Anemia prior to or during COVID-19 is a risk factor for rehospitalization after SARS-CoV-2 clearance

Patrick Lenehan[1]*, Eshwan Ramudu[1]*, AJ Venkatakrishnan[1], Gabriela Berner[1], Reid McMurry[1], John C. O'Horo[2], Andrew D. Badley[2], William Morice II[2,3], John Halamka[2], Venky Soundararajan[1+]

[1] nference, One Main Street, Cambridge, MA 02142, USA

[2] Mayo Clinic, Rochester, MN 55902, USA

[3] Mayo Clinic Laboratories, Rochester, MN 55902, USA

* Joint first authors

+ Address correspondence to Venky Soundararajan (venky@nference.net; 857-207-2169)

Abstract

Background: As the number of new and recovering COVID-19 cases continues to rise, it has become evident that patients can experience symptoms and complications after viral clearance. Clinical biomarkers characterizing patients who are likely to experience these prolonged effects are unknown.

Methods: We conducted a retrospective study to compare longitudinal lab test measurements (hemoglobin, hematocrit, estimated glomerular filtration rate, serum creatinine, and blood urea nitrogen) in patients rehospitalized after PCR-confirmed SARS-CoV-2 clearance (n=49) versus patients not rehospitalized after viral clearance (n=173).

Results: Compared to patients who were not rehospitalized after PCR-confirmed viral clearance, those who were rehospitalized had lower median hemoglobin levels in the year prior to COVID-19 diagnosis (cohen's D = -0.74; p=0.01) and during the active infection window (cohen's D = -1.02; $p=2.4x10^{-7}$). Patients hospitalized after viral clearance were also more likely to be diagnosed with moderate or severe anemia during both intervals (pre-COVID: OR=5.91; p=0.03; active infection: OR=3.13; $p=1.37x10^{-8}$).

Conclusions: The diagnosis of moderate or severe anemia in the year prior to COVID-19 diagnosis and during active SARS-CoV-2 infection can aid in the identification of patients who are likely to be rehospitalized after viral clearance. Whether interventions to mitigate anemia in COVID-19 patients improve long term outcomes should be further investigated.

Trial Registration: This study was not affiliated with any clinical trials.

Funding: This study was funded by nference.

Brief Summary

Moderate or severe anemia in the year prior to or during COVID-19 infection is associated with rehospitalization after PCR-confirmed clearance of SARS-CoV-2.

Introduction

Since the first diagnosed case of COVID-19 in December 2019, over 80 million people have been infected with SARS-CoV-2 worldwide resulting in over 1.8 million deaths (1). While significant progress has been made in understanding the pathogenesis of COVID-19, including the rapid development and clinical rollout of multiple vaccine candidates (2, 3, 4, 5, 6, 7) along with detailed characterizations of the SARS-CoV-2 entry receptor ACE2 (8, 9, 10, 11, 12), there are still few options available for effective treatment of patients with severe COVID-19. Further, as the pandemic has progressed, there have been reports of long-lasting effects of COVID-19 even in patients who did not experience a severe disease course during their active period of infection (13, 14, 15). However, the clinical, molecular, and demographic biomarkers characterizing patients who are more likely to experience these lasting effects after clearing SARS-CoV-2 are not yet known.

The need to answer such questions during the rapidly evolving COVID-19 pandemic has emphasized the requirement for tools facilitating real-time analysis of patient data as it is obtained and stored in large electronic health records (EHR) systems. Specifically, clinical research efforts to understand the features defining COVID-19 patients, or subsets thereof, fundamentally require reliable systems that enable (1) conversion of unstructured information (e.g., patients notes written by healthcare professionals) into structured formats suitable for downstream analysis and (2) temporal alignment and integration of such unstructured data with the already structured information available in EHR databases (e.g., lab test results, disease diagnosis codes).

With these requirements in mind, we have previously reported the development of augmented curation methods that enable the rapid creation and comparison of defined cohorts of COVID-19 patients within a large EHR system (16, 17). Here we expand on our prior textual sentiment-based analysis (16) to understand the clinical features of patients likely to experience lasting effects of COVID-19, and we find that lab tests support our previously derived hypothesis

that anemia and kidney malfunction during active SARS-CoV-2 infection may serve as biomarkers of patients who are more likely to be rehospitalized after PCR-confirmed viral clearance.

Results

Longitudinal analysis of laboratory measurements provides a framework to test hypotheses derived from unstructured electronic health records

Using NLP-based extraction of phenotypes from a large EHR system, we previously reported that COVID-19 patients who are rehospitalized after viral clearance (as assessed by PCR) were more likely to experience anemia and acute kidney injury (AKI) in the year prior to their diagnosis and during their PCR-positive phase of COVID-19 compared to patients who were not rehospitalized after clearance of SARS-CoV-2 (16). Here, we sought to assess whether diagnostic lab tests for AKI and anemia corroborate these phenotypic associations.

As was previously described (16), we split the cohort of hospitalized COVID-19 patients with confirmed viral clearance into two groups: (1) post-clearance hospitalized ("PCH"; n=49) and (2) post-clearance non-hospitalized ("PCNH"; n=173), where viral clearance was defined by two consecutive negative SARS-CoV-2 PCR tests following a positive test (**Figure 1A-B**). A demographic summarization of these two cohorts is provided in **Table 1**. We then compared a set of selected lab test results between these cohorts during two time windows: (1) the year prior to COVID-19 diagnosis ("pre-COVID phase") and (2) the time during which each patient was SARS-CoV-2 positive according to their PCR results ("SARS-CoV-2^+ phase"). Given our previous EHR-based findings, we considered both anemia-related and kidney function lab tests including hemoglobin, hematocrit, estimated glomerular filtration rate (eGFR), serum creatinine, and serum blood urea nitrogen (BUN) levels (**Figure 1C**).

Patients rehospitalized after viral clearance are more likely to experience anemia and kidney dysfunction before and during their SARS-CoV-2 infection

For each patient, we first considered the median values of each lab test over the designated interval. Histograms showing the number of measurements per patient in each time

period for the selected tests are shown in **Figures S1-S2**. Consistent with our previous findings, we found that patients in the PCH cohort showed significantly lower median hemoglobin and hematocrit levels in both the pre-COVID phase (cohen's D = -0.74, p=0.01; cohen's D = -0.76, p=0.01) and the SARS-CoV-2$^+$ phase (cohen's D = -1.02, p=2.4x10^{-7}; cohen's D = -1.09, p=1.1x10^{-7}) (**Table 2**, **Figures 2A-D**). In the SARS-CoV-2$^+$ phase, median eGFR was lower (cohen's D = -0.52; p=0.01) in the PCH cohort, but renal function test results were similar between the cohorts in the pre-COVID phase (**Table 2**, **Figure S3**).

We also tested whether extreme (i.e. minimum or maximum) values of a given lab test over the designated periods varied between hospitalized and non-hospitalized patients, as a measure of central tendency (e.g., median) may fail to capture a single occurrence of phenotypes such as anemia or AKI. Specifically, we compared the patient-level minimum values of hemoglobin, hematocrit, and eGFR, and maximum values of serum creatinine and BUN in each time period. Interestingly, we found that the PCH cohort tended to have lower minimum values of hemoglobin, hematocrit, and eGFR in both the pre-COVID phase (cohen's D = -0.71, p=0.01; cohen's D = -0.71, p=0.01; cohen's D = -0.65, p=0.03) and the SARS-CoV-2$^+$ phase (cohen's D = -1.07, p=1.4x10^{-7}; cohen's D = -1.08, p=1.4x10^{-7}; cohen's D = -0.54, p=0.01), while also showing higher maximum serum BUN (cohen's D = 0.69; p=8.4x10^{-4}) during the SARS-CoV-2$^+$ phase (**Tables 3-4**, **Figure 3 and Figure S4**).

We next asked whether these differences in hemoglobin and hematocrit during the SARS-CoV-2$^+$ phase could be attributed to cohort imbalances in blood draw frequency or index infection severity. This was particularly important as patients admitted to the ICU during index infection had slightly lower median hemoglobin measurements in the SARS-CoV-2$^+$ phase (cohen's D = -0.26, p=0.03; **Figure S5**) and the rate of ICU admission during index infection was higher in the PCH cohort compared to the PCNH cohort (46% [26/57] vs. 20% [23/117], OR = 2.3, p=0.012; **Figure S6**). We thus evaluated the association between post-clearance rehospitalization and the

following independent variables via logistic regression during the two defined time intervals: hemoglobin level (median or minimum), number of blood draws, sex, and ICU admission status. While rehospitalization was inversely associated with both median and minimum hemoglobin levels in the pre-COVID phase (median: β=-0.404, p=0.005; minimum: β=-0.344, p=0.005) and the SARS-CoV-2^{+} phase (median: β=-0.473, p=3.0x10^{-6}; minimum: β=-0.394, p=1.0x10^{-5}), none of the other potential confounders were associated with rehospitalization in either time interval (**Table 5**).

Taken together, these analyses corroborate our prior textual sentiment-based EHR findings, suggesting that patients who are rehospitalized after SARS-CoV-2 clearance are more likely to experience anemia and kidney malfunction both prior to and during SARS-CoV-2 infection.

Post-clearance rehospitalized patients have lower hemoglobin and hematocrit before and during SARS-CoV-2 infection regardless of sex

Males and females have different normal ranges of hemoglobin and hematocrit. While our logistic regression analyses failed to identify a significant association between post-clearance rehospitalization and sex (**Table 5**), we sought to more thoroughly evaluate this potential confounding factor by performing sex-split subanalyses of anemia-related lab tests similar to those described previously. When split by sex, patient-level median values of hemoglobin and hematocrit during the SARS-CoV-2^{+} phase were still significantly lower in both the female (cohen's D = -1.52; p= 7.1x10^{-7}; cohen's D = -1.66, p=3.6x10^{-7}) and male (cohen's D = -0.71; p=0.006; cohen's D = -0.73, p=0.006) PCH cohorts versus their PCNH counterparts (**Tables 6-7**, **Figures 4A-D**). Further, pre-COVID median measurements of hemoglobin and hematocrit were lower in both the female (cohen's D = -0.89, p=0.05; cohen's D = -1.02, p=0.06) and male (cohen's D = -0.68, p=0.04; cohen's D = -0.64, p=0.05) PCH cohorts (**Tables 6-7**, **Figures 4E-H**). Similarly, in our analysis of extreme values, the minimum measurements of hemoglobin and hematocrit

during the SARS-CoV-2$^+$ phase were lower in both female (cohen's D = -1.72; p=3.0x10^{-7}; cohen's D = -1.76, p=3.0x10^{-7}) and male (cohen's D = -0.68; p=0.01; cohen's D = -0.67, p=0.02) PCH patients (**Tables 8-9**, **Figures 4I-L**). These trends were observed in the pre-COVID phase among females (cohen's D = -0.84; p=0.04; cohen's D = -0.97, p=0.02) and males (cohen's D = -0.64; p=0.04; cohen's D = -0.58, p=0.04) as well (**Tables 8-9**, **Figures 4M-P**).

Post-clearance rehospitalized patients are more likely to have experienced moderate or severe anemia before COVID-19 diagnosis and during active infection

Since anemia can be clinically diagnosed based on the lab tests we have considered, we next evaluated whether outright anemia occurs more frequently in the PCH cohort than the PCNH cohort. To do so, we classified each patient in a binary fashion for each time window based on whether the median of their pre-COVID phase measurements or the minimum of their SARS-CoV-2$^+$ phase measurements met the criteria for laboratory-diagnosed anemia (see **Methods** and **Figure 5A**). Anemia was indeed observed more frequently in the PCH cohort during both the pre-COVID phase (16/22 [73%] vs. 23/65 [35%]; OR=2.06; p=0.003) and the SARS-CoV-2$^+$ phase (44/46 [96%] vs. 123/167 [74%]; OR=1.30; p=9.12x10^{-4}) (**Figures 5B-C**).

Anemia can be further characterized as mild, moderate, or severe based on hemoglobin measurements, where a hemoglobin level <10 g/dL is consistent with moderate or severe disease for both males and females. We thus performed a similar follow-up analysis in which patients were classified based on whether the median of their pre-COVID hemoglobin measurements or the minimum of their SARS-CoV-2$^+$ measurements was <10g/dL. Moderate or severe anemia was observed more frequently in the PCH cohort during both the pre-COVID phase (4/22 [18%] vs. 2/65 [3%]; OR=5.91; p=0.034) and the SARS-CoV-2$^+$ phase (31/46 [67%] vs. 36/167 [22%]; OR=3.13; p=1.37x10^{-8}) (**Figures 5D-E**).

Again it was important to assess whether this strong enrichment for the diagnosis of moderate/severe anemia in PCH patients during the SARS-CoV-2$^+$ phase was confounded by

index infection severity. To this end, an altered logistic regression analysis in which the anemia term was encoded as a binary variable showed that the occurrence of moderate/severe anemia, but not ICU admission status, was significantly associated with post-clearance rehospitalization. This conclusion was further corroborated by a split cohort subanalysis showing that moderate/severe anemia was more common in PCH patients thans PCNH patients regardless of ICU admission during index infection (rates in ICU admitted patients: 77% [20/26] vs. 30% [16/53], OR=2.55, $p=1.11x10^{-4}$; rates in non-ICU admitted patients: 55% [11/20] vs. 18% [20/114], OR=3.32, $p=7.78x10^{-4}$) (**Figures S7A-B**).

Given these findings, we evaluated whether therapeutic interventions to treat moderate or severe anemia prior to or during COVID-19 infection provided any protective effect against post clearance hospitalization. We classified all patients with moderate or severe anemia on the basis of whether they received any anemia-targeted intervention during the respective time interval (see **Methods**). Patients who received one or more of these interventions were hospitalized after viral clearance at similar rates to patients who did not receive them in both the pre-COVID phase (3/5 [60%] vs. 1/1 [100%]; OR=0.75; p=1.0) and SARS-CoV-2^+ phase (24/50 [48%] vs. 7/17 [41%]; OR=1.072; p=0.78) (**Figures 5F-G**). Importantly, both of these analyses were limited due to the small total number of patients with moderate or severe anemia in these time periods (n=6 for pre-COVID, n = 67 for SARS-CoV-2^+).

Taken together, we conclude the following about COVID-19 patients who are rehospitalized after viral clearance compared to those who are not: (1) they tend to have lower hemoglobin and hematocrit levels in the year prior to COVID-19 diagnosis and during active COVID-19 infection, (2) they are more likely to be diagnosed with moderate or severe anemia during both intervals; and (3) it is unclear whether anemia-mitigating interventions in the hospital setting impact the risk of post clearance hospitalization.

Prior diagnosis of acute kidney injury is modestly associated with rehospitalization status in COVID-19 patients

Given that AKI is not defined by static lab measurements, but rather by changes in measurements (e.g. serum creatinine) over time, we next classified patients on the basis of whether they experienced laboratory-confirmed AKI during the pre-COVID and/or SARS-CoV-2$^+$ intervals. Specifically, we referred to the creatinine-related components of the KDIGO (Kidney Disease: Improving Global Outcomes) criteria for diagnosis and staging of AKI in adults (18): stage 1 AKI is characterized by an increase in serum creatinine by ≥0.3 mg/dL within 48 hours or an increase in serum creatinine to ≥1.5x the baseline value (known or assumed to have occurred within the last 7 days); stage 2 AKI is characterized by an increase in serum creatinine to ≥2x the baseline value; and stage 3 AKI is characterized by an increase in serum creatinine to ≥3x the baseline value or to ≥4 mg/dL (**Figure S8A**).

We then used our longitudinal lab testing data to identify all instances meeting these criteria for each patient, and we compared the prevalence of AKI in the PCH and PCNH cohorts during the two time intervals. Similar fractions of patients in the two cohorts experienced any stage AKI during the pre-COVID phase (2/16 [13%] vs 6/53 [11%]; OR=1.10; p=1.00) (**Figure S8B**), but the rate of AKI was slightly higher in the PCH cohort during the SARS-CoV-2$^+$ phase (15/30 [50%] vs. 23/92 [25%]; OR=2.00; p=0.01) (**Figure S8C**).

Similar to our previous analysis of anemia severity, we also asked whether the rate of moderate/severe AKI varied between these two cohorts. Here we found that stage 2+ AKI was slightly more common in the PCH cohort during both the pre-COVID phase (3/16 [19%] vs. 3/53 [6%]; OR=3.31; p=0.13) and SARS-CoV-2$^+$ phase (6/30 [20%] vs. 7/92 [8%]; OR=2.63; p=0.08), although these associations did not reach statistical significance (**Figure S8D-E**). Intriguingly

when split by sex, we found that male PCH patients were more likely to experience stage 2+ AKI in both the pre-COVID phase (3/10 [30%] vs. 1/31 [3%]; OR=9.30; p=0.04) and SARS-CoV-2$^+$ phase (5/17 [29%] vs. 1/53 [2%]; OR=15.59; p=0.003) (**Figure S8F-G**), but this analysis is limited by the total cohort sizes (n=41 for pre-COVID comparison; n=70 for SARS-CoV-2$^+$ phase comparison).

Taken together, these results suggest that AKI diagnosis during the SARS-CoV-2$^+$ phase may be associated with subsequent hospitalization, particularly in male patients. However, the validity of this association must be further tested in larger patient cohorts. Unlike the findings from anemia-related tests, it seems that the occurrence of AKI in the year prior to COVID-19 diagnosis is not associated with post viral clearance hospitalization.

Discussion

Approximately one year after the first confirmed case, the COVID-19 pandemic continues to ravage communities across the globe. While efforts early in the pandemic rightly focused on the acute lung inflammation caused by SARS-CoV-2, the subsequent realization that COVID-19 may have more lasting effects has mandated a better understanding of factors that predispose patients to experience long-term COVID-19 related complications. We have previously sought to address this knowledge gap using state-of-the-art NLP models deployed on a complete EHR system (16), and here we have expanded this effort to include the longitudinal analysis of laboratory measurements both prior to COVID-19 diagnosis and during active SARS-CoV-2 infection.

It is important to note that this study has several limitations. First, this analysis considers only patients within one EHR system; while this system does contain patient data from multiple sites of clinical care in distinct geographic locations (Minnesota, Arizona, Florida), there are still likely underlying biases in important factors such as patient demographics and tendencies around the ordering of laboratory tests by clinicians. Such biases could prevent the studied cohort and their associated data points from serving as true representative samples of all COVID-19 patients. Second, the analyzed cohort was relatively small (n = 222) as most patients diagnosed with COVID-19 do not subsequently receive two confirmatory negative PCR tests, and only a subset of these 222 patients possessed data for the lab tests of interest. Finally, the definition of the SARS-CoV-2^{+} window is imperfect as the true date of viral clearance for a given patient would likely precede their first negative PCR test by an unknown amount of time.

Consistent with our previous conclusions (16), this lab test analysis suggests that anemia and renal function in the pre-COVID and SARS-CoV-2^{+} phases are associated with the risk of post viral clearance hospitalization. While the pathophysiologic basis for these associations are not yet clear, the findings do merit consideration in the context of clinical care of COVID-19

patients. Indeed, pre-existing conditions are already integrated in the clinical decision-making algorithms around COVID-19 as the Center for Diseases Control (CDC) has designated various chronic conditions as risk factors for severe COVID-19 infection (e.g. cancer, chronic kidney disease, chronic obstructive pulmonary disease, and cardiovascular diseases such as heart disease, obesity, and diabetes). However, there is much less known regarding factors or conditions that place people at risk for subsequent complications such as rehospitalization after viral clearance. Once identified, such factors and conditions should similarly be incorporated into the clinical decision-making process when treating COVID-19 patients.

Our finding that lower hemoglobin and hematocrit levels, and the outright diagnosis of moderate or severe anemia, prior to or during active SARS-CoV-2 infection is associated with post viral clearance hospitalization has not been previously reported. And while sickle cell disease is considered a risk factor for severe COVID-19, anemia itself is not considered to be such a risk factor. While we did not find evidence that administration of potential anemia-mitigating interventions was associated with lower risks of hospitalization after viral clearance, this does not rule out a mechanistic role for anemia in long-term COVID-19 complications. Indeed, we were not able to account for patient compliance in our analyses of interventions, and the patient cohorts available for these analyses were limited in size. Further, even if mitigation of anemia does not impact subsequent hospitalization, these strong associations between anemia and COVID-19 are interesting in light of several previous lines of research.

First, several groups have reported an association between blood groups and susceptibility to or severity of COVID-19 infection (19, 20), suggesting that individuals with type O blood may be at lower risk for contracting COVID-19 or experiencing respiratory failure in the context of COVID-19. Whether this association reflects a direct or indirect interaction between SARS-CoV-2 and erythrocytes is not known, but it could certainly be relevant to pursue whether

blood type is also associated with the occurrence or severity of anemia in the setting of COVID-19.

Second, fatigue has been commonly reported as both an acute symptom and a lasting effect of COVID-19 (17, 21, 22), but the mechanisms underlying this phenotype have not been established. It is worth noting that 167 of the 222 (75%) COVID-19 patients in this study had at least a mild anemia during their SARS-CoV-2$^+$ phase, and 67 of the 222 (30%) patients had a moderate or severe anemia (defined as hemoglobin < 10 g/dL) during this interval. It would be worthwhile to perform a longitudinal follow-up on these patients to determine whether they continue to experience anemia in the months following SARS-CoV-2 clearance, and whether the presence of such a post-COVID anemia is associated with reports of fatigue.

Our findings regarding renal function tests and acute kidney injury may also be of clinical interest. Indeed, chronic kidney disease (CKD) has been recognized as a risk factor for severe COVID-19 infection. The fact that both median and extreme lab measurements suggest poorer renal function in the PCH cohort is consistent with this established risk factor, and suggests that the severity of one's CKD may have implications for their likelihood of hospitalization after viral clearance. Our finding that PCH males were more likely to experience moderate/severe AKI during the SARS-CoV-2$^+$ phase than their PCNH counterparts is interesting, but the small sample size available for analysis here necessitates further validation of this association.

Along with our previous analysis (16), this study illustrates the value of deploying sophisticated platforms across EHR systems that enable the integrated analysis of diverse data types including sentiment-laden text and laboratory test measurements. Taken together, these studies provide the first example of leveraging augmented curation methods to first identify phenotypes that distinguish defined clinical cohorts and to then cross check these phenotypic associations through a hypothesis-driven analysis of the most relevant lab tests. This framework

can be effectively scaled for other clinical research efforts not only in COVID-19 but also in any other disease areas of interest.

Methods

Study design

This was a case-control study. The primary outcome was hospitalization status within 90 days of PCR-confirmed SARS-CoV-2 clearance. The exposure variables were anemia and kidney dysfunction as assessed through selected lab measurements detailed below.

Selection of study participants

Cases and controls were selected from a cohort of 22,223 patients who presented to the Mayo Clinic Health System (including tertiary medical centers in Minnesota, Arizona, and Florida) and received at least one positive SARS-CoV-2 PCR test between the start of the COVID-19 pandemic and October 27, 2020 (see **Figure 1B**). Post clearance hospitalized ("PCH") cases (n=49) were defined as patients who were hospitalized for COVID-19, had two documented negative SARS-CoV-2 PCR tests following their last positive test result, and were subsequently admitted to the hospital within 90 days of clearance. Post clearance non-hospitalized ("PCNH") controls (n=173) were defined as those who were hospitalized for COVID-19, had two documented negative SARS-CoV-2 PCR tests following their last positive test result, and were not hospitalized within 90 days of clearance. The definitions of the cases and controls were also described previously (16).

Analysis of potential confounding variables

As shown in **Table 1**, there were no statistically significant differences between these groups in age, relative cleared date (defined as time to second negative SARS-CoV-2 PCR test after first positive test), race, ethnicity, or sex. However, we did note that a majority of PCNH cases were male as compared to PCH counterparts (59% vs. 49%). This potential confounding factor was addressed by performing (1) sex-split subgroup analyses (**Tables 6-9**) and (2) multivariate logistic regression (**Table 5**; see *Statistics* below).

Although hospitalization during index infection was required for inclusion in both the PCH and PCNH cohorts, this criterion does not necessarily ensure comparable severities of index infection. To better assess potential differences in index infection severity, we compared the rates of ICU admission and found this to be significantly higher in the PCH cohort compared to the PCNH cohort (26/49 [53%] vs. 57/173 [33%], OR=2.3, p=0.01; **Figure S6**). Further, patients admitted to the ICU during index infection had slightly lower hemoglobin measurements during the SARS-CoV-2$^+$ phase than patients not admitted to the ICU (cohen's D = -0.26, p=0.03; **Figure S5**). Thus, we considered ICU admission as a potential confounding factor in our analyses. We addressed this by performing (1) subgroup analyses to determine whether differences between the PCH and PCNH cohorts were observed both in patients who were and were not admitted to the ICU (**Figure S7**) and (2) multivariate logistic regression (**Table 5**; see *Statistics* below).

We observed that patients in the PCH cohort were more likely to experience anemia in both the pre-COVID and SARS-CoV-2$^+$ phases than patients in the PCNH cohort. Because hospitalized patients can experience anemia due to repeated blood draws for laboratory testing, we also considered the number of blood draws per patient as a potential confounding variable. To address this, we performed multivariate logistic regression (**Table 5**; see *Statistics* below).

Definition of the considered time intervals: pre-COVID phase and SARS-CoV-2$^+$ phase

Laboratory results were assessed (1) during the year prior to COVID-19 diagnosis, referred to throughout this manuscript as the "pre-COVID phase" and (2) during the period in which a patient was positive for SARS-CoV-2 by PCR, referred to throughout this manuscript as the "SARS-CoV-2$^+$ phase." A diagnosis of COVID-19 was conferred by a positive SARS-CoV-2 PCR test, and clearance was defined as two consecutive negative SARS-CoV-2 PCR tests occurring after a positive test.

Selection and summarization of laboratory measurements

The primary exposure variables were anemia and AKI. The selected laboratory measurements related to anemia included hemoglobin and hematocrit, and laboratory measurements related to AKI included serum creatinine, serum blood urea nitrogen (BUN), and estimated glomerular filtration rate (eGFR). For a given lab test, we considered the median, maximum, and minimum measurements for each patient during the specified time windows (i.e. the pre-COVID and SARS-CoV-2^{+} phases). Given the directionality of these tests (i.e. anemia is defined by low hemoglobin and hematocrit, while kidney dysfunction is characterized by increases in serum creatinine and BUN but a decrease in eGFR), we were primarily interested in comparing the patient-level minimum values of hemoglobin, hematocrit, and eGFR, and maximum values of serum creatinine and BUN in each time period.

Classification of patients using clinical diagnostic criteria for anemia and AKI

We classified patients in a binary fashion for each time window based on whether their lab tests were consistent with the clinical diagnosis of anemia or acute kidney injury. Classifications were defined according to the Mayo Clinic reference ranges for anemia and the KDIGO (Kidney Disease: Improving Global Outcomes) criteria for AKI (18) as follows:

- *Anemia (mild, moderate, or severe)*: for males, median hemoglobin < 13.5 g/dL or median hematocrit < 38.3%. For females, median hemoglobin < 12.0 g/dL or median hematocrit < 35.5%.
- *Anemia (moderate or severe)*: for both males and females, median hemoglobin < 10.0 g/dL.
- *Acute kidney injury (stage 1, 2, or 3)*: increase in serum creatinine by ≥0.3 mg/dL within 48 hours or an increase in serum creatinine to ≥1.5x the baseline value which is known or

assumed to have occurred in the prior 7 days. The baseline was defined as the minimum value among all serum creatinine tests for the given patient in the prior 7 days.

- *Acute kidney injury (stage 2 or 3)*: increase in serum creatinine to ≥2x the baseline value which is known or assumed to have occurred in the prior 7 days, or a serum creatinine value of ≥4 mg/dL. The baseline was defined as the minimum value among all serum creatinine tests for the given patient in the prior 7 days.

Classification of patients based on administration of potential anemia mitigating interventions

To investigate the effects of anemia-targeted interventions in the pre-COVID and SARS-CoV-2$^+$ phases, we identified patients from this cohort who received (1) blood transfusions per medical procedure documentation and/or (2) one or more of the following therapies per medication administration records: oral or intravenous iron, multivitamin or prenatal vitamins containing iron, oral or intramuscular vitamin B12 (cyanocobalamin), and intravenous or subcutaneous darbepoetin or erythropoietin.

Quantification of number of blood draws per patient

To test whether trends in anemia-related measurements could be explained by differences in the number of blood draws received in the pre-COVID phase or SARS-CoV-2$^+$ phase, we counted the number of blood draws in these time intervals for each patient. All tests with a documented source of “Blood”, “Plasma”, or “Serum” were collected for each patient; for a given patient on a given day, we then took the count of the most frequently obtained test as the number of blood draws for that patient on that day. For example, if the record for Patient *P* on Day *D* contained 5 serum sodium measurements, 3 hemoglobin measurements, and 1 plasma IL-6 measurement, then we inferred that Patient *P* received 5 blood draws on Day *D*.

Statistics

Laboratory values were assessed within each time interval as patient-wise medians, minima, or maxima. To perform statistical comparisons between the PCH and PCNH cohorts, one-sided Mann-Whitney U-tests and Cohen's D were applied to continuous outcome measures, generating a *p*-value and an effect size measurement. The distribution of patient-wise median, minimum, and maximum values obtained for each laboratory measurement among this cohort were assessed with a Kolmogorov-Smirnov (KS) Test of Normality (**Figures S9-S11**). While hemoglobin and hematocrit measurements showed a normal distribution (KS Test p-value > 0.05), the other lab tests did not ($p < 0.05$ in each case); as such, the non-parametric Mann-Whitney U test was chosen for statistical comparisons. A one-sided test was used because these analyses were performed as follow-up to our previous EHR-based analysis which found a higher prevalence of anemia and kidney injury in the PCH cohort (16), providing a pre-supposed direction of change for each tested laboratory measurement. For each set of comparisons performed, p-values were corrected using a Benjamini-Hochberg (BH) correction for multiple hypothesis testing. Differences were considered statistically significant and biologically relevant if the BH-corrected p-value was ≤ 0.05 and the cohen's D magnitude was ≥ 0.5. Fisher exact tests were applied to categorical outcome measures, generating a *p*-value and an Odds Ratio. All of the tests described above were applied using the SciPy package (23) in Python (version 3.5).

To address potential confounding variables that may have contributed to our anemia-related observations, we performed multivariate logistic regressions for both the pre-COVID and SARS-CoV-2^{+} phases (**Table 5**). For each regression, the binary dependent variable was defined as post viral clearance hospitalization status (i.e. assignment to the PCH cohort vs. PCNH cohort), and the independent variables included one anemia metric along with sex (binary), number of

blood draws (continuous) in the given time interval, and ICU admission status during that time interval (binary). Stated explicitly, the logistic regression equation was as follows: $log(P_{PCH}/(1 - P_{PCH})) = \beta_0 + \beta_1*(Anemia\ Term) + \beta_2*(Sex) + \beta_3*(Blood\ Draw\ Count) + \beta_4*(ICU\ Admission\ Status)$. Metrics used for the "Anemia Term" included the median or minimum hemoglobin measurement from the given time interval (continuous) or the diagnosis status of moderate/severe anemia as defined above (binary). For each time interval, three regressions were performed (i.e. one regression per anemia metric), yielding a coefficient (log odds ratios) and p-value (calculated using the log likelihood ratio test) for each independent variable. We performed separate regressions for each anemia term rather than including all three terms in one model because these values are strongly correlated with each other, and multicollinearity of independent variables can negatively impact the estimation of logistic regression coefficients. Of note, data regarding ICU admission status was not available prior to April 2020, so this feature was omitted from the pre-COVID regression analyses. Binary variables were assigned as follows: moderate/severe anemia: 0 = no anemia, 1 = anemia; sex: 0 = female, 1 = male; ICU admission during interval: 0 = not admitted to ICU, 1 = admitted to ICU. All regressions were performed using the Statsmodels package in python (24).

Study Approval

This retrospective study was reviewed and approved by the Mayo Clinic Institutional Review Board (IRB 20-003278) as a minimal risk study. Subjects were excluded if they did not have a research authorization on file.

.

Author Contributions

PL, ER, AV, RM, ADB, and VS contributed to the study design and methodology. ER, GB, and JCO were responsible for data curation. ER performed the formal statistical analyses, which were reviewed by PL, AV, RM, and VS. PL drafted the manuscript with inputs from ER. AV, RM, JCO, ADB, WH, JH, and VS provided critical revisions, which were incorporated into the final manuscript by PL. All authors approved the submitted version of the manuscript. PL and ER were considered co-first authors as they performed the majority of analysis and writing. PL was responsible for creating detailed specifications for all analyses and sub-analyses performed (which were executed by ER) and for writing the manuscript, and so is listed first among the co-first authors.

References

1. COVID-19 Map - Johns Hopkins Coronavirus Resource Center. Accessed December 31, 2020. https://coronavirus.jhu.edu/map.html

2. Folegatti PM et al. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *Lancet*. 2020;396(10249):467-478.

3. Jackson LA et al. An mRNA Vaccine against SARS-CoV-2 - Preliminary Report. *N Engl J Med*. 2020;383(20). doi:10.1056/NEJMoa2022483

4. Corbett KS et al. Evaluation of the mRNA-1273 Vaccine against SARS-CoV-2 in Nonhuman Primates. *N Engl J Med*. 2020;383(16):1544-1555.

5. Mercado NB et al. Single-shot Ad26 vaccine protects against SARS-CoV-2 in rhesus macaques. *Nature*. 2020;586(7830):583-588.

6. Bos R et al. Ad26 vector-based COVID-19 vaccine encoding a prefusion-stabilized SARS-CoV-2 Spike immunogen induces potent humoral and cellular immune responses. *NPJ Vaccines*. 2020;5:91.

7. Mulligan MJ et al. Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. *Nature*. 2020;586(7830):589-593.

8. Venkatakrishnan AJ et al. Knowledge synthesis of 100 million biomedical documents augments the deep expression profiling of coronavirus receptors. Published online May 28, 2020. doi:10.7554/eLife.58040

9. Anand P et al. SARS-CoV-2 strategically mimics proteolytic activation of human ENaC. Published online May 26, 2020. doi:10.7554/eLife.58603

10. Ziegler CGK et al. SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. *Cell*. 2020;181(5). doi:10.1016/j.cell.2020.04.035

11. Zhao Y et al. Single-Cell RNA Expression Profiling of ACE2, the Receptor of SARS-CoV-2. *Am J Respir Crit Care Med*. 2020;202(5):756.

12. Singh M et al. A Single-Cell RNA Expression Map of Human Coronavirus Entry Factors. *Cell Rep*. 2020;32(12):108175.

13. Yelin D et al. Long-term consequences of COVID-19: research needs. *Lancet Infect Dis*. 2020;20(10):1115.

14. del Rio C et al. Long-term Health Consequences of COVID-19. *JAMA*. 2020;324(17):1723-1724.

15. Carfì A et al. Persistent Symptoms in Patients After Acute COVID-19. *JAMA*. 2020;324(6):603-605.

16. Pawlowski C et al. Pre-existing conditions are associated with COVID patients' hospitalization, despite confirmed clearance of SARS-CoV-2 virus. *medRxiv*. Published online November 3, 2020:2020.10.28.20221655.

17. Wagner T et al. Augmented curation of clinical notes from a massive EHR system reveals symptoms of impending COVID-19 diagnosis. Published online July 7, 2020. doi:10.7554/eLife.58227

18. Khwaja A. KDIGO Clinical Practice Guidelines for Acute Kidney Injury. *NEC*. 2012;120(4):c179-c184.

19. Latz CA et al. Blood type and outcomes in patients with COVID-19. *Ann Hematol*. 2020;99(9):2113-2118.

20. Ellinghaus D et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N Engl J Med*. 2020;383(16). doi:10.1056/NEJMoa2020283

21. Pascarella G et al. COVID-19 diagnosis and management: a comprehensive review. *J Intern Med*. 2020;288(2). doi:10.1111/joim.13091

22. Townsend L et al. Persistent fatigue following SARS-CoV-2 infection is common and independent of severity of initial infection. *PLoS One*. 2020;15(11):e0240784.

23. Virtanen P et al. Author Correction: SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17(3):352.

24. Seabold S, Perktold J. Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of the 9th Python in Science Conference*. Published online 2010. doi:10.25080/majora-92bf1922-011

Acknowledgements

We thank Murali Aravamudan for his thoughtful review and feedback on this manuscript. We also thank Andrew Danielsen, Jason Ross, Jeff Anderson, and Sankar Ardhanari for their support that enabled the rapid completion of this study. The authors acknowledge funding from nference for this study.

Competing Interests

PL, ER, AV, GB, RM, and VS are employees of nference and have financial interests in the company. JO, WM, and JH have financial conflicts of interest in technology used in the research and with Mayo Clinic may stand to gain financially from the successful outcome of the research. ADB is a consultant for Abbvie, is on scientific advisory boards for nference and Zentalis, is founder and President of Splissen therapeutics, and has financial conflicts of interest in technology used in the research and with Mayo Clinic may stand to gain financially from the successful outcome of the research.

Data Availability

After publication, the data will be made available to others upon reasonable requests to the corresponding author. A proposal with a detailed description of study objectives and the statistical analysis plan will be needed for evaluation of the reasonability of requests. Deidentified data will be provided after approval from the corresponding author and Mayo Clinic.

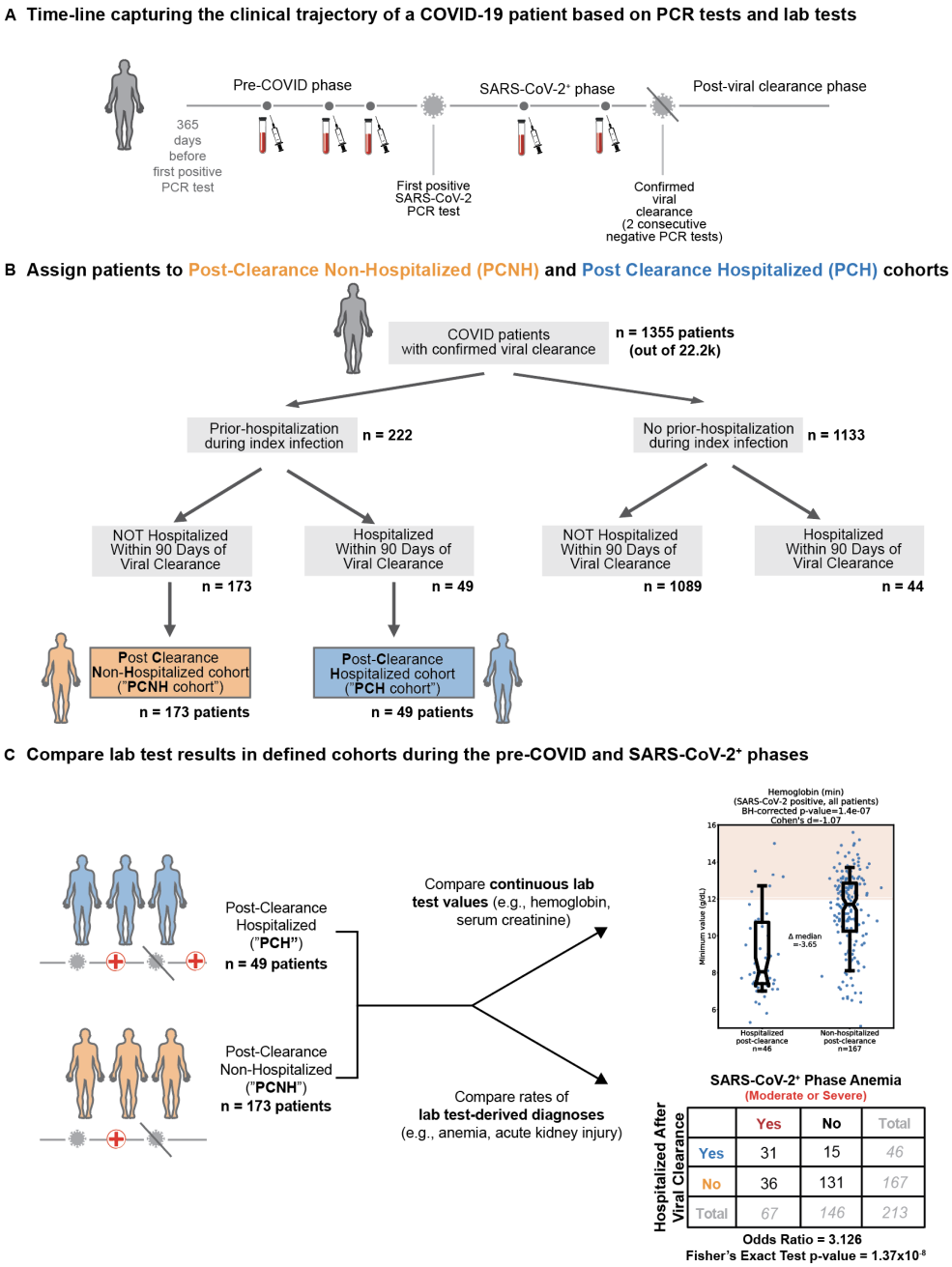


Figure 1. Schematic summarizing cohort creation and lab test analyses. (A) Time intervals ("phases") were defined relative to SARS-CoV-2 PCR testing results. (B) Of the 1355 patients who were diagnosed by PCR with COVID-19 and subsequently confirmed to have cleared SARS-CoV-2 with two consecutive negative tests, we created two cohorts: (1) patients who were hospitalized during their index infection and not hospitalized after confirmed viral clearance (post-clearance non-hospitalized, or "PCNH"; n=173), and (2) patients who were hospitalized during their index infection and rehospitalized within 90 days of confirmed viral clearance (post clearance hospitalized, or "PCH"; n=49). (C) A defined set of anemia and kidney related lab test measurements were compared between the PCH and PCNH cohorts in the pre-COVID and SARS-CoV-2^{+} intervals.

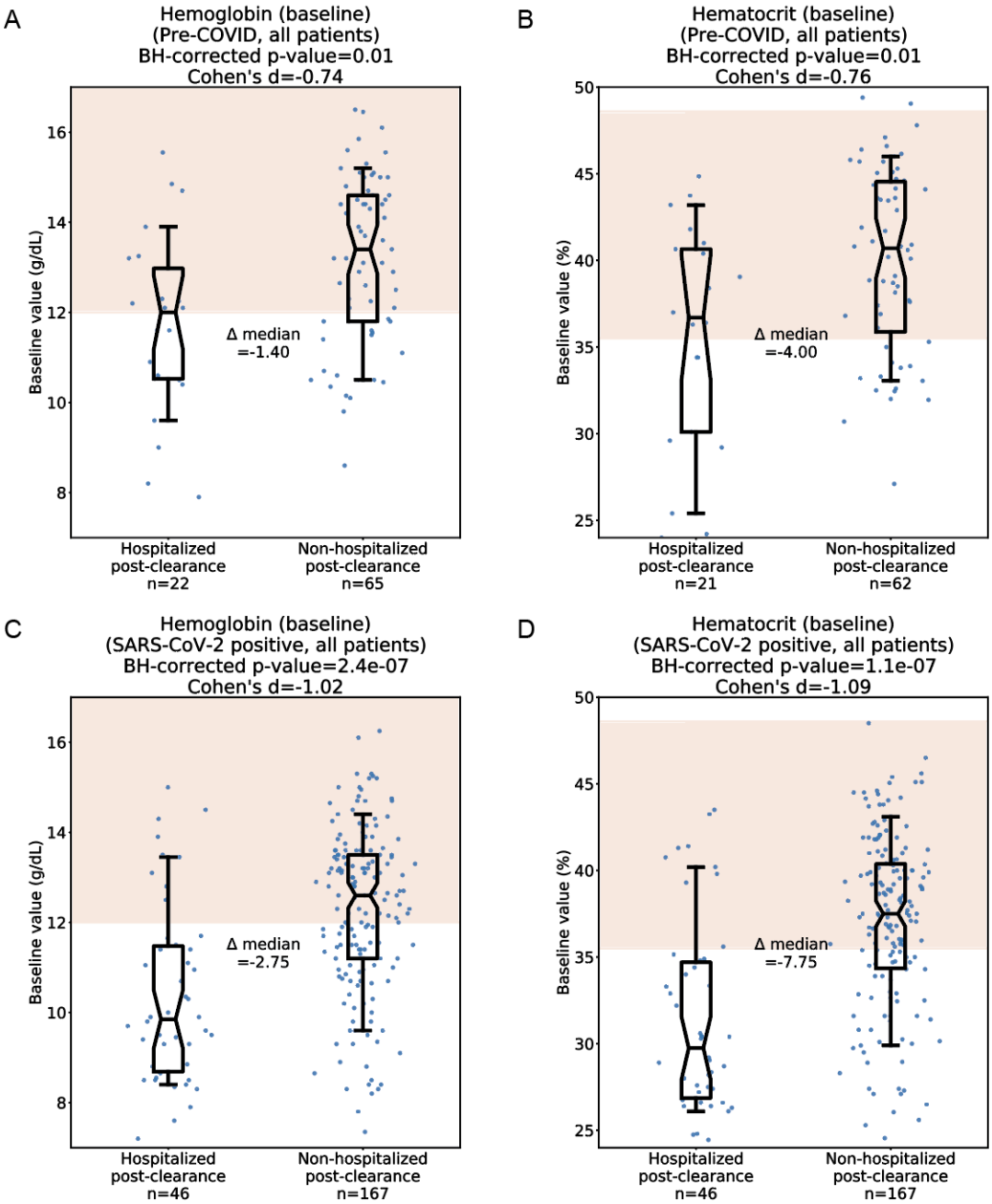


Figure 2. Comparison of median hemoglobin and hematocrit values during the pre-COVID and SARS-CoV-2^{+} phases. (A) Pre-COVID median hemoglobin in the PCH (n=22) and PCNH (n=65) cohorts. (B) Pre-COVID median hematocrit in the PCH (n=21) and PCNH (n=62) cohorts. (C) SARS-CoV-2^{+} median hemoglobin in the PCH (n=46) and PCNH (n=167) cohorts. (D) SARS-CoV-2^{+} median hematocrit in the PCH (n=46) and PCNH (n=167) cohorts. Red shading indicates normal ranges for hemoglobin and hematocrit, spanning from the lower limit of normal for females (12 g/dL hemoglobin, 35.5% hematocrit) to the upper limit of normal for males (17.5 g/dL hemoglobin, 48.6% hematocrit)**.** For each comparison, statistics shown include the number of patients analyzed, Cohen's D, BH-corrected Mann Whitney U test p-value, and the difference of medians between the two cohorts. Box and whisker plots depict median and IQR along with the 10th and 90th percentiles.

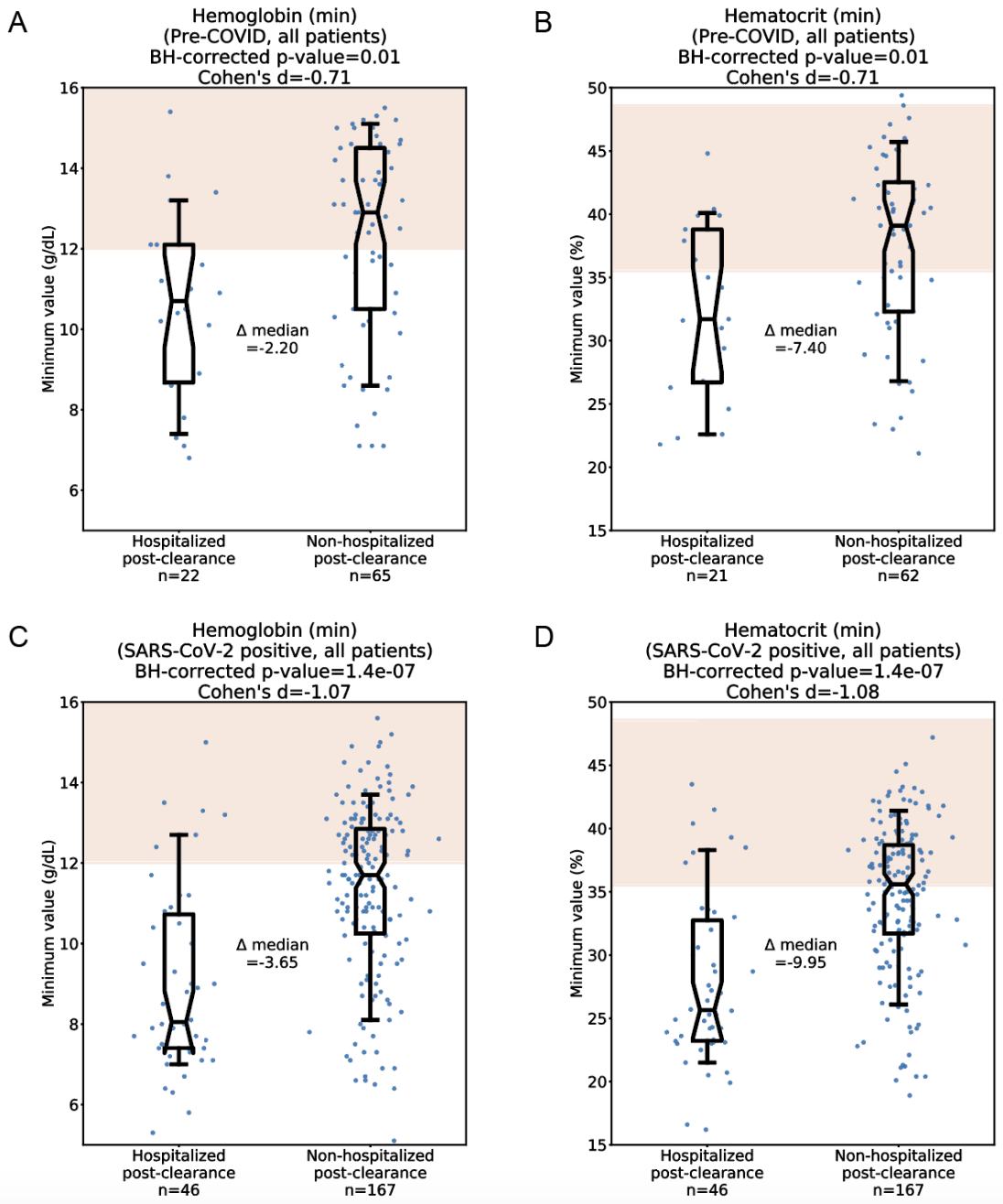


Figure 3. Comparison of minimum values for hemoglobin and hematocrit in the pre-COVID and SARS-CoV-2^{+} intervals. (A) Pre-COVID minimum hemoglobin in the PCH (n=22) and PCNH (n=65) cohorts. (B) Pre-COVID minimum hematocrit in the PCH (n=21) and PCNH (n=62) cohorts. (C) SARS-CoV-2^{+} minimum hemoglobin in the PCH (n=46) and PCNH (n=167) cohorts. (D) SARS-CoV-2^{+} minimum hematocrit in the PCH (n=46) and PCNH (n=167) cohorts. Red shading indicates normal ranges for hemoglobin and hematocrit as described in Figure 1. For each comparison, statistics shown include the number of patients analyzed, Cohen's D, BH-corrected Mann Whitney U test p-value, and the difference of medians between the two cohorts. Box and whisker plots depict median and IQR along with the 10th and 90th percentiles.

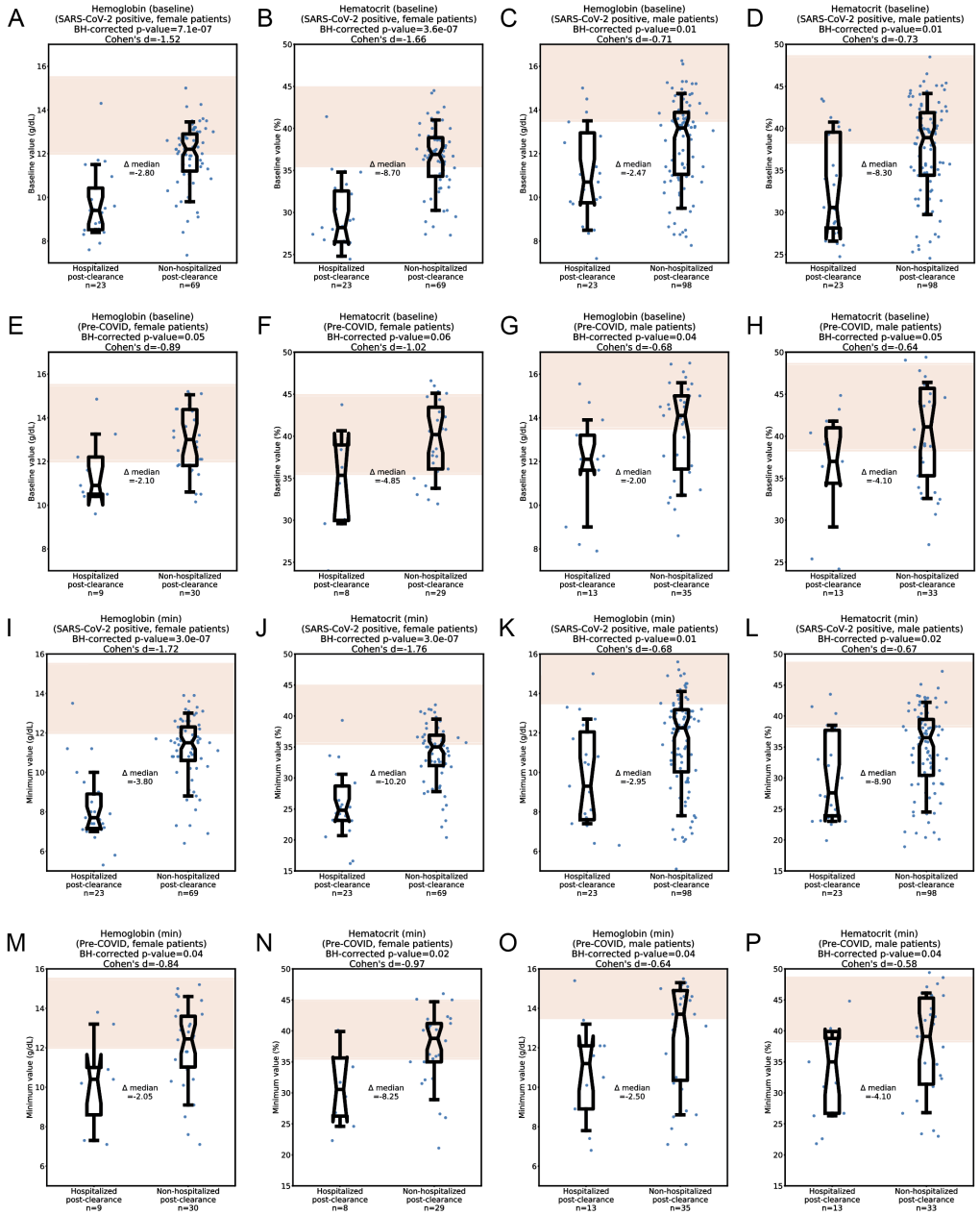


Figure 4. Sex-split analysis of anemia-related lab tests in the pre-COVID and SARS-CoV-2$^+$ phases. (A-H) Median values of hemoglobin and hematocrit in the SARS-CoV-2$^+$ phase (A-D) and pre-COVID phase (E-H), split to show female patients (A-B, E-F) and male patients (C-D, G-H) separately. (I-P) Minimum values of hemoglobin and hematocrit in the SARS-CoV-2$^+$ phase (I-L) and pre-COVID phase (M-P), split to show female patients (I-J, M-N) and male patients (K-L, O-P) separately. Red shading indicates normal ranges for hemoglobin and hematocrit depending on sex (females: 12.0-15.5 g/dL, 35.5-44.9%; males: 12.0-15.5 g/dL, 38.3-48.6%). For each comparison, statistics shown include the number of patients analyzed, Cohen's D, BH-corrected Mann Whitney U test p-value, and the difference of medians between the two cohorts. Box and whisker plots depict median and IQR along with the 10th and 90th percentiles.

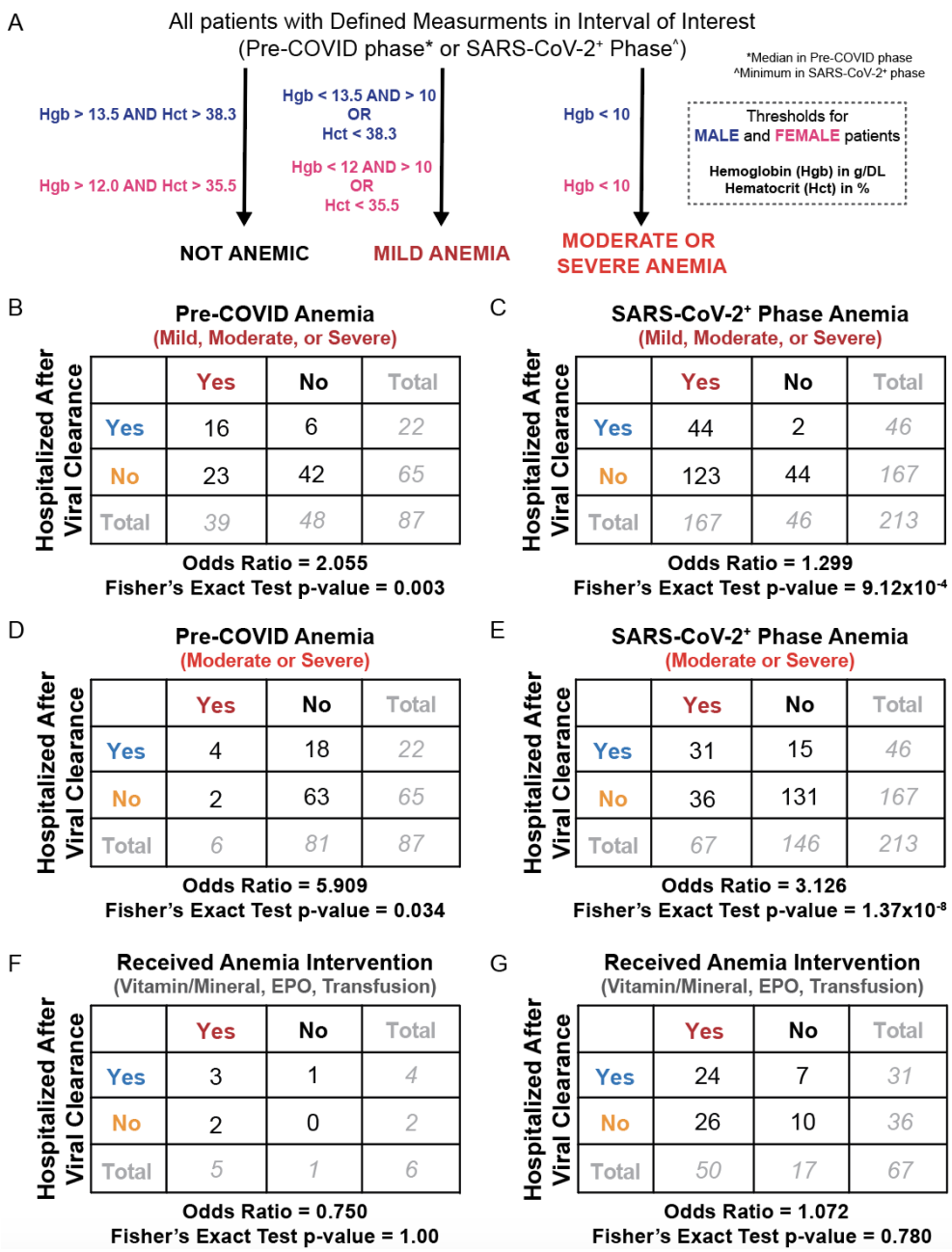


Figure 5. Comparison of outright anemia prevalence in the PCH and PCNH cohorts. (A) Schematic illustrating how patients are classified as having no anemia, mild anemia, or severe/moderate anemia. (B-C) Comparison of mild, moderate, or severe anemia frequency in the PCH and PCNH cohorts during the pre-COVID (n=87) and SARS-CoV-2$^+$ (n=213) phases. (D-E) Comparison of moderate or severe anemia frequency in the PCH and PCNH cohorts during the pre-COVID (n=87) and SARS-CoV-2$^+$ (n=213) phases. (F-G) Among patients with moderate or severe anemia in the specified time interval, comparison of administration rates for potential anemia-mitigating interventions between the PCH and PCNH cohorts in the pre-COVID (n=6) and SARS-CoV-2$^+$ (n=67) intervals. Contingency tables show the counts of patients in each intersecting category. Below each contingency table, the associated odds ratio and Fisher Exact test p-value is shown.

Demographic or Clinical Characteristic	Hospitalized Post-Clearance ("PCH Cohort")	Non-Hospitalized Post-Clearance ("PCNH Cohort")	BH-adjusted p-value
Total number of patients, *N*	49	173	
Age *years (standard deviation)*	58.7 (18.5)	58.5 (16.7)	0.959
Relative Cleared Date, *days (standard deviation)*	21.7 (13.5)	21.7 (12.8)	0.993
Sex			
- Female, *n (%)*	25 (51.0%)	71 (41.0%)	0.697
- Male, *n (%)*	24 (49.0%)	102 (59.0%)	0.697
Race			
- White, *n (%)*	32 (65.3%)	104 (60.1%)	0.849
- Asian, *n (%)*	1 (2.0%)	18 (10.4%)	0.697
- Black, *n (%)*	4 (8.2%)	22 (12.7%)	0.843
- Other, *n (%)*	12 (24.5%)	29 (16.8%)	0.697
Ethnicity			
- Hispanic, *n (%)*	9 (18.4%)	39 (22.5%)	0.849
- Non-Hispanic, *n (%)*	39 (79.6%)	127 (73.4%)	0.843
- Other, *n (%)*	1 (2.0%)	7 (4.0%)	0.843
Comorbidities			
- Asthma, *n (%)*	5 (10.2%)	9 (5.2%)	0.476
- Cancer, *n (%)*	8 (16.3%)	21 (12.1%)	0.536
- CKD, *n (%)*	9 (18.4%)	15 (8.7%)	0.271
- COPD, *n (%)*	4 (8.2%)	5 (2.9%)	0.333
- Obesity, *n (%)*	7 (14.3%)	16 (9.2%)	0.536
- Sleep apnea, *n (%)*	6 (12.2%)	15 (8.7%)	0.536
- Type 1 DM, *n (%)*	1 (2.0%)	2 (1.2%)	0.536
- Type 2 DM, *n (%)*	11 (22.4%)	31 (17.9%)	0.536

Table 1. Demographics and clinical characteristics of study cohorts, including patients who were and who were not rehospitalized after PCR-confirmed clearance of SARS-CoV-2. Each demographic variable or clinical characteristic was tested for difference in proportion with a Fisher Exact test or a difference in magnitude (for continuous variables) using a Mann-Whitney U test, and p-values were corrected for multiple testing using the Benjamini-Hochberg (BH) correction.

Test	Units	Phase	N (PCH)	Mean of Medians (PCH)	N (PCNH)	Mean of Medians (PCNH)	Cohen's D	BH-Adj P-val
Hct	%	SARS-CoV-2$^+$	46	31.27	167	37.01	-1.09	1.12E-07
Hgb	g/dL	SARS-CoV-2$^+$	46	10.37	167	12.27	-1.02	2.44E-07
Hgb	g/dL	Pre-COVID	22	11.75	65	13.18	-0.74	0.0141
Hct	%	Pre-COVID	21	35.75	62	40.03	-0.76	0.0141
eGFR	mL/min/BSA	SARS-CoV-2$^+$	35	60.96	107	71.26	-0.52	0.0141
BUN	mg/dL	SARS-CoV-2$^+$	30	27.01	92	20.82	0.41	0.0141
BUN	mg/dL	Pre-COVID	14	29.15	47	18.33	0.77	0.0422
Cr	mg/dL	Pre-COVID	14	2.15	53	1.22	0.49	0.0998
Cr	mg/dL	SARS-CoV-2$^+$	30	1.31	92	1.14	0.13	0.0998
eGFR	mL/min/BSA	Pre-COVID	15	61.17	50	68.66	-0.44	0.1873

Table 2. Analysis of median values for all selected lab tests in pre-COVID and SARS-CoV-2$^+$ phases, including both male and female patients. Lab tests were only considered for statistical analysis if there were at least 5 patients contributing data points in both the PCH and PCNH cohorts. Entries are sorted in order of statistical significance by the BH-adjusted Mann Whitney U-test p-value. Abbreviations are defined as follows: Hgb - hemoglobin; Hct - hematocrit; eGFR - estimated glomerular filtration rate; Cr - creatinine, BUN - blood urea nitrogen; g/DL - grams per deciliter; mL/min/BSA - milliliters per minute normalized for body surface area.

Test	Units	Phase	N (PCH)	Mean of Minima (PCH)	N (PCNH)	Mean of Minima (PCNH)	Cohen's D	BH-Adj P-val
Hgb	g/dL	SARS-CoV-2$^+$	46	9.05	167	11.37	-1.07	1.35E-07
Hct	%	SARS-CoV-2$^+$	46	27.97	167	34.51	-1.08	1.35E-07
Hct	%	Pre-COVID	21	32.49	62	37.59	-0.71	0.0073
Hgb	g/dL	Pre-COVID	22	10.59	65	12.38	-0.71	0.0079
eGFR	mL/min/BSA	SARS-CoV-2$^+$	35	46.94	107	59.75	-0.54	0.0117
eGFR	mL/min/BSA	Pre-COVID	15	49.07	50	61.34	-0.65	0.0337
BUN	mg/dL	Pre-COVID	14	21.51	47	15.45	0.56	0.1156
Cr	mg/dL	Pre-COVID	14	1.61	53	1.10	0.37	0.1546
BUN	mg/dL	SARS-CoV-2$^+$	30	16.81	92	14.54	0.21	0.2383
Cr	mg/dL	SARS-CoV-2$^+$	30	1.04	92	0.93	0.12	0.3934

Table 3. Analysis of minimum values for all selected lab tests in pre-COVID and SARS-CoV-2$^+$ phases, including both male and female patients. Lab tests were only considered for statistical analysis if there were at least 5 patients contributing data points in both the PCH and PCNH cohorts. Entries are sorted in order of statistical significance by the BH-adjusted Mann Whitney U-test p-value. Abbreviations are defined as follows: Hgb - hemoglobin; Hct - hematocrit; eGFR - estimated glomerular filtration rate; Cr - creatinine, BUN - blood urea nitrogen; g/DL - grams per deciliter; mL/min/BSA - milliliters per minute normalized for body surface area.

Test	Units	Phase	N (PCH)	Mean of Maxima (PCH)	N (PCNH)	Mean of Maxima (PCNH)	Cohen's D	BH-Adj P-val
Hct	%	SARS-CoV-2$^+$	46	36.46	167	40.52	-0.76	0.0002
Hgb	g/dL	SARS-CoV-2$^+$	46	12.27	167	13.59	-0.68	0.0006
BUN	mg/dL	SARS-CoV-2$^+$	30	46.93	92	29.53	0.69	0.0008
BUN	mg/dL	Pre-COVID	14	42.49	47	22.89	0.80	0.0066
Cr	mg/dL	SARS-CoV-2$^+$	30	2.00	92	1.50	0.26	0.0170
Hgb	g/dL	Pre-COVID	22	12.94	65	13.88	-0.57	0.0448
eGFR	mL/min/BSA	SARS-CoV-2$^+$	35	70.74	107	80.38	-0.53	0.0448
Hct	%	Pre-COVID	21	39.79	62	42.20	-0.51	0.0652
Cr	mg/dL	Pre-COVID	14	2.67	53	1.38	0.57	0.0652
eGFR	mL/min/BSA	Pre-COVID	15	70.53	50	75.00	-0.27	0.4442

Table 4. Analysis of maximum values for all selected lab tests in pre-COVID and SARS-CoV-2$^+$ phases, including both male and female patients. Lab tests were only considered for statistical analysis if there were at least 5 patients contributing data points in both the PCH and PCNH cohorts. Entries are sorted in order of statistical significance by the BH-adjusted Mann Whitney U-test p-value. Abbreviations are defined as follows: Hgb - hemoglobin; Hct - hematocrit; eGFR - estimated glomerular filtration rate; Cr - creatinine, BUN - blood urea nitrogen; g/DL - grams per deciliter; mL/min/BSA - milliliters per minute normalized for body surface area.

Time Interval	Anemia Metric Tested in Regression	Anemia	Blood Draw Count During Interval	Sex	ICU Admission During Interval
Pre-COVID	Median Hemoglobin (continuous)	β = -0.404 **p = 0.005**	β = -0.009 p = 0.568	β = 0.364 p = 0.495	*NA*
Pre-COVID	Minimum Hemoglobin (continuous)	β = -0.344 **p = 0.005**	β = -0.019 p = 0.334	β = 0.410 p = 0.444	*NA*
Pre-COVID	Moderate/severe anemia (binary)	β = 1.91 **p = 0.039**	β = -0.002 p = 0.848	β = 0.022 p = 0.966	*NA*
SARS-CoV-2$^+$	Median Hemoglobin (continuous)	β = -0.473 **p = 3.0x10^{-6}**	β = 0.006 p = 0.183	β = -0.021 p = 0.956	β = 0.741 p = 0.066
SARS-CoV-2$^+$	Minimum Hemoglobin (continuous)	β = -0.394 **p = 1.0x10^{-5}**	β = 0.003 p = 0.570	β = -0.162 p = 0.662	β = 0.541 p = 0.174
SARS-CoV-2$^+$	Moderate/severe anemia (binary)	β = 1.778 **p = 9.0x10^{-6}**	β = 0.003 p = 0.468	β = -0.298 p = 0.422	β = 0.583 p = 0.142

Table 5. Logistic regression analyses to assess the association between post-viral clearance hospitalization and anemia or potential confounding variables (number of blood draws in the given interval, sex, ICU admission status during the given interval). For each time interval (pre-COVID and SARS-CoV-2$^+$), three regression analyses were conducted using different anemia-related laboratory metrics: median hemoglobin (continuous value), minimum hemoglobin (continuous value), and the diagnosis of moderate or severe anemia as defined by a median (pre-COVID) or minimum (SARS-CoV-2$^+$) hemoglobin measurement < 10 mg/dL. For all six regressions, the coefficient (β) and associated p-value (p) are shown for each independent variable assessed. The coefficient represents the log-odds ratio, and p-values were calculated using the log likelihood ratio test. An association between an independent variable and post clearance hospitalization is considered significant if $p < 0.05$ (shown in bold). The association between post-viral clearance hospitalization and ICU admission during the pre-COVID interval was not analyzed because this information was not available for our cohort prior to April 2020. Binary variables were assigned as follows: moderate/severe anemia: 0 = no anemia, 1 = anemia; sex: 0 = female, 1 = male; ICU admission during interval: 0 = not admitted to ICU, 1 = admitted to ICU.

Test	Units	Phase	N (PCH)	Mean of Medians (PCH)	N (PCNH)	Mean of Medians (PCNH)	Cohen's D	BH-Adj P-val
Hct	%	SARS-CoV-2$^+$	23	29.33	69	36.37	-1.66	3.57E-07
Hgb	g/dL	SARS-CoV-2$^+$	23	9.68	69	11.92	-1.52	7.12E-07
Hgb	g/dL	Pre-COVID	9	11.54	30	12.93	-0.89	0.0495
Hct	%	Pre-COVID	8	34.65	29	39.64	-1.02	0.0574
BUN	mg/dL	SARS-CoV-2$^+$	13	22.27	39	19.03	0.20	0.1134
Cr	mg/dL	SARS-CoV-2$^+$	13	0.79	39	1.40	-0.34	0.3370
eGFR	mL/min/BSA	Pre-COVID	6	76.50	18	73.75	0.32	0.3525
eGFR	mL/min/BSA	SARS-CoV-2$^+$	16	71.53	41	73.54	-0.11	0.3645

Table 6. Sex-split analysis of median values for all selected lab tests in female patients during the pre-COVID and SARS-CoV-2$^+$ phases. Lab tests were only considered for statistical analysis if there were at least 5 patients contributing data points in both the PCH and PCNH cohorts. Entries are sorted in order of statistical significance by the BH-adjusted Mann Whitney U-test p-value. Abbreviations are defined as follows: Hgb - hemoglobin; Hct - hematocrit; eGFR - estimated glomerular filtration rate; Cr - creatinine, BUN - blood urea nitrogen; g/DL - grams per deciliter; mL/min/BSA - milliliters per minute normalized for body surface area.

Test	Units	Phase	N (PCH)	Mean of Medians (PCH)	N (PCNH)	Mean of Medians (PCNH)	Cohen's D	BH-Adj P-val
Hgb	g/dL	SARS-CoV-2$^+$	23	11.05	98	12.52	-0.71	0.0059
Hct	%	SARS-CoV-2$^+$	23	33.20	98	37.45	-0.73	0.0059
eGFR	mL/min/BSA	SARS-CoV-2$^+$	19	52.05	66	69.85	-0.90	0.0059
Cr	mg/dL	SARS-CoV-2$^+$	17	1.71	53	0.95	1.09	0.0059
BUN	mg/dL	Pre-COVID	10	35.56	27	20.71	0.98	0.0089
BUN	mg/dL	SARS-CoV-2$^+$	17	30.64	53	22.13	0.62	0.0191
Hgb	g/dL	Pre-COVID	13	11.90	35	13.40	-0.68	0.0408
Hct	%	Pre-COVID	13	36.43	33	40.37	-0.64	0.0468
eGFR	mL/min/BSA	Pre-COVID	9	50.94	32	65.80	-0.81	0.0544
Cr	mg/dL	Pre-COVID	10	2.69	31	1.19	0.99	0.1065

Table 7. Sex-split analysis of median values for all selected lab tests in male patients during the pre-COVID and SARS-CoV-2$^+$ phases. Lab tests were only considered for statistical analysis if there were at least 5 patients contributing data points in both the PCH and PCNH cohorts. Entries are sorted in order of statistical significance by the BH-adjusted Mann Whitney U-test p-value. Abbreviations are defined as follows: Hgb - hemoglobin; Hct - hematocrit; eGFR - estimated glomerular filtration rate; Cr - creatinine, BUN - blood urea nitrogen; g/DL - grams per deciliter; mL/min/BSA - milliliters per minute normalized for body surface area.

Test	Units	Phase	N (PCH)	Mean of Minima (PCH)	N (PCNH)	Mean of Minima (PCNH)	Cohen's D	BH-Adj P-val
Hgb	g/dL	SARS-CoV-2$^+$	23	8.23	69	11.18	-1.72	2.97E-07
Hct	%	SARS-CoV-2$^+$	23	25.59	69	34.07	-1.76	2.97E-07
Hct	%	Pre-COVID	8	31.12	29	37.26	-0.97	0.0243
Cr	mg/dL	SARS-CoV-2$^+$	13	0.60	39	1.09	-0.40	0.0250
Hgb	g/dL	Pre-COVID	9	10.28	30	12.10	-0.84	0.0409
eGFR	mL/min/BSA	Pre-COVID	6	56.17	18	65.67	-0.66	0.1286
eGFR	mL/min/BSA	SARS-CoV-2$^+$	16	54.81	41	61.51	-0.28	0.2336
BUN	mg/dL	SARS-CoV-2$^+$	13	12.97	39	13.95	-0.07	0.3957

Table 8. Sex-split analysis of minimum values for all selected lab tests in female patients during the pre-COVID and SARS-CoV-2$^+$ phases. Lab tests were only considered for statistical analysis if there were at least 5 patients contributing data points in both the PCH and PCNH cohorts. Entries are sorted in order of statistical significance by the BH-adjusted Mann Whitney U-test p-value. Abbreviations are defined as follows: Hgb - hemoglobin; Hct - hematocrit; eGFR - estimated glomerular filtration rate; Cr - creatinine, BUN - blood urea nitrogen; g/DL - grams per deciliter; mL/min/BSA - milliliters per minute normalized for body surface area.

Test	Units	Phase	N (PCH)	Mean of Minima (PCH)	N (PCNH)	Mean of Minima (PCNH)	Cohen's D	BH-Adj P-val
Hgb	g/dL	SARS-CoV-2$^+$	23	9.86	98	11.51	-0.68	0.0134
eGFR	mL/min/BSA	SARS-CoV-2$^+$	19	40.32	66	58.65	-0.78	0.0134
Hct	%	SARS-CoV-2$^+$	23	30.34	98	34.83	-0.67	0.0153
BUN	mg/dL	Pre-COVID	10	26.62	27	16.53	0.97	0.0171
Cr	mg/dL	SARS-CoV-2$^+$	17	1.38	53	0.81	1.01	0.0171
Hgb	g/dL	Pre-COVID	13	10.81	35	12.61	-0.64	0.0396
Hct	%	Pre-COVID	13	33.32	33	37.88	-0.58	0.0396
eGFR	mL/min/BSA	Pre-COVID	9	44.33	32	58.91	-0.71	0.0813
BUN	mg/dL	SARS-CoV-2$^+$	17	19.75	53	14.98	0.60	0.0835
Cr	mg/dL	Pre-COVID	10	1.97	31	1.08	1.02	0.0958

Table 9. Sex-split analysis of minimum values for all selected lab tests in male patients during the pre-COVID and SARS-CoV-2$^+$ phases. Lab tests were only considered for statistical analysis if there were at least 5 patients contributing data points in both the PCH and PCNH cohorts. Entries are sorted in order of statistical significance by the BH-adjusted Mann Whitney U-test p-value. Abbreviations are defined as follows: Hgb - hemoglobin; Hct - hematocrit; eGFR - estimated glomerular filtration rate; Cr - creatinine, BUN - blood urea nitrogen; g/DL - grams per deciliter; mL/min/BSA - milliliters per minute normalized for body surface area.

Supplementary Material

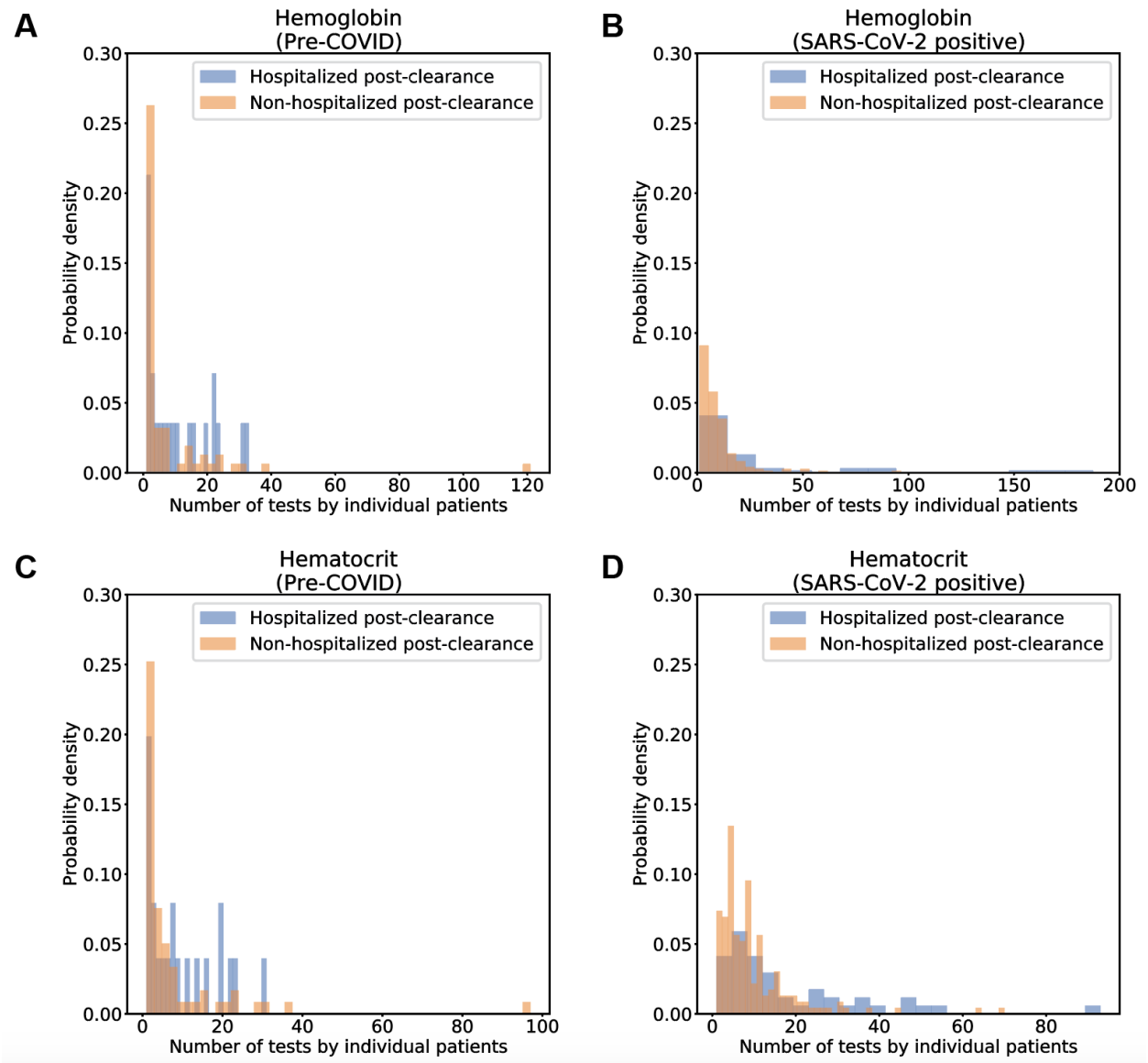


Figure S1. Histograms depicting the number of measurements per patient for the selected set of anemia-related lab tests in the pre-COVID and SARS-CoV-2^{+} intervals. (A) Number of hemoglobin tests per patient in the pre-COVID phase. (B) Number of hemoglobin tests per patient in the SARS-CoV-2^{+} phase. (C) Number of hematocrit tests per patient in the pre-COVID phase. (D) Number of hematocrit tests per patient in the SARS-CoV-2^{+} phase. Counts are shown separately for the PCH (blue) and PCNH (orange) cohorts.

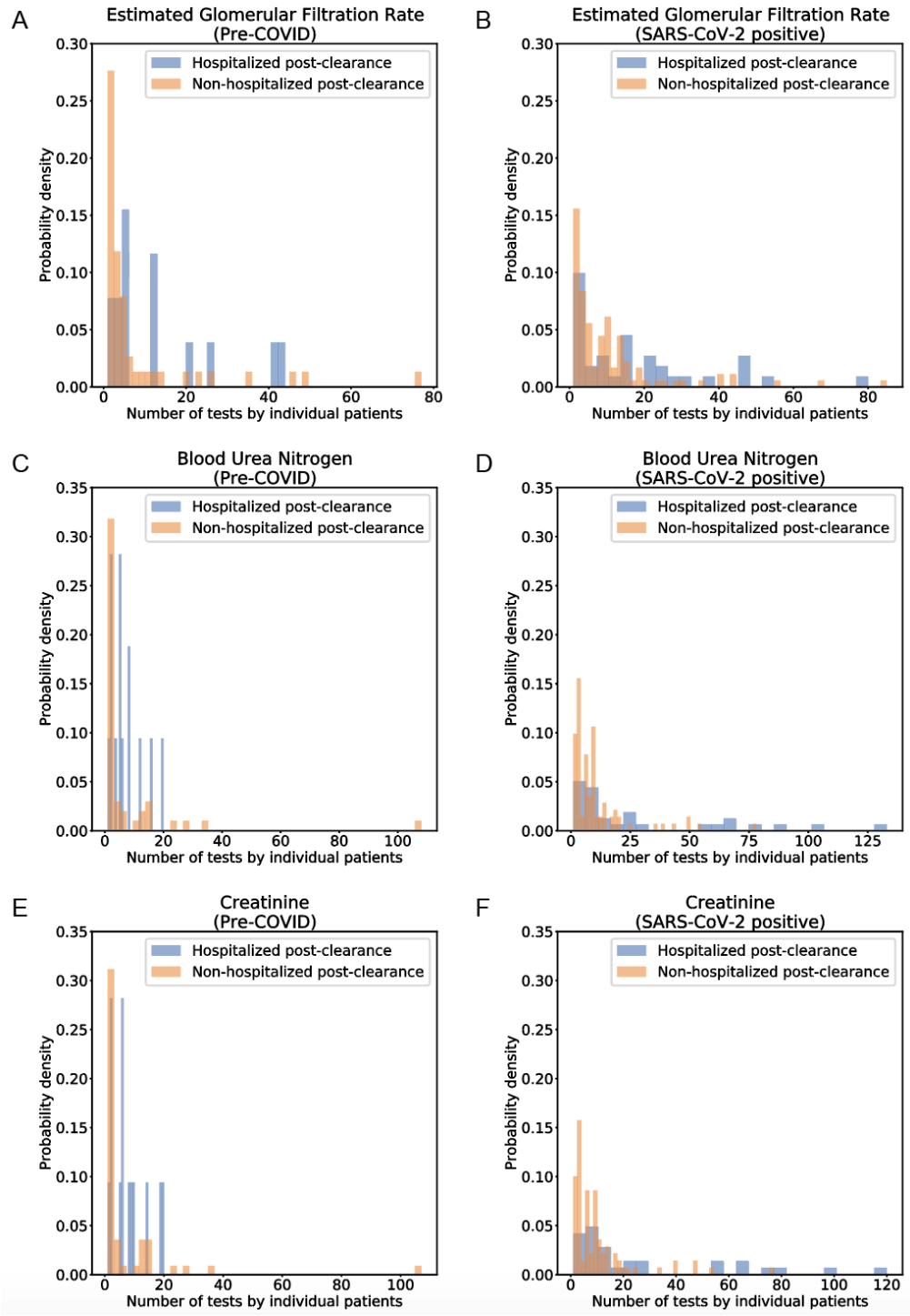


Figure S2. Histograms depicting the number of measurements per patient for the selected set of renal function tests in the pre-COVID and SARS-CoV-2^{+} intervals. (A) Number of eGFR tests per patient in the pre-COVID phase. (B) Number of eGFR tests per patient in the SARS-CoV-2^{+} phase. (C) Number of BUN tests per patient in the pre-COVID phase. (D) Number of BUN tests per patient in the SARS-CoV-2^{+} phase. (C) Number of serum creatinine tests per patient in the pre-COVID phase. (D) Number of serum creatinine tests per patient in the SARS-CoV-2^{+} phase.Counts are shown separately for the PCH (blue) and PCNH (orange) cohorts.

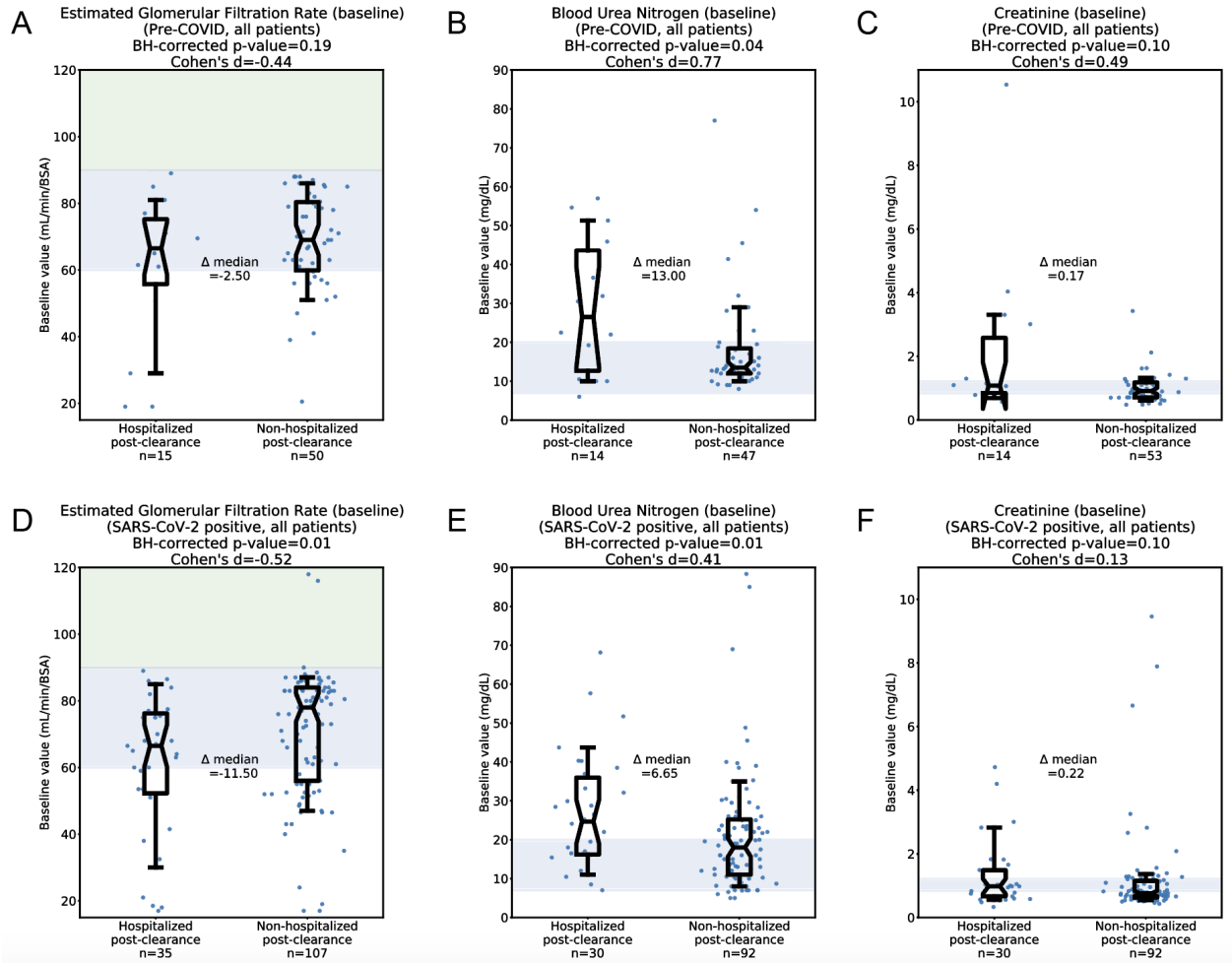


Figure S3. Comparison of median values for renal function tests during the pre-COVID and SARS-CoV-2$^+$ phases. (A) Pre-COVID median eGFR in the PCH (n=15) and PCNH (n=50) cohorts. (B) Pre-COVID median BUN in the PCH (n=14) and PCNH (n=47) cohorts. (C) Pre-COVID median serum creatinine in the PCH (n=14) and PCNH (n=53) cohorts. (D) SARS-CoV-2$^+$ median eGFR in the PCH (n=35) and PCNH (n=107) cohorts. (E) SARS-CoV-2$^+$ median BUN in the PCH (n=30) and PCNH (n=92) cohorts. (F) SARS-CoV-2$^+$ median serum creatinine in the PCH (n=30) and PCNH (n=92) cohorts. Shaded regions correspond to normal ranges for each test. For eGFR, the blue shading (60-90 mL/min/BSA) indicates moderately reduced levels which can be considered normal in older patients, while the green shading (>90 mL/min/BSA) indicates the normal range for younger patients. Normal ranges shown for BUN and serum creatinine are 7-20 mg/dL and 0.84-1.21 mg/dL, respectively. For each comparison, statistics shown include the number of patients analyzed, Cohen's D, BH-corrected Mann Whitney U test p-value, and the difference of medians between the two cohorts. Box and whisker plots depict median and IQR along with the 10th and 90th percentiles.

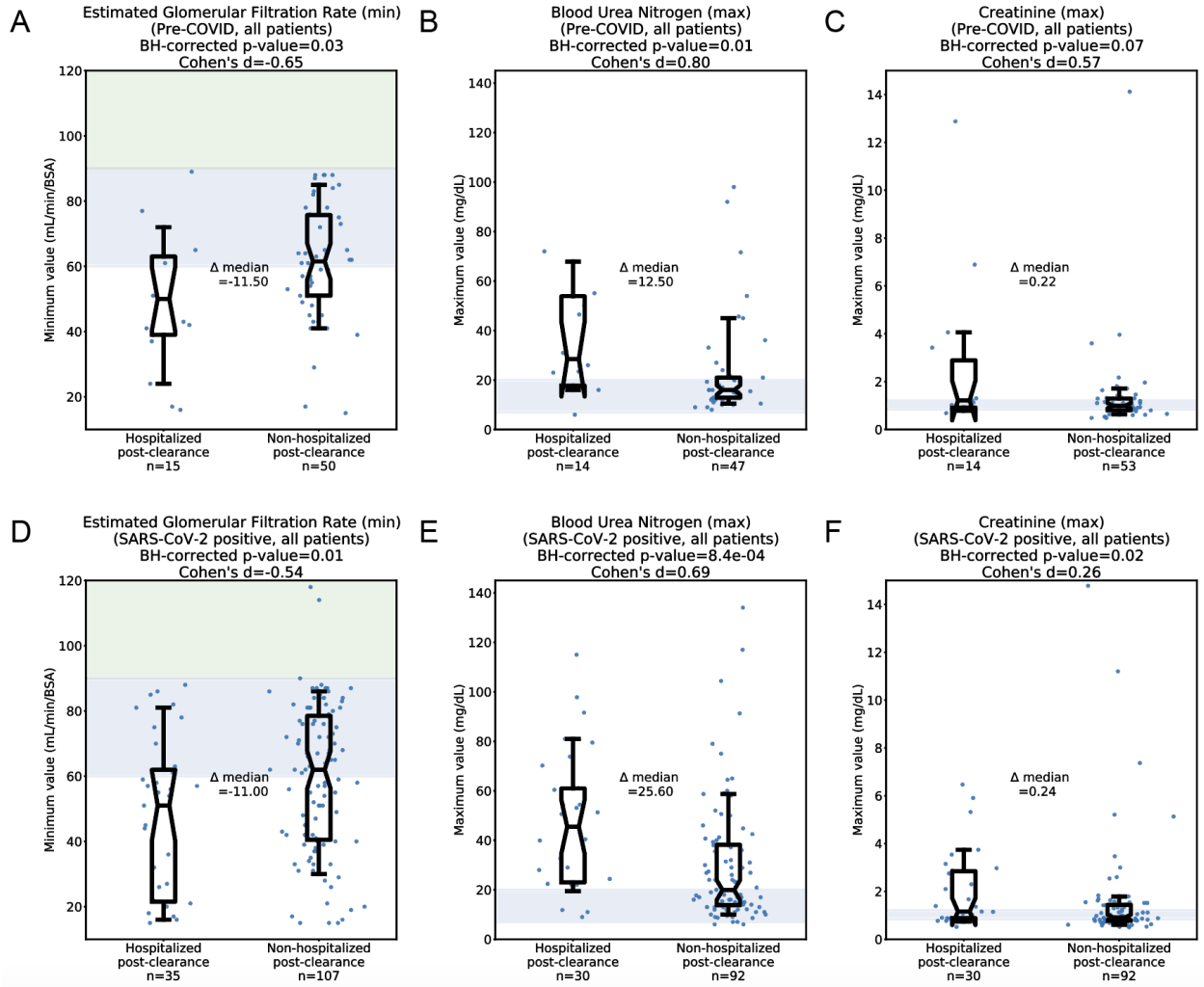


Figure S4. Comparison of "extreme" (minimum or maximum) values for renal function tests during the pre-COVID and SARS-CoV-2$^+$ phases. (A) Pre-COVID minimum eGFR in the PCH (n=15) and PCNH (n=50) cohorts. (B) Pre-COVID maximum BUN in the PCH (n=14) and PCNH (n=47) cohorts. (C) Pre-COVID maximum serum creatinine in the PCH (n=14) and PCNH (n=53) cohorts. (D) SARS-CoV-2$^+$ minimum eGFR in the PCH (n=35) and PCNH (n=107) cohorts. (E) SARS-CoV-2$^+$ maximum BUN in the PCH (n=30) and PCNH (n=92) cohorts. (F) SARS-CoV-2$^+$ maximum serum creatinine in the PCH (n=30) and PCNH (n=92) cohorts. Shaded regions correspond to normal ranges for each test. For eGFR, the blue shading (60-90 mL/min/BSA) indicates moderately reduced levels which can be considered normal in older patients, while the green shading (>90 mL/min/BSA) indicates the normal range for younger patients. Normal ranges shown for BUN and serum creatinine are 7-20 mg/dL and 0.84-1.21 mg/dL, respectively. For each comparison, statistics shown include the number of patients analyzed, Cohen's D, BH-corrected Mann Whitney U test p-value, and the difference of medians between the two cohorts. Box and whisker plots depict median and IQR along with the 10th and 90th percentiles.

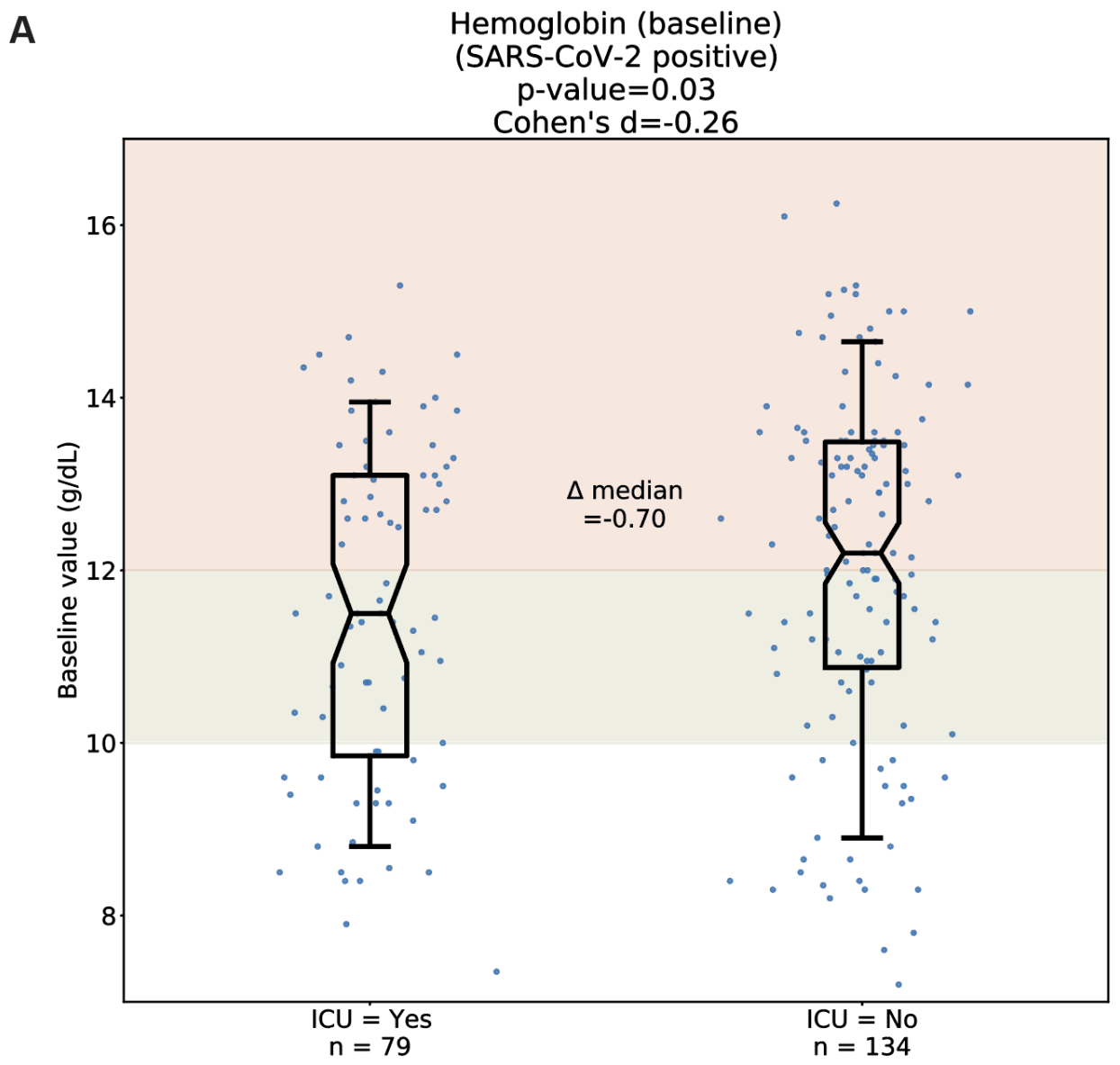


Figure S5. Comparison of hemoglobin levels in patients based on ICU admission status during COVID-19 index infection. Each dot corresponds to the median ("baseline") hemoglobin level of an individual patient during the SARS-CoV-2^{+} phase. The 213 patients with hemoglobin measurements in the SARS-CoV-2^{+} phase (out of 222 total patients) were divided based on whether they were admitted to the ICU during their index infection. Red shading indicates normal ranges for hemoglobin and hematocrit; as these ranges are lower for females than males, the shaded range here spans from the lower limit of normal for females (12 g/dL hemoglobin, 35.5% hematocrit) to the upper limit of normal for males (17.5 g/dL hemoglobin, 48.6% hematocrit). Green shading indicates mild anemia, defined as a hemoglobin level greater than 10 g/dL and less than the sex-dependent lower limit of normal. Statistics shown include the number of patients per group, Cohen's D, Mann Whitney U test p-value, and the difference of medians between the two cohorts. Box and whisker plots depict median and IQR along with the 10th and 90th percentiles.

A

Hospitalized After Viral Clearance	ICU Admission During COVID-19 Index Infection: Yes	No	Total
Yes	26	57	83
No	23	116	139
Total	49	173	222

ICU Admission Rates

PCH Cohort: 46% (26 / 57)

PCNH Cohort: 20% (23/117)

Rate Ratio = 2.3

Odds Ratio = 2.3

Fisher's Exact Test p-value = 0.0123

Figure S6. Comparison of rates of admission to the intensive care unit (ICU) during index COVID-19 infection between the PCH (n=49) and PCNH (n=173) cohorts. Contingency table shows the counts of patients in each intersecting category. On the right, ICU admission rates along with the associated rate ratio, odds ratio, and Fisher Exact test p-value is shown. ICU admission rate is higher in the PCH cohort than in the PCNH cohort.

A

SARS-CoV-2⁺ Phase Anemia
(Moderate or Severe)

Hospitalized After Viral Clearance	Yes	No	Total
Yes	20	6	*26*
No	16	37	*53*
Total	*36*	*43*	*79*

Odds Ratio = 2.548
Fisher's Exact Test p-value = 1.11×10^{-4}

Considering 79 patients who **were admitted to the ICU** during index COVID-19 infection

B

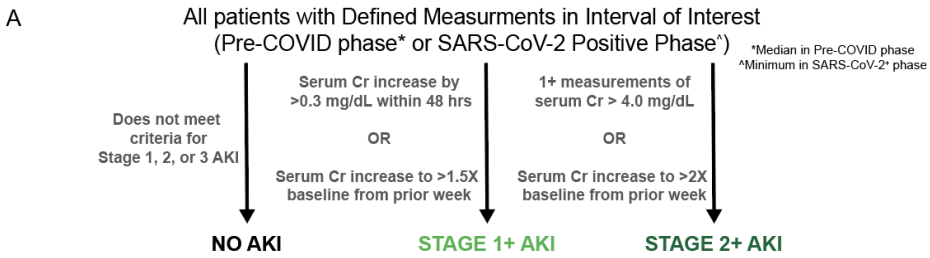
SARS-CoV-2⁺ Phase Anemia
(Moderate or Severe)

Hospitalized After Viral Clearance	Yes	No	Total
Yes	11	9	*20*
No	20	94	*114*
Total	*31*	*103*	*134*

Odds Ratio = 3.315
Fisher's Exact Test p-value = 7.78×10^{-4}

Considering 134 patients who **were NOT admitted to the ICU** during index COVID-19 infection

Figure S7. Comparison of rates of moderate and severe anemia in the PCH versus PCNH cohorts split by index infection ICU admission status. (A) Rates of moderate or severe anemia in patients who were admitted to the ICU during their index COVID-19 infection (n=79). (B) Rates of moderate or severe anemia in patients who were not admitted to the ICU during their index COVID-19 infection (n=134). Contingency tables show the counts of patients in each intersecting category. Below each contingency table, the associated odds ratio and Fisher Exact test p-value is shown. Regardless of ICU admission status, patients who experienced moderate or severe anemia during the SARS-CoV-2⁺ phase were more likely to be hospitalized after viral clearance.



B

Pre-COVID
Stage 1+ AKI

Hospitalized After Viral Clearance	Yes	No	Total
Yes	2	14	16
No	6	47	53
Total	8	61	69

Odds Ratio = 1.104
Fisher's Exact Test p-value = 1.000

C

SARS-CoV-2+Phase
Stage 1+ AKI

Hospitalized After Viral Clearance	Yes	No	Total
Yes	15	15	30
No	23	69	92
Total	38	84	122

Odds Ratio = 2.000
Fisher's Exact Test p-value = 0.013

D

Pre-COVID
Stage 2+ AKI

Hospitalized After Viral Clearance	Yes	No	Total
Yes	3	13	16
No	3	50	53
Total	6	63	69

Odds Ratio = 3.312
Fisher's Exact Test p-value = 0.132

E

SARS-CoV-2+ Phase
Stage 2+ AKI

Hospitalized After Viral Clearance	Yes	No	Total
Yes	6	24	30
No	7	85	92
Total	13	109	122

Odds Ratio = 2.629
Fisher's Exact Test p-value = 0.084

F

Pre-COVID MALES
Stage 2+ AKI

Hospitalized After Viral Clearance	Yes	No	Total
Yes	3	7	10
No	1	30	31
Total	4	37	41

Odds Ratio = 9.300
Fisher's Exact Test p-value = 0.039

G

SARS-CoV-2+ Phase MALES
Stage 2+ AKI

Hospitalized After Viral Clearance	Yes	No	Total
Yes	5	12	17
No	1	52	53
Total	6	64	70

Odds Ratio = 15.588
Fisher's Exact Test p-value = 0.003

Figure S8. Comparison of AKI frequency in the PCH and PCNH cohorts during the pre-COVID and SARS-CoV-2$^+$ phases. (A) Schematic illustrating how patients are classified as having no AKI, stage 1 AKI, or stage 2/3 AKI, based on the KDIGO criteria. (B) Association between pre-COVID AKI (any stage) and post viral clearance hospitalization status in all patients (n=69). (C) Association between SARS-CoV-2$^+$ phase AKI and post viral clearance hospitalization status in all patients (n=122). (D) Association between pre-COVID Stage 2 or 3 AKI and post viral clearance hospitalization status in all patients (n=69). (E) Association between SARS-CoV-2$^+$ phase Stage 2 or 3 AKI and post viral clearance hospitalization status in all patients (n=122). (F) Association between pre-COVID Stage 2 or 3 AKI and post viral clearance hospitalization status in only male patients (n=41). (F) Association between SARS-CoV-2$^+$ phase Stage 2 or 3 AKI and post viral clearance hospitalization status in only male patients (n=70). Contingency tables show the counts of patients in each intersecting category. Below each contingency table, the associated odds ratio and Fisher Exact test p-value is shown.

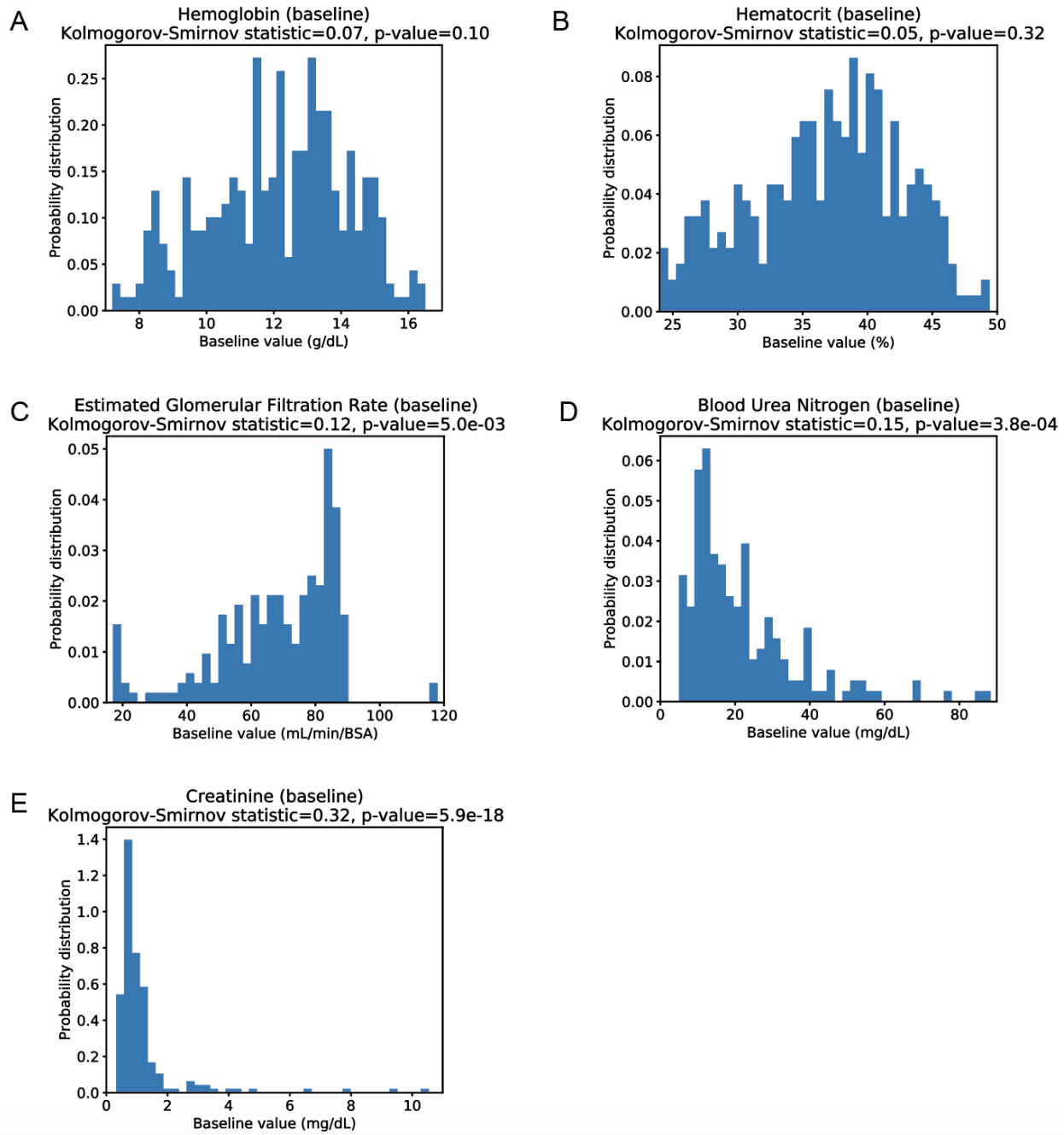


Figure S9. Distributions and normality test of patient-wise median values for all considered lab tests. Lab tests evaluated include (A) hemoglobin, (B) hematocrit, (C) eGFR, (D) BUN, and (E) serum creatinine. Above each plot, the Kolmogorov-Smirnov statistic and associated p-value are shown; a p-value below 0.05 indicates that the data do not follow a normal distribution.

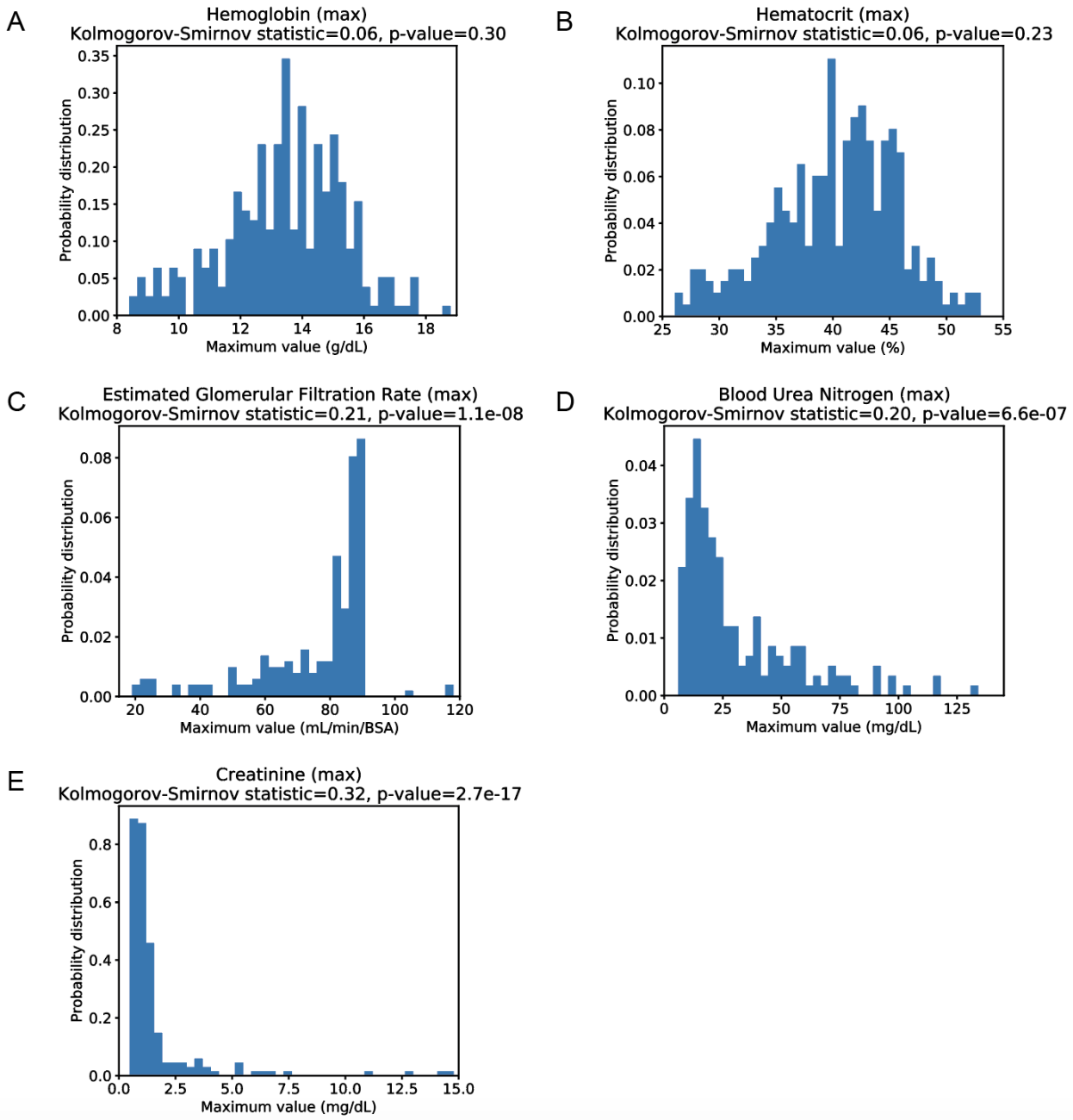


Figure S10. Distributions and normality test of patient-wise maximum values for all considered lab tests. Lab tests evaluated include (A) hemoglobin, (B) hematocrit, (C) eGFR, (D) BUN, and (E) serum creatinine. Above each plot, the Kolmogorov-Smirnov statistic and associated p-value are shown; a p-value below 0.05 indicates that the data do not follow a normal distribution.

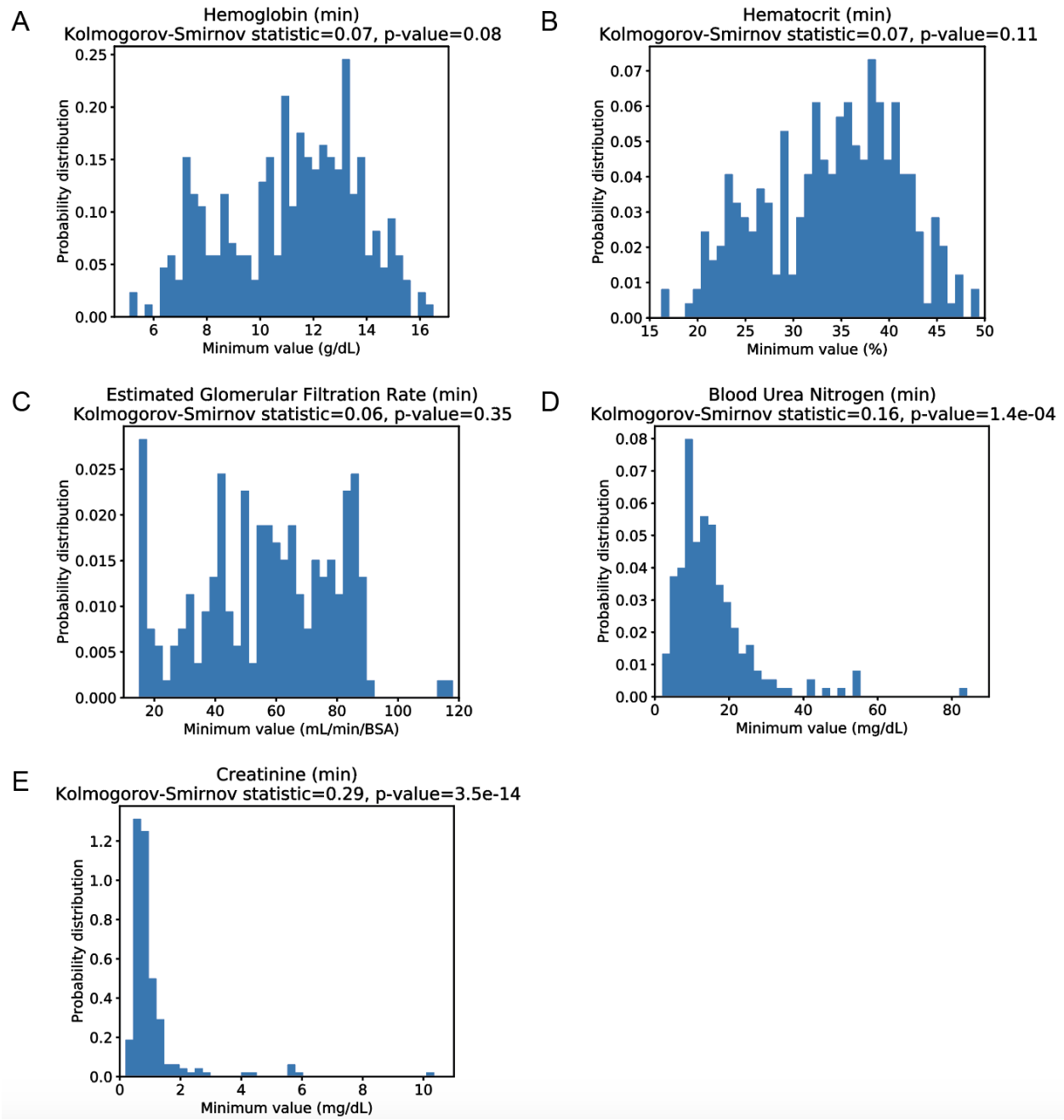


Figure S11. Distributions and normality test of patient-wise minimum values for all considered lab tests. Lab tests evaluated include (A) hemoglobin, (B) hematocrit, (C) eGFR, (D) BUN, and (E) serum creatinine. Above each plot, the Kolmogorov-Smirnov statistic and associated p-value are shown; a p-value below 0.05 indicates that the data do not follow a normal distribution.