalency table, where phenylephrine and vasopressin were converted to a norepinephrine equivalency(26). SOFA scores were derived every 24 hours from the time of intubation, and the worst score within that 24-hour data period was selected for each patient .(14)

Inclusion exclusion criteria

We included patients with positive results on viral RNA detection by real-time reverse transcriptase polymerase chain reaction (RT-PCR) test from nasopharyngeal swabs specimens and treated with mechanical ventilation at the ICU in NYP-WCM and NYP-LMH. We excluded patients who were less than 18 years old. Since our aim was to identify clinically meaningful organ dysfunction progression patterns of intubated patients, trajectories with low quality (20 (5.7%) patients missing over 50% SOFA

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

records) and outlier trajectories (10 (2.9%) patients with unchanged or heavily fluctuated within the 7-day window after intubation) were excluded from the analysis (Supplemental Appendix 3 and Figure E-1).

Subphenotyping model description

SOFA scores were derived every 24 hours and post intubation 7 day SOFA trajectories were constructed for analysis. Missing values within a trajectory were imputed based on the last observation carried forward (LOCF) strategy.

A two-staged subphenotyping method was performed to derive SOFA trajectory subphenotypes (Figure 1). In the first stage, we used baseline SOFA to group patients with a similar upfront risk of death(17), as additive organ dysfunction has previously been identified to be associated with poor outcomes in COVID19(8). We partitioned the patients into three baseline severity strata (mild, intermediate, and severe) according to their SOFA scores within the first 24 hours after intubation. The SOFA score cut-offs were set to 0-10, 11-12, and 13-24 in order to obtain a balanced distribution of patients across the three strata. In the second stage, we identified the subphenotypes with similar 7-day SOFA progression patterns. Dynamic Time Warping (DTW)(19) was adopted to evaluate the similarities between pairwise patient SOFA trajectories within each baseline stratum and then hierarchical agglomerative clustering (HAC)(27) was performed on these similarities to derive the similar patient clusters as trajectory subphenotypes. DTW can account for the differences among the evolution heterogeneity among the temporal curves and is thus able to evaluate their similarity more robustly.(19) The optimal numbers of subphenotypes were determined by clear

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

separation illustrated by clustergram according to the McClain index(28). DTW was implemented with Python 3.7 based on tslearn package 0.3.1 and HAC was performed with Python 3.7 based on scikit-learn package 0.22.2.

To validate these findings, we replicated these subphenotypes from the NYP-LMH cohort.

Clinical outcomes

We analyzed 30-day all-cause mortality as the primary outcome for patients within each phenotype. Successful extubation or need for tracheostomy within 30 days after intubation were secondary outcomes.

Statistical analysis

We examined the associations between patient characteristics and clinical variables and the identified trajectory subphenotypes, to see if there are early markers that can discriminate between them. Patient characteristics we investigated included demographics, comorbidities, medications prescribed within the window from 3-day before to 5-day after intubation, and blood type(29). Laboratory test results included: complete blood count, basic metabolic panel, liver function tests, coagulation profile and inflammatory markers including d-dimer, fibrinogen, ferritin, erythrocyte sedimentation rate, lactic acid, troponin, lactate dehydrogenase, creatine kinase, procalcitonin and C-reactive protein. Vital signs included: GCS, mean arterial pressure and temperature, urine output. Respiratory variables included: P/F ratio, FiO2, Pao2, PaCO2, PH, PEEP,

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

peak inspiratory pressure, plateau pressure, driving pressure, static compliance, minute ventilation, ventilatory ratio, and tidal volume indexed to ideal body weight at day 1 and day 3 post-intubation.

Univariate statistical tests were performed in those association analyses. Specifically, one-way analysis of variance (ANOVA, with Tukey HSD post hoc test), Kruskal–Wallis test (with Dunn post hoc test), student's t-test, Mann-Whitney test, Chi-square test, and Fisher's exact test have been used whenever appropriate. The p-values were then corrected for multiple testing using false discovery rate (FDR) estimation. Analysis of covariance (ANCOVA) for the between-strata/subphenotypes comparisons was also applied based on the generalized linear model (GLM) with adjustment on age at baseline. All statistical tests were performed with Python 3.7 based on statsmodels package 0.11.1.

Subphenotype prediction modeling

We trained a random forest model with the trajectory subphenotypes as targets and the patient clinical characteristics at specific time points after intubation as input predictors to define if these trajectory subphenotypes can be predicted early. Our implementation was with Python 3.7 based on scikit-learn package 0.22.2. Candidate predictors included demographics, comorbidities, medications prescribed around the intubation event, SOFA subscores, laboratory tests, vital signs, and respiratory variables as described above. Prediction performances were measured by area under the receiver operating characteristics (AUC-ROC). The importance of predictors was visualized as a heatmap to demonstrate their contributions on subphenotype prediction.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

IRB approval

The study is approved by the IRB of Weill Cornell Medicine with protocol number 20-04021909.

Results

Patients and baseline severity strata

A total of 318 mechanically ventilated COVID-19 patients from the NYP-WCMC cohort were included for analysis, consisting of 100 females (31.45%) and an average age of 62.78 $\pm$ 14.34. One day post-intubation the mean SOFA score for this cohort is 11.89 $\pm$ 2.56. A total of 84 mechanically ventilated COVID-19 patients from the NYP-LMH were included as a validation cohort, consisting of 33 (39.29%) females and an average age of 66.06 $\pm$ 13.06. One day post-intubation the mean SOFA score is 12.51 $\pm$ 2.25. The clinical characteristics of both cohorts are summarized in Table 1.

For the NYP-WCMC cohort, patients were first partitioned into mild, intermediate, and severe strata based on the SOFA scores within one day after intubation, consisting of 76 (23.29%), 116 (36.48%), and 126 (39.62%) patients, respectively; while for the NYP-LMH validation cohort, the three strata consist of 10 (11.90%), 35 (41.67%), and 39 (46.43%) patients, respectively. As shown in Table 1, the patients in both NYP-WCMC and NYP-LMH cohorts exhibit additive patterns of post intubation baseline organ

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

dysfunction according to the SOFA subscores. Specifically, CNS and respiration dysfunction were present in the mild stratum; the intermediate stratum had additional cardiovascular dysfunction on top of CNS and respiratory dysfunction compared to the mild stratum; and the severe stratum had renal dysfunction in addition to all other organ failure. Liver and coagulation dysfunction were rare in all strata. Patients in the severe stratum were generally older and were more likely to suffer from chronic comorbidities at baseline.

SOFA trajectory subphenotypes

The clustergrams built upon the pairwise SOFA trajectory distance matrix derived by DTW are shown in Supplemental Figure E-2. The optimal number of subphenotypes within each stratum as determined by the McClain Index(28) are shown in Supplemental Table E-1, suggesting two being the best choice across all strata in both cohorts. Figure 2 demonstrates the individual averaged SOFA curves for patients in the two subphenotypes across all strata: a worsening subphenotype of which SOFA score increased within the 7-day observation window, and a recovering subphenotype of which SOFA score improved. The clinical characteristics of these subphenotypes were summarized in Table 2. Overall, there was no marked difference in terms of demographics, comorbidity burden, and pattern of organ dysfunction (distribution of SOFA subscores and total score) between the worsening and recovering subphenotypes within each baseline severity stratum at baseline. This suggests that, though the subphenotypes varied in 7-day organ dysfunction progression patterns, they have similar clinical status immediately after intubation. We further investigated

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

medications prescribed within each subphenotype and didn't find significant signal as well (Supplemental Table E-3). In addition, clinical characteristics and medications of the subphenotypes re-derived in the NYP-LMH validation cohort were summarized in Supplemental Tables E-2 and 4.

30-days clinical outcomes

Statistics of 30-day post-intubation clinical primary and secondary outcomes (mortality, extubation, and tracheostomy) of subphenotypes were illustrated in Figure 2A and Supplemental Figure E-3A. The worsening subphenotypes, across baseline strata, suffered from a significantly higher risk of mortality within the 30-day window after intubation (worsening vs recovering, mortality proportion: mild stratum, 29.7% vs. 10.3%, p=0.033; intermediate stratum, 29.3% vs. 8.0%, p=0.002; severe stratum, 53.7% vs. 22.2%, p<0.001). The recovering subphenotypes, across all baseline strata, showed significantly higher extubation proportions within the 30-day window compared to the worsening subphenotypes (recovering vs. worsening, extubation proportion: mild stratum, 76.9% vs. 27.0%, p<0.001; intermediate stratum, 54.7% vs. 31.7%, p=0.018; severe stratum 50.0% vs. 14.8%, p<0.001). There was no significant difference of 30-day tracheostomy detected between the subphenotypes. Importantly, the recovering subphenotype within the severe baseline stratum had a lower mortality risk compared to the worsening subphenotypes at mild and intermediate baseline strata.

The trajectory subphenotypes derived in the NYP-LMH validation cohort had similar trends in all three clinical outcomes within the 30-day window after intubation (see Figure 2A and Supplemental Figure E-3B). Across all baseline strata, the

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

worsening subphenotypes accounted for higher risk of mortality (worsening vs recovering, mortality proportion: mild stratum, 57.1% vs. 0.0%, p=0.200; intermediate stratum, 31.8% vs. 7.7%, p=0.211; severe stratum, 83.3% vs. 17.4%, p<0.001), while the recovering subphenotypes showed higher extubation proportion within 30-days after intubation (recovering vs. worsening, extubation proportion: mild stratum, 33.3% vs. 14.3%, p=0.490; intermediate stratum, 69.2% vs. 27.3%, p=0.015; severe stratum, 48.1% vs. 9.1%, p=0.017).

Correlation of subphenotypes with early-stage markers

Vital signs, laboratory variables, and respiratory variables were evaluated to identify early-stage markers predictive of the two-stage classification. First of all, the three baseline strata of the NYP-WCMC cohort were observed to be well separated by a series of clinical variables in addition to the differential organ dysfunction pattern noted above (Supplemental Table E-5). For instance, ANOVA(or Kruskal–Wallis test) results showed significantly increased laboratory values like procalcitonin, ferritin, lactate dehydrogenase (LDH), and creatinine, and decreased carbon dioxide within the severe stratum, at day 1 post-intubation; Vital sign such as the Glasgow Coma Scale (GCS) and urine output volume and respiratory variables such as peak inspiratory pressure (PIP) were also observed to separate the strata; the mild stratum was associated with significantly decreased levels of troponin, creatinine, and glucose. Detailed statistical analyses are described in Supplemental Table E-5. Statistics of these clinical variables across baseline strata within the NYP-LMH validation cohort showed similar signals and were detailed in Supplemental Table E-6.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

We further evaluated the clinical variable values at day 1 and day 3 post-intubation to compare the worsening and recovering subphenotypes within each baseline stratum. Markers, which significantly separated the worsening and recovering subphenotypes (p-value<0.05), varied across different baseline strata (Supplemental Table E-7). Specifically, on day 1, laboratory tests including white blood cell (WBC) count, procalcitonin, CRP, creatinine, neutrophil count, and globulin, and respiratory variables such as minute ventilation and ratio of tidal volume to predicted body weight (PBW) were found to significantly differentiate the trajectory subphenotypes within the mild stratum. Markers that significantly separated the trajectory subphenotypes within the intermediate stratum on day 1 included AST, creatinine, LDH and FiO2. The severe stratum had more markers separating the worsening and recovering subphenotypes on day 1, including laboratory tests like lymphocyte count, platelet count, triglycerides, procalcitonin level, ferritin, troponin, LDH, creatinine, and AST, respiratory variables such as PIP, PaO2, PaO2/FiO2 ratio (P/F ratio), and SpO2. In addition, more markers separating subphenotypes emerged within each stratum on day 3 after intubation. For instance, additional day 3 markers including GCS, platelet, sodium, and glucose were found within the mild stratum. For the intermediate stratum, GCS, albumin, WBC count, lymphocyte percentage, CRP, sodium, neutrophil percentage and count, carbon dioxide level, urine output volume, and P/F ratio were found to be additional markers on day 3. The additional markers found within the severe stratum on day 3 included vital signs GCS and temperature, lymphocyte percentage, hemoglobin level, albumin level, potassium level, ESR, CK, carbon dioxide level, and lactic acid, positive end-expiratory

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

pressure (PEEP), plateau pressure, minute ventilation, static compliance, driving pressure, FiO2, and ventilator ratio.

As shown in Supplemental Table E-8, most markers identified within the NYP-WCMC cohort showed consistent signals within the NYP-LMH subphenotypes, even though some significance vanishes, as the confidence intervals were wide.

Subphenotype prediction models

We trained random forest models for predicting the worsening and recovering trajectory subphenotypes within each baseline stratum according to the early stage marker values. Overall, as shown in Supplemental Figure E-4, within the mild, intermediate, and severe strata, the prediction models achieved the AUC-ROCs of 0.70 (95% confidence interval [CI], [0.59, 0.80]), 0.67 (95% CI, [0.62, 0.72]), and 0.73 (95% CI, [0.67, 0.79]) respectively, with the predictor values evaluated at day 1 post-intubation. AUC-ROCs of the models increased to 0.77 (95% CI, [0.66, 0.89]), 0.77 (95% CI, [0.74, 0.81]), and 0.80 (95% CI, [0.75, 0.85]), with the predictor values evaluated at day 3 post-intubation; and to 0.83 (95% CI, [0.72, 0.94]), 0.91 (95% CI, [0.88, 0.94]), and 0.88 (95% CI, [0.80, 0.95]), with the predictor values evaluated at day 5 post-intubation.

Importance of the predictors were illustrated as heatmaps, where color intensity represents the normalized importance of specific predictors (Supplemental Figure E-5). Generally, predictor importance varied as the progress of time. Models trained on day 1-3 after intubation were observed to involve more contributions from the laboratory tests, vital signs, respiratory variables than other predictors; SOFA subscores, especially

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

cardiovascular, CNS, and renal subscores showed relatively higher importance over models trained on day 4 or 5 data within the intermediate and severe strata. Age contributed to day 1-3 prediction to some extent, while other demographics, medications and comorbidities showed weak importance in prediction.

Correlation to blood type

We finally assessed ABO/RH blood type distribution across the subphenotypes (Supplemental Tables E-9 and 10). Overall, there is no significant signal detected.

Discussion:

In this study, we identified novel trajectory subphenotypes of COVID-19 patients with an objective machine learning approach. The subphenotypes we identified are based on organ dysfunction trajectory over 7-days following intubation, which is different from existing data-driven subphenotyping methods that focus on patient data at a specific timestamp (12, 30, 31). The use of novel methodology, in addition to the robust size of our cohort, ensure that the identified trajectory based subphenotypes are less likely to suffer from cognitive bias(13) and are likely to be temporally stable(32). More concretely, we adopted a divide and conquer approach to identify the subphenotypes. Prior research has identified that additive organ dysfunction is predictive of increased mortality in COVID-19 associated ARDS(8). Therefore, we divided the patients into three different baseline strata (mild, intermediate and severe) according to additive

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

SOFA based organ dysfunction. We identified two salient trajectory subphenotypes within each stratum.

Importantly, the baseline demographics, comorbidities and pattern of organ dysfunction did not differ between the worsening and recovering subphenotypes at each stratum. This suggests the existence of differential progression pathways that are irrespective of baseline risk factors for severe disease. This finding is unique compared to other subphenotyping projects as we are including a more complete picture of the disease course(12, 30, 31). It also highlights the temporal heterogeneity of COVID-19 and the importance of avoiding prognostication based on early post intubation clinical characteristics. We found that the worsening subphenotypes in the baseline mild and intermediate strata showed an even higher risk of death compared to the recovering subphenotype within the baseline severe stratum (Figure 3). Indeed, there is an urgent need to understand the pathophysiology of progressive non-pulmonary organ dysfunction in this disease.

We assessed the differences between a broad range of laboratory tests, vital signs, and respiratory variables in the worsening and recovering subphenotypes. Importantly, basic laboratory tests and inflammatory markers were differentially associated with the worsening and recovering subphenotypes over time, which suggests that there is value in clinically following markers such as D-dimer, LDH, ferritin, procalcitonin and C-reactive protein. In the mild stratum, markers from the regular blood panel such as total white blood cell count and neutrophil counts, while inflammatory markers, such as ferritin and LDH, differentiate in the severe stratum. Laboratory tests, vitals and inflammatory markers in the intermediate stratum were less able to

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

distinguish between the worsening and recovering subphenotypes. However, differences emerged over a longer time horizon (e.g., day 3). This further highlights the dynamic nature of COVID-19 and the difficulty in early prognosis in the critically ill population, despite severely deranged baseline organ dysfunction and inflammatory markers.

We built multivariable prediction models for the identified trajectory subphenotypes from patient baseline characteristics and early-stage clinical feature values. Models were built on at successive time points (day 1, 2, 3, 4, and 5) after intubation. Predictive performances measured by AUC-ROC improved as the number of days increased. The predictors' importance to differentiating worsening and recovering subphenotypes showed varying patterns (Figure E-5). Importantly, aside from age and BMI, demographics, baseline comorbidities, and medications prescribed around intubation did not contribute to discriminating the subphenotypes in any of the strata. The persistence or development of renal failure, predicts subphenotype assignment later in the severe and intermediate strata, respectively. The persistence of vasodilatory shock in the intermediate stratum predicts the worsening subphenotype. While the development of thrombocytopenia, is discriminative late in the severe stratum. Interestingly, over the course of the first 7 days following intubation, liver failure remained rare. At different points in the course, inflammatory markers such as creatine kinase and D-dimer predicted worsening and recovering subphenotypes.

Our study was conducted on the two NYP system hospitals. Woresning and recovering SOFA subphenotypes, clinical characteristics, and outcomes from the validation cohort was consistent with the original subphenotypes. Although, due to the

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

limited size of NYP-LMH validation cohort, statistical significance of some markers vanished, most of the results reflected the development cohort's findings. This consistency ensures the existence of the worsening and recovering trajectory subphenotypes at each baseline stratum of the critically ill COVID-19 patients.

Limitations

While this study presents a step forward in the efforts to parse the progression heterogeneity of critically ill patients with COVID-19, several limitations remain. The first limitation could be SOFA's inadequacy in tabulating organ dysfunction in COVID-19 associated respiratory failure. For example, COVID-19 is associated with a different pattern of hypercoagulability compared to sepsis, which is reflected in the preserved platelet count in most this cohort's patients.(33) Also despite elevations in liver chemistries in many patients, hyperbilirubinemia was rare. Despite this limitation, SOFA trajectory subphenotypes predicted mortality and will allow for future comparisons with other diseases.

Second, we did not use the progression of inflammatory markers such as C-reactive protein, D-dimer or ferritin, which are known risk factors for this disease, to identify the subphenotypes. Nor did we stratify patients based the severity of respiratory failure alone. Instead, we chose to see how these factors interacted with traditional organ dysfunction, as most patients with COVID19 die from multisystem organ failure and not refractory respiratory failure(8, 9).

Third, differentiating trajectory subphenotypes in this critically ill population was difficult, as AUC-ROC metrics of prediction modeling using data at day 1 post-intubation were around 0.7. By restricting our analysis to a very high-risk population, we

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

decreased the discriminative power of many of our biomarkers to predict outcomes. All patients were high risk. However, we have added to our understanding of patients with critical COVID-19, by documenting the natural history of organ dysfunction in this population. Future research efforts, with novel biomarkers, are needed to predict worsening and recovering subphenotypes at an earlier time point in those with respiratory failure.

Fourth, the surge conditions in New York City during the study period could affect the study. Care may have been influenced by the surge conditions during this difficult time. However, all patients were cared for in a critical care environment and despite the massive patient burden, the all cause 30-day mortality was 25.9%.

Conclusions

In a population of critically ill patients with COVID-19 respiratory failure, there are distinct worsening and recovering organ dysfunction trajectory subphenotypes. Worsening status was predictive of poor outcomes in all strata regardless of baseline severity. These findings highlight the importance of supportive care for sequential organ failure in addition to respiratory failure in this disease. Trajectory based subphenotypes offer a potential road map for understanding the evolution of critical illness in COVID-19. We call for further analysis.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Acknowledgments

We thank the following Weill Cornell Medicine medical students for their contributions to the COVID-19 Registry through medical chart abstraction: Zara Adamou BA, Haneen Aljayyousi BA, Mark N. Alshak BA (student leader), Bryan K. Ang BA, Elena Beideck BS, Orrin S. Belden BS, Sharmi Biswas MD, Anthony F. Blackburn BS, Joshua W. Bliss PharmD, Kimberly A. Bogardus BA, Chelsea D. Boydstun BA, Clare A. Burchenal MPH, Eric T. Caliendo BS, John K. Chae BA, David L. Chang BS, Frank R. Chen BS, Kenny Chen BA, Andrew Cho PhD, Alice Chung BA, Alisha N. Dua MRes, Andrew Eidelberg BS, Rahmi S. Elahjji BA, Mahmoud Eljaby MMSc, Emily R. Eruysal BS, Kimberly N. Forlenza MSc, Rana Khan Fowlkes BA, Rachel L. Friedlander BA, Gary George BS, Shannon Glynn BS, Leora Haber BA, Janice Havasy BS, Alex Huang BA, Hao Huang BS, Jennifer H. Huang BS, Sonia Iosim BS, Mitali Kini BS, Rohini V. Kopparam BS, Jerry Y. Lee BA, Mark Lee BS BA, Aretina K. Leung BA, Han A. Li BA (student leader), Bethina Liu AB, Charalambia Louka BS, Brienne Lubor BS, Dianne Lumaquin BS, Matthew L. Magruder BA, Ruth Moges MSc, Prithvi M. Mohan BS, Max F. Morin BS, Sophie Mou BA, J. J. Nario BS, Yuna Oh BS, Noah Rossen BA, Emma M. Schatoff PhD, Pooja D. Shah BA, Sachin P. Shah BA, Daniel Skaf BS, Shoran Tamura BS, Ahmed Toure BA, Camila M. Villasante BA, Gal Wald BA, Graham T. Wehmeyer BS (student leader), Samuel Williams BA, Ashley Wu BS, Andrew L. Yin BA, Lisa Zhang BA

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

References

1. Fan E, Beitler JR, Brochard L, Calfee CS, Ferguson ND, Slutsky AS, Brodie D. COVID-19-associated acute respiratory distress syndrome: is a different approach to management warranted? *Lancet Respir Med*.
2. Pan C, Chen L, Lu C, Zhang W, Xia J-A, Sklar MC, Du B, Brochard L, Qiu H. Lung Recruitability in COVID-19–associated Acute Respiratory Distress Syndrome: A Single-Center Observational Study. *Am J Respir Crit Care Med* 2020; 201: 1294-1297.
3. Bos LD, Paulus F, Vlaar APJ, Beenen LFM, Schultz MJ. Subphenotyping ARDS in COVID-19 Patients: Consequences for Ventilator Management. *Ann Am Thorac Soc*.
4. Schenck EJ, Hoffman K, Goyal P, Choi J, Torres L, Rajwani K, Tam CW, Ivascu N, Martinez FJ, Berlin DA. Respiratory Mechanics and Gas Exchange in COVID-19 Associated Respiratory Failure. *Ann Am Thorac Soc* 2020.
5. Ziehr DR, Alladina J, Petri CR, Maley JH, Moskowitz A, Medoff BD, Hibbert KA, Thompson BT, Hardin CC. Respiratory Pathophysiology of Mechanically Ventilated Patients with COVID-19: A Cohort Study. *Am J Respir Crit Care Med* 2020; 201: 1560-1564.
6. Bhatraju PK, Ghassemieh BJ, Nichols M, Kim R, Jerome KR, Nalla AK, Greninger AL, Pipavath S, Wurfel MM, Evans L, Kritek PA, West TE, Luks A, Gerbino A, Dale CR, Goldman JD, O'Mahony S, Mikacenic C. Covid-19 in Critically Ill Patients in the Seattle Region — Case Series. *N Engl J Med* 2020; 382: 2012-2022.
7. Cummings MJ, Baldwin MR, Abrams D, Jacobson SD, Meyer BJ, Balough EM, Aaron JG, Claassen J, Rabbani LE, Hastie J, Hochman BR, Salazar-Schicchi J, Yip NH, Brodie D, O'Donnell MR. Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. *Lancet* 2020; 395: 1763-1770.
8. Gupta S, Hayek SS, Wang W, Chan L, Mathews KS, Melamed ML, Brenner SK, Leonberg-Yoo A, Schenck EJ, Radbel J, Reiser J, Bansal A, Srivastava A, Zhou Y, Sutherland A, Green A, Shehata AM, Goyal N, Vijayan A, Velez JCQ, Shaefi S, Parikh CR, Arunthamakun J, Athavale AM, Friedman AN, Short SAP, Kibbelaar ZA, Abu Omar S, Admon AJ, Donnelly JP, Gershengorn HB, Hernán MA, Semler MW, Leaf DE, Investigators ftS-C. Factors Associated With Death in Critically Ill Patients With Coronavirus Disease 2019 in the US. *JAMA Intern Med* 2020.
9. Grasselli G, Greco M, Zanella A, Albano G, Antonelli M, Bellani G, Bonanomi E, Cabrini L, Carlesso E, Castelli G, Cattaneo S, Cereda D, Colombo S, Coluccello A, Crescini G, Forastieri Molinari A, Foti G, Fumagalli R, Iotti GA, Langer T, Latronico N, Lorini FL, Mojoli F, Natalini G, Pessina CM, Ranieri VM, Rech R, Scudeller L, Rosano A, Storti E, Thompson BT, Tirani M, Villani PG, Pesenti A, Cecconi M, Network ftC-LI. Risk Factors Associated With Mortality Among Patients With COVID-19 in Intensive Care Units in Lombardy, Italy. *JAMA Intern Med* 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

10. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li H, Wu X, Xu J, Tu S, Zhang Y, Chen H, Cao B. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020; 395: 1054-1062.
11. Sinha P, Delucchi KL, McAuley DF, O'Kane CM, Matthay MA, Calfee CS. Development and validation of parsimonious algorithms to classify acute respiratory distress syndrome phenotypes: a secondary analysis of randomised controlled trials. *Lancet Respir Med* 2020; 8: 247-257.
12. Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med* 2014; 2: 611-620.
13. Bos LDJ, Sinha P, Dickson RP. The perils of premature phenotyping in COVID: a call for caution. *Eur Respir J* 2020: 2001768.
14. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med* 1996; 22: 707-710.
15. Karakike E, Kyriazopoulou E, Tsangaris I, Routsi C, Vincent J-L, Giamarellos-Bourboulis EJ. The early change of SOFA score as a prognostic marker of 28-day sepsis mortality: analysis through a derivation and a validation cohort. *Crit Care* 2019; 23: 387.
16. Minne L, Abu-Hanna A, de Jonge E. Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review. *Crit Care* 2008; 12: R161.
17. Ferreira FL, Bota DP, Bross A, Mélot C, Vincent J-L. Serial Evaluation of the SOFA Score to Predict Outcome in Critically Ill Patients. *JAMA* 2001; 286: 1754-1758.
18. Weng C, Shah NH, Hripcsak G. Deep phenotyping: Embracing complexity and temporality-Towards scalability, portability, and interoperability. *J Biomed Inform* 2020; 105: 103433.
19. Müller M. Dynamic Time Warping. Information Retrieval for Music and Motion. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. p. 69-84.
20. Forestier G, Lalys F, Riffaud L, Trelhu B, Jannin P. Classification of surgical processes using dynamic time warping. *J Biomed Inform* 2012; 45: 255-264.
21. Zhang X, Chou J, Liang J, Xiao C, Zhao Y, Sarva H, Henchcliffe C, Wang F. Data-Driven Subtyping of Parkinson's Disease Using Longitudinal Clinical Records: A Cohort Study. *Sci Rep* 2019; 9: 797.
22. Griffin KM, Karas MG, Ivascu NS, Lief L. Hospital Preparedness for COVID-19: A Practical Guide from a Critical Care Perspective. *Am J Respir Crit Care Med* 2020; 201: 1337-1344.
23. Sholle ET, Kabariti J, Johnson SB, Leonard JP, Pathak J, Varughese VI, Cole CL, Campion TR, Jr. Secondary Use of Patients' Electronic Records (SUPER): An Approach for Meeting Specific Data Needs of Clinical and Translational Researchers. *AMIA Annu Symp Proc* 2017; 2017: 1581-1588.
24. Goyal P, Choi JJ, Pinheiro LC, Schenck EJ, Chen R, Jabri A, Satlin MJ, Campion TR, Nahid M, Ringel JB, Hoffman KL, Alshak MN, Li HA, Wehmeyer GT, Rajan M, Reshetnyak E, Hupert N, Horn EM, Martinez FJ, Gulick RM, Safford MM.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Clinical Characteristics of Covid-19 in New York City. *N Engl J Med* 2020; 382: 2372-2374.

25. Rice TW, Wheeler AP, Bernard GR, Hayden DL, Schoenfeld DA, Ware LB. Comparison of the Spo2/Fio2 Ratio and the Pao2/Fio2 Ratio in Patients With Acute Lung Injury or ARDS. *Chest* 2007; 132: 410-417.
26. Khanna A, English SW, Wang XS, Ham K, Tumlin J, Szerlip H, Busse LW, Altaweel L, Albertson TE, Mackey C, McCurdy MT, Boldt DW, Chock S, Young PJ, Krell K, Wunderink RG, Ostermann M, Murugan R, Gong MN, Panwar R, Hästbacka J, Favory R, Venkatesh B, Thompson BT, Bellomo R, Jensen J, Kroll S, Chawla LS, Tidmarsh GF, Deane AM. Angiotensin II for the Treatment of Vasodilatory Shock. *N Engl J Med* 2017; 377: 419-430.
27. Jain AK, Murty MN, Flynn PJ. Data clustering: a review. *ACM Comput Surv* 1999; 31: 264–323.
28. McClain JO, Rao VR. CLUSTISZ: A Program to Test for the Quality of Clustering of a Set of Objects. *J Mark Res* 1975; 12: 456-460.
29. Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, Fernández J, Prati D, Baselli G, Asselta R, Grimsrud MM, Milani C, Aziz F, Kässens J, May S, Wendorff M, Wienbrandt L, Uellendahl-Werth F, Zheng T, Yi X, de Pablo R, Chercoles AG, Palom A, Garcia-Fernandez A-E, Rodriguez-Frias F, Zanella A, Bandera A, Protti A, Aghemo A, Lleo A, Biondi A, Caballero-Garralda A, Gori A, Tanck A, Carreras Nolla A, Latiano A, Fracanzani AL, Peschuck A, Julià A, Pesenti A, Voza A, Jiménez D, Mateos B, Nafria Jimenez B, Quereda C, Paccapelo C, Gassner C, Angelini C, Cea C, Solier A, Pestaña D, Muñiz-Diaz E, Sandoval E, Paraboschi EM, Navas E, García Sánchez F, Ceriotti F, Martinelli-Boneschi F, Peyvandi F, Blasi F, Téllez L, Blanco-Grau A, Hemmrich-Stanisak G, Grasselli G, Costantino G, Cardamone G, Foti G, Aneli S, Kurihara H, ElAbd H, My I, Galván-Femenia I, Martín J, Erdmann J, Ferrusquía-Acosta J, Garcia-Etxebarria K, Izquierdo-Sanchez L, Bettini LR, Sumoy L, Terranova L, Moreira L, Santoro L, Scudeller L, Mesonero F, Roade L, Rühlemann MC, Schaefer M, Carrabba M, Riveiro-Barciela M, Figuera Basso ME, Valsecchi MG, Hernandez-Tejero M, Acosta-Herrera M, D'Angiò M, Baldini M, Cazzaniga M, Schulzky M, Cecconi M, Wittig M, Ciccarelli M, Rodríguez-Gandía M, Bocciolone M, Miozzo M, Montano N, Braun N, Sacchi N, Martínez N, Özer O, Palmieri O, Faverio P, Preatoni P, Bonfanti P, Omodei P, Tentorio P, Castro P, Rodrigues PM, Blandino Ortiz A, de Cid R, Ferrer R, Gualtierotti R, Nieto R, Goerg S, Badalamenti S, Marsal S, Matullo G, Pelusi S, Juzenas S, Aliberti S, Monzani V, Moreno V, Wesse T, Lenz TL, Pumarola T, Rimoldi V, Bosari S, Albrecht W, Peter W, Romero-Gómez M, D'Amato M, Duga S, Banales JM, Hov JR, Folseraas T, Valenti L, Franke A, Karlsen TH. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N Engl J Med* 2020.
30. Seymour CW, Kennedy JN, Wang S, Chang C-CH, Elliott CF, Xu Z, Berry S, Clermont G, Cooper G, Gomez H, Huang DT, Kellum JA, Mi Q, Opal SM, Talisa V, van der Poll T, Visweswaran S, Vodovotz Y, Weiss JC, Yealy DM, Yende S, Angus DC. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA* 2019; 321: 2003-2017.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

31. Knox DB, Lanspa MJ, Kuttler KG, Brewer SC, Brown SM. Phenotypic clusters within sepsis-associated multiple organ dysfunction syndrome. *Intensive Care Med* 2015; 41: 814-822.
32. Camporota L, Vasques F, Sanderson B, Barrett NA, Gattinoni L. Identification of pathophysiological patterns for triage and respiratory support in COVID-19. *Lancet Respir Med* 2020.
33. Helms J, Tacquard C, Severac F, Leonard-Lorant I, Ohana M, Delabranche X, Merdji H, Clere-Jehl R, Schenck M, Fagot Gandet F, Fafi-Kremer S, Castelain V, Schneider F, Grunebaum L, Anglés-Cano E, Sattler L, Mertes P-M, Meziani F, Group CT. High risk of thrombosis in patients with severe SARS-CoV-2 infection: a multicenter prospective cohort study. *Intensive Care Med* 2020; 46: 1089-1098.

Tables

Table 1. Clinical characteristics of the studied cohorts.

Variable	NYP-WCMC cohort				NYP-LMH validation cohort			
	All	Mild stratum	Intermediate stratum	Severe stratum	All	Mild stratum	Intermediate stratum	Severe stratum
# of patients (%)	**318**	**76 (23.90)**	**116 (36.48)**	**126 (39.62)**	**84**	**10 (11.90)**	**35 (41.67)**	**39 (46.43)**
Demographics								
Age, Mean (SD)	62.78 (14.34)	61.47 (16.51)	60.53 (14.14)	65.64 (12.52)	66.06 (13.06)	61.00 (17.10)	61.63 (11.46)	71.33 (11.07)
Sex female, n (%)	100 (31.45%)	23 (30.26%)	38 (32.76%)	39 (30.95%)	33 (39.29%)	4 (40.00%)	19 (54.29%)	10 (25.64%)
CAUCASIAN, n (%)	91 (28.62%)	20 (26.32%)	39 (33.62%)	32 (25.40%)	7 (8.33%)	0 (0.00%)	4 (11.43%)	3 (7.69%)
AFRICAN AMERICAN, n (%)	27 (8.49%)	3 (3.95%)	5 (4.31%)	19 (15.08%)	7 (8.33%)	0 (0.00%)	0 (0.00%)	7 (17.95%)
ASIAN/PACIFIC ISLANDER, n (%)	33 (10.38%)	11 (14.47%)	9 (7.76%)	13 (10.32%)	32 (38.10%)	5 (50.00%)	12 (34.29%)	15 (38.46%)
MULTI-RACIAL, n (%)	86 (27.04%)	21 (27.63%)	34 (29.31%)	31 (24.60%)	10 (11.90%)	2 (20.00%)	5 (14.29%)	3 (7.69%)
BMI, Mean (SD)	29.53 (8.40)	29.23 (9.06)	30.75 (9.17)	28.59 (7.01)	28.70 (7.70)	26.67 (3.94)	30.03 (9.94)	28.03 (5.61)
Comorbidities								
Coronary Artery Disease, n (%)	49 (15.41%)	7 (9.21%)	17 (14.66%)	25 (19.84%)	11 (13.10%)	1 (10.00%)	1 (2.86%)	9 (23.08%)
Cerebrovascular accident (Stroke), n (%)	20 (6.29%)	3 (3.95%)	7 (6.03%)	10 (7.94%)	4 (4.76%)	0 (0.00%)	0 (0.00%)	4 (10.26%)
Heart Failure, n (%)	21 (6.60%)	3 (3.95%)	9 (7.76%)	9 (7.14%)	3 (3.57%)	0 (0.00%)	1 (2.86%)	2 (5.13%)
Hypertension, n (%)	167 (52.52%)	35 (46.05%)	57 (49.14%)	75 (59.52%)	50 (59.52%)	5 (50.00%)	17 (48.57%)	28 (71.79%)
Diabetes Mellitus, n (%)	94 (29.56%)	17 (22.37%)	30 (25.86%)	47 (37.30%)	35 (41.67%)	4 (40.00%)	12 (34.29%)	19 (48.72%)
Pulmonary Disease, n (%)	63 (19.81%)	15 (19.74%)	22 (18.97%)	26 (20.63%)	15 (17.86%)	2 (20.00%)	4 (11.43%)	9 (23.08%)
Renal Disease, n (%)	26 (8.18%)	5 (6.58%)	5 (4.31%)	16 (12.70%)	7 (8.33%)	0 (0.00%)	2 (5.71%)	5 (12.82%)
Cirrhosis, n (%)	5 (1.57%)	3 (3.95%)	0 (0.00%)	2 (1.59%)	1 (1.19%)	0 (0.00%)	0 (0.00%)	1 (2.56%)
Hepatitis, n (%)	4 (1.26%)	1 (1.32%)	0 (0.00%)	3 (2.38%)	2 (2.38%)	0 (0.00%)	1 (2.86%)	1 (2.56%)
HIV, n (%)	4 (1.26%)	1 (1.32%)	2 (1.72%)	1 (0.79%)	1 (1.19%)	0 (0.00%)	1 (2.86%)	0 (0.00%)
Active Cancer, n (%)	21 (6.60%)	3 (3.95%)	2 (1.72%)	16 (12.70%)	2 (2.38%)	0 (0.00%)	0 (0.00%)	2 (5.13%)

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Transplant, n (%)	14 (4.40%)	5 (6.58%)	3 (2.59%)	6 (4.76%)	1 (1.19%)	0 (0.00%)	0 (0.00%)	1 (2.56%)
Inflammatory Bowel Disease, n (%)	7 (2.20%)	2 (2.63%)	2 (1.72%)	3 (2.38%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Rheumatologic Disease, n (%)	15 (4.72%)	4 (5.26%)	3 (2.59%)	8 (6.35%)	3 (3.57%)	0 (0.00%)	2 (5.71%)	1 (2.56%)
Other Immunosuppressed State, n (%)	12 (3.77%)	4 (5.26%)	1 (0.86%)	7 (5.56%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Baseline SOFA scores								
Cardiovascular, Mean (SD)	3.02 (1.35)	1.32 (1.34)	3.41 (0.88)	3.69 (0.70)	3.45 (1.03)	1.40 (1.02)	3.57 (0.80)	3.87 (0.40)
Central nervous system, Mean (SD)	3.72 (0.68)	3.34 (1.13)	3.72 (0.47)	3.94 (0.24)	3.39 (0.74)	2.60 (1.36)	3.37 (0.48)	3.62 (0.54)
Coagulation, Mean (SD)	0.15 (0.47)	0.12 (0.40)	0.04 (0.20)	0.28 (0.64)	0.13 (0.40)	0.00 (0.00)	0.11 (0.40)	0.18 (0.45)
Liver, Mean (SD)	0.24 (0.56)	0.20 (0.46)	0.14 (0.43)	0.37 (0.67)	0.20 (0.48)	0.10 (0.30)	0.14 (0.42)	0.28 (0.55)
Renal, Mean (SD)	0.94 (1.32)	0.16 (0.54)	0.35 (0.67)	1.96 (1.44)	1.36 (1.35)	0.50 (0.67)	0.37 (0.64)	2.46 (1.08)
Respiration, Mean (SD)	3.81 (0.58)	3.45 (0.89)	3.89 (0.45)	3.97 (0.25)	3.98 (0.22)	4.00 (0.00)	3.94 (0.33)	4.00 (0.00)
SOFA score, Mean (SD)	11.89 (2.56)	8.58 (1.84)	11.55 (0.58)	14.20 (1.46)	12.51 (2.25)	8.60 (2.11)	11.51 (0.50)	14.41 (1.08)
30-day Clinical Outcomes								
Extubation, n (%)	138 (43.40%)	40 (52.63%)	54 (46.55%)	44 (34.92%)	31 (36.90%)	2 (20.00%)	15 (42.86%)	14 (35.90%)
Mortality, n (%)	77 (24.21%)	14 (18.42%)	18 (15.52%)	45 (35.71%)	26 (30.95%)	4 (40.00%)	8 (22.86%)	14 (35.90%)
Tracheostomy, n (%)	41 (12.89%)	10 (13.16%)	18 (15.52%)	13 (10.32%)	6 (7.14%)	0 (0.00%)	3 (8.57%)	3 (7.69%)

Abbreviation: BMI=body mass index, HIV=Human Immunodeficiency Virus, NYP-WCMC=New York Presbyterian Hospital-Weill Cornell Medical Center, NYP-LMH=New York Presbyterian-Lower Manhattan Hospital, SD=standard deviation, SOFA=Sequential Organ Failure Assessment

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 2. Clinical characteristics of the trajectory subphenotypes in NYP-WCMC cohort

	Mild stratum (SOFA 0-10, n=76)			Intermediate stratum (SOFA 11-12, n=116)			Severe stratum (SOFA 13-24, n=126)		
Variable	**Worsening**	**Recovering**	**p-value†**	**Worsening**	**Recovering**	**p-value†**	**Worsening**	**Recovering**	**p-value†**
Total #	**37**	**39**	**-**	**41**	**75**	**-**	**54**	**72**	**-**
Demographics									
Age, Mean (SD)	61.08 (14.95)	61.85 (17.86)	0.240	63.80 (13.90)	58.73 (13.95)	0.059	65.72 (11.05)	65.58 (13.52)	0.951
Sex female, n (%)	9 (24.32%)	14 (35.90%)	0.323	13 (31.71%)	25 (33.33%)	1.000	17 (31.48%)	22 (30.56%)	1.000
CAUCASIAN, n (%)	9 (24.32%)	11 (28.21%)		14 (34.15%)	25 (33.33%)		16 (29.63%)	16 (22.22%)	
AFRICAN AMERICAN, n (%)	1 (2.70%)	2 (5.13%)	0.927	2 (4.88%)	3 (4.00%)	0.883	8 (14.81%)	11 (15.28%)	0.846
ASIAN/PACIFIC ISLANDER, n (%)	5 (13.51%)	6 (15.38%)		3 (7.32%)	6 (8.00%)		4 (7.41%)	9 (12.50%)	
MULTI-RACIAL, n (%)	12 (32.43%)	9 (23.08%)		10 (24.39%)	24 (32.00%)		13 (24.07%)	18 (25.00%)	
BMI, Mean (SD)	29.42 (10.01)	29.07 (8.21)	0.435	29.99 (7.25)	31.18 (10.09)	0.416	29.79 (7.01)	27.71 (6.89)	0.018
Comorbidities									
Coronary Artery Disease, n (%)	5 (13.51%)	2 (5.13%)	0.248	5 (13.51%)	2 (5.13%)	0.248	11 (20.37%)	14 (19.44%)	0.824
Cerebrovascular accident (Stroke), n (%)	0 (0.00%)	3 (7.69%)	0.241	0 (0.00%)	3 (7.69%)	0.241	2 (3.70%)	8 (11.11%)	0.189
Heart Failure, n (%)	2 (5.41%)	1 (2.56%)	0.604	2 (5.41%)	1 (2.56%)	0.604	4 (7.41%)	5 (6.94%)	1.000
Hypertension, n (%)	15 (40.54%)	20 (51.28%)	0.479	15 (40.54%)	20 (51.28%)	0.479	35 (64.81%)	40 (55.56%)	0.248
Diabetes Mellitus, n (%)	6 (16.22%)	11 (28.21%)	0.275	6 (16.22%)	11 (28.21%)	0.275	24 (44.44%)	23 (31.94%)	0.130
Pulmonary Disease, n (%)	7 (18.92%)	8 (20.51%)	1.000	7 (18.92%)	8 (20.51%)	1.000	14 (25.93%)	12 (16.67%)	0.184
Renal Disease, n (%)	0 (0.00%)	5 (12.82%)	0.055	0 (0.00%)	5 (12.82%)	0.055	8 (14.81%)	8 (11.11%)	0.589
Cirrhosis, n (%)	1 (2.70%)	2 (5.13%)	1.000	1 (2.70%)	2 (5.13%)	1.000	1 (1.85%)	1 (1.39%)	1.000
Hepatitis, n (%)	0 (0.00%)	1 (2.56%)	1.000	0 (0.00%)	1 (2.56%)	1.000	2 (3.70%)	1 (1.39%)	0.572
HIV, n (%)	0 (0.00%)	1 (2.56%)	1.000	0 (0.00%)	1 (2.56%)	1.000	0 (0.00%)	1 (1.39%)	1.000
Active Cancer, n (%)	2 (5.41%)	1 (2.56%)	0.604	2 (5.41%)	1 (2.56%)	0.604	10 (18.52%)	6 (8.33%)	0.102
Transplant, n (%)	1 (2.70%)	4 (10.26%)	0.359	1 (2.70%)	4 (10.26%)	0.359	5 (9.26%)	1 (1.39%)	0.081
Inflammatory Bowel Disease, n (%)	0 (0.00%)	2 (5.13%)	0.494	0 (0.00%)	2 (5.13%)	0.494	1 (1.85%)	2 (2.78%)	1.000
Rheumatologic Disease, n (%)	0 (0.00%)	4 (10.26%)	0.116	0 (0.00%)	4 (10.26%)	0.116	3 (5.56%)	5 (6.94%)	1.000
Other Immunosuppressed State, n (%)	2 (5.41%)	2 (5.13%)	1.000	2 (5.41%)	2 (5.13%)	1.000	5 (9.26%)	2 (2.78%)	0.129
Baseline SOFA scores									

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Cardiovascular, Mean (SD)	1.41 (1.26)	1.23 (1.40)	0.220	3.27 (0.86)	3.48 (0.88)	0.061	3.65 (0.72)	3.72 (0.67)	0.286
Central nervous system, Mean (SD)	3.41 (1.03)	3.28 (1.22)	0.358	3.71 (0.45)	3.73 (0.47)	0.342	3.94 (0.23)	3.93 (0.25)	0.379
Coagulation, Mean (SD)	0.05 (0.32)	0.18 (0.45)	0.033	0.05 (0.22)	0.04 (0.20)	0.415	0.31 (0.74)	0.25 (0.55)	0.499
Liver, Mean (SD)	0.27 (0.50)	0.13 (0.40)	0.059	0.17 (0.44)	0.12 (0.43)	0.148	0.37 (0.75)	0.36 (0.61)	0.369
Renal, Mean (SD)	0.24 (0.67)	0.08 (0.35)	0.103	0.46 (0.63)	0.29 (0.69)	0.023	1.94 (1.45)	1.97 (1.44)	0.466
Respiration, Mean (SD)	3.68 (0.70)	3.23 (1.00)	0.021	3.85 (0.52)	3.91 (0.41)	0.330	3.93 (0.38)	4.00 (0.00)	0.052
SOFA score, Mean (SD)	9.05 (1.45)	8.13 (2.04)	0.009	11.51 (0.50)	11.57 (0.61)	0.164	14.15 (1.57)	14.24 (1.37)	0.253

† p-value calculated by Chi-square test/Fisher's exact test, or student's t-test/Mann-Whitney test where appropriate.
** False discovery rate corrected p-value < 0.05
Abbreviation: BMI=body mass index, HIV=Human Immunodeficiency Virus, NYP-WCMC=New York Presbyterian Hospital-Weill Cornell Medical Center, SD=standard deviation, SOFA=Sequential Organ Failure Assessment

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figures

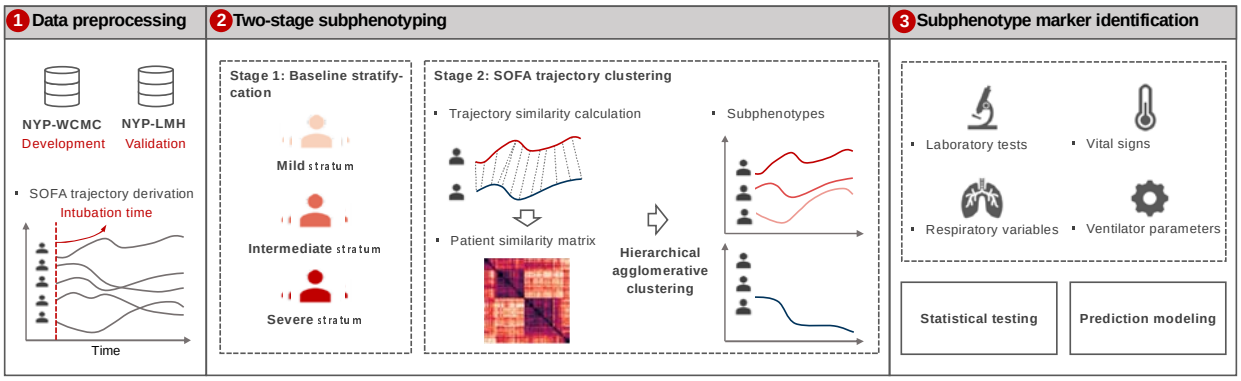

Figure 1. A schematic of the analysis plan. Intubated patients of two cohorts, New NYP-WCMC and NYP-LMH cohorts were analyzed, as development and validation cohorts, respectively. 7-day post-intubation SOFA trajectories were constructed. A two-stage subphenotyping model was then performed on the top of the SOFA trajectories. Statistical testing and prediction modeling were finally performed to identify markers at early-stage after intubation for separating the identified trajectory subphenotypes.

Abbreviations: NYP-WCMC=New York Presbyterian Hospital-Weill Cornell Medical Center, NYP-LMH=New York Presbyterian-Lower Manhattan Hospital, SOFA=Sequential Organ Failure Assessment.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

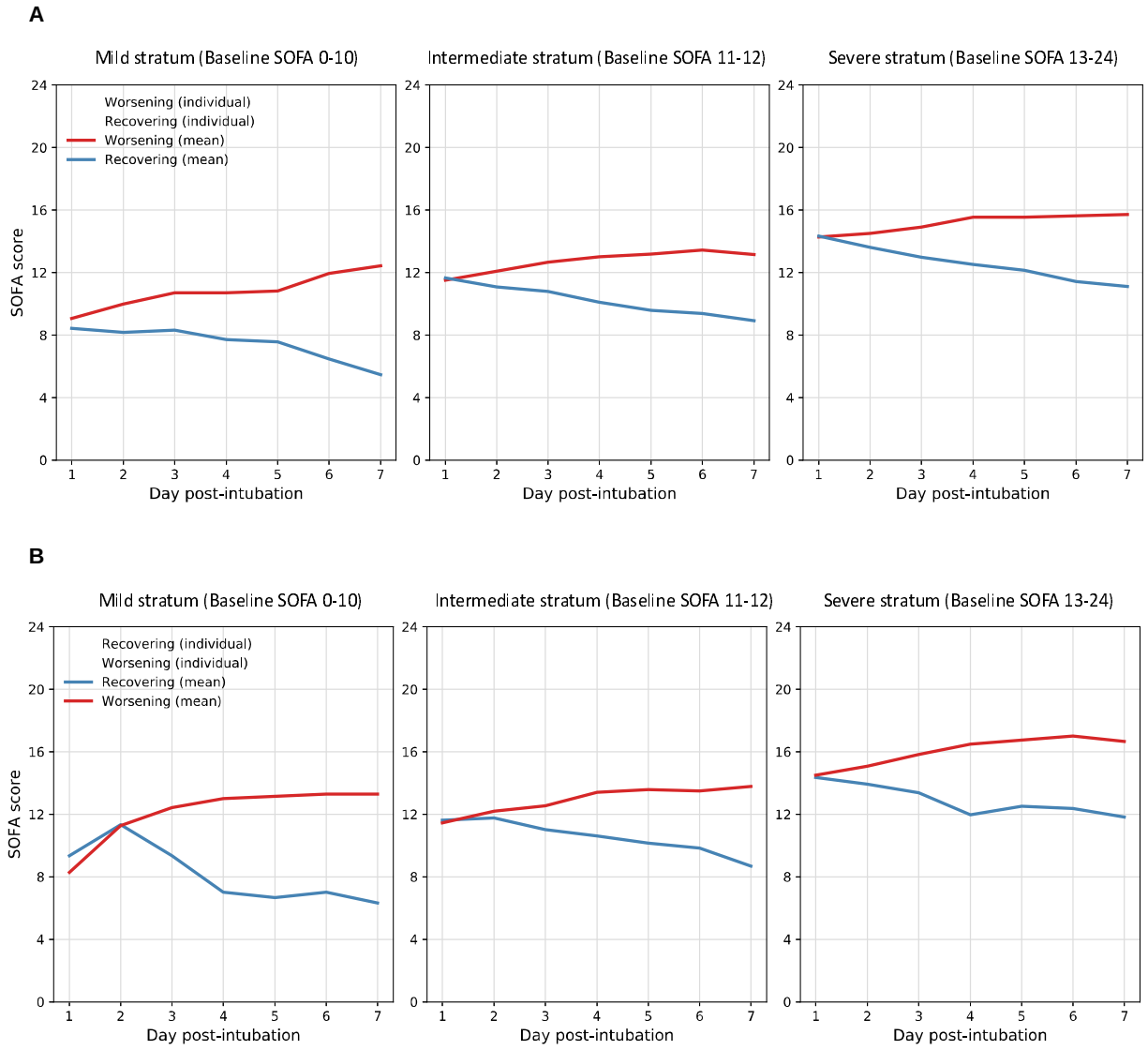

Figure 2. Averaged Sequential Organ Failure Assessment (SOFA) trajectories of the identified subphenotypes. Solid curves are mean SOFA trajectories of the subphenotypes, while dashed curves are individual SOFA trajectories of the patients. *(A)* SOFA trajectories of subphenotypes derived in NYP-WCMC cohort. *(B)* SOFA trajectories of subphenotypes derived in NYP-LMH validation cohort.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

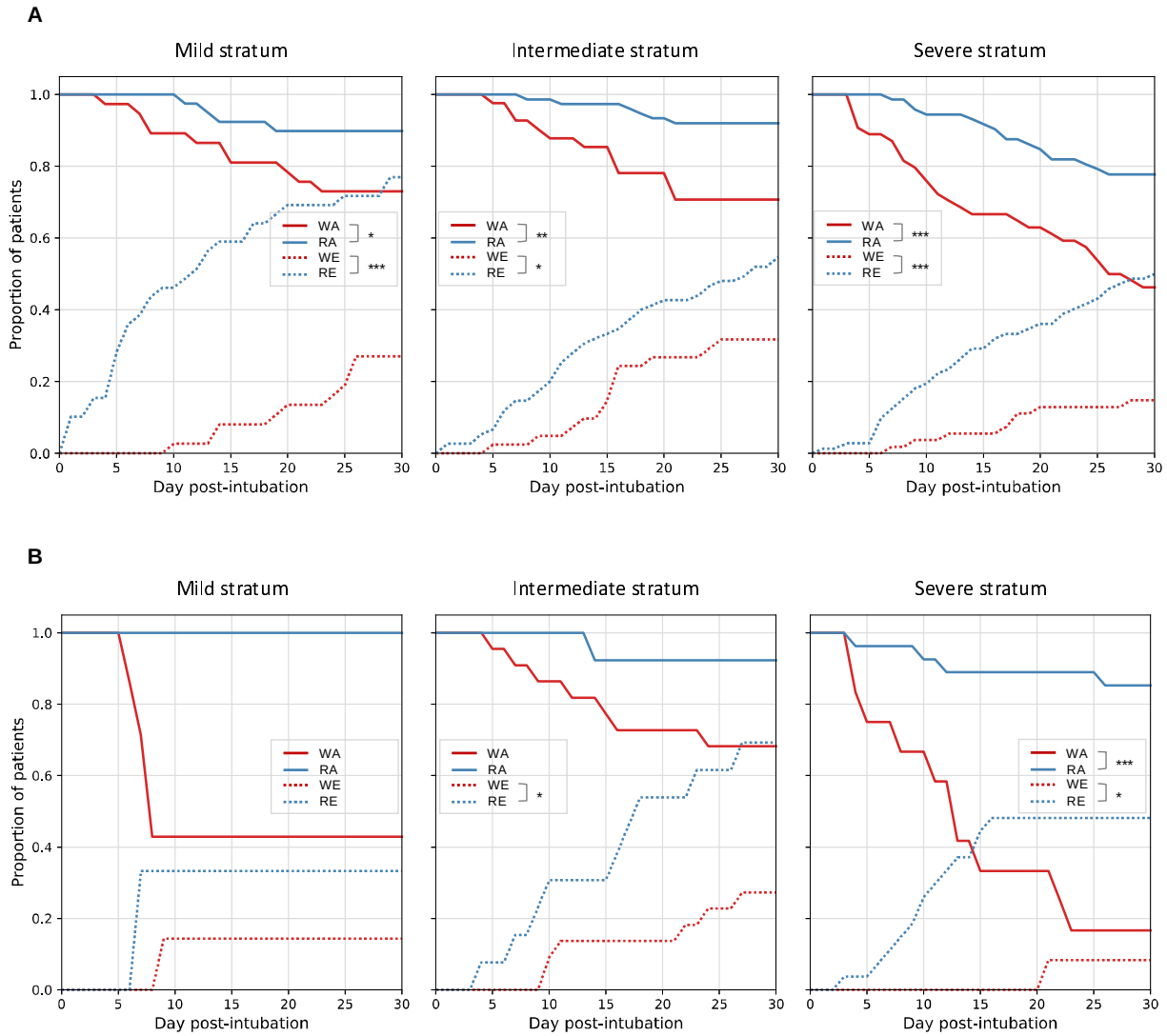

Figure 3. 30-day outcomes (extubation, mortality, and tracheostomy) of the trajectory subphenotypes. Chi-square/Fisher's exact tests were applied to compare 30-day outcomes between the worsening and recovering subphenotypes for each baseline strata. *(A)* 30-day outcomes of subphenotypes derived in NYP-WCMC cohort. *(B)* 30-day outcomes of subphenotypes derived in NYP-LMH validation cohort.

* denoting testing significance passed p-value < 0.05;

** denoting testing significance passed p-value < 0.01;

medRxiv preprint doi: https://doi.org/10.1101/2020.07.16.20155382; this version posted July 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

*** denoting testing significance passed p-value < 0.001.

Abbreviations: WA=worsening subphenotype alive; RA=recovering subphenotype alive; WE=worsening subphenotype extubated; RE=recovering subphenotype extubated.