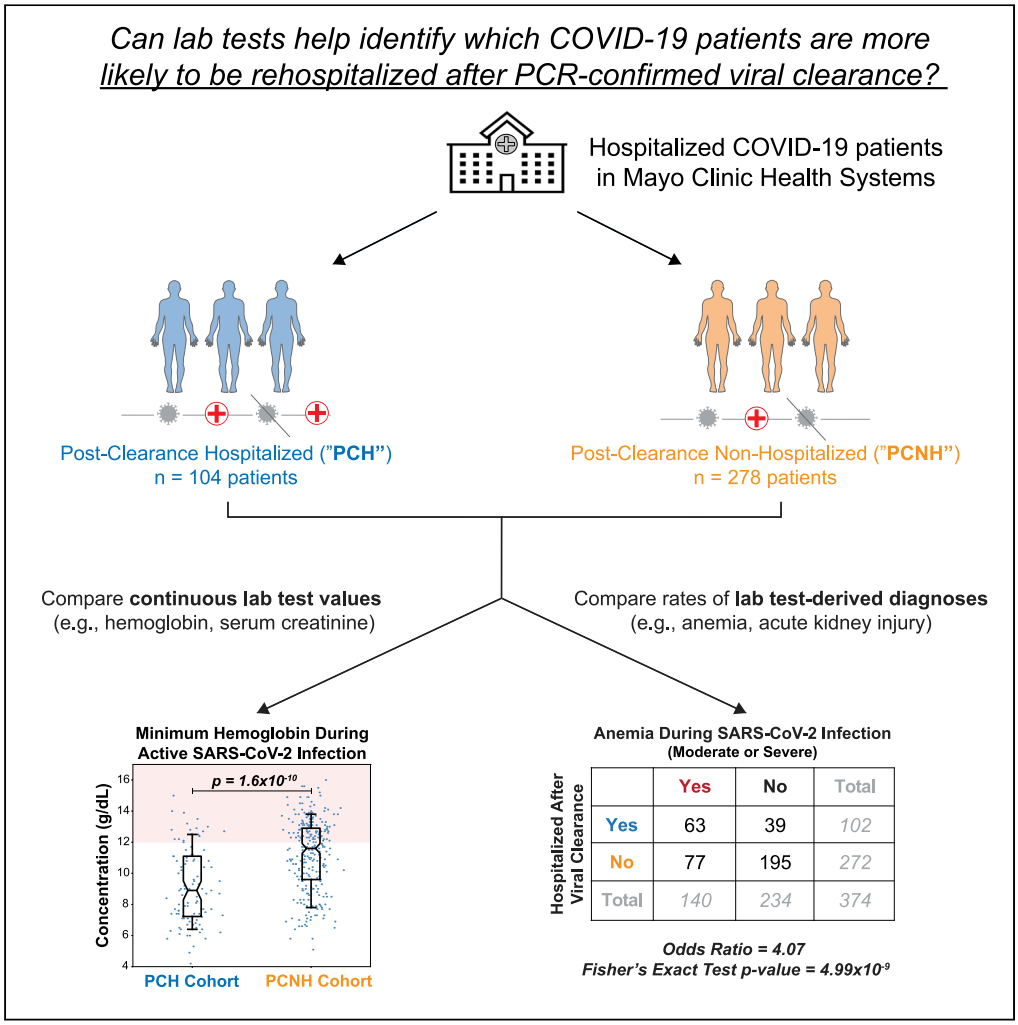
iScience

CellPress
OPEN ACCESS

Article

Anemia during SARS-CoV-2 infection is associated with rehospitalization after viral clearance



Patrick J. Lenehan, Eshwan Ramudu, A.J. Venkatakrishnan, ..., William Morice II, John Halamka, Venky Soundararajan

venky@nference.net

Highlights

Patients rehospitalized after SARS-CoV-2 clearance have distinct laboratory test profiles

Rehospitalized patients have lower hemoglobin before and during SARS-CoV-2 infection

Rehospitalized patients are more likely to experience anemia during active infection

Lenehan et al., iScience 24, 102780
July 23, 2021 © 2021
https://doi.org/10.1016/j.isci.2021.102780

iScience

CellPress
OPEN ACCESS

Article

Anemia during SARS-CoV-2 infection is associated with rehospitalization after viral clearance

Patrick J. Lenehan,[1,4] Eshwan Ramudu,[1,4] A.J. 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] and Venky Soundararajan[1,5,*]

SUMMARY

Patients with COVID-19 can experience symptoms and complications after viral clearance. It is important to identify clinical features of patients who are likely to experience these prolonged effects. We conducted a retrospective study to compare longitudinal laboratory 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 = 104) versus patients not rehospitalized after viral clearance (n = 278). Rehospitalized patients had lower median hemoglobin levels in the year prior to COVID-19 diagnosis (Cohen's D = −0.50; p = 1.2×10^{-3}) and during their active SARS-CoV-2 infection (Cohen's D = −0.71; p = 4.6×10^{-8}). Rehospitalized patients were also more likely to be diagnosed with moderate or severe anemia during their active infection (Odds Ratio = 4.07; p = 4.99×10^{-9}). These findings suggest that anemia-related laboratory tests should be considered in risk stratification algorithms for patients with COVID-19.

INTRODUCTION

Since the first diagnosed case of COVID-19 in December 2019, over 170 million people have been infected with SARS-CoV-2 worldwide resulting in over 3.5 million deaths (Johns Hopkins Coronavirus Resource Center, n.d.). Although significant progress has been made in understanding the pathogenesis of COVID-19, including the rapid development and clinical rollout of multiple vaccines (Bos et al., 2020; Corbett et al., 2020; Folegatti et al., 2020; Jackson et al., 2020; Mercado et al., 2020; Mulligan et al., 2020), along with detailed characterizations of the SARS-CoV-2 entry receptor ACE2 (Anand et al., 2020; Singh et al., 2020; Venkatakrishnan et al., 2020; Zhao et al., 2020; Ziegler et al., 2020), there are still few options available for effective treatment of patients with severe COVID-19. Furthermore, 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 infection period (Carfì et al., 2020; del Rio et al., 2020; Yelin et al., 2020). However, the clinical, molecular, and demographic biomarkers characterizing patients who are more likely to experience these lasting effects after clearing SARS-CoV-2 ("long COVID") 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 patients with COVID-19, 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., laboratory 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 patients with COVID-19 within a large EHR system (Pawlowski et al., 2021; Wagner et al., 2020). For example, we have used natural language processing (NLP) to train disease diagnosis models that classify mentions of phenotypes in EHR notes as positive (i.e., Patient has Disease X), negative (i.e., Patient does not have Disease X), or other

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

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

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

[4]These authors contributed equally

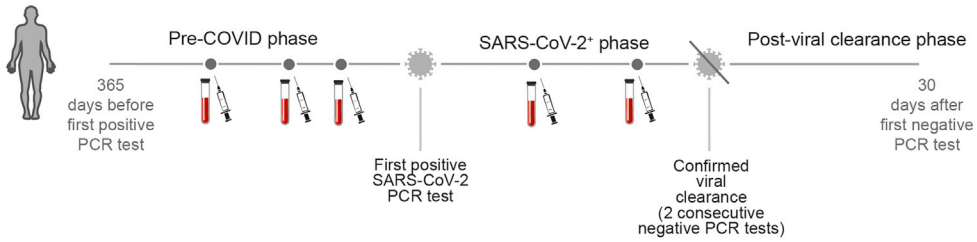
[5]Lead contact

*Correspondence: venky@nference.net

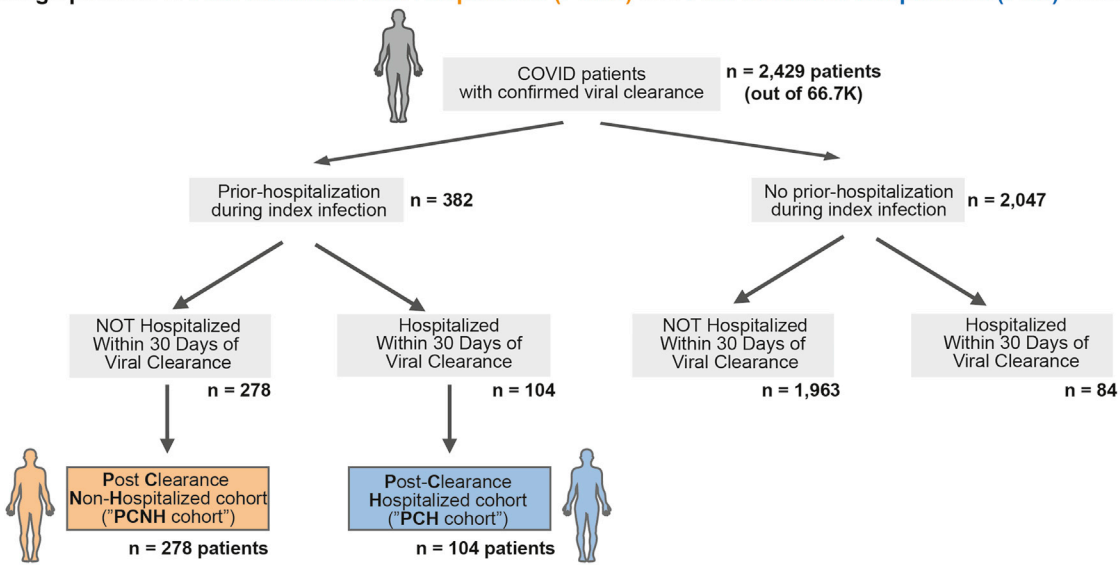
https://doi.org/10.1016/j.isci.2021.102780

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

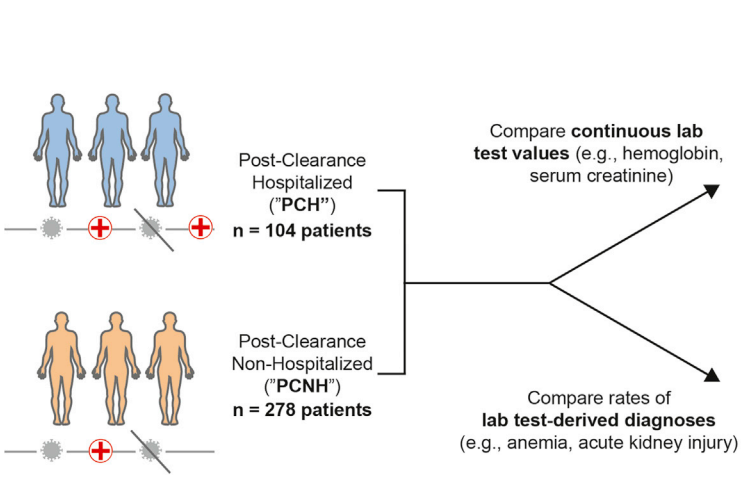
A Time-line capturing the clinical trajectory of a COVID-19 patient based on PCR tests and lab tests

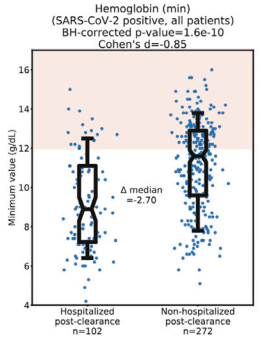


B Assign patients to Post-Clearance Non-Hospitalized (PCNH) and Post Clearance Hospitalized (PCH) cohorts



C Compare lab test results in defined cohorts during the pre-COVID and SARS-CoV-2⁺ phases





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

Hospitalized After Viral Clearance	Yes	No	Total
Yes	63	39	*102*
No	77	195	*272*
Total	*140*	*234*	*374*

Odds Ratio = 4.07
Fisher's Exact Test p-value = 4.99×10^{-9}

Figure 1. Schematic summarizing cohort creation and laboratory test analyses

(A) Time intervals ("phases") were defined relative to SARS-CoV-2 PCR testing results.

(B) Of the 2,429 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 = 278), 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 = 104).

(C) A defined set of anemia and kidney-related laboratory test measurements were compared between the PCH and PCNH cohorts in the pre-COVID and SARS-CoV-2^{+} intervals.

(e.g., Patient is suspected to have Disease X or has a family history of Disease X). Using this textual sentiment-based curation model, we found that diagnoses of anemia and acute kidney injury (AKI) were recorded more frequently in the notes of patients with COVID-19 who were rehospitalized after PCR-confirmed SARS-CoV-2 clearance compared with patients who were not rehospitalized after viral clearance (Pawlowski et al., 2021). These findings applied to notes both in the year prior to COVID-19 diagnosis and during active SARS-CoV-2 infection.

Anemia, defined as a deficiency of red blood cells or hemoglobin in circulation, has several etiologies including vitamin or mineral deficiencies, chronic inflammation, drug- or infection-induced hemolysis, bone marrow suppression, and blood loss (Turner et al., 2021). AKI is generally caused by reduced blood flow to the kidney (pre-renal AKI); direct damage to the kidney itself by drugs, infectious agents, or excessive inflammation (intrinsic AKI); or obstruction of outflow from the renal tubular system (post-renal AKI) (Makris and Spanou, 2016). Both anemia and AKI are common in critically ill patients (Case et al., 2013; Girling et al., 2020; Mohsenin, 2017; Roubinian et al., 2019; Walsh et al., 2006; Warner et al., 2020) and have

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

Demographic or clinical characteristic	Hospitalized post-clearance ("PCH cohort")	Non-hospitalized post-clearance ("PCNH cohort")	BH-adjusted p value
Total number of patients, *N*	104	278	
Age			
Years (standard deviation)	59.2 (18.4%)	59.3 (17.5%)	1.00
≥60 years old, *n (%)*	61 (58.7%)	145 (52.2%)	(0.30)
Sex			
Female, *n (%)*	50 (48.1%)	118 (42.4%)	0.64
Male, *n (%)*	54 (51.9%)	160 (57.6%)	0.64
Race			
White, *n (%)*	78 (75.0%)	184 (66.2%)	0.43
Asian, *n (%)*	5 (4.8%)	22 (7.9%)	0.64
Black, *n (%)*	6 (5.8%)	38 (13.7%)	0.19
Other, *n (%)*	15 (14.4%)	34 (12.2%)	0.73
Ethnicity			
Hispanic, *n (%)*	16 (15.4%)	53 (19.1%)	0.69
Non-Hispanic, *n (%)*	85 (81.7%)	218 (78.4%)	0.73
Other, *n (%)*	3 (2.9%)	7 (2.5%)	1.00
Relative Cleared Date, *days (standard deviation)*	39.5 (39.4%)	32.4 (29.1%)	0.64
ICU Admission, *n (%)*	48 (46%)	73 (26%)	0.004*

Each demographic variable or clinical characteristic was tested for difference in proportion with a Fisher's exact test or a difference in magnitude (for continuous variables) using a Mann-Whitney U test, and p values shown without parentheses were corrected for multiple testing using the Benjamini-Hochberg (BH) correction. Statistically significant differences ($p < 0.05$) are denoted with an asterisk (*).

Table 2. Analysis of median values for all selected laboratory tests in pre-COVID and SARS-CoV-2$^+$ phases, including both male and female patients

Test	Units	Phase	N (PCH)	Mean of medians (PCH)	N (PCNH)	Mean of medians (PCNH)	Cohen's D	BH-adjusted p-value
Hgb	g/dL	SARS-CoV-2$^+$	102	10.70	272	12.13	−0.71	4.57×10^{-8}
Hct	%	SARS-CoV-2$^+$	102	32.65	272	36.61	−0.69	8.46×10^{-8}
BUN	mg/dL	SARS-CoV-2$^+$	102	26.99	272	20.85	0.42	8.87×10^{-6}
eGFR	mL/min/BSA	SARS-CoV-2$^+$	83	60.55	188	69.72	−0.45	1.21×10^{-3}
Hgb	g/dL	Pre-COVID	55	11.50	136	12.60	−0.50	1.23×10^{-3}
Hct	%	Pre-COVID	55	35.55	132	38.60	−0.48	2.46×10^{-3}
Cr	mg/dL	SARS-CoV-2$^+$	103	1.51	272	1.20	0.20	5.53×10^{-3}
BUN	mg/dL	Pre-COVID	56	25.73	133	19.80	0.46	1.33×10^{-2}
Cr	mg/dL	Pre-COVID	57	1.95	142	1.38	0.31	1.93×10^{-2}
eGFR	mL/min/BSA	Pre-COVID	47	56.65	109	65.92	−0.46	2.45×10^{-2}

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.

been suggested as biomarkers for mortality and disease severity in patients with COVID-19 (Chan et al., 2021; Faghih Dinevari et al., 2021; Hariyanto and Kurniawan, 2020; Nadim et al., 2020; Oh et al., 2021). However, our previously mentioned NLP analysis was the first to associate anemia with the long COVID syndrome and rehospitalization after SARS-CoV-2 clearance. Here we complement this work by evaluating whether diagnostic laboratory tests for anemia and AKI corroborate these phenotypic associations in a larger patient cohort.

RESULTS

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

We split the set of hospitalized COVID-19 patients with confirmed viral clearance (n = 382) into two groups: (1) post-clearance hospitalized ("PCH"; n = 104) and (2) post-clearance non-hospitalized ("PCNH"; n = 278), where viral clearance was defined by two consecutive negative SARS-CoV-2 PCR tests following a positive test (see STAR Methods and Figures 1A and 1B). A demographic summarization of these two cohorts is provided in Table 1. We then compared a set of selected laboratory 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 NLP-based findings (Pawlowski et al., 2021), we considered both anemia-related and kidney function laboratory tests including hemoglobin, hematocrit, estimated glomerular filtration rate (eGFR), serum creatinine, and serum blood urea nitrogen (BUN) levels (Figure 1C).

Rehospitalized patients show pathologic alterations in hemoglobin, hematocrit, eGFR, and BUN before COVID-19 diagnosis and during active infection

For each patient, we first considered the median values of each laboratory test over the designated interval. Consistent with our previous augmented curation-derived findings, PCH patients had significantly lower median hemoglobin and hematocrit levels in both the pre-COVID phase (Cohen's D = −0.50, $p = 1.2 \times 10^{-3}$; Cohen's D = −0.48, $p = 2.5 \times 10^{-3}$) and the SARS-CoV-2$^+$ phase (Cohen's D = −0.71, $p = 4.6 \times 10^{-8}$; Cohen's D = −0.69, $p = 8.5 \times 10^{-8}$) (Table 2, Figures 2A–2D). Furthermore, PCH patients had lower median eGFR and higher median BUN levels during the pre-COVID phase (Cohen's D = −0.46, p = 0.02; Cohen's D = −0.45, $p = 1.2 \times 10^{-3}$) and the SARS-CoV-2$^+$ phase (Cohen's D = 0.46, p = 0.01; Cohen's D = 0.42, $p = 8.9 \times 10^{-6}$) (Table 2, Figure S1).

We also tested whether extreme (i.e., minimum or maximum) values of a given laboratory test over the designated periods varied between PCH and PCNH patients, as a measure of central tendency (e.g., median) may fail to capture a single occurrence of phenotypes such as anemia or AKI. PCH patients had lower minimum values of hemoglobin, hematocrit, and eGFR in both the pre-COVID phase (Cohen's D = −0.49,

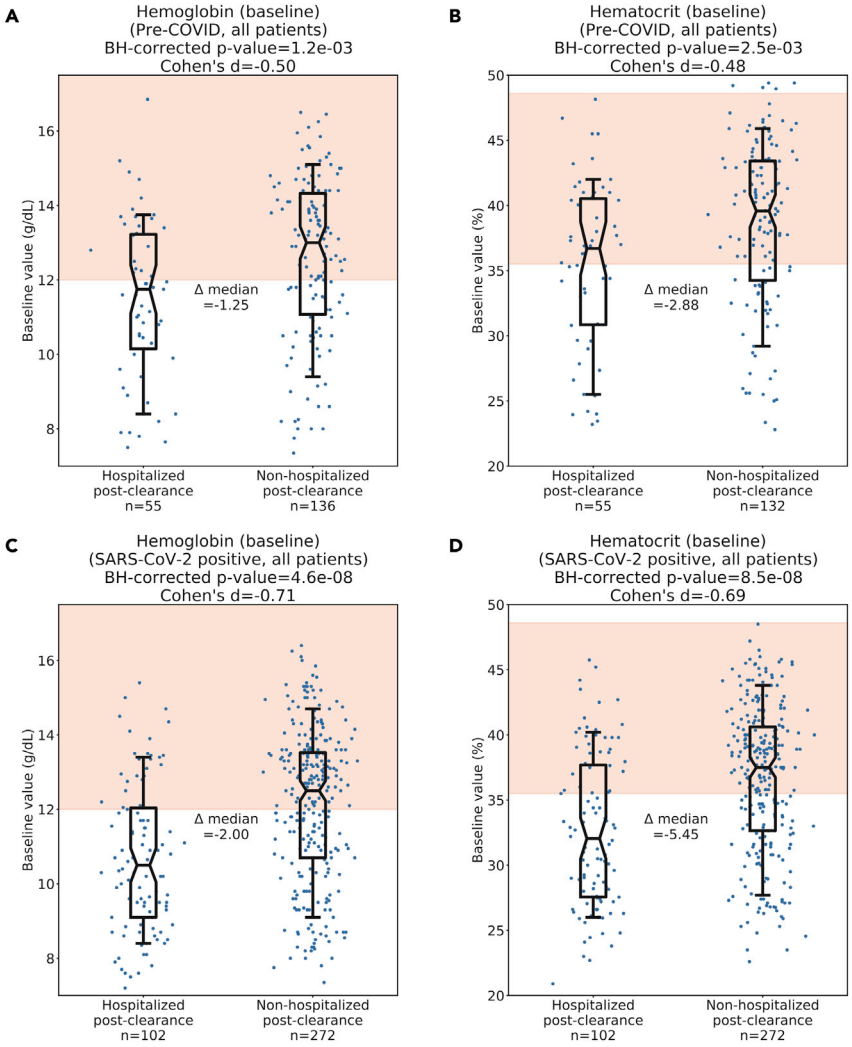


Figure 2. Comparison of median hemoglobin and hematocrit values during the pre-COVID and SARS-CoV-2$^+$ phases

(A) Pre-COVID median hemoglobin in the PCH (n = 55) and PCNH (n = 136) cohorts.

(B) Pre-COVID median hematocrit in the PCH (n = 55) and PCNH (n = 132) cohorts.

(C) SARS-CoV-2$^+$ median hemoglobin in the PCH (n = 102) and PCNH (n = 272) cohorts.

(D) SARS-CoV-2$^+$ median hematocrit in the PCH (n = 102) and PCNH (n = 272) 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 interquartile range (IQR) along with the 10th and 90th percentiles.

$p = 2.8 \times 10^{-3}$; Cohen's D = −0.45, $p = 3.0 \times 10^{-3}$; Cohen's D = −0.57, $p = 3.0 \times 10^{-3}$) and the SARS-CoV-2$^+$ phase (Cohen's D = −0.85, $p = 1.6 \times 10^{-10}$; Cohen's D = −0.79, $p = 1.2 \times 10^{-9}$; Cohen's D = −0.51, $p = 4.4 \times 10^{-4}$) (Table 3; Figures 3 and S2). They also had higher maximum serum BUN levels during both the pre-COVID phase (Cohen's D = 0.50, $p = 6.6 \times 10^{-4}$) and the SARS-CoV-2$^+$ phase (Cohen's D = 0.60, $p = 5.2 \times 10^{-8}$) (Table 4 and Figure S2).

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 have pathologically altered anemia-related and renal function laboratory tests both prior to and during SARS-CoV-2 infection.

Table 3. Analysis of minimum values for all selected laboratory tests in pre-COVID and SARS-CoV-2$^+$ phases, including both male and female patients

Test	Units	Phase	N (PCH)	Mean of minima (PCH)	N (PCNH)	Mean of minima (PCNH)	Cohen's D	BH-adjusted p-value
Hgb	g/dL	SARS-CoV-2$^+$	102	9.18	272	11.17	−0.85	1.61×10^{-10}
Hct	%	SARS-CoV-2$^+$	102	28.98	272	34.03	−0.79	1.22×10^{-9}
eGFR	mL/min/BSA	SARS-CoV-2$^+$	83	45.70	188	57.56	−0.51	4.38×10^{-4}
Hgb	g/dL	Pre-COVID	55	10.33	136	11.65	−0.49	2.82×10^{-3}
eGFR	mL/min/BSA	Pre-COVID	47	44.72	109	56.68	−0.57	2.99×10^{-3}
Hct	%	Pre-COVID	55	32.26	132	35.82	−0.45	3.01×10^{-3}
BUN	mg/dL	SARS-CoV-2$^+$	102	16.01	272	14.48	0.14	1.04×10^{-1}
Cr	mg/dL	Pre-COVID	57	1.48	142	1.17	0.23	1.20×10^{-1}
Cr	mg/dL	SARS-CoV-2$^+$	103	1.10	272	0.97	0.12	2.54×10^{-1}
BUN	mg/dL	Pre-COVID	56	16.99	133	14.62	0.27	4.23×10^{-1}

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.

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

As males and females have different normal ranges of hemoglobin and hematocrit, we performed sex-split subanalyses of anemia-related laboratory tests similar to those described above. Patient-level median hemoglobin and hematocrit during the pre-COVID phase were significantly lower in both the female (Cohen's D = −0.66, p = 0.01; Cohen's D = −0.67, p = 0.01) and male (Cohen's D = −0.42, p = 0.02; Cohen's D = −0.37, p = 0.05) PCH cohorts versus their PCNH counterparts (Tables 5 and 6, Figures 4A–4D). These trends were even stronger during the SARS-CoV-2$^+$ phase among both the female (Cohen's D = −0.85; $p = 7.0 \times 10^{-6}$; Cohen's D = −0.91, $p = 7.0 \times 10^{-6}$) and male (Cohen's D = −0.60; $p = 8.2 \times 10^{-4}$; Cohen's D = −0.53, $p = 1.8 \times 10^{-3}$) cohorts (Tables 5 and 6, Figures 4E–4H).

Similarly, in our analysis of extreme values, minimum hemoglobin and hematocrit measurements during the pre-COVID phase tended to be lower in both female (Cohen's D = −0.60; p = 0.01; Cohen's D = −0.58, p = 0.01) and male (Cohen's D = −0.42; p = 0.05; Cohen's D = −0.38, p = 0.07) PCH patients (Table 7 and 8, Figures 4I–4L). These trends were again even stronger in the SARS-CoV-2$^+$ phase among both female (Cohen's D = −1.02; $p = 7.5 \times 10^{-7}$; Cohen's D = −0.97, $p = 1.8 \times 10^{-6}$) and male (Cohen's D = −0.74; $p = 1.1 \times 10^{-4}$; Cohen's D = −0.66, $p = 2.7 \times 10^{-4}$) patients (Tables 7 and 8, Figures 4M–4P).

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

We next evaluated whether outright anemia occurred more frequently in the PCH cohort than the PCNH cohort (see STAR Methods and Figure 5A). Anemia was indeed observed more frequently in the PCH cohort during both the pre-COVID phase (39/55 [71%] versus 61/136 [45%]; Odds Ratio [OR] = 2.98; $p = 1.32 \times 10^{-3}$) and the SARS-CoV-2$^+$ phase (93/102 [91%] versus 202/272 [74%]; OR = 3.57; $p = 2.02 \times 10^{-4}$) (Figures 5B and 5C). Furthermore, the prevalence of moderate or severe anemia was higher in the PCH cohorts during both the pre-COVID phase (13/55 [31%] versus 19/136 [14%]; OR = 1.90; p = 0.13) and the SARS-CoV-2$^+$ phase (63/102 [62%] versus 77/272 [28%]; OR = 4.07; $p = 4.99 \times 10^{-9}$), although the former observation did not reach statistical significance (Figures 5D and 5E). To determine whether this association is specific to COVID-19, we performed a similar analysis among patients who were hospitalized within 1 week of influenza diagnosis since 2003 (see STAR Methods). Anemia was indeed more frequently observed in the year prior to flu diagnosis and during the initial influenza-associated hospitalization among patients who were subsequently rehospitalized compared with those who were not, with the strongest trend observed for moderate to severe anemia during the flu-positive phase (49/127 [39%] versus 158/754 [21%]; OR = 2.37; $p = 3.81 \times 10^{-5}$) (Figure S3).

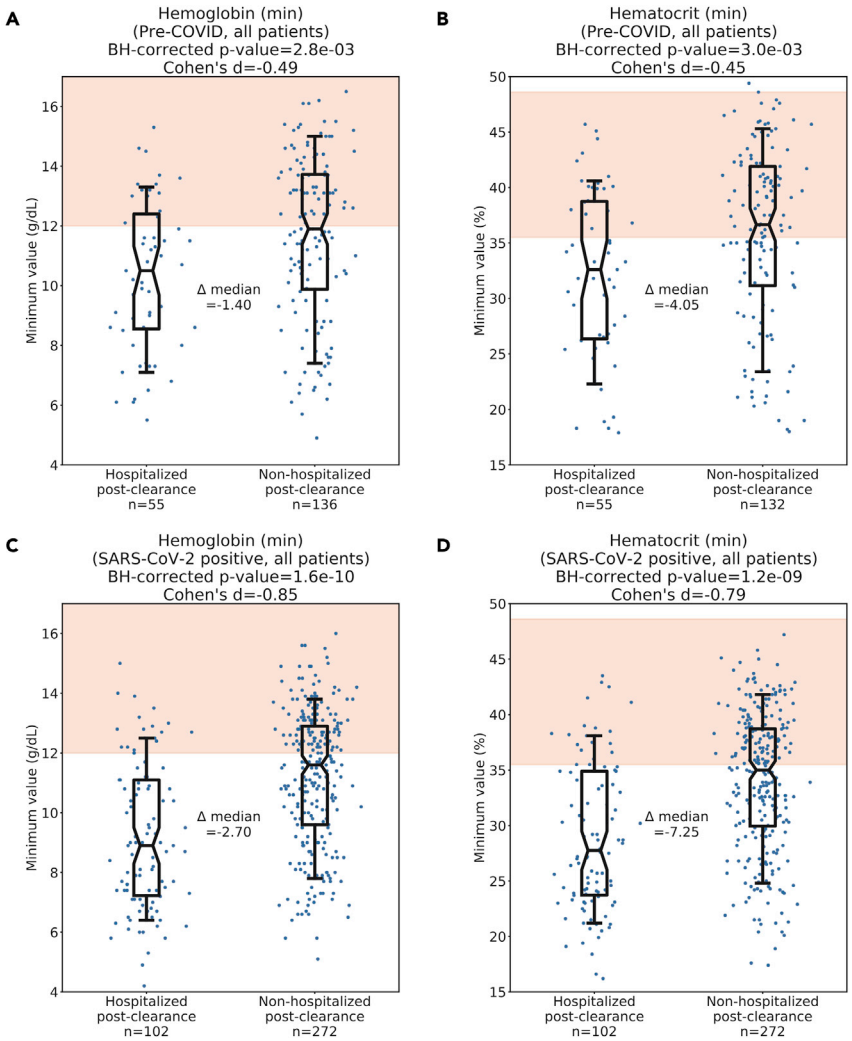


Figure 3. Comparison of minimum values for hemoglobin and hematocrit in the pre-COVID and SARS-CoV-2$^+$ intervals

(A) Pre-COVID minimum hemoglobin in the PCH (n = 55) and PCNH (n = 136) cohorts.

(B) Pre-COVID minimum hematocrit in the PCH (n = 55) and PCNH (n = 132) cohorts.

(C) SARS-CoV-2$^+$ minimum hemoglobin in the PCH (n = 102) and PCNH (n = 272) cohorts.

(D) SARS-CoV-2$^+$ minimum hematocrit in the PCH (n = 102) and PCNH (n = 272) 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.

Similarly, we assessed whether laboratory-diagnosed AKI occurs more frequently in the PCH cohort based on the creatinine-related components of the KDIGO (Kidney Disease: Improving Global Outcomes) criteria for diagnosis and staging of AKI in adults (Figure S4A) (Khwaja, 2012). Any stage AKI was indeed more common in the PCH cohort than the PCNH cohort in both the pre-COVID phase (30/69 [43%] versus 45/198 [23%]; OR = 2.61; $p = 1.69 \times 10^{-3}$) (Figure S4B) and the SARS-CoV-2$^+$ phase (46/103 [45%] versus 64/272 [24%]; OR = 2.62; $p = 1.19 \times 10^{-4}$) (Figure S4C). Furthermore, stage 2+ AKI was more common in the PCH cohort during both the pre-COVID phase (21/69 [30%] versus 21/198 [11%]; OR = 3.67; $p = 2.24 \times 10^{-4}$) and the SARS-CoV-2$^+$ phase (23/103 [22%] versus 24/272 [9%]; OR = 2.96; $p = 7.91 \times 10^{-4}$) (Figures S4D and S4E). It is intriguing that, when split by sex, we found that this trend was driven by males, as PCH males were more likely to experience stage 2+ AKI in both the pre-COVID phase (14/37 [38%] versus 13/109 [12%]; OR = 4.44; $p = 1.06 \times 10^{-3}$) and the SARS-CoV-2$^+$ phase (15/53 [28%] versus 14/158 [9%]; OR = 3.19; $p = 9.15 \times 10^{-4}$) (Figures S4F and S4G), whereas this was not true when comparing PCH and PCNH females (data not shown).

Table 4. Analysis of maximum values for all selected laboratory tests in pre-COVID and SARS-CoV-2$^+$ phases, including both male and female patients

Test	Units	Phase	N (PCH)	Mean of maxima (PCH)	N (PCNH)	Mean of maxima (PCNH)	Cohen's D	BH-adjusted p-value
BUN	mg/dL	SARS-CoV-2$^+$	102	45.21	272	30.04	0.60	5.23×10^{-8}
Hct	%	SARS-CoV-2$^+$	102	37.50	272	40.32	−0.48	1.50×10^{-4}
Cr	mg/dL	SARS-CoV-2$^+$	103	2.22	272	1.56	0.30	1.50×10^{-4}
BUN	mg/dL	Pre-COVID	56	40.23	133	27.77	0.50	6.59×10^{-4}
Cr	mg/dL	Pre-COVID	57	2.58	142	1.67	0.37	1.04×10^{-3}
Hgb	g/dL	SARS-CoV-2$^+$	102	13.40	272	13.75	−0.10	3.85×10^{-3}
Hgb	g/dL	Pre-COVID	55	12.97	136	13.53	−0.30	1.78×10^{-2}
Hct	%	Pre-COVID	55	40.05	132	41.36	−0.24	6.26×10^{-2}
eGFR	mL/min/BSA	SARS-CoV-2$^+$	83	72.51	188	80.14	−0.41	9.20×10^{-2}
eGFR	mL/min/BSA	Pre-COVID	47	67.13	109	74.68	−0.38	2.42×10^{-1}

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.

Anemia is strongly associated with post-clearance rehospitalization independent of intensive care unit admission status and other covariates

To test whether the previous observations were affected by potential confounding demographic or clinical covariates, we performed a series of logistic regression analyses. First, we evaluated the association between post-clearance rehospitalization and the following independent variables during the pre-COVID and SARS-CoV-2$^+$ phases, separately: minimum hemoglobin, maximum BUN, sex, age, number of blood draws, and intensive care unit (ICU) admission status (for the SARS-CoV-2$^+$ phase only) (Table 9). The consideration of ICU admission status as a covariate was particularly important because ICU admission was more common among PCH than PCNH patients (Table 1). Although none of these variables in the pre-COVID phase showed a significant association with rehospitalization status, minimum hemoglobin during the SARS-CoV-2$^+$ phase was singularly associated with rehospitalization ($\beta = -0.29$, $p = 4.2 \times 10^{-5}$).

We then modified our logistic regression analysis by replacing the minimum hemoglobin and maximum BUN terms with binary labels of moderate/severe anemia and stage 2+ AKI, respectively (Table 10). Among the tested pre-COVID variables, only stage 2+ AKI was modestly associated with rehospitalization status ($\beta = 0.93$, $p = 0.09$). ICU admission during the SARS-CoV-2$^+$ phase was also modestly associated with post-clearance rehospitalization ($\beta = 0.70$, $p = 0.06$), suggesting that patients with more severe courses

Table 5. Sex-split analysis of median values for all selected laboratory tests in female patients during the pre-COVID and SARS-CoV-2$^+$ phases

Test	Units	Phase	N (PCH)	Mean of medians (PCH)	N (PCNH)	Mean of medians (PCNH)	Cohen's D	BH-adjusted p-value
Hgb	g/dL	SARS-CoV-2$^+$	49	10.21	114	11.66	−0.85	7.04×10^{-6}
Hct	%	SARS-CoV-2$^+$	49	31.16	114	35.67	−0.91	7.04×10^{-6}
BUN	mg/dL	SARS-CoV-2$^+$	49	22.26	114	17.81	0.36	1.76×10^{-3}
Hgb	g/dL	Pre-COVID	24	11.24	63	12.42	−0.66	6.63×10^{-3}
Hct	%	Pre-COVID	24	34.76	60	38.33	−0.67	7.62×10^{-3}
eGFR	mL/min/BSA	SARS-CoV-2$^+$	36	64.01	77	70.75	−0.36	4.67×10^{-2}
Cr	mg/dL	Pre-COVID	25	1.31	64	1.20	0.06	5.31×10^{-2}
eGFR	mL/min/BSA	Pre-COVID	23	60.83	45	70.19	−0.52	1.08×10^{-1}
Cr	mg/dL	SARS-CoV-2$^+$	50	1.25	114	1.05	0.15	1.34×10^{-1}
BUN	mg/dL	Pre-COVID	24	21.77	61	17.34	0.37	1.52×10^{-1}

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.

Table 6. Sex-split analysis of median values for all selected laboratory tests in male patients during the pre-COVID and SARS-CoV-2$^+$ phases

Test	Units	Phase	N (PCH)	Mean of medians (PCH)	N (PCNH)	Mean of medians (PCNH)	Cohen's D	BH-adjusted p-value
BUN	mg/dL	SARS-CoV-2$^+$	53	31.36	158	23.05	0.55	5.40×10^{-4}
Hgb	g/dL	SARS-CoV-2$^+$	53	11.15	158	12.47	−0.60	8.20×10^{-4}
Cr	mg/dL	SARS-CoV-2$^+$	53	1.75	158	1.30	0.28	1.75×10^{-3}
Hct	%	SARS-CoV-2$^+$	53	34.03	158	37.29	−0.53	1.79×10^{-3}
eGFR	mL/min/BSA	SARS-CoV-2$^+$	47	57.89	111	69.00	−0.53	6.05×10^{-3}
Hgb	g/dL	Pre-COVID	31	11.70	73	12.75	−0.42	1.88×10^{-2}
BUN	mg/dL	Pre-COVID	32	28.70	72	21.89	0.52	1.88×10^{-2}
Hct	%	Pre-COVID	31	36.16	72	38.83	−0.37	5.19×10^{-2}
eGFR	mL/min/BSA	Pre-COVID	24	52.65	64	62.91	−0.49	6.35×10^{-2}
Cr	mg/dL	Pre-COVID	32	2.44	78	1.52	0.48	7.43×10^{-2}

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.

of initial COVID-19 illness are more likely to experience subsequent hospitalization. However, moderate/severe anemia during the SARS-CoV-2$^+$ phase was again the most strongly associated variable with rehospitalization ($\beta = 1.16$, $p = 9.0 \times 10^{-5}$).

Finally, we performed a split cohort subanalysis to specifically test whether the robust association between anemia and rehospitalization status applied to both patients who were and were not admitted to the ICU during their index infection. Indeed, the rate of moderate/severe anemia was significantly higher in PCH patients than in PCNH patients when considering only those who were admitted to the ICU (38/48 [79%] versus 26/69 [38%]; OR = 6.18, $p = 1.09 \times 10^{-5}$) or only those who were not admitted to the ICU (25/54 [46%] versus 51/203 [25%]; OR = 2.56, $p = 4.02 \times 10^{-3}$) (Figures S5A and S5B).

DISCUSSION

Approximately 18 months after the first confirmed case, the COVID-19 pandemic continues to impact communities across the globe. Although 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 (Pawlowski et al., 2021), 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.

This laboratory test analysis shows that anemia and renal function in the pre-COVID and SARS-CoV-2$^+$ phases are associated with the risk of post-viral-clearance rehospitalization. Our logistic regression analyses suggest that AKI and static renal laboratory measurements are not independently associated with rehospitalization, with ICU admission during COVID-19 infection representing a likely confounding factor contributing to the observed trends. Indeed, AKI has previously been reported as a common morbidity in ICU patients (Case et al., 2013; Girling et al., 2020; Mohsenin, 2017) and was observed frequently among ICU-admitted patients with COVID-19 in our cohort, with 78% (94/121) and 39% (47/121) of ICU-admitted patients experiencing stage 1+ and stage 2+ AKI, respectively. On the other hand, hemoglobin levels and the outright diagnosis of moderate or severe anemia are robustly associated with post-clearance rehospitalization independent of sex, age, number of blood draws, and ICU admission status.

Although the pathophysiologic foundations for these associations are not clear, the findings do merit consideration in the context of COVID-19 clinical care. Indeed, pre-existing conditions are already integrated in the clinical decision-making algorithms around COVID-19, as the Centers for Disease Control and Prevention (CDC) has designated various chronic conditions as risk factors for severe COVID-19

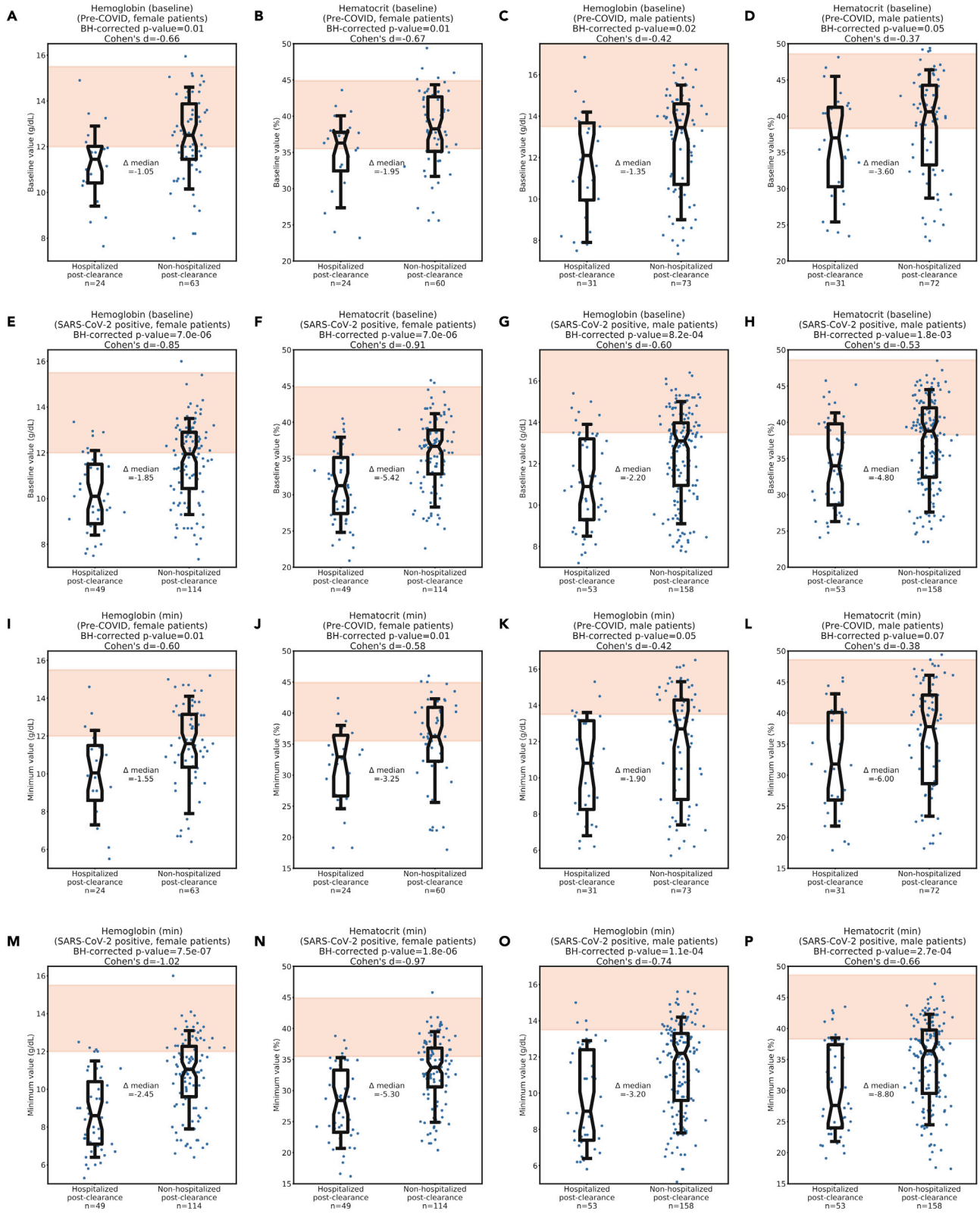

A
Hemoglobin (baseline)
(Pre-COVID, female patients)
BH-corrected p-value=0.01
Cohen's d=-0.66
Baseline value (g/dL)
Δ median
=-1.05
Hospitalized post-clearance n=24
Non-hospitalized post-clearance n=63
B
Hematocrit (baseline)
(Pre-COVID, female patients)
BH-corrected p-value=0.01
Cohen's d=-0.67
Baseline value (%)
Δ median
=-1.95
Hospitalized post-clearance n=24
Non-hospitalized post-clearance n=60
C
Hemoglobin (baseline)
(Pre-COVID, male patients)
BH-corrected p-value=0.02
Cohen's d=-0.42
Baseline value (g/dL)
Δ median
=-1.35
Hospitalized post-clearance n=31
Non-hospitalized post-clearance n=73
D
Hematocrit (baseline)
(Pre-COVID, male patients)
BH-corrected p-value=0.05
Cohen's d=-0.37
Baseline value (%)
Δ median
=-3.60
Hospitalized post-clearance n=31
Non-hospitalized post-clearance n=72
E
Hemoglobin (baseline)
(SARS-CoV-2 positive, female patients)
BH-corrected p-value=7.0e-06
Cohen's d=-0.85
Baseline value (g/dL)
Δ median
=-1.85
Hospitalized post-clearance n=49
Non-hospitalized post-clearance n=114
F
Hematocrit (baseline)
(SARS-CoV-2 positive, female patients)
BH-corrected p-value=7.0e-06
Cohen's d=-0.91
Baseline value (%)
Δ median
=-5.42
Hospitalized post-clearance n=49
Non-hospitalized post-clearance n=114
G
Hemoglobin (baseline)
(SARS-CoV-2 positive, male patients)
BH-corrected p-value=8.2e-04
Cohen's d=-0.60
Baseline value (g/dL)
Δ median
=-2.20
Hospitalized post-clearance n=53
Non-hospitalized post-clearance n=158
H
Hematocrit (baseline)
(SARS-CoV-2 positive, male patients)
BH-corrected p-value=1.8e-03
Cohen's d=-0.53
Baseline value (%)
Δ median
=-4.80
Hospitalized post-clearance n=53
Non-hospitalized post-clearance n=158
I
Hemoglobin (min)
(Pre-COVID, female patients)
BH-corrected p-value=0.01
Cohen's d=-0.60
Minimum value (g/dL)
Δ median
=-1.55
Hospitalized post-clearance n=24
Non-hospitalized post-clearance n=63
J
Hematocrit (min)
(Pre-COVID, female patients)
BH-corrected p-value=0.01
Cohen's d=-0.58
Minimum value (%)
Δ median
=-3.25
Hospitalized post-clearance n=24
Non-hospitalized post-clearance n=60
K
Hemoglobin (min)
(Pre-COVID, male patients)
BH-corrected p-value=0.05
Cohen's d=-0.42
Minimum value (g/dL)
Δ median
=-1.90
Hospitalized post-clearance n=31
Non-hospitalized post-clearance n=73
L
Hematocrit (min)
(Pre-COVID, male patients)
BH-corrected p-value=0.07
Cohen's d=-0.38
Minimum value (%)
Δ median
=-6.00
Hospitalized post-clearance n=31
Non-hospitalized post-clearance n=72
M
Hemoglobin (min)
(SARS-CoV-2 positive, female patients)
BH-corrected p-value=7.5e-07
Cohen's d=-1.02
Minimum value (g/dL)
Δ median
=-2.45
Hospitalized post-clearance n=49
Non-hospitalized post-clearance n=114
N
Hematocrit (min)
(SARS-CoV-2 positive, female patients)
BH-corrected p-value=1.8e-06
Cohen's d=-0.97
Minimum value (%)
Δ median
=-5.30
Hospitalized post-clearance n=49
Non-hospitalized post-clearance n=114
O
Hemoglobin (min)
(SARS-CoV-2 positive, male patients)
BH-corrected p-value=1.1e-04
Cohen's d=-0.74
Minimum value (g/dL)
Δ median
=-3.20
Hospitalized post-clearance n=53
Non-hospitalized post-clearance n=158
P
Hematocrit (min)
(SARS-CoV-2 positive, male patients)
BH-corrected p-value=2.7e-04
Cohen's d=-0.66
Minimum value (%)
Δ median
=-8.80
Hospitalized post-clearance n=53
Non-hospitalized post-clearance n=158

Figure 4. Sex-split analysis of anemia-related laboratory tests in the pre-COVID and SARS-CoV-2$^+$ phases

(A–H) Median values of hemoglobin and hematocrit in the pre-COVID phase (A–D) and SARS-CoV-2$^+$ phase (E–H), split to show female patients (A and B, E and F) and male patients (C and D, G and H) separately.

(I–P) Minimum values of hemoglobin and hematocrit in the pre-COVID phase (I–L) and SARS-CoV-2$^+$ phase (M–P), split to show female patients (I and J, M and N) and male patients (K and L, O and 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.

infection (e.g., cancer, chronic kidney disease, chronic obstructive pulmonary disease, and cardiovascular diseases such as heart disease, obesity, and diabetes) (CDC, 2020). 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 patients with COVID-19.

Our finding that lower hemoglobin levels and the outright diagnosis of moderate or severe anemia is associated with post-viral-clearance rehospitalization has not been previously reported. Although this analysis certainly does not establish a causal role for anemia in post-clearance hospitalization, the robust association warrants further studies to determine whether anemia mitigating therapies (e.g., vitamin or mineral supplementation, erythropoietin administration, or blood transfusion) engender long-term benefits in select patients with COVID-19. Furthermore, it is interesting that fatigue has been commonly reported as both an acute symptom and a lasting effect of COVID-19 (Pascarella et al., 2020; Townsend et al., 2020; Wagner et al., 2020), but the mechanisms underlying this phenotype have not been established. Of the 374 hospitalized patients with COVID-19 in this study, 295 (79%) had at least mild anemia during their SARS-CoV-2$^+$ phase, and 140 of the 374 patients (37%) had 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.

Along with our previous analysis (Pawlowski et al., 2021), 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 exemplify the value of leveraging augmented curation methods to first identify phenotypes that distinguish defined clinical cohorts and then cross-checking these phenotypic associations through a hypothesis-driven analysis of related laboratory 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.

Table 7. Sex-split analysis of minimum values for all selected laboratory tests in female patients during the pre-COVID and SARS-CoV-2$^+$ phases

Test	Units	Phase	N (PCH)	Mean of minima (PCH)	N (PCNH)	Mean of minima (PCNH)	Cohen's D	BH-adjusted p-value
Hgb	g/dL	SARS-CoV-2$^+$	49	8.78	114	10.82	−1.02	7.50×10^{-7}
Hct	%	SARS-CoV-2$^+$	49	27.73	114	33.14	−0.97	1.76×10^{-6}
Hgb	g/dL	Pre-COVID	24	10.02	63	11.44	−0.60	1.37×10^{-2}
Hct	%	Pre-COVID	24	31.57	60	35.50	−0.58	1.37×10^{-2}
eGFR	mL/min/BSA	Pre-COVID	23	47.91	45	61.53	−0.72	1.37×10^{-2}
eGFR	mL/min/BSA	SARS-CoV-2$^+$	36	49.00	77	58.17	−0.41	3.79×10^{-2}
BUN	mg/dL	SARS-CoV-2$^+$	49	13.67	114	12.29	0.14	1.68×10^{-1}
Cr	mg/dL	Pre-COVID	25	0.98	64	1.04	−0.05	2.75×10^{-1}
BUN	mg/dL	Pre-COVID	24	13.17	61	12.46	0.10	3.53×10^{-1}
Cr	mg/dL	SARS-CoV-2$^+$	50	0.93	114	0.84	0.10	4.26×10^{-1}

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.

Table 8. Sex-split analysis of minimum values for all selected laboratory tests in male patients during the pre-COVID and SARS-CoV-2$^+$ phases

Test	Units	Phase	N (PCH)	Mean of minima (PCH)	N (PCNH)	Mean of minima (PCNH)	Cohen's D	BH-adjusted p-value
Hgb	g/dL	SARS-CoV-2$^+$	53	9.55	158	11.43	−0.74	1.10×10^{-4}
Hct	%	SARS-CoV-2$^+$	53	30.14	158	34.68	−0.66	2.70×10^{-4}
eGFR	mL/min/BSA	SARS-CoV-2$^+$	47	43.17	111	57.14	−0.59	2.50×10^{-3}
Hgb	g/dL	Pre-COVID	31	10.57	73	11.82	−0.42	5.26×10^{-2}
eGFR	mL/min/BSA	Pre-COVID	24	41.67	64	53.27	−0.52	5.89×10^{-2}
Hct	%	Pre-COVID	31	32.80	72	36.09	−0.38	6.51×10^{-2}
Cr	mg/dL	SARS-CoV-2$^+$	53	1.26	158	1.06	0.16	7.68×10^{-2}
Cr	mg/dL	Pre-COVID	32	1.87	78	1.28	0.43	1.41×10^{-1}
BUN	mg/dL	SARS-CoV-2$^+$	53	18.18	158	16.07	0.19	1.43×10^{-1}
BUN	mg/dL	Pre-COVID	32	19.85	72	16.45	0.35	2.79×10^{-1}

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.

Limitations of the study

This study has a few important limitations to consider. First, this analysis considers patients within only one EHR system; although this system does contain patients from multiple sites of clinical care in distinct

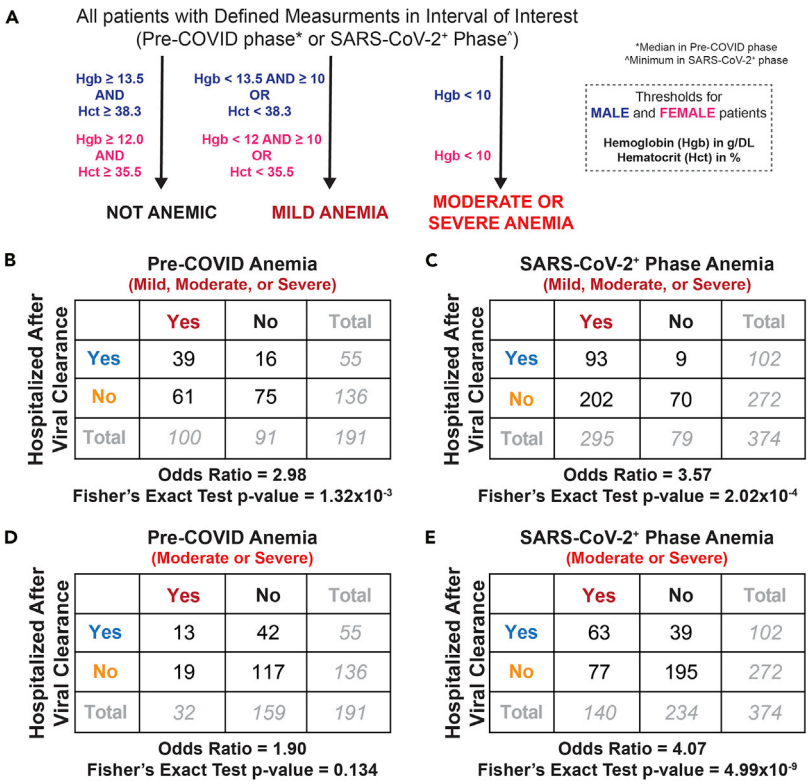


Figure 5. Comparison of outright anemia prevalence in the PCH and PCNH cohorts

(A) Schematic illustrating how patients are classified as having no anemia, mild anemia, or severe/moderate anemia.

(B and C) Comparison of mild, moderate, or severe anemia frequency in the PCH and PCNH cohorts during the pre-COVID (n = 191) and SARS-CoV-2$^+$ (n = 374) phases.

(D and E) Comparison of moderate or severe anemia frequency in the PCH and PCNH cohorts during the pre-COVID (n = 191) and SARS-CoV-2$^+$ (n = 374) phases. Contingency tables show the counts of patients in each intersecting category. Below each contingency table, the associated odds ratio and Fisher's exact test p value is shown.

Table 9. Logistic regression analyses to assess the association between post-viral-clearance hospitalization and minimum hemoglobin, maximum BUN, or potential confounding variables during the pre-COVID and SARS-CoV-2$^+$ phases

	Minimum hemoglobin	Maximum BUN	Sex	Age	Blood draw count during interval	ICU admission during interval
Pre-COVID	β = −0.108	β = 0.011	β = 0.086	β = 0.006	β = −0.013	NA
Phase	p = 0.578	p = 0.560	p = 1.00	p = 0.982	p = 0.878	
SARS-CoV-2$^+$	β = −0.289	β = 0.006	β = −0.173	β = −0.003	β = −0.003	β = 0.590
Phase	**p = 4.2x10^{-5}**	p = 0.852	p = 0.986	p = 1.00	p = 0.948	p = 0.188

Confounding variables considered include sex, age, the number of blood draws in the given interval, and ICU admission status during the given interval. For each regression (row), the coefficient (β) and associated Bonferroni-adjusted p value (p) are shown for each independent variable (column) assessed. The coefficient represents the log-odds ratio. p Values were calculated using the log likelihood ratio test and adjusted using the Bonferroni correction. 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: sex: 0 = female, 1 = male; ICU admission during interval: 0 = not admitted to ICU, 1 = admitted to ICU.

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 would prevent the studied cohort and their associated data points from serving as true representative samples of all patients with COVID-19. Second, the analyzed cohort was relatively small (n = 382) as most patients diagnosed with COVID-19 do not subsequently receive two confirmatory negative PCR tests, and only a subset of these 382 patients possessed data for each laboratory test 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.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human subjects

Table 10. Logistic regression analyses to assess the association between post-viral-clearance hospitalization and the diagnosis of moderate to severe anemia, the diagnosis of stage 2+ AKI, or potential confounding variables during the pre-COVID and SARS-CoV-2$^+$ phases

	Moderate/ Severe anemia	Stage 2+ AKI	Sex	Age	Blood draw count during interval	ICU admission during interval
Pre-COVID	β = 0.316	β = 0.934	β = 0.011	β = 0.008	β = −0.007	NA
Phase	p = 0.956	p = 0.085	p = 1.00	p = 0.914	p = 0.989	
SARS-CoV-2$^+$	β = 1.16	β = 0.453	β = −0.250	β = −0.001	β = −0.001	β = 0.697
Phase	**p = 9.00x10^{-5}**	p = 0.739	p = 0.900	p = 1.00	p = 1.00	p = 0.060

Confounding variables considered include sex, age, the number of blood draws in the given interval, and ICU admission status during the given interval. For each regression (row), the coefficient (β) and associated Bonferroni-adjusted p value (p) are shown for each independent variable (column) assessed. The coefficient represents the log-odds ratio. p Values were calculated using the log likelihood ratio test and adjusted using the Bonferroni correction. 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.

- METHOD DETAILS
  - Study design
  - Selection of study participants
  - Definition of the considered time intervals: pre-COVID phase and SARS-CoV-2^{+} phase
  - Selection and summarization of laboratory measurements
  - Consideration of potential confounding variables
  - Classification of patients using clinical diagnostic criteria for anemia and AKI
  - Quantification of number of blood draws per patient
  - Assessment of anemia in the context of rehospitalization of influenza patients
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Comparison of lab test results and clinical diagnoses between PCH and PCNH cohorts
  - Logistic regression analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102780.

ACKNOWLEDGMENTS

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. This study was funded by nference.

AUTHOR CONTRIBUTIONS

P.J.L., E.R., A.J.V., R.M., A.D.B., and V.S. contributed to the study design and methodology. E.R., G.B., and J.C.O. were responsible for data curation. E.R. performed the formal statistical analyses, which were reviewed by P.J.L., A.J.V., R.M., and V.S. P.J.L. drafted the manuscript with inputs from E.R. A.J.V., R.M., J.C.O., A.D.B., W.M., J.H., and V.S. provided critical revisions, which were incorporated into the final manuscript by P.J.L. All authors approved the submitted version of the manuscript.

DECLARATION OF INTERESTS

P.J.L., E.R., A.J.V., G.B., R.M., and V.S. are employees of nference and have financial interests in the company. J.C.O. receives personal fees from Elsevier and Bates College, and receives small grants from nference, Inc, outside the submitted work. A.D.B. is a consultant for Abbvie, is on scientific advisory boards for nference and Zentalis, and is founder and President of Splissen therapeutics. J.C.O., W.M., J.H., and A.D.B. are employees of the Mayo Clinic. The Mayo Clinic may stand to gain financially from the successful outcome of the research.

Received: January 19, 2021
Revised: June 1, 2021
Accepted: June 22, 2021
Published: July 23, 2021

REFERENCES

Anand, P., Puranik, A., Aravamudan, M., Venkatakrishnan, A.J., and Soundararajan, V. (2020). SARS-CoV-2 strategically mimics proteolytic activation of human ENaC. eLife. https://doi.org/10.7554/eLife.58603.

Bos, R., Rutten, L., van der Lubbe, J.E.M., Bakkers, M.J.G., Hardenberg, G., Wegmann, F., Zuijdgeest, D., de Wilde, A.H., Koornneef, A., Verwilligen, A., et al. (2020). Ad26 vector-based COVID-19 vaccine encoding a prefusion-stabilized SARS-CoV-2 Spike immunogen induces potent humoral and cellular immune responses. NPJ Vaccin. *5*, 91.

Carfì, A., Bernabei, R., and Landi, F.; for the Gemelli Against COVID-19 Post-Acute Care Study Group (2020). Persistent symptoms in patients after acute COVID-19. JAMA *324*, 603–605.

Case, J., Khan, S., Khalid, R., and Khan, A. (2013). Epidemiology of acute kidney injury in the intensive care unit. Crit. Care Res. Pract. *2013*, 479730.

CDC (2020). COVID-19 and your health [WWW document]. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html.

Chan, L., Chaudhary, K., Saha, A., Chauhan, K., Vaid, A., Zhao, S., Paranjpe, I., Somani, S., Richter, F., Miotto, R., et al.; Mount Sinai COVID Informatics Center (MSCIC) (2021). AKI in hospitalized patients with COVID-19. J. Am. Soc. Nephrol. *32*, 151–160.

Corbett, K.S., Flynn, B., Foulds, K.E., Francica, J.R., Boyoglu-Barnum, S., Werner, A.P., Flach, B., O'Connell, S., Bock, K.W., Minai, M., et al. (2020). Evaluation of the mRNA-1273 vaccine against SARS-CoV-2 in nonhuman primates. N. Engl. J. Med. *383*, 1544–1555.

del Rio, C., Collins, L.F., and Malani, P. (2020). Long-term health consequences of COVID-19. JAMA *324*, 1723–1724.

Faghih Dinevari, M., Somi, M.H., Sadeghi Majd, E., Abbasalizad Farhangi, M., and Nikniaz, Z. (2021). Anemia predicts poor outcomes of

COVID-19 in hospitalized patients: a prospective study in Iran. BMC Infect. Dis. *21*, 170.

Folegatti, P.M., Ewer, K.J., Aley, P.K., Angus, B., Becker, S., Belij-Rammerstorfer, S., Bellamy, D., Bibi, S., Bittaye, M., Clutterbuck, E.A., , et al.Oxford COVID Vaccine Trial Group. (2020). 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 *396*, 467–478.

Girling, B.J., Channon, S.W., Haines, R.W., and Prowle, J.R. (2020). Acute kidney injury and adverse outcomes of critical illness: correlation or causation? Clin. Kidney J. *13*, 133–141.

Hariyanto, T.I., and Kurniawan, A. (2020). Anemia is associated with severe coronavirus disease 2019 (COVID-19) infection. Transfus. Apher. Sci. *59*, 102926.

Jackson, L.A., Anderson, E.J., Rouphael, N.G., Roberts, P.C., Makhene, M., Coler, R.N., McCullough, M.P., Chappell, J.D., Denison, M.R., Stevens, L.J., et al. (2020). An mRNA vaccine against SARS-CoV-2 - preliminary report. N. Engl. J. Med. *383*. https://doi.org/10.1056/NEJMoa2022483.

Johns Hopkins coronavirus resource center [WWW document], n.d. URL https://coronavirus.jhu.edu/map.html (accessed 5.31.21).

Khwaja, A. (2012). KDIGO clinical practice guidelines for acute kidney injury. NEC *120*, c179–c184.

Makris, K., and Spanou, L. (2016). Acute kidney injury: definition, pathophysiology and clinical phenotypes. Clin. Biochem. Rev. *37*, 85–98.

Mercado, N.B., Zahn, R., Wegmann, F., Loos, C., Chandrashekar, A., Yu, J., Liu, J., Peter, L., McMahan, K., Tostanoski, L.H., et al. (2020). Single-shot Ad26 vaccine protects against SARS-CoV-2 in rhesus macaques. Nature *586*, 583–588.

Mohsenin, V. (2017). Practical approach to detection and management of acute kidney injury in critically ill patient. J. Intensive Care Med. *5*, 1–8.

Mulligan, M.J., Lyke, K.E., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Neuzil, K., Raabe, V., Bailey, R., Swanson, K.A., et al. (2020). Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. Nature *586*, 589–593.

Nadim, M.K., Forni, L.G., Mehta, R.L., Connor, M.J., Jr., Liu, K.D., Ostermann, M., Rimmelé, T., Zarbock, A., Bell, S., Bihorac, A., et al. (2020). COVID-19-associated acute kidney injury: consensus report of the 25th Acute Disease Quality Initiative (ADQI) Workgroup. Nat. Rev. Nephrol. *16*, 747–764.

Oh, S.M., Skendelas, J.P., Macdonald, E., Bergamini, M., Goel, S., Choi, J., Segal, K.R., Vivek, K., Nair, S., and Leff, J. (2021). On-admission anemia predicts mortality in COVID-19 patients: a single center, retrospective cohort study. Am. J. Emerg. Med. *48*, 140–147.

Pascarella, G., Strumia, A., Piliego, C., Bruno, F., Del Buono, R., Costa, F., Scarlata, S., and Agrò, F.E. (2020). COVID-19 diagnosis and management: a comprehensive review. J. Intern. Med. *288*. https://doi.org/10.1111/joim.13091.

Pawlowski, C., Venkatakrishnan, A.J., Ramudu, E., Kirkup, C., Puranik, A., Kayal, N., Berner, G., Anand, A., Barve, R., O'Horo, J.C., et al. (2021). Pre-existing conditions are associated with COVID-19 patients' hospitalization, despite confirmed clearance of SARS-CoV-2 virus. EClinicalMedicine *34*, 100793.

Roubinian, N.H., Murphy, E.L., Mark, D.G., Triulzi, D.J., Carson, J.L., Lee, C., Kipnis, P., Kleinman, S., Liu, V.X., and Escobar, G.J. (2019). Long-term outcomes among patients discharged from the hospital with moderate anemia: a retrospective cohort study. Ann. Intern. Med. *170*, 81–89.

Seabold, S., and Perktold, J. (2010). Statsmodels: econometric and statistical modeling with Python. Proceedings of the 9th Python in science conference. https://doi.org/10.25080/majora-92bf1922-011.

Singh, M., Bansal, V., and Feschotte, C. (2020). A single-cell RNA expression map of human coronavirus entry factors. Cell Rep. *32*, 108175.

Townsend, L., Dyer, A.H., Jones, K., Dunne, J., Mooney, A., Gaffney, F., O'Connor, L., Leavy, D., O'Brien, K., Dowds, J., et al. (2020). Persistent fatigue following SARS-CoV-2 infection is common and independent of severity of initial infection. PLoS One *15*, e0240784.

Turner, J., Parsi, M., and Badireddy, M. (2021). Anemia. In StatPearls (Treasure Island, FL: StatPearls Publishing), https://www.ncbi.nlm.nih.gov/books/NBK499994/.

Venkatakrishnan, A.J., Puranik, A., Anand, A., Zemmour, D., Yao, X., Wu, X., Chilaka, R., Murakowski, D.K., Standish, K., Raghunathan, B., et al. (2020). Knowledge synthesis of 100 million biomedical documents augments the deep expression profiling of coronavirus receptors. eLife. https://doi.org/10.7554/eLife.58040.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al.; SciPy 1.0 Contributors (2020). Author Correction: SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 352.

Wagner, T., Shweta, F.N.U., Murugadoss, K., Awasthi, S., Venkatakrishnan, A.J., Bade, S., Puranik, A., Kang, M., Pickering, B.W., O'Horo, J.C., et al. (2020). Augmented curation of clinical notes from a massive EHR system reveals symptoms of impending COVID-19 diagnosis. eLife. https://doi.org/10.7554/eLife.58227.

Walsh, T.S., Saleh, E.-E.-D., Lee, R.J., and McClelland, D.B. (2006). The prevalence and characteristics of anaemia at discharge home after intensive care. Intensive Care Med. *32*, 1206–1213.

Warner, M.A., Hanson, A.C., Frank, R.D., Schulte, P.J., Go, R.S., Storlie, C.B., and Kor, D.J. (2020). Prevalence of and recovery from anemia following hospitalization for critical illness among adults. JAMA Netw. Open *3*, e2017843.

Yelin, D., Wirtheim, E., Vetter, P., Kalil, A.C., Bruchfeld, J., Runold, M., Guaraldi, G., Mussini, C., Gudiol, C., Pujol, M., et al. (2020). Long-term consequences of COVID-19: research needs. Lancet Infect. Dis. *20*, 1115.

Zhao, Y., Zhao, Z., Wang, Y., Zhou, Y., Ma, Y., and Zuo, W. (2020). Single-cell RNA expression profiling of ACE2, the receptor of SARS-CoV-2. Am. J. Respir. Crit. Care Med. *202*, 756.

Ziegler, C.G.K., Allon, S.J., Nyquist, S.K., Mbano, I.M., Miao, V.N., Tzouanas, C.N., Cao, Y., Yousif, A.S., Bals, J., Hauser, B.M., et al. (2020). 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 *181*. https://doi.org/10.1016/j.cell.2020.04.035.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Python version 3.5	https://www.python.org/downloads/release/python-350/	N/A
Python software package: SciPy version 1.6.3	https://www.scipy.org/scipylib/download.html	N/A
Python software package: statsmodels version 0.10.0	https://www.statsmodels.org	N/A
R version 4.0.3	https://www.r-project.org/	N/A
R software package: stats version 4.0.3	https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html	N/A
Python and R scripts for all statistical analyses described in manuscript, including comparison of individual lab test measurements between PCH and PCNH cohorts as well as logistic regression analyses	This study	Data S1

RESOURCE AVAILABILITY

Lead contact

Further information and requests for information should be directed to and will be fulfilled by the lead contact, Venky Soundararajan (venky@nference.net).

Materials availability

This study did not generate new reagents.

Data and code availability

- Data: The data supporting this study has not been deposited in a public repository because it contains personally identifiable information from human subjects which are protected by national privacy regulations, but a de-identified version of this data may be made available from the lead contact on request. A proposal with detailed description of study objectives and statistical analysis plan will be needed for evaluation of the reasonability of requests. Deidentified data will be provided after approval from the lead contact and the Mayo Clinic's standard IRB process for such requests.
- Code: Original code from this analysis is available in Data S1.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human subjects

The total cohort included 382 individuals. Each individual was part of one of the following two cohorts, on the basis of whether they were rehospitalized after PCR-confirmed clearance of SARS-CoV-2: (1) post clearance hospitalized ("PCH"; n=104), or (2) post clearance non-hospitalized ("PCNH"; n=278). More details describing the participant selection algorithm are provided in the Method details and are illustrated in Figure 1. Demographic and clinical characteristics of the analyzed cohorts (including age, sex, race, ethnicity, time to PCR-confirmed SARS-CoV-2 clearance, and ICU admission status during the index COVID-19 infection) are provided in Table 1.

This 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. The IRB approved was titled: Study of COVID-19 patient characteristics with augmented curation of Electronic Health Records (EHR) to inform strategic and operational decisions with the Mayo Clinic. The study was deemed exempt by the Mayo Clinic Institutional Review Board and waived from consent. The following resource provides further information on the Mayo Clinic Institutional Review Board and adherence to basic

ethical principles underlying the conduct of research, and ensuring that the rights and well-being of potential research subjects are adequately protected (https://www.mayo.edu/research/institutional-review-board/overview).

METHOD DETAILS

Study design

This was a case-control study. The primary outcome was rehospitalization status within 30 days of PCR-confirmed SARS-CoV-2 clearance. The exposure variables were anemia and kidney dysfunction as assessed through selected laboratory measurements detailed below.

Selection of study participants

Cases and controls were selected from a cohort of 66,689 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 December 12, 2020 (see Figure 1B). Post clearance hospitalized ("PCH") cases (n=104) 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 30 days of clearance. Post clearance non-hospitalized ("PCNH") controls (n=278) 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 30 days of clearance. Demographic and clinical features of the PCH and PCNH cohorts are summarized in Table 1.

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." COVID-19 diagnosis 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. The estimated viral clearance date was taken as the date of the first negative PCR test in this sequence of two consecutive negative tests.

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). The majority of eGFR measurements (~96%) were estimated by creatinine; these tests had a maximum recorded value of 90 mL/min/BSA, which corresponds to the lower limit of normal. The remaining 4% of eGFR measurements were estimated by cystatin C levels; for these tests, a value above 90 mL/min/BSA was possible and was indeed recorded in 5 of 31 cases.

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). Histograms showing the number of measurements per patient in each time period for the selected tests are shown in Figures S6 and S7. 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 patient-level maximum values of serum creatinine and BUN in each time period.

Consideration 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 the time to the first negative SARS-CoV-2 PCR test in a series of two consecutive negative tests after the last positive test), race, ethnicity, or sex. However, we did note that a higher fraction of PCNH cases were male as compared with PCH counterparts (58% versus 52%). This potential confounding factor was addressed by performing (1) sex-split subgroup analyses (see Tables 5, 6, 7, and 8) and (2) multivariate logistic regression (see Tables 9 and 10 and *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 with the PCNH cohort (48/104 [46%] versus 73/278 [26%], $p=4.0x10^{-3}$; Table 1). Further, patients admitted to the ICU during index infection had slightly lower median hemoglobin measurements during the SARS-CoV-2^+ phase than patients not admitted to the ICU (cohen's D = -0.27, p=0.01; Figure S8). 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 (see Figure S5) and (2) multivariate logistic regression (see Tables 9 and 10 and *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 (see Tables 9 and 10 and *Statistics* below).

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 (Khwaja, 2012) as follows (see also Figures 5A and S4A):

- *Anemia (mild, moderate, or severe)*: for males, hemoglobin < 13.5 g/dL or hematocrit < 38.3%. For females, median < 12.0 g/dL or hematocrit < 35.5%. Patient-level median values were considered for the pre-COVID phase, and patient-level minimum values were considered for the SARS-CoV-2^+ phase.
- *Anemia (moderate or severe)*: for both males and females, hemoglobin < 10.0 g/dL. Patient-level median values were considered for the pre-COVID phase, and patient-level minimum values were considered for the SARS-CoV-2^+ phase.
- *Acute kidney injury (stage 1, 2, or 3)*: increase in serum creatinine by $\geq$0.3 mg/dL within 48 hours or an increase in serum creatinine to $\geq$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 $\geq$2x the baseline value which is known or assumed to have occurred in the prior 7 days, or a serum creatinine value of $\geq$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.

Using these classifications, we then tested whether the experience of anemia or acute kidney injury during the pre-COVID or SARS-CoV-2^+ intervals was associated with subsequent rehospitalization by computing odds ratios and Fisher's exact test p-values.

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 first 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*.

Assessment of anemia in the context of rehospitalization of influenza patients

To assess whether the observed association between anemia and rehospitalization is specific to COVID-19, we repeated a subset of our analyses on a cohort of influenza patients. We identified patients in the Mayo

Clinic health system who were hospitalized between seven days prior to and seven days after a positive influenza diagnostic test in any year between 2003 and 2019. This hospitalization within one week of an influenza diagnosis was defined as the "index hospitalization." The group of hospitalized influenza patients was split into two cohorts based on whether they were rehospitalized within 30 days of discharge from their index hospitalization. We defined two time periods for each patient: (i) the pre-Flu phase, defined as the one year prior to the first positive influenza test for a given patient; and (ii) the Flu-positive phase, defined as the duration of the index hospitalization. Each patient was classified in a binary fashion for each time window based on whether their lab tests were consistent with the clinical diagnosis of anemia, as described previously for our COVID-19 analyses. Using these classifications, we then tested whether the experience of anemia during the pre-Flu or Flu-positive phases was associated with subsequent rehospitalization by computing odds ratios and Fisher's exact test p-values.

QUANTIFICATION AND STATISTICAL ANALYSIS

Comparison of lab test results and clinical diagnoses between PCH and PCNH cohorts

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 distributions 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). As these measurements did not follow a normal distribution (KS Test p-value $<$ 0.05), the non-parametric Mann-Whitney U test was chosen for statistical comparisons. A one-sided test was used because these comparisons 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 (Pawlowski et al., 2021), 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 $\leq$ 0.05 and the cohen's D magnitude was $\geq$ 0.4. These tests were applied using the SciPy package v1.6.3(Virtanen et al., 2020) in Python (version 3.5). Code corresponding to these analyses is provided in Data S1.

To assess categorical outcome measures (i.e., contingency tables of anemia or AKI status versus rehospitalization status), we computed odds ratios and Fisher's exact test p-values. These tests were applied using the fisher.test function from the stats package in R (version 4.0.3). Corresponding code is provided in Data S1.

Logistic regression analyses

To address potential confounding variables that may be related to the observed trends in laboratory measurements, we performed multivariate logistic regressions for both the pre-COVID and SARS-CoV-2^+ phases (see Tables 9 and 10). For each regression, the binary dependent variable was defined as post viral clearance rehospitalization status (i.e. assignment to the PCH cohort versus PCNH cohort), and the independent variables included one anemia metric and one renal function metric along with sex (binary), age (continuous), 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:

$$log(P_{PCH}/(1 - P_{PCH})) = \beta_0 + \beta_1*(Anemia\ Metric) + \beta_2*(Renal\ Function\ Metric) + \beta_3*(Sex) + \beta_4*(Age) + \beta_5*(Blood\ Draw\ Count) + \beta_6*(ICU\ Admission\ Status).$$

In one set of logistic regressions, the anemia and renal function metrics employed were minimum hemoglobin (continuous) and maximum BUN (continuous), respectively (see Table 9). In another set, the anemia and renal function metrics employed were diagnosis status for moderate or severe anemia (binary) and diagnosis status for stage 2+ AKI (binary), respectively (see Table 10). Binary variables were assigned as follows: anemia: 0 = no anemia or mild anemia, 1 = moderate or severe anemia; AKI: 0 = no AKI or stage 1 AKI, 1 = stage 2+ AKI; sex: 0 = female, 1 = male; ICU admission during interval: 0 = not admitted to ICU, 1 = admitted to ICU. 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.

For each time interval, we performed separate regression analyses for the continuous and binarized anemia and AKI terms rather than including all terms in one model because these metrics are strongly correlated with each other, and multicollinearity of independent variables can negatively impact the estimation of logistic regression coefficients. Each regression yielded a coefficient (log odds ratios) and a p-value (calculated using the log likelihood ratio test) for each independent variable. P-values were adjusted within the output of each model using the Bonferroni correction: $p_{Bonf} = 1-(1-p)^n$. Here, *n* corresponds to the number of independent variables tested (5 for pre-COVID models, 6 for SARS-CoV-2^+ models).

Logistic regressions were performed using the Statsmodels package v0.10.0 (Seabold and Perktold, 2010) in Python (version 3.5). The corresponding code is provided in Data S1.