Supplementary Material*

Garibaldi BT, Fiksel J, Muschelli J, et al. Patient trajectories among persons hospitalized for COVID-19: a cohort study. Ann Intern Med. 2020. doi:10.7326/M20-3905

* This supplementary material was provided by the authors to give readers further details on their article. The material was reviewed but not copyedited.

Supplement. Natural Language Processing

In order to identify symptoms at presentation, we first created a meta-lexicon of four symptom categories (organized into 11 sub-categories) based on the guidelines provided by the CDC, WHO, and clinical findings. For each symptom category, we generated a set of synonym terms using the Unified Medical Language System (UMLS) Metathesaurus,[1] and we iteratively worked with domain experts to revise the symptom categories and synonyms. Table S4 includes the list of symptom categories and the search terms.

We then selected relevant clinical note types for each patient, including H&P, Critical Care Notes, Progress Notes, and ED Notes, focusing specifically on the notes created within 48 hours before and after admission. Next, we pre-processed the note text and extracted only the relevant narrative parts, particularly the chief complaint and history of the present illness sections. A total of 7628 notes were used in the NLP analysis. 231 patients had a total of 458 notes prior to "admission status" (6% of total notes).

We then used a COVID-19-customized version of MedTagger,[2] together with our in-house Python tools to (a) identify phrases and synonyms of particular symptoms within the text narratives, (b) determine if these symptom mentions are negated, possible, or positive in their context, (c) classify symptoms into the predefined 11 categories, and (d) map them to their corresponding UMLS Concept Unique Identifiers (CUIs). These NLP pipelines use a combination of machine learning models, including Conditional random fields (CRFs),[3] and contextual rule-based methods, including regular expressions. Finally, we selected only the positive symptom mentions in the notes and aggregated all presenting symptoms for each patient.

To evaluate the performance of our NLP methods, two abstractors manually reviewed over 100 notes from 20 randomly selected patients. For each patient, each symptom was labeled as present or not-present (same label set as the NLP output), resulting in 220 manually labeled symptoms. To account for the issue of inaccuracies in electronic health records due to auto-populated sections containing negative findings, we considered a symptom as present if it was explicitly mentioned as present in any of the notes in the 48-hour admission window even if other notes or the same note contradicted that assertion. This method yielded an inter-rater agreement of 97%. The discrepancies between the two abstractors arose due to abstractor mistakes. As there are many notes to review per patient in the 48-hour admission window, one of the observers could have missed a single phrase in a note where a symptom was mentioned to be positive resulting in a discrepancy between observers. The 3% symptom-level disagreement arose entirely from these mistakes. To resolve these discrepancies, the observers met to evaluate the phrases used to support the presence or absence of the symptoms in question. As all of the discrepancies were related to observer mistakes, the adjudication process was simple and resulted in a gold standard symptom list. Comparing the created gold standard to the labels generated by the NLP methods, we achieved the following results:

Metric	Value (95% CI)
Sensitivity:	0.956 (0.913 – 0.998)

Specificity:	0.931 (0.887 – 0.974)
Negative Predictive Value:	0.968 (0.937 – 0.999)
False Positive Rate:	0.069 (0.026 – 0.113)
False Negative Rate:	0.044 (0.002 – 0.087)

References

1. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Research 2004;32:5.
2. Wen A, Fu S, Moon S, et al. Desiderata for delivering NLP to accelerate healthcare AI advancement and a Mayo Clinic NLP-as-a-service implementation. npj Digital Medicine 2019;2:130.
3. Lafferty J MA, Pereira F. Conditional random fields: probabilistic models for segmenting and labeling sequence data. Proceedings of the Eighteenth International Conference on Machine Learning 2001:8.

Supplement Table 1. Meta-lexicon of Symptom Categories and Synonym Terms for NLP Text Mining

Symptom	Lexicon
	Respiratory
Cough	acute cough(ing), cough(ing), cough(ing) effortful, cough(ing) non-productive, cough(ing) unproductive, cough(ing) without sputum, coughing with no sputum, dry cough(ing), non-productive cough(ing), unproductive cough(ing)
Shortness of Breath	breath shortness, breathing difficult(ies), breathing difficulty, breathing fast, breathing rate increased, breathing shortness, breathless(ness), difficulty breathing, distress, respiratory, distressed breathing, distressed respiration, distressed respiratory, dyspn(o)ea, elevated RR, fast breathing, gasp(ing), high RR, hyperpnea, increased respiratory rate, infantile respiratory distress, panting, polypnea, rapid breathing, rapid respiration, rate of respiration, increased, respiration difficult, respiration rate increased, respiration; rapid, respiratory difficulty, respiratory distress, respiratory rate high, respiratory rate increased, respiratory; distress, short(ness) of breath, sob, tachypn(o)ea
	Constitutional
Fever	elevated body temperature, elevated temp(erature), febrile, febris, fever(s), fevered, fevering, feverish, high body temperature, high temp(erature), hyperthermia, pyrexia(l)
Muscular/body aches	body ache(s), general muscular aches and pains, generalized (acute) body aches, generalized chronic body aches, muscle aches, muscle discomfort, muscle pains, muscular aches, muscular discomfort, muscular pains, myalgia(s), myalgic, myodynia
Sore throat	painful swallowing, difficulty swallowing, pharyngeal discomfort, pharyngeal pain, pharynx discomfort, pharynx pain, sore throat, throat discomfort, throat pain, throat soreness
Chills	chilliness, chills, rigor(s), shiver(s), shivering
Repeated shaking with chills	body shaking chills, chills shaking, recurrent chills, repeated shaking (with) chills, shaking chill episodes, shaking chills
Headache	aching head, cephalalgia, H/A, head aching, head pain, headache, pain in head
	Gastrointestinal
Diarrhea	diarh(o)ea, diarrh(o)ea, loose bowel movement, loose stool, watery stool
Nausea/Vomiting	feel(ing) sick, emesis, feeling bilious, feeling queasy, nausea(ted), nauseating, nauseous, threw up, throw(s) up, vomit(ing)
	Loss of taste and/or smell
Loss of taste or smell	ageusia, anosmia, abnormal taste/smell, altered taste/smell, change in taste/smell, decreased taste/smell, diminished taste/smell, impaired taste/smell, lack of taste/smell, lack of sense of taste/smell, loss of taste/smell, loss of the sense of taste/smell, no taste/smell, taste/smell impaired, taste/smell; loss

Supplement Table 2. Characteristics of Patients on Admission Stratified by Hospital of Admission

Characteristic	Site 1 N = 107	Site 2 N = 201	Site 3 N = 193	Site 4 N = 246	Site 5 N = 85
Age	59 (50, 71)	60 (48, 71)	54 (39, 67)	71 (59, 83)	70 (56, 84)
Male	59 (55%)	102 (51%)	91 (47%)	153 (62%)	38 (45%)
BMI*	31 (26, 36)	30 (26, 35)	31 (26, 36)	27 (24, 32)	28 (24, 34)
Admission from nursing facility*	12 (11%)	41 (20%)	8 (4.1%)	103 (42%)	8 (9.5%)
Charlson index					
0	41 (38%)	83 (41%)	72 (37%)	71 (29%)	33 (39%)
1-2	45 (42%)	93 (46%)	86 (45%)	119 (48%)	31 (36%)
3-4	13 (12%)	21 (10%)	22 (11%)	46 (19%)	18 (21%)
≥5	8 (7.5%)	4 (2.0%)	13 (6.7%)	10 (4.1%)	3 (3.5%)
Race and ethnicity*					
American Indian or Alaska Native	0 (0%)	1 (0.5%)	0 (0%)	1 (0.4%)	0 (0%)
Asian	3 (2.8%)	18 (9.0%)	9 (4.7%)	16 (6.5%)	2 (2.4%)
Black	35 (33%)	98 (49%)	85 (45%)	78 (32%)	40 (48%)
Hispanic	26 (24%)	27 (13%)	43 (23%)	35 (14%)	4 (4.8%)
Non-Hispanic white	38 (36%)	47 (23%)	45 (24%)	103 (42%)	31 (37%)
Other	5 (4.7%)	10 (5.0%)	8 (4.2%)	12 (4.9%)	7 (8.3%)
Severe illness or death upon admit	2 (1.9%)	15 (7.5%)	5 (2.6%)	22 (8.9%)	1 (1.2%)
Severe illness or death in first 12 hours	17 (16%)	25 (12%)	30 (16%)	44 (18%)	4 (4.7%)
Severe illness or death in first 24 hours	23 (21%)	33 (16%)	36 (19%)	49 (20%)	8 (9.4%)
Severe illness or death in first 48 hours	31 (29%)	43 (21%)	43 (22%)	58 (24%)	10 (12%)
DNR/DNI upon admit	0 (0%)	10 (5.0%)	0 (0%)	31 (13%)	11 (13%)
DNR/DNI in first hour	0 (0%)	13 (6.5%)	4 (2.1%)	38 (15%)	12 (14%)
DNR/DNI in first 2 hours	3 (2.8%)	16 (8.0%)	4 (2.1%)	48 (20%)	13 (15%)
DNR/DNI in first 6 hours	7 (6.5%)	21 (10%)	6 (3.1%)	51 (21%)	16 (19%)
DNR/DNI in first 12 hours	8 (7.5%)	21 (10%)	7 (3.6%)	54 (22%)	18 (21%)
DNR/DNI in first 24 hours	11 (10%)	23 (11%)	9 (4.7%)	61 (25%)	20 (24%)
Maximum illness severity					
Mild/Moderate	62 (58%)	132 (66%)	136 (70%)	136 (55%)	57 (67%)
Severe	36 (34%)	48 (24%)	40 (21%)	41 (17%)	13 (15%)
Death	9 (8.4%)	21 (10%)	17 (8.8%)	69 (28%)	15 (18%)

*Data unavailable for BMI in 16 patients, admission source in 1 patient, race and ethnicity in 5 patients

This table shows the key demographic characteristics grouped by hospital of admission.

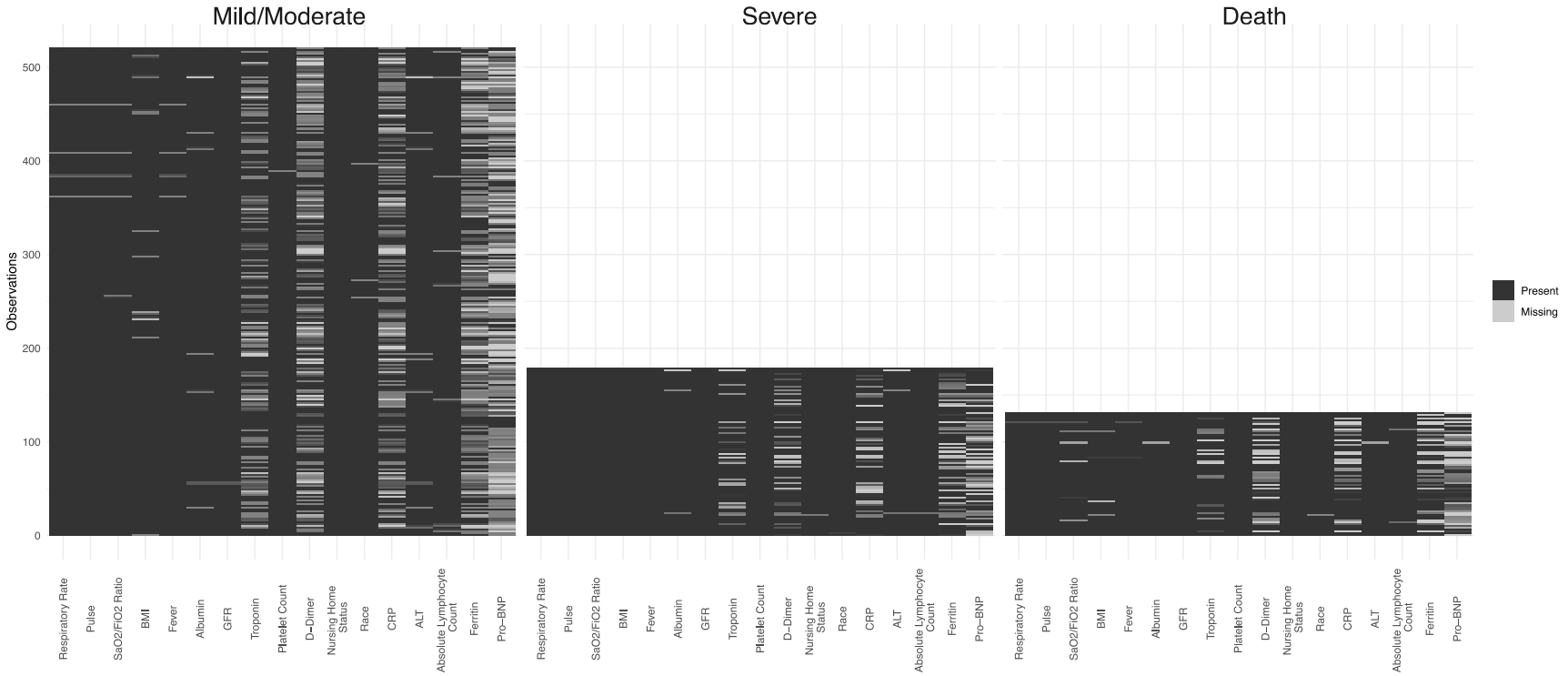
Supplement Table 3. Medications Administered in the Hospital to Patients With Observed Outcomes

Medication	Overall (N = 832)	Outcome observed (N = 825)		
		Mild/Moderate (N = 523)	Severe (N = 171)	Death (N = 131)
Antibiotics	639 (77%)	350 (67%)	159 (93%)	123 (94%)
ACE inhibitors	80 (9.6%)	46 (8.8%)	27 (16%)	6 (4.6%)
Antifungals	80 (9.6%)	9 (1.7%)	37 (22%)	27 (21%)
ARBs	74 (8.9%)	48 (9.2%)	21 (12%)	5 (3.8%)
Baricitinib	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Corticosteroids	132 (16%)	42 (8.0%)	53 (31%)	33 (25%)
DAS-181 or placebo[a]	6 (0.7%)	3 (0.6%)	2 (1.2%)	1 (0.8%)
Heparin or enoxaparin[b]	574 (69%)	364 (70%)	132 (77%)	74 (56%)
Hydroxychloroquine	384 (46%)	215 (41%)	98 (57%)	69 (53%)
Ibrutinib	1 (0.1%)	1 (0.2%)	0 (0%)	0 (0%)
Oseltamivir	2 (0.2%)	1 (0.2%)	1 (0.6%)	0 (0%)
Remdesivir[a]	8 (1.0%)	4 (0.8%)	4 (2.3%)	0 (0%)
Ritonavir	1 (0.1%)	0 (0%)	1 (0.6%)	0 (0%)
Rituximab	1 (0.1%)	0 (0%)	1 (0.6%)	0 (0%)
Sarilumab	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Statins	276 (33%)	156 (30%)	68 (40%)	47 (36%)
Tocilizumab	39 (4.7%)	3 (0.6%)	22 (13%)	11 (8.4%)

[a] Patients were enrolled in a randomized trial. [b]Heparin and enoxaparin does not distinguish between prophylactic or therapeutic dosing.

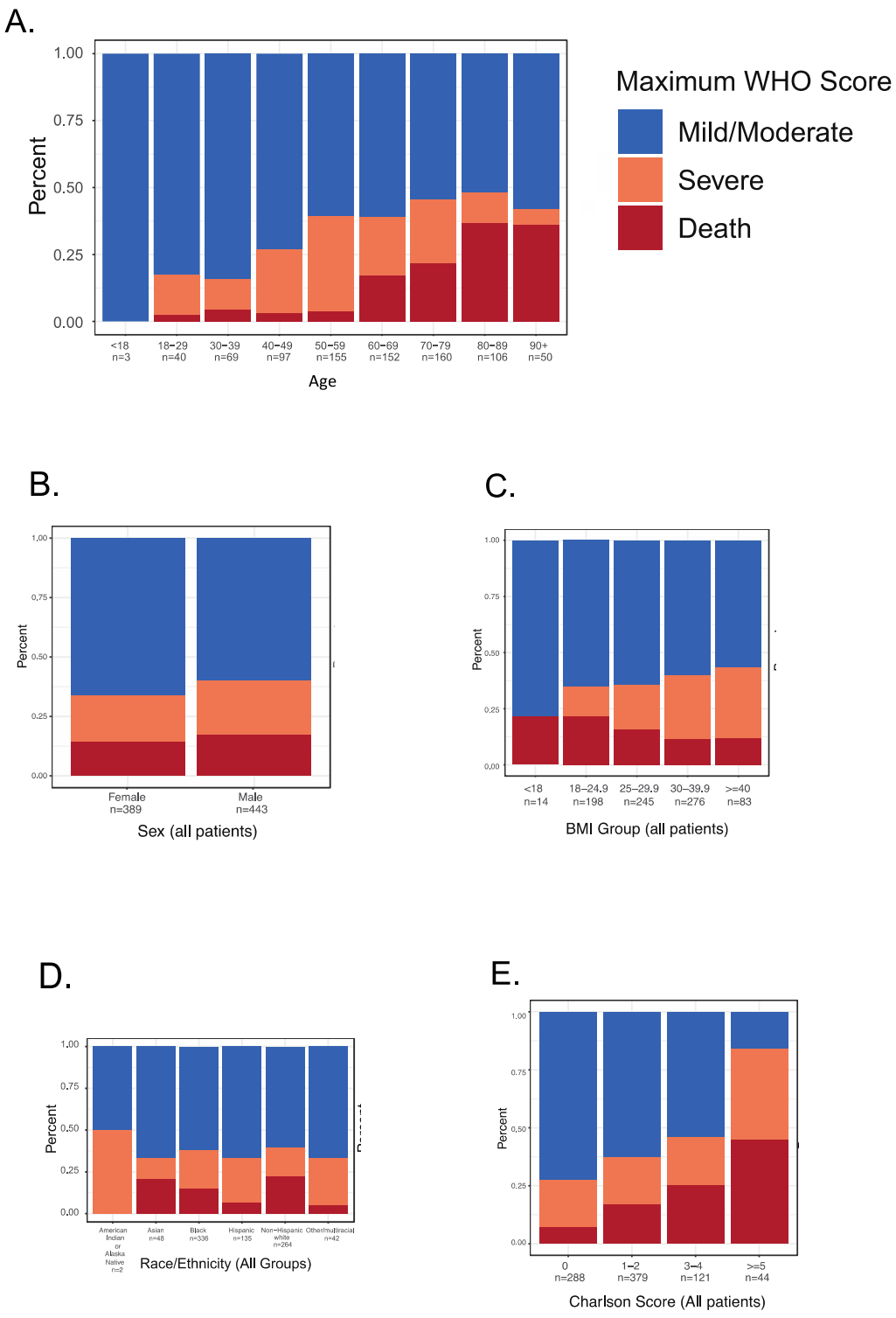
This table summarizes the medications received by patients in the cohort overall and grouped by maximum WHO score.

Supplement Figure 1



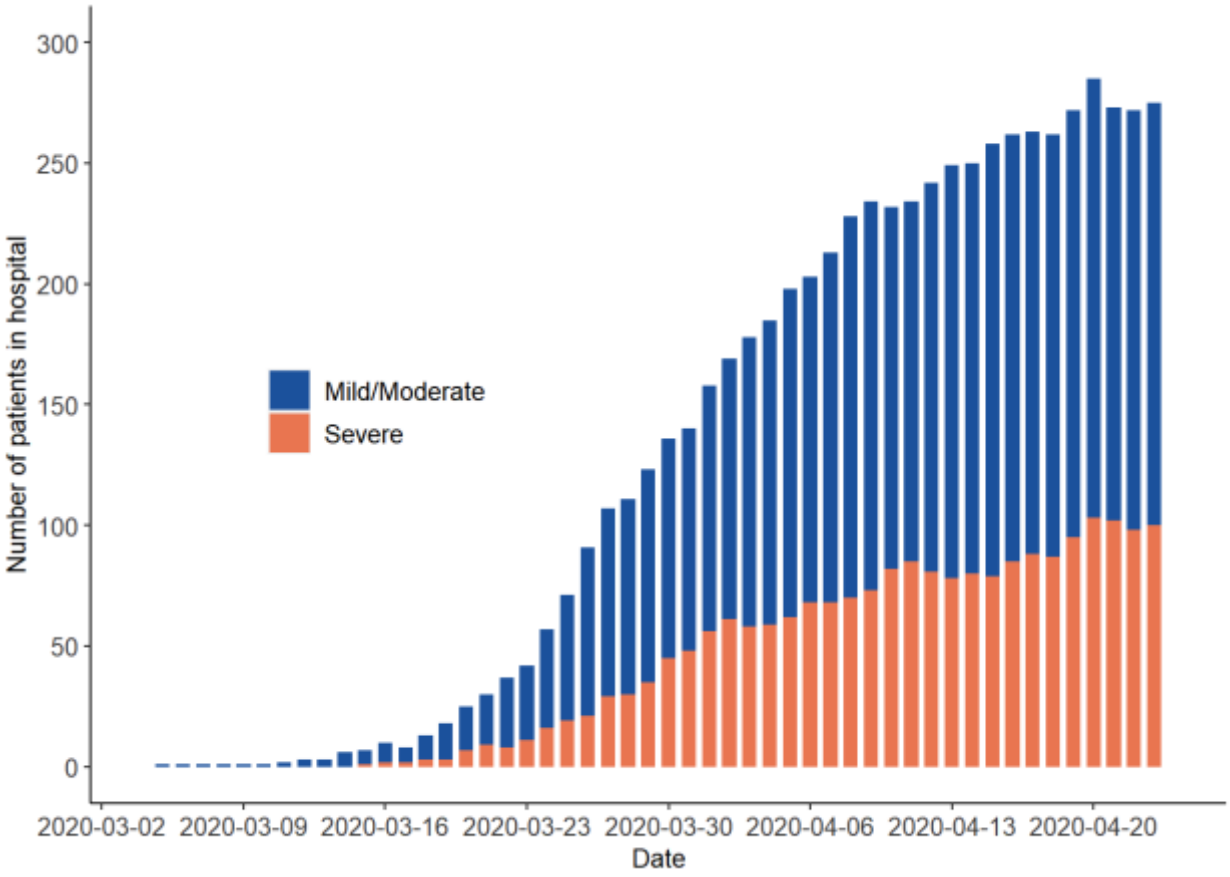
Legend for Supplement Figure 1. Plot of Missingness of Key Measurements. This figure shows the degree of missingness for key vital signs and laboratory values, grouped by WHO peak disease state. Each row represents a unique patient. A white line in a column indicates that the specified column value is missing for that patient. The plot shows that some lab values such as ferritin and pro-BNP had a high degree of missingness.

Supplement Figure 2



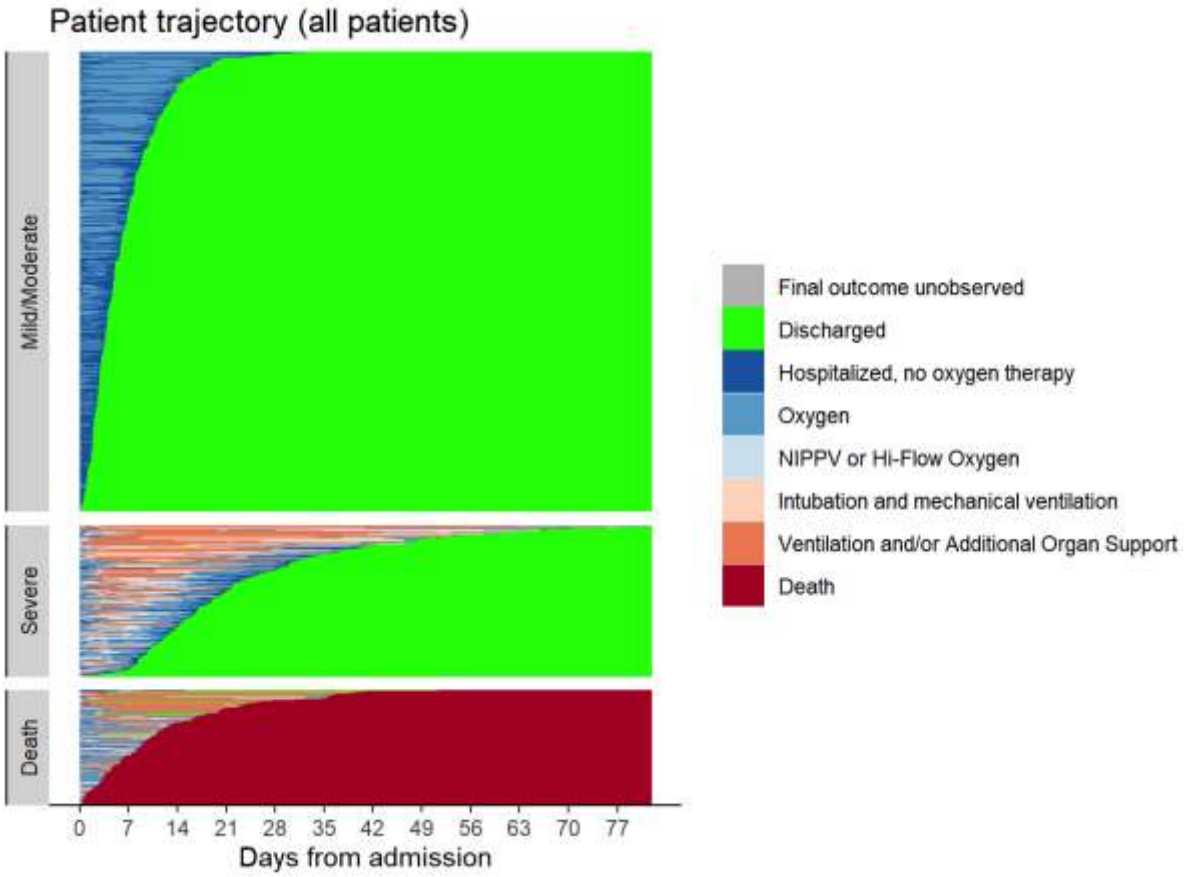
Legend for Supplement Figure 2. WHO Maximum Disease State Distribution by Specific Demographic and Clinical Characteristics. This figure shows the distribution of WHO disease severity states grouped according to specific characteristics. A. WHO maximum disease state grouped by age (n=832). B. WHO maximum disease state grouped by sex (n=832) C. WHO maximum disease state grouped by BMI (n=816) D. WHO maximum disease state grouped by race/ethnicity (n=824) E. WHO maximum disease state grouped by Charlson Comorbidity Index (n=832)

Supplement Figure 3



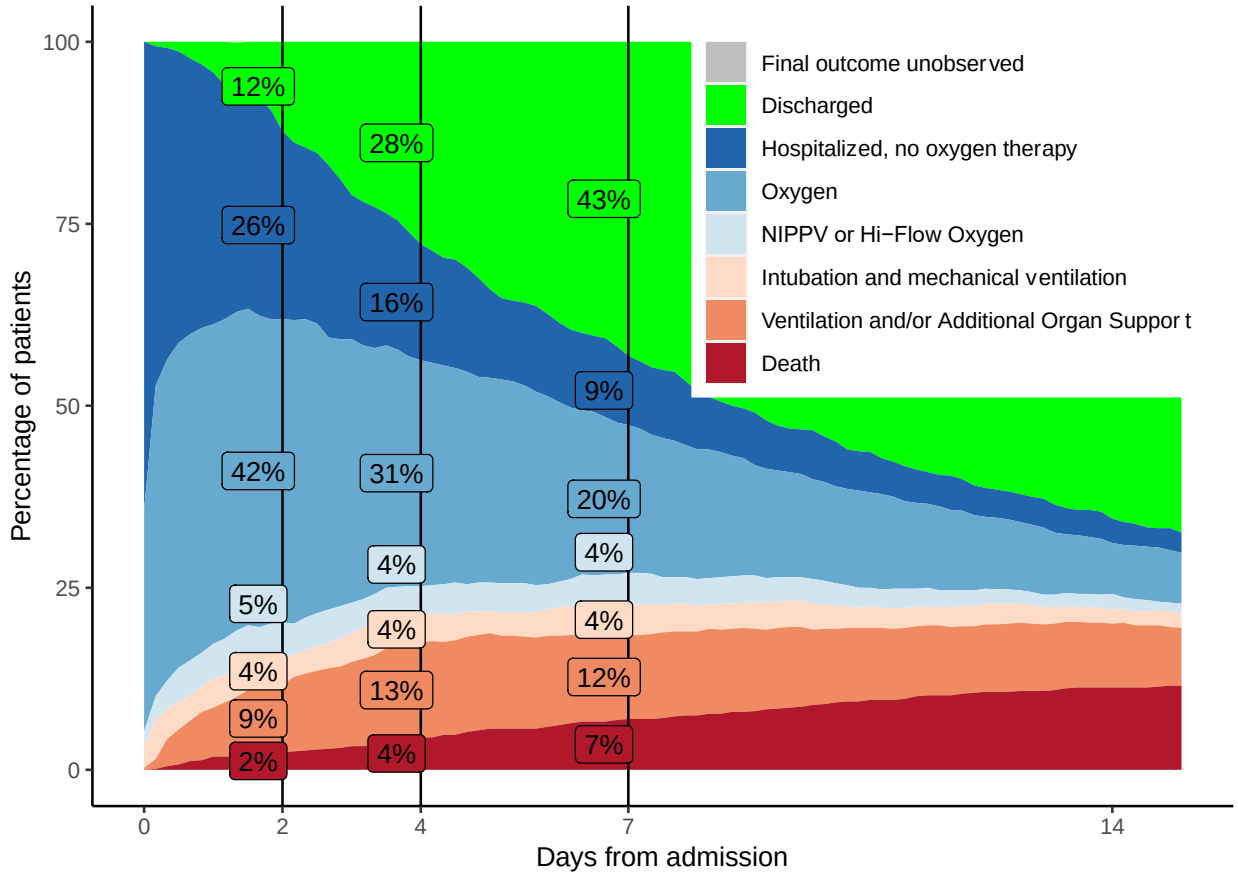
Legend for Supplement Figure 3. COVID-19 Admissions and Census for the Johns Hopkins Medical System. The Johns Hopkins Medical System in the Baltimore-Washington DC region is comprised of 5 hospitals with 2,513 licensed beds, including 354 ICU beds, and serves a population of 7 million. The first case was admitted on March 3, 2020 and the last cases for this cohort were admitted on April 24; data was censored on May 4, 2020. The figure shows the number of COVID-19 positive patients admitted across the health system each day during the study period, and the numbers with mild disease compared to severe disease. Discharges and deaths are not shown here (see Appendix Figure 3).

Supplement Figure 4



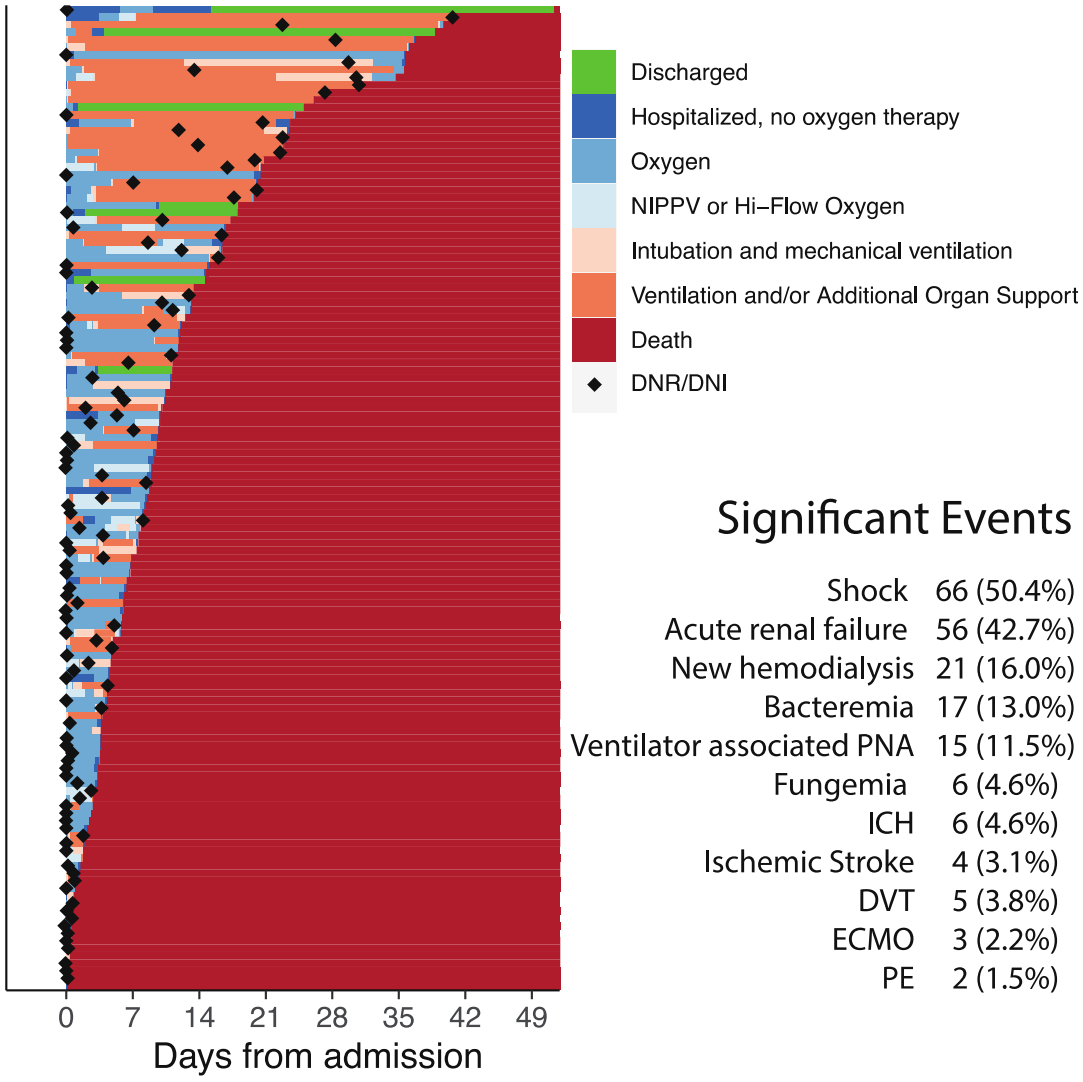
Legend for Supplement Figure 4. Patient Trajectory According to WHO Ordinal Severity Scale. Patient trajectories illustrating WHO COVID-19 disease state transitions plotted by days from admission and grouped by peak WHO disease state with individual colors showing specific transitions. Patients who have been discharged (represented in bright green), patients who remain hospitalized and have not yet achieved a final outcome (represented in gray), and patients who have died (represented in red) are illustrated. Each horizontal line is an individual patient. Among the 832 hospitalized patients, 694 (83.4)%) were discharged, 131 (15.8%) died, and 7 (0.84%) remained hospitalized at day of censoring (June 24, 2020). Median length of stay was 7.0 days overall (IQR 3.15, 13.19) and differed by peak illness state (mild/moderate - 4.9 days [IQR 2.6,8.6]; severe - 19.0 days [IQR 11.9, 30.0] and death - 8.49 days [IQR 3.62, 16.2]). Among 179 patients who were mechanically ventilated the median time on ventilation was 10.53 days (IQR 3.9,21.2). 102 (57.0%) were discharged (median ventilation time 11.63 days, IQR 6.68,22.90), 70 (39.1%) died (median ventilation time 7.1 days, IQR 1.26,16.6), and 7 (3.9%) remained hospitalized, of whom 3 (1.6%) remained intubated (median ventilation time 66.5 days, IQR 64.6,75.7) and 4 were extubated (median ventilation time 35.7 days, IQR 30.0,40.9).

Supplement Figure 5



Legend for Supplement Figure 5. Proportion of Hospitalized Patients in Each WHO COVID-19 Disease State and Discharged by Days From Admission. This figure illustrates the fraction of the 832 patients who are in each WHO disease state at a given time post admission to the hospital up to day 14. Note that both discharge (green) and death (red) are cumulative. The critical time points of day 2, day 4 and day 7 are illustrated with a vertical line at those time points. The percentage of patients in each of the WHO disease states is shown in the color corresponding to that state. For example, at day 2, 4 and 7, 16%, 25% and 27%, respectively of hospitalized patients had achieved severe status.

Supplement Figure 6



Legend for Supplement Figure 6. Patient Trajectories Among Those Who Died (*n* = 131). This figure illustrates the trajectory of each patient who died. Each horizontal line represents a single patient. The black diamonds represent the day that a code status including either a do not resuscitate (DNR) or do not intubate (DNI) was entered into the medical record. Twenty-five (22.1%) had a DNI/DNR order on admission and 55 (48.7%%) had a DNR/DNI order placed within 24 hours of admission. Fifty-eight patients (44%) died without receiving mechanical ventilation of which 55 (95%) were DNR/DNI. Causes of death among those not ventilated included hypoxic respiratory failure (n=41, 71%), septic shock (n=8, 14%), acute aspiration (n=3, 5%), stroke/ICH (n=3, 5%), progressive renal failure (n=2, 3%), and heart failure (n=1, 2%).