SARS-CoV-2 Reinfection is Preceded by Unique Biomarkers and Related to Initial Infection Timing and Severity: an N3C RECOVER EHR-Based Cohort Study

Emily Hadley[1], Yun Jae Yoo[2], Saaya Patel[2], Andrea Zhou[3], Bryan Laraway[4], Rachel Wong[2], Alexander Preiss[1], Rob Chew[1], Hannah Davis[5], Christopher G Chute[6], Emily R Pfaff[7], Johanna Loomba[3], Melissa Haendel[8], Elaine Hill[9], Richard Moffitt[10], and the N3C and RECOVER consortia

On behalf of the N3C and RECOVER consortia

[1]RTI International, Durham, NC, US
[2]Stony Brook University, Stony Brook, NY, US
[3]University of Virginia, Charlottesville, VA, US
[4]TISLab, Aurora, CO, US
[5]RECOVER Patient Led Research Collaborative (PLRC), US
[6]Johns Hopkins University, Baltimore, MD, US
[7]University of North Carolina, US
[8]University of Colorado Anschutz Medical Campus, Denver, CO, US
[9]University of Rochester Medical Center, Rochester, NY, US
[10]Emory University, Atlanta, GA, US

Corresponding Authors:
Emily Hadley, ehadley@rti.org, RTI International, 3040 E Cornwallis Rd, Research Triangle, NC 27709
Richard Moffitt (ramoffi@emory.edu), Emory University, 201 Dowman Drive, Atlanta, GA, 30322

Abstract:

Although the COVID-19 pandemic has persisted for over 2 years, reinfections with SARS-CoV-2 are not well understood. We use the electronic health record (EHR)-based study cohort from the National COVID Cohort Collaborative (N3C) as part of the NIH Researching COVID to Enhance Recovery (RECOVER) Initiative to characterize reinfection, understand development of Long COVID after reinfection, and compare severity of reinfection with initial infection. We validate previous findings of reinfection incidence (5.9%), the occurrence of most reinfections during the Omicron epoch, and evidence of multiple reinfections. We present novel findings that Long COVID diagnoses occur closer to the index date for infection or reinfection in the Omicron BA epoch. We report lower albumin levels leading up to reinfection and a statistically significant association of severity between first infection and reinfection (chi-squared value: 9446.2, p-value: 0) with a medium effect size (Cramer's V: 0.18, DoF = 4).

INTRODUCTION

Throughout the COVID-19 pandemic, hundreds of millions of SARS-CoV-2 cases have been confirmed worldwide.[1] However, an infection with SARS-CoV-2 does not confer lasting immunity, particularly in the context of immunologic escape displayed by new variants.[2] Reports of SARS-CoV-2 reinfection are well documented, and whole genome sequencing analysis has confirmed reinfections from SARS-CoV-2 variants that are genetically distinct from an initial SARS-CoV-2 infection.[3] Reinfections are concerning because they may interfere with the development of herd immunity.[4]

In this work, we seek to contribute to the growing literature on SARS-CoV-2 reinfections with novel findings from a large cohort of more than 1.5 million individuals in the electronic health record (EHR)-based N3C Data Enclave. We first characterize reinfection by describing incidence, attributes, and biomarkers. We then explore the severity of reinfection as measured by hospitalization and consider the relationship of reinfection and Long COVID. Finally, we discuss findings and suggest opportunities for further research.

Incidence estimates of reinfections among persons who experienced a SARS-CoV-2 infection are low, ranging from 0.2% to 5.5%.[5,6] A review of laboratory studies found that the time from primary SARS-CoV-2 infection to reinfection can range from 19 to 293 days.[7] Guidelines generally suggest that a new positive COVID-19 antigen or PCR test should be considered a reinfection if it occurred at least 60 to 90 days after initial infection.[5,6,8–11] A few studies document cases of two or three infections, noting that third infections were mainly associated with the transmission of the Omicron variant.[10,11] In this work, we aim to validate these findings with analyses from a larger cohort.

Biomarkers are an important tool for characterizing a disease. Existing research has explored the relationship of severity of COVID-19 with biomarkers such as laboratory indicators of inflammation, dysregulated coagulation and end-organ dysfunction.[12–14] Although studies of reinfection have been less common, one study that characterized patients with suspected reinfection showed increased rates of metabolic failure and similar rates of renal and hepatic failure with reinfection compared to their index encounter, but did not further analyze these findings using laboratory biomarkers.[13] Most studies of biomarkers related to COVID-19 infection are limited to the time period during infection, with limited insight into the trajectories of laboratory indicators between infection and reinfection. In this work, we aim to use biomarkers captured in EHR data between the index date and reinfection to add to the characterization of the disease in reinfection and population-level trends in biomarkers.

Considerable interest exists regarding the severity of reinfection as compared to initial infection. Hospitalization can be an indicator of disease severity because more severe disease often requires treatment. Studies looking at reinfection and hospitalization have generally found that rates of hospitalization following reinfection were similar to or lower than rates of hospitalization following initial SARS-CoV-2 infection.[15,16] One study found that the reduced risk of hospitalization following reinfection persisted when disaggregated by age.[17] No study has clearly

disaggregated by severity of hospitalization, such as considering the distinction between an emergency department (ED) visit, an inpatient hospitalization, and an inpatient hospitalization requiring intensive care. In this work, we aim to assess the association between the severity of initial infection and severity of first reinfection to contribute more granular findings disaggregated by severity of hospitalization.

Less attention has been given to the relationship of reinfections to post-acute sequelae of SARS-CoV-2 infection (PASC) or Long COVID.[18] PASC is understood as complications resulting from SARS-CoV-2 that persist or occur de novo for at least 4 weeks post-infection, and Long COVID is the clinical diagnosis for these conditions. Long COVID is associated with commonly reported symptoms including fatigue that interferes with daily life, fever, cough, sleep problems, difficulty breathing, and difficulty thinking.[19] Existing work suggests that reinfection can increase the risk of post-acute sequelae in the pulmonary organ systems.[20] Additional knowledge about the relationship of reinfections and Long COVID could help inform interested parties who may be concerned that reinfections could contribute to the incidence of Long COVID. In this work, we contribute novel findings on the relationship of reinfection and Long COVID.

METHODS

This study uses individual EHR data stored in the N3C Data Enclave as part of the NIH Researching COVID to Enhance Recover (RECOVER) Initiative. The RECOVER Initiative seeks to understand, treat, and prevent PASC. For more information on RECOVER, visit https://recovercovid.org. The N3C Data Enclave provides access to harmonized EHRs from more than 75 health sites with data from over 16 million patients.[21,22] We used N3C data from version 87 (8/2/2022), which has 63 contributing sites, for the current investigation. The N3C Data Enclave's Palantir Foundry platform (2021, Denver, CO), a secure analytics platform, was used for data access and analysis.

Key Definitions

We describe the following key definitions for the study cohort, reinfection, COVID-19 variant epochs, and Long COVID.

Study Cohort Definition, Inclusion, and Exclusion Criteria

The study inclusion criteria include (1) having an International Classification of Diseases-10-Clinical Modification (ICD-10) COVID-19 diagnosis code (U07.1) or a positive SARS-CoV-2 PCR or antigen test between March 1, 2020, and May 1, 2022; the earliest of these events was considered the COVID-19 index date; (2) reinfection events (if any) occurring before July 1, 2022; (3) being 18 years of age or older; (4) having at least one recorded healthcare visit in the two years prior to index; (5) having at least one recorded healthcare visit more than 60 days after the COVID-19 index date; (6) being from a hospital partner with data that has been updated in the last two months prior to July 1, 2022; (7) being from a hospital partner with at minimum 100 hospitalizations related to a first known COVID-19 infection and at

minimum 25 hospitalizations related to a COVID-19 reinfection. A total of 1,597,490 individuals met these criteria.

Definition of Reinfection

A COVID-19 reinfection was defined as a positive SARS-CoV-2 PCR or antigen test that occurred 60 or more days after a COVID-19 infection index date. The date of the test was considered the first COVID-19 reinfection index date. Subsequent reinfections were defined as a new positive SARS-CoV-2 PCR or antigen test that occurred 60 or more days after each reinfection index date. Although a threshold of 90 days for reinfection post index date is common in the literature, other findings suggest that nearly all patients stop shedding SARS-CoV-2 within 60 days of infection and many stop shedding much sooner than that.[23–26] Based on these findings and support from the RECOVER clinician advisory panel, 60 days was selected as a more appropriate threshold.

Definition of the COVID-19 Variant Epoch

We define the following COVID-19 variant epochs based on the patient's COVID-19 diagnosis code (U07.1) or a positive SARS-CoV-2 PCR or antigen test date: Wild-type COVID-19 (March 01, 2020-November 30, 2020), Alpha/Beta/Gamma variant (December 01, 2020-May 31, 2021), Delta variant (June 01, 2021-October 31, 2022), Omicron variant (November 01, 2021-March 11, 2022), and Omicron BA variant (March 12, 2022-August 01, 2022).[27]

Definition of Long COVID

Patients with a Long COVID diagnosis were identified with the U09.9 or B94.8 ICD-10-CM diagnosis codes. The U09.9 code was implemented in October 2021 for providers to use in a clinical setting with patients experiencing ongoing conditions after a COVID-19 infection, commonly understood as Long COVID. Many hospital sites appear to have rapidly adopted the use of U09.9 once it became available.[28] The B94.8 code is not specific to COVID-19 and instead represents sequelae of other specified infectious and parasitic diseases. This code was rarely used prior to the pandemic, but it started seeing considerably more use in November 2020. The use of this code is understood to represent Long COVID diagnoses prior to the availability of U09.9.[28] For the purposes of the Long COVID analysis, we limited the study cohort to individuals at sites that had at least 250 uses of either the U09.9 code after October 1, 2021 or the B94.8 code after November 1, 2020. Eligible reinfections for U09.9 or B94.8 had to occur after the respective dates of use of the codes. This subcohort included 1,568,810 individuals.

Statistical Analysis

In this work, we perform three main analyses focused on describing characterizing reinfection, understanding reinfection severity, and exploring the relationship of reinfection and long COVID. All analysis and visualization were done in the N3C enclave using SQL, Python (v3.6), and R (v3.6), including ggplot2, survival, and survminer packages.

Characterization of Reinfection

We used two approaches to characterize reinfection. The first is a cohort summary where we calculate summary statistics related to reinfection and disaggregate by demographic characteristics. Chi-square tests were used for categorical variables and Student's t-test or ANOVA were used for continuous variables (**Table 1, Supplementary Table 2, 3**). Time to reinfection analysis was based on Kaplan-Meier curves from the *survival* package in R. This analysis was performed using the date of the COVID-19 index date (the date of earliest diagnosis or positive test) and the date of event (first reinfection date) as endpoints.

The second approach to characterizing reinfection is with biomarkers. We explored the trajectories of various biomarkers around COVID-19 initial and subsequent index dates from patients with and without a reinfection. Biomarker measurements included laboratory values of ferritin, fibrinogen, C-reactive protein, procalcitonin, white blood cell count, absolute lymphocyte count, absolute neutrophil count, erythrocyte sedimentation rate, albumin, D-dimer, alanine transaminase (ALT), aspartate transaminase (AST), and serum creatinine. Units were harmonized across data partners, and clinically infeasible values were excluded according to standard N3C data quality protocols.[22] Measurements were taken from 100 days prior to and 180 days after the COVID-19 index date in patients with and without reinfection. The same time frame was used for collection around the first reinfection index date in individuals with at least one reinfection. Laboratory values were reported separately for hospitalized and nonhospitalized patients. The median laboratory value of each biomarker, with upper (75%) and lower (25%) quartiles, was binned by 7-day intervals and visualized according to time from COVID-19 infection or reinfection index date. For patients with more than one measurement of the same laboratory test in a day, values were averaged.

Severity of Reinfection

Severity of infection was measured with records of COVID-associated hospitalization, which was defined as an inpatient visit with a start date 1 day prior to 16 days after the COVID-19 index date with a COVID diagnosis code used during the visit. A COVID-associated ED visit was defined as an ED visit with a start date 1 day prior to 16 days after the COVID-19 index date. These thresholds were intended to capture hospitalizations and ED visits that are related to COVID. Severity of infection is assessed with hospitalization metrics. Four levels of severity are considered: mild infection that does not require an ED visit or hospitalization, mild infection that requires an ED visit, moderate infection that requires hospitalization, and severe infection that requires hospitalization and use of extracorporeal membrane oxygenation (ECMO), invasive mechanical ventilation (IMV), or vasopressors. Vasopressors were included in addition to the more intensive ECMO and IMV because some hospitals may be limited in their ability to provide ECMO or IMV. We compared the severity of the first COVID-19 infection versus severity of first reinfection using a pivot table with row, column, and table percentages along with a chi-square test for association. Death after initial infection or reinfection is also included in the table.

Reinfections and Long COVID

The subcohort of individuals described in Section 2.1.4 was used for analysis of reinfections and Long COVID. Kaplan-Meier curves were calculated to explore the differences in time to Long COVID diagnosis following initial infection versus reinfection. Time to event analysis was performed using the initial COVID-19 index date and with the first reinfection index date (for those with one or more reinfections) to the first B94.8 or U09.9 diagnosis.

RESULTS

The study cohort included 1,597,490 adults (age: mean 48.85 years, standard deviations (SD) 18.27; 63.11% female) (**Table 1, Supplementary Table 1**). The study cohort contains data from 38 health facility data partners. Women make up nearly two-thirds of the study cohort and a larger proportion of individuals with reinfections. The skew in sex can be attributed to the inclusion requirement for at least two visits prior to COVID diagnosis (**Supplementary Figure 1**). The sex ratio of individuals hospitalized with COVID-19 was more balanced (**Supplementary Table 2**).

Characterization of Reinfection

Table 1 describes the study cohort and highlights differences between the subgroup with no reinfections and the subgroup with at least one reinfection. A total of 5.9% of the study cohort had at least one documented reinfection. A documented reinfection was defined as a positive SARS-CoV-2 PCR or antigen test that occurred 60 or more days after a COVID-19 infection index date. Home COVID-19 tests administered outside a healthcare setting were not included in the dataset. The subgroup of reinfected patients tends to be younger and more likely to have documented race and ethnicity information. Although the majority of individuals (N=89,357) had only one reinfection, a small group (N=102) had three or more reinfections. This group was characterized by non-Hispanic, male, White race, and older age groups compared to other groups (**Supplementary Table 1**). Supplementary Table 1 shows Table 1 further disaggregated by number of reinfections.

Figure 1 illustrates three approaches for understanding the occurrence of reinfection as it relates to COVID-19 variants. **Figure 1A** shows the percentage of patients at risk that had a reinfection in a given month. Distinct colors are used to indicate the epoch of initial infection and the size of the dot illustrates the number of persons with a reinfection. This figure is useful for appreciating the varying likelihood of reinfection while accounting for individuals who passed away following their first infection since these individuals are no longer considered at risk. This figure shows the largest increase in reinfections in the Omicron epoch among individuals with initial infections during the WT COVID and Alpha, Beta, Gamma periods which overlap in the Omicron epoch, and a smaller increase among those first infected in the Delta epoch. The difference between these variants is smaller for reinfections in the Omicron BA epoch.

Table 1: Descriptive characteristics of reinfected and non-reinfected COVID-19 positive patients

Category	Variable	No Reinfection (N=1505855)	Reinfected Patients (N=91635)	Total (N=1597490)	P value
Age, Mean (SD)	Age	49.14(18.26)	44.31(17.71)	48.85(18.27)	<0.0001
Sex (N, %)	Female	946166 (62.83)	61989 (67.65)	1008155 (63.11)	
	Male	559280 (37.14)	29617 (32.32)	588897 (36.86)	<0.0001
	No sex Information	409 (0.03)	29 (0.03)	438 (0.03)	
Race (N, %)	White	1141563 (75.81)	69852 (76.23)	1211415 (75.83)	
	Black	199891 (13.27)	12871 (14.05)	212762 (13.32)	
	Asian	24615 (1.63)	1022 (1.12)	25637 (1.6)	
	Others	33996 (2.26)	2222 (2.42)	36218 (2.27)	<0.0001
	No race Information	105790 (7.03)	5668 (6.19)	111458 (6.98)	
Ethnicity (N, %)	Not Hispanic	1197960 (79.55)	73706 (80.43)	1271666 (79.6)	
	Hispanic	106956 (7.1)	7925 (8.65)	114881 (7.19)	<0.0001
	No Ethnicity information	200939 (13.34)	10004 (10.92)	210943 (13.2)	

Figure 1B is a KM-curve that shows time to event between initial COVID-19 infection and first COVID-19 reinfection by COVID-19 variant. This figure is useful for understanding the days to infection, including demonstrating that in many cases, reinfection occurred more than 100 days after the initial COVID-19 infection. **Figure 1B** also shows that this analysis is not particularly sensitive to the decision to use a 60-day threshold rather than a 90-day threshold because few reinfections occur in the 60- to 90-day window for any variant. We recognize that this figure cannot appropriately account for individuals who passed away following their first infection; because of this violation of proportional hazards, we have chosen to not report odds ratios because they may be misinterpreted. **Figure 1C** most clearly details the relationship between the variant of initial infection and the variant of reinfection, highlighting that reinfections in the

Omicron time period were particularly common among individuals initially infected in the
Wild-type COVID-19 and the Alpha, Beta, and Gamma epochs.

Figure 1
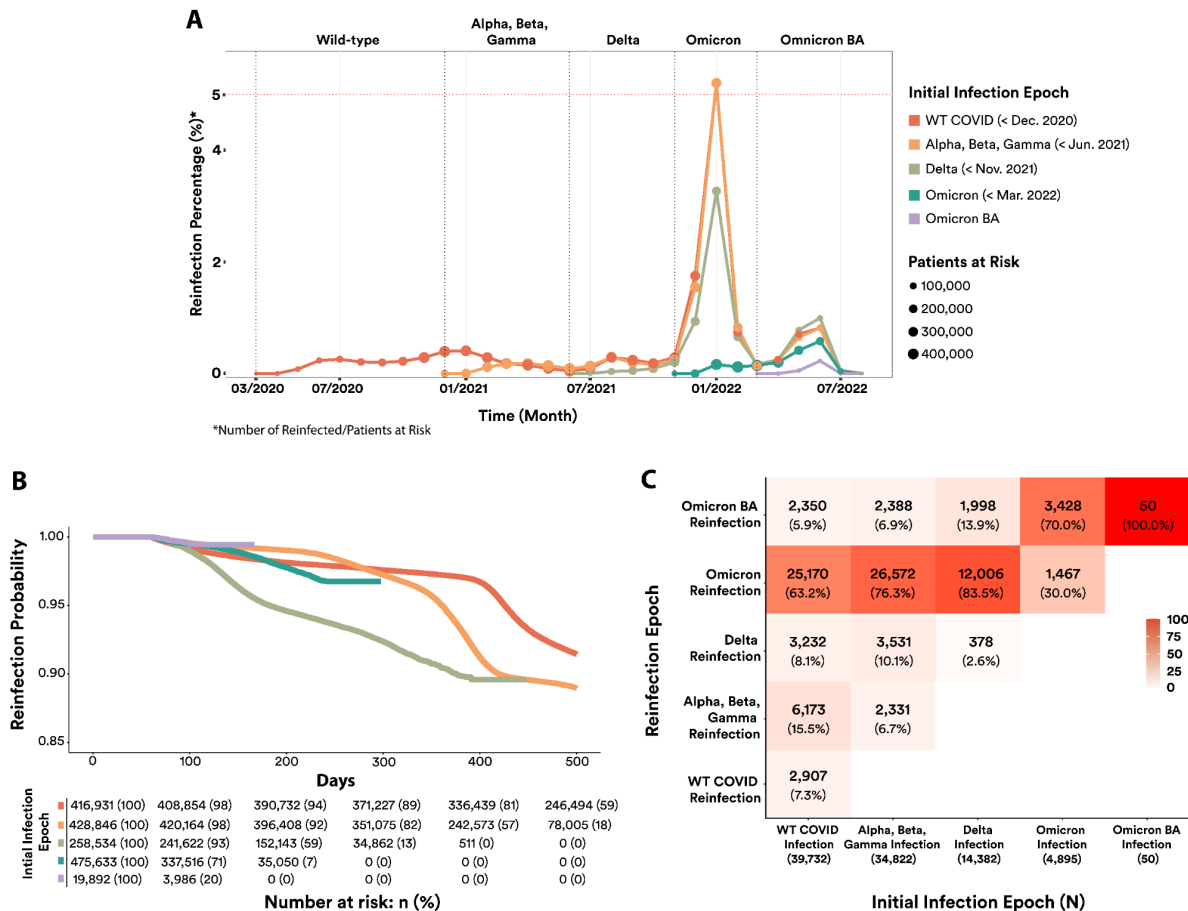
We evaluated biomarker trends in **Figure 2**, comparing the median laboratory values of patients
with and without a reinfection. Comparisons were made between the index date of initial
infection and subsequent reinfection date. Biomarkers of hepatic inflammation (ALT and AST)
were less elevated during acute reinfection compared to initial COVID-19 infection and
normalized over a similar time period. However, albumin trends show that among patients with
reinfection, albumin levels were persistently lower after initial COVID-19 infection and that levels
were also lower prior to the reinfection date.

Figure 2
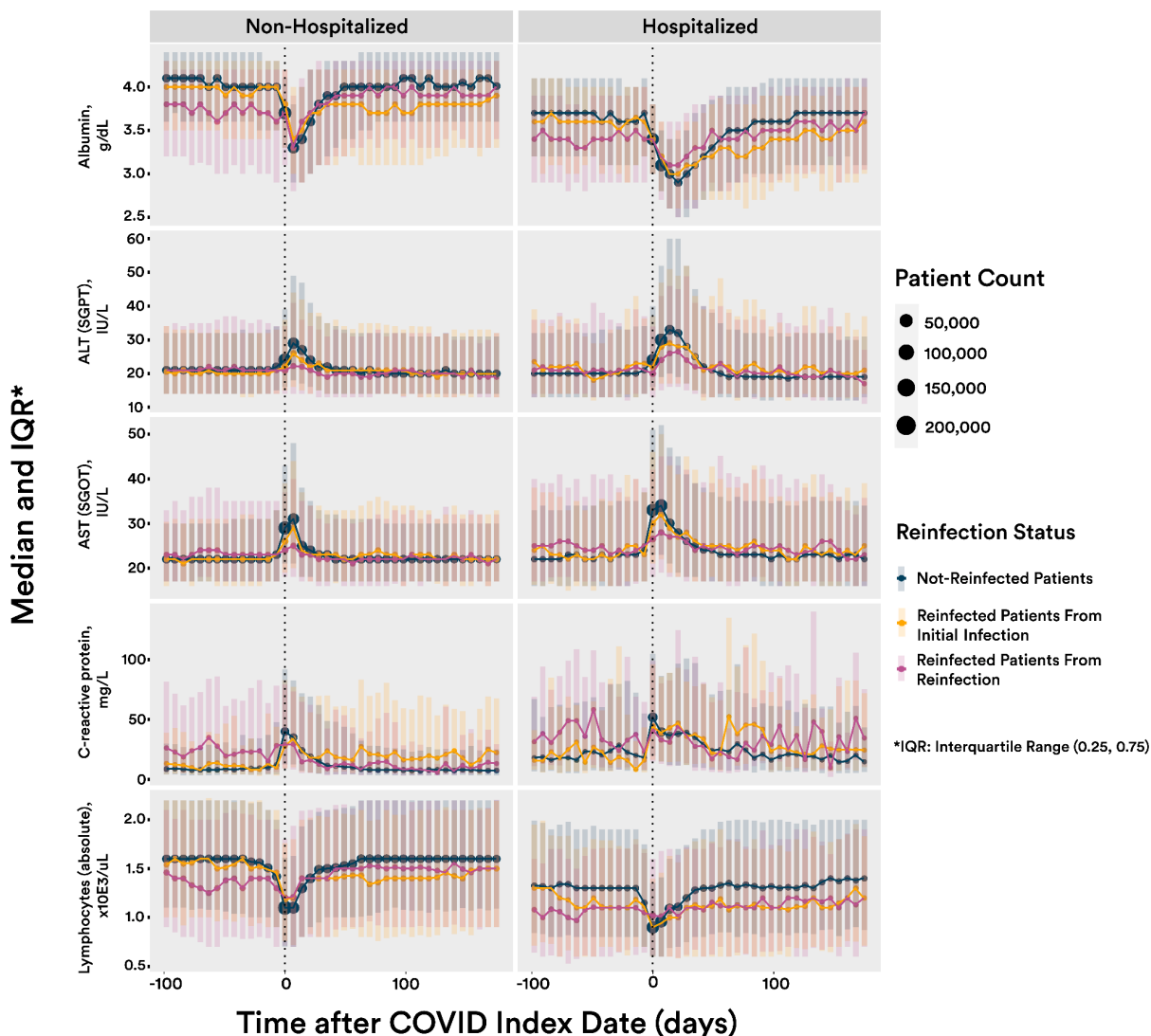
Severity of Reinfection

We explored the characteristics of severity of infection and hospitalization in **Table 2.** The "No Documented Reinfection" column highlights that the large majority of individuals, even those with ED visits or hospitalization, do not have documentation of a reinfection. The shaded portion of Table 2 summarizes results from individuals with a documented reinfection. We performed a chi-square test of independence for both the entirety of Table 2 (with "No Reinfection" included) and for the shaded portion of Table 2. For the entire table, we found a statistically significant difference (chi-squared value: 9049.2, p-value: 0), although with a negligible effect size (Cramer's V: 0.04, DoF = 4). For the shaded portion of the table, we found a statistically significant difference (chi-squared value: 9446.2, p-value: 0) with a medium effect size (Cramer's V: 0.18, DoF = 4). These results, particularly those among individuals who experience

reinfection, suggest that the severity of reinfection may not be independent from the severity of initial infection.

Table 2: Comparison of Severity of First and Reinfection with SARS-CoV-2

Count *Row (R)%* *Column (C)%* *Table (T)%* *among those with a reinfection**	**Severity of Reinfection**						**Total**
Severity of Initial Infection	*No Documented Reinfection*	*Mild with No ED Visit or Hospital-ization Around Reinfection Index*	*Mild with ED Visit Around Reinfection Index*	*Moderate with Hospitaliza-tion Around Reinfection Index*	*Severe with ECMO or IMV or Vasopressor During Hospitalization Around Reinfection Index*	*Death Within 60 Days of Reinfection Index*	
Mild with No ED Visit or Hospitalization Around COVID Index	1288792	71126 R: 88.8% C: 90.9% T: 77.6%	4970 R: 6.2% C: 67.5% T: 5.4%	2961 R: 3.7% C: 65.3% T: 3.2%	414 R: 0.5% C: 64.7% T: 0.4%	622 R: 0.8% C: 72.0% T: 0.7%	1368885 (80093 multiple infections)
Mild with ED Visit Around COVID Index	115740	4357 R: 64.9% C: 5.6% T: 4.8%	1820 R: 27.1% C: 24.7% T: 2.0%	432 R: 6.4% C: 9.5% T: 0.5%	67 R: 1.0% C: 10.5% T: <0.1%	39 R: 0.6% C: 4.5% T: <0.1%	122455 (6715 multiple infections)
Moderate with Hospitalization Around COVID Index	85959	2496 R: 57.3% C: 3.2% T: 2.7%	523 R: 12.0% C: 7.1% T: 0.6%	1034 R: 23.8% C: 22.8% T: 1.1%	124 R: 2.9% C: 19.4% T: 0.1%	176 R: 4.0% C: 20.4% T: 0.2%	90312 (4353 multiple infections)
Severe with ECMO or IMV or Vasopressor During Hospitalization Around COVID Index	9357	251 R: 52.9% C: 0.3% T: 0.2%	52 R: 12.0% C: 0.7% T: <0.1%	109 R: 23.0% C: 2.4% T: 0.1%	35 R: 7.4% C: 5.5% T: <0.1%	27 R: 5.7% C: 3.1% T: <0.1%	9831 (474 multiple infections)
Death Within 60 Days of COVID Index	6007	N/A	N/A	N/A	N/A	N/A	6007
Total	1505855	78230	7365	4536	640	864	1597490

*The shaded region indicates the values used for calculation of row, column, and table percentages among those with a SARS-CoV-2 reinfection. Individuals without a reinfection are not included in this calculation.

Among individuals with a documented reinfection, the majority (77.6%) experienced both mild first and reinfections that did not require an ED visit or hospitalization. A smaller percentage of individuals (12.6%) required either an ED visit or hospitalization or passed away following their

reinfection as compared to individuals who required an ED visit or hospitalization or passed away after first infection (14.4%). Among those with either an ED visit or hospitalization during their first infection, a substantial proportion also required an ED visit or hospitalization or passed away following their first reinfection. This included nearly half (47.1%) of those who experienced a severe first infection.

Among individuals who experienced a severe initial infection, 7.4% experienced a severe reinfection and another 5.7% passed away following reinfection. More individuals with a severe initial infection (23.0%) had a moderate reinfection which required hospitalization but did not require use of ECMO or IMV. The largest proportion of individuals who experienced a severe reinfection had mild initial infections with no ED visit or hospitalization (64.7%). Individuals with mild initial infections also made up the majority of those with a mild infection requiring an ED visit or moderate hospitalization for reinfection. The second largest group of individuals with a mild reinfection requiring an ED visit was those who also had a first infection that was mild and required an ED visit (27.1%). A similar proportion of individuals with a moderate initial infection requiring hospitalization (23.8%) also required hospitalization during reinfection.

We provide additional detail in Supplementary Table 2 with a breakdown of demographic details for patients who experience hospitalization for COVID-19. In Supplementary Table 3, we provide Table 2 disaggregated by age, noting that disclosure requirements may limit the robustness of conclusions.

Reinfection and Diagnosis of Long COVID

Finally, **Figure 3** shows the Kaplan-Meier curves for time to Long COVID diagnosis after initial infection versus first reinfection. For most variant epochs, incidence of new Long COVID diagnoses after reinfection is lower than after initial COVID-19 infection. The exception is Omicron BA where incidence of new Long COVID diagnoses after reinfection in Omicron BA is higher than after initial COVID-19 in Omicron BA. The rate of Long COVID diagnoses following the first infection was largest in the Delta epoch and smallest in the Omicron BA epoch. This contrasts with the incidence of Long COVID diagnoses after first reinfection which is largest in the Omicron BA epoch and smallest following reinfection during the Delta epoch.

Figure 3

Comparison of U09.9/B94.8 Rates Between Variants

Variants are classified based on the date of COVID infection

Epoch
- Delta (< Nov. 2021)
- Delta Reinfection (< Nov. 2021)
- Omicron (< Mar. 2022)
- Omicron Reinfection (< Mar. 2022)
- Omicron BA
- Omicron BA Reinfection

Y-axis: Non U09.09/B94.8 Probability (1.00, 0.99, 0.98, 0.97)

X-axis: 0, 100, 200, 300, 400

$p < 0.0001$

Epoch — Number at risk: n (%)

255,396 (100)	235,539 (92)	146,644 (57)	32,788 (13)	497 (0)
6,842 (100)	5,713 (83)	3,780 (55)	973 (14)	32 (0)
464,898 (100)	322,982 (69)	33,780 (7)	0 (0)	0 (0)
63,055 (100)	21,891 (35)	1,417 (2)	0 (0)	0 (0)
19,281 (100)	3,870 (20)	0 (0)	0 (0)	0 (0)
9,407 (100)	287 (3)	0 (0)	0 (0)	0 (0)

DISCUSSION

The overall proportion of individuals with at least one documented reinfection in the cohort (5.7%) was slightly larger than the upper bound of incidence of reinfection noted in the literature (5.5%).[5] This is likely an underestimate given the advent of home testing, particularly in more recent epochs when home testing has been even more accessible. Similar to existing findings, the large majority of reinfections occurred during the Omicron epoch. Figure 1 suggests that this finding holds regardless of the epoch of the initial COVID-19 infection. The large number of reinfections during Omicron makes it challenging to draw conclusions about comparing reinfections between variants because there may be other factors, such as adherence to mask or social distance policies, that impacted likelihood of exposure and subsequent reinfections. However, it is notable that in Figure 1, initial infection during Delta appeared to be more protective against reinfection during Omicron than initial infection during WT COVID or Alpha, Beta, Gamma. This difference disappeared during Omicron BA. Previous studies have documented reinfections up to 293 days after initial COVID-19 infection.[7] This study finds evidence of reinfections that occur more than 300 days and up to 500 days after initial

COVID-19 infection. This work also validates the occurrence of multiple reinfections, including a small subset of individuals with three or more reinfections.

The mean age of reinfected individuals is close to 5 years lower than the mean age of those without reinfection. This aligns with literature that suggests that reinfections are more commonly reported for younger individuals: one study found that the reinfection rate was highest among those aged 18 to 29 years old[29] while public health data collected from September 2021 through September 2022 in the state of Washington found disproportionately large numbers of individuals with reinfections in the 18- to 34- and 35- to 49-year old age groups.[30] One possible explanation is that younger age groups are less likely to use COVID-19 preventative measures like social distancing and more likely to engage with others for work and leisure, leading to multiple exposures and reinfections.[31] Another possible explanation is that younger individuals are less likely to be vaccinated and may be more susceptible to reinfection.[32] A third possible explanation is that older individuals were more likely to die following initial infection, so more reinfections occurred among younger individuals.[33] A fourth possible explanation is that young adults have high rates of asymptomatic and paucisymptomatic infection which may be less protective of reinfection.[34] More research is needed to explore these potential explanations.

Women make up nearly two-thirds of the study cohort and a larger proportion of those with a documented reinfection. Some studies have also noted larger proportions of women than men with reinfections and suggested that there may be relevant differences in immune response by sex.[35–37] However, Supplemental Figure 1 suggests that the inclusion criteria requiring at least two visits prior to COVID-19 index date contributes to this imbalance. Previous research also suggests that women are more likely to utilize health services and demonstrate health-seeking behavior.[38,39] The sex imbalance in these findings may be more likely associated with differential healthcare utilization rather than biological differences.

We explore the characterization of reinfection with biomarkers. Biomarkers have been well studied in acute COVID-19 infection and several laboratory markers have been associated with higher severity of infection and mortality.[40,41] Our work extends knowledge of biomarkers to reinfections. In finding that biomarkers of hepatic inflammation were less elevated during acute reinfection compared to acute initial infection, we contribute a novel result that has not yet been reported. Another new finding is that albumin appears to be lower leading up to reinfection. Previous studies have suggested that hypoalbuminemia is common in COVID-19 patients and dynamic monitoring of serum albumin can be a useful tool for evaluating the prognosis of COVID-19 infections. We suggest that further work may explore if lower albumin levels may be a predictor of COVID-19 reinfection.

Similar to previous studies measuring severity of reinfection through hospitalization, we find that most individuals did not require an ED visit or hospitalization for either first infection or reinfection. We also find that the total proportion of individuals requiring an ED visit or hospitalization or passing away is similar between first infection and reinfection. However, when we further interrogate these results, we contribute novel findings that individuals who were hospitalized for an acute initial infection are potentially at much greater risk for hospitalization

during reinfection. The effect size is medium when we consider the results among those with reinfections, noting that we cannot account for what would have happened among those who passed away following their first infection.

We first consider the varying degrees of severity of initial infection among individuals who also have a reinfection. Nearly two thirds of individuals who had a mild initial infection that did include an ED visit (64.9%) did not require an ED visit or hospitalization or pass away after reinfection, but a substantial proportion (27.1%) recorded an ED visit for their reinfection. This proportion is larger than the respective proportion of individuals who had a moderate or severe hospitalization during first infection and an ED visit following reinfection. Further analysis could investigate if the group of individuals who visited an ED for both first and reinfection reflects ED-seeking behavior that may result from convenience of the ED, limited access to other healthcare options, or health insurance status.[42]

More than a quarter of individuals with either a moderate or severe first infection requiring hospitalization also experienced hospitalization for reinfection. Although this is concerning, a promising finding is that the proportion experiencing a severe reinfection requiring ECMO, IMV, or vasopressors during hospitalization was small (2.9% of those with moderate first infection and 7.4% of those with severe first infection). One possible explanation is that experiencing hospitalization during reinfection may be related to experiencing hospitalization during first infection, but the hospitalization may be less severe. Another possible explanation is that clinical thresholds for this treatment or clinical behavior may have changed over time. We also observe that the proportion of individuals who pass away following reinfection is higher than the proportion who experience a severe reinfection among those with an initial moderate infection and similar for those who have a severe initial infection. We suggest further analysis to better understand attributes, such as comorbidities, that are predictive of moderate and severe reinfections.

This study contributes novel findings of the relationship of reinfection with Long COVID diagnosis. The largest proportion of Long COVID diagnoses occur among individuals with a first reinfection in the Omicron BA epoch. Long COVID diagnoses also occurred much closer to the index date following both initial or first reinfection in the Omicron BA epoch as compared to earlier Delta and Omicron epochs. The rate of Long COVID diagnoses has been increasing for reinfections in more recent variants. A number of possible explanations exist for these associative findings. One is that there may be a biological explanation where reinfection may be associated with increased risk of post-acute sequelae. This has been suggested in other literature.[20] Another explanation is that physician diagnosing behavior has changed and physicians are more likely to have adopted use of either the new U09.9 Long COVID diagnosis code or the existing B94.8 code. This work has also not accounted for other factors like the impact of vaccination status or use of outpatient therapeutics like paxlovid. Future analysis could explore a causal relationship between reinfection and Long COVID and account for these other factors.

A limitation of this study is the reliance on EHR data. EHR limitations are well documented and include selection bias based on varying rates of healthcare utilization, concerns about fitness for purpose and drawing inappropriate conclusions, and data quality and missing data challenges.[43] EHR studies also differ from clinical studies, where patients may be followed more closely. A major limitation of this particular analysis is that we are limited to EHR collected at specific hospitals; we cannot join patient records between hospitals. Strategies we have used to address these limitations include hospitalization inclusion criteria for sites and visit inclusion criteria for individuals that promote more detailed, robust, and higher quality data.

A second limitation is that it is not feasible to include results of home COVID-19 tests. Individuals may be testing positive for COVID-19 reinfections that are not documented in this dataset. This may result in an underestimate of the number of individuals with reinfections. The varying availability of home tests over the duration considered for this project may result in an uneven impact of this limitation. To address this limitation, we have attempted to maintain a focus on behaviors that require healthcare interaction, such as Long COVID diagnosis, biomarkers, and hospitalization. We do not suggest generalizability of results to situations that would not involve healthcare settings. We also limited analysis to individuals who had a COVID diagnosis prior to May 1, 2022, in an effort to focus on a window in time when home tests were less common. Future work could explore which time periods best capture testing for reinfection.

Reinfections are well documented in an EHR-based cohort from the RECOVER initiative and align with overall incidence rates in the literature. This work validates existing characterization of reinfection as most common in the Omicron epoch and contributes a novel characterization of lower albumin levels after initial COVID-19 infection and leading up to reinfection. The severity of reinfection appears to be associated with the severity of initial infection, and Long COVID diagnoses appear to occur more often following and closer to the index date of reinfection. We describe a number of opportunities for further research with these findings to better understand COVID-19 reinfections.

ACKNOWLEDGMENTS

University of Mississippi Medical Center — U54GM115428: Mississippi Center for Clinical and Translational Research (CCTR) • University of Nebraska Medical Center — U54GM115458: Great Plains IDeA-Clinical & Translational Research • University of North Carolina at Chapel Hill — UL1TR002489: North Carolina Translational and Clinical Science Institute • University of Oklahoma Health Sciences Center — U54GM104938: Oklahoma Clinical and Translational Science Institute (OCTSI) • University of Rochester — UL1TR002001: UR Clinical & Translational Science Institute • University of Southern California — UL1TR001855: The Southern California Clinical and Translational Science Institute (SC CTSI) • University of Vermont — U54GM115516: Northern New England Clinical & Translational Research (NNE-CTR) Network • University of Virginia — UL1TR003015: iTHRIV Integrated Translational health Research Institute of Virginia • University of Washington — UL1TR002319: Institute of Translational Health Sciences • University of Wisconsin-Madison — UL1TR002373: UW Institute for Clinical and Translational Research • Vanderbilt University Medical Center — UL1TR002243: Vanderbilt Institute for Clinical and Translational Research • Virginia Commonwealth University — UL1TR002649: C. Kenneth and Dianne Wright Center for Clinical and Translational Research • Wake Forest University Health Sciences — UL1TR001420: Wake Forest Clinical and Translational Science Institute • Washington University in St. Louis — UL1TR002345: Institute of Clinical and Translational Sciences • Weill Medical College of Cornell University — UL1TR002384: Weill Cornell Medicine Clinical and Translational Science Center • West Virginia University — U54GM104942: West Virginia Clinical and Translational Science Institute (WVCTSI)

Submitted: Icahn School of Medicine at Mount Sinai — UL1TR001433: ConduITS Institute for Translational Sciences • The University of Texas Health Science Center at Tyler — UL1TR003167: Center for Clinical and Translational Sciences (CCTS) • University of California, Davis — UL1TR001860: UCDavis Health Clinical and Translational Science Center • University of California, Irvine — UL1TR001414: The UC Irvine Institute for Clinical and Translational Science (ICTS) • University of California, Los Angeles — UL1TR001881: UCLA Clinical Translational Science Institute • University of California, San Diego — UL1TR001442: Altman Clinical and Translational Research Institute • University of California, San Francisco — UL1TR001872: UCSF Clinical and Translational Science Institute
Pending: Arkansas Children's Hospital — UL1TR003107: UAMS Translational Research Institute • Baylor College of Medicine — None (Voluntary) • Children's Hospital of Philadelphia — UL1TR001878: Institute for Translational Medicine and Therapeutics • Cincinnati Children's Hospital Medical Center — UL1TR001425: Center for Clinical and Translational Science and Training • Emory University — UL1TR002378: Georgia Clinical and Translational Science Alliance • HonorHealth — None (Voluntary) • Loyola University Chicago — UL1TR002389: The Institute for Translational Medicine (ITM) • Medical College of Wisconsin — UL1TR001436: Clinical and Translational Science Institute of Southeast Wisconsin • MedStar Health Research Institute — UL1TR001409: The Georgetown-Howard Universities Center for Clinical and Translational Science (GHUCCTS) • MetroHealth — None (Voluntary) • Montana State University — U54GM115371: American Indian/Alaska Native CTR • NYU Langone Medical Center — UL1TR001445: Langone Health's Clinical and Translational Science Institute • Ochsner Medical Center — U54GM104940: Louisiana Clinical and Translational Science (LA CaTS) Center • Regenstrief Institute — UL1TR002529: Indiana Clinical and Translational Science Institute • Sanford Research — None (Voluntary) • Stanford University — UL1TR003142: Spectrum: The Stanford Center for Clinical and Translational Research and Education • The Rockefeller University — UL1TR001866: Center for Clinical and Translational Science • The Scripps Research Institute — UL1TR002550: Scripps Research Translational Institute • University of Florida — UL1TR001427: UF Clinical and Translational Science Institute • University of New Mexico Health Sciences Center — UL1TR001449: University of New Mexico

REFERENCES

1. WHO Coronavirus (COVID-19) Dashboard. Accessed October 18, 2022. https://covid19.who.int/

2. Pinto LM, Nanda V, Sunavala A, Rodriques C. Reinfection in COVID-19: A scoping review. *Armed Forces Med J India*. 2021;77:S257-S263.

3. Falahi S, Kenarkoohi A. COVID-19 reinfection: prolonged shedding or true reinfection? *New Microbes New Infect*. 2020;38:100812.

4. Wang J, Kaperak C, Sato T, Sakuraba A. COVID-19 reinfection: a rapid systematic review of case reports and case series. *J Investig Med*. 2021;69(6):1253-1255.

5. Medić S, Anastassopoulou C, Lozanov-Crvenković Z, et al. Risk and severity of SARS-CoV-2 reinfections during 2020–2022 in Vojvodina, Serbia: A population-level observational study. *The Lancet Regional Health – Europe*. 2022;20. doi:10.1016/j.lanepe.2022.100453

6. Lawandi A, Warner S, Sun J, et al. Suspected severe acute respiratory syndrome Coronavirus 2 (SARS-COV-2) reinfections: Incidence, predictors, and healthcare use among patients at 238 US healthcare facilities, 1 June 2020 to 28 February 2021. *Clin Infect Dis*. 2022;74(8):1489-1492.

7. Ren X, Zhou J, Guo J, et al. Reinfection in patients with COVID-19: a systematic review. *Glob Health Res Policy*. 2022;7(1):12.

8. Toro-Huamanchumo CJ, Hilario-Gomez MM, Pinedo-Castillo L, et al. Clinical and epidemiological features of patients with COVID-19 reinfection: a systematic review. *New Microbes New Infect*. 2022;48:101021.

9. Rahman S, Rahman MM, Miah M, Begum MN. COVID-19 reinfections among naturally infected and vaccinated individuals. *Sci Rep*. Published online 2022. https://www.nature.com/articles/s41598-022-05325-5

10. Pulliam JRC, van Schalkwyk C, Govender N, et al. Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. *Science*. 2022;376(6593):eabn4947.

11. Swift MD, Hainy CM, Sampathkumar P, Breeher LE. Multiple SARS-CoV-2 Reinfections: A Case Series of Thrice-Infected Individuals. *Mayo Clin Proc*. 2022;97(5):1021-1023.

12. Chen CH, Lin SW, Shen CF, Hsieh KS, Cheng CM. Biomarkers during COVID-19: Mechanisms of Change and Implications for Patient Outcomes. *Diagnostics (Basel)*. 2022;12(2). doi:10.3390/diagnostics12020509

13. Terpos E, Ntanasis-Stathopoulos I, Elalamy I, et al. Hematological findings and complications of COVID-19. *Am J Hematol*. 2020;95(7):834-847.

14. Bivona G, Agnello L, Ciaccio M. Biomarkers for Prognosis and Treatment Response in COVID-19 Patients. *Ann Lab Med*. 2021;41(6):540-548.

15. Sheehan MM, Reddy AJ, Rothberg MB. Reinfection Rates Among Patients Who Previously Tested Positive for Coronavirus Disease 2019: A Retrospective Cohort Study. *Clin Infect Dis*. 2021;73(10):1882-1886.

16. Shaheen NA, Sambas R, Alenezi M, Alharbi NK, Aldibasi O, Bosaeed M. COVID-19 reinfection: A multicenter retrospective study in Saudi Arabia. *Ann Thorac Med*. 2022;17(2):81-86.

17. Marinov GK, Mladenov M, Rangachev A, Alexiev I. SARS-CoV-2 reinfections during the first three major COVID-19 waves in Bulgaria. *PLoS One*. 2022;17(9):e0274509.

18. Desimmie BA, Raru YY, Awadh HM, He P, Teka S, Willenburg KS. Insights into SARS-CoV-2 Persistence and Its Relevance. *Viruses*. 2021;13(6). doi:10.3390/v13061025

19. CDC. Long COVID or Post-COVID Conditions. Centers for Disease Control and Prevention. Published September 21, 2022. Accessed October 18, 2022. https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html

20. Al-Aly Z, Bowe B, Xie Y. Outcomes of SARS-CoV-2 reinfection. *Research Square*. Published online June 17, 2022. doi:10.21203/rs.3.rs-1749502/v1

21. National COVID Cohort Collaborative (N3C). National Center for Advancing Translational Sciences. Published May 12, 2020. Accessed October 18, 2022. https://ncats.nih.gov/n3c

22. Bradwell KR, Wooldridge JT, Amor B. Harmonizing units and values of quantitative data elements in a very large nationally pooled electronic health record (EHR) dataset. *Journal of the*. Published online 2022. https://academic.oup.com/jamia/advance-article-abstract/doi/10.1093/jamia/ocac054/6569865

23. Hossain ME, Lister D, Bartolo C, et al. Prolonged Viral Shedding in Patients with Mild to Moderate COVID-19 Disease: A Regional Perspective. *Infect Dis*. 2021;14:11786337211010428.

24. Cao HR, Zhu XY, Zhou L, et al. Factors associated with delayed viral shedding in COVID-19 infected patients: A retrospective small-scale study. *Respir Med*. 2021;178:106328.

25. Lo IL, Lio CF, Cheong HH, et al. Evaluation of SARS-CoV-2 RNA shedding in clinical specimens and clinical characteristics of 10 patients with COVID-19 in Macau. *Int J Biol Sci*. 2020;16(10):1698-1707.

26. He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med*. 2020;26(5):672-675.

27. CDC. CDC museum COVID-19 timeline. Centers for Disease Control and Prevention. Published August 16, 2022. Accessed November 3, 2022. https://www.cdc.gov/museum/timeline/covid19.html

28. Pfaff ER, Madlock-Brown C, Baratta JM, et al. Coding Long COVID: Characterizing a new

disease through an ICD-10 lens. *medRxiv*. Published online September 2, 2022. doi:10.1101/2022.04.18.22273968

29. Eythorsson E, Runolfsdottir HL, Ingvarsson RF, Sigurdsson MI, Palsson R. Rate of SARS-CoV-2 Reinfection During an Omicron Wave in Iceland. *JAMA Netw Open*. 2022;5(8):e2225320.

30. 421-024-ReportedReinfections.pdf. https://doh.wa.gov/sites/default/files/2022-02/421-024-ReportedReinfections.pdf

31. Bielecki M, Züst R, Siegrist D, et al. Social Distancing Alters the Clinical Course of COVID-19 in Young Adults: A Comparative Cohort Study. *Clin Infect Dis*. 2021;72(4):598-603.

32. US Coronavirus vaccine tracker. USAFacts. Published November 15, 2022. Accessed November 16, 2022. https://usafacts.org/visualizations/covid-vaccine-tracker-states

33. Mueller AL, McNamara MS, Sinclair DA. Why does COVID-19 disproportionately affect older people? *Aging* . 2020;12(10):9959-9981.

34. Letizia AG, Ge Y, Vangeti S, et al. SARS-CoV-2 seropositivity and subsequent infection risk in healthy young adults: a prospective cohort study. *Lancet Respir Med*. 2021;9(7):712-720.

35. Slezak J, Bruxvoort K, Fischer H, Broder B, Ackerson B, Tartof S. Rate and severity of suspected SARS-Cov-2 reinfection in a cohort of PCR-positive COVID-19 patients. *Clin Microbiol Infect*. 2021;27(12):1860.e7-1860.e10.

36. Bechmann N, Barthel A, Schedl A, et al. Sexual dimorphism in COVID-19: potential clinical and public health implications. *Lancet Diabetes Endocrinol*. 2022;10(3):221-230.

37. Abdelrahman MM, Abd-Elrahman NM, Bakheet TM. Persistence of symptoms after improvement of acute COVID19 infection, a longitudinal study. *J Med Virol*. 2021;93(10):5942-5946.

38. Thompson AE, Anisimowicz Y, Miedema B, Hogg W, Wodchis WP, Aubrey-Bassler K. The influence of gender and other patient characteristics on health care-seeking behaviour: a QUALICOPC study. *BMC Fam Pract*. 2016;17:38.

39. *Utilization of Ambulatory Medical Care by Women: United States, 1997–98*. US Department of Health and Human Services; 2001. https://www.cdc.gov/nchs/data/series/sr_13/sr13_149.pdf

40. Zheng Y, Zhang Y, Chi H, et al. The hemocyte counts as a potential biomarker for predicting disease progression in COVID-19: a retrospective study. *Clin Chem Lab Med*. 2020;58(7):1106-1115.

41. Huang I, Pranata R, Lim MA, Oehadian A, Alisjahbana B. C-reactive protein, procalcitonin, D-dimer, and ferritin in severe coronavirus disease-2019: a meta-analysis. *Ther Adv Respir Dis*. 2020;14:1753466620937175.

42. Uscher-Pines L, Pines J, Kellermann A, Gillen E, Mehrotra A. Emergency department visits for nonurgent conditions: systematic literature review. *Am J Manag Care*. 2013;19(1):47-59.

43. Bots SH, Groenwold RHH, Dekkers OM. Using electronic health record data for clinical research: a quick guide. *Eur J Endocrinol*. 2022;186(4):E1-E6.