VSD Study 1344

Long COVID-19: Changes in Healthcare Utilization Following Infection with SARS-CoV-2

Protocol Version 2.0, December 7, 2022

Principal Investigator: Sara Tartof, PhD MPH, Department of Research & Evaluation, Division of Epidemiology

Co-Investigators: Lei Qian, PhD; Bruno Lewin, MD; Debbie Malden, DPhil, MSc; Lina Sy, MPH, Joshua Williams, MD, Simon Hambidge, MD, PhD, Jonathan Alpern, MD, Matthew Daley, PhD, Jennifer Nelson, PhD, David McClure, PhD, Ousseny Zerbo, PhD, Michelle Henninger, PhD, Candace Fuller, PhD, Eric Weintraub, MPH, Sharon Saydah, PhD

BACKGROUND AND SIGNIFICANCE

As of November 2022, more than 600 million people worldwide have been infected with COVID-19 and over 6.5 million have died (**https://covid19.who.int/**). To date, the United States has experienced the greatest absolute burden of COVID-19-associated cases and deaths worldwide, with more than 12.7 million cases and 260,000 deaths as of December 1st, 2020. Moreover, reported figures likely underestimate the true burden of COVID-19 due to either asymptomatic infection or mild illness. As such, it is estimated that for every case of COVID-19 in the US there are approximately 7-8 unreported infections.

Recent results from follow up studies have indicated that a substantial proportion of patients may continue to experience a range of lingering symptoms for many months following the initial acute phase of infection[1-6]. These have been termed “long haulers”, or persons with “long COVID”. Currently, there is no consensus definition of what defines the “long haul”, i.e. post-acute COVID-19 sequelae or “long COVID-19”. In the COVID Symptom Study, in which more than 4 million people in the US, UK and Sweden entered their symptoms after a COVID-19 diagnosis, post-acute COVID-19 was defined as the presence of symptoms extending beyond 3 weeks from the initial onset and chronic COVID-19 as extending beyond 12 weeks.

Estimates of the proportion of patients with COVID-19 who experience post-COVID conditions (PCC) range widely. One study, which has not yet been published, using an app created by King’s College London and Massachusetts General Hospital, found that of more than 4,000 patients with COVID-19 about 10% of those age 18 to 49 still experienced symptoms 4 weeks after becoming sick, 4.5% of all ages had symptoms lasting more than eight weeks, and 2.3% had them for more than 12 weeks.[7]

Other studies from the U.S. and Europe of non-hospitalized COVID patients have reported approximately 25-30% of patients still experience symptoms after 90 days. A study conducted by CDC found that 35% of symptomatic respondents reported residual symptoms at least 2 weeks after testing positive[8]. A study of hospitalized patients in Italy found that in patients who had recovered from COVID-

19, 87.4% reported persistence of at least one symptom, particularly fatigue and dyspnea, a mean of 60 days after initial symptom onset[1].

While other viral infections have been associated with persistence of symptoms[9], what differentiates COVID-19 persistence is the wide-ranging symptoms involving multiple organ systems. PCC can affect many parts of the body, including the heart, lungs, and the digestive and nervous systems[1,2,4,10]. The most commonly reported symptoms after acute COVID-19 infection are fatigue and dyspnea[4]. However, additional reported symptoms include cardiac inflammation, abdominal pain and diarrhea, anosmia, a condition resembling chronic fatigue syndrome, "brain fog" characterized by difficulty with concentration and memory, psychiatric disorders such as anxiety and depression[11], and dysautonomia.

The PCC outcomes are diverse and complex and may vary by severity of the acute phase of COVID-19 infection or pre-existing health status prior to infection. While it is possible that PCC impacts all age groups, most studies have been restricted to adult populations. Furthermore, the majority of evidence around PCC has been gathered from follow-up studies of previously hospitalized patients that often lack information on symptom history before acute COVID-19 illness, or details on symptom severity[1]. Furthermore, most are single-center studies with information on a relatively small number of hospitalized patients, and many lack unbiased comparison groups of patients discharged for other reasons, limiting the interpretation of the reported associations. More studies are needed to better characterize the duration, frequency, and types of PCC.

The current study will address these limitations by evaluating PCC in a very large, multisite, demographically and clinically diverse population of all ages with COVID-19 identified in both outpatient and inpatient settings.

Two-Stage Approach

We propose a 2-stage approach that will maximize our potential to publish important findings on PCC as quickly as possible, while also diving more deeply into research questions that may incur a longer timeline.

For Phase I, we will conduct a more rapid multi-site interrupted time series (ITS) study to assess changes in high-level utilization counts in the 3 months (and ultimately 6 months and longer, when more follow-up time has accrued) post-index date versus a pre-index date time period in 2019[12]. To adjust for secular confounding (e.g. impact of COVID-19 pandemic), we will compare pre-post utilization among patients with a COVID-19 positive laboratory test versus a COVID-test-negative matched population. We will further stratify by characteristics of interest to explore sub-groups at highest risk of PCC.

For Phase II, we propose a retrospective cohort study to investigate the association between prior COVID-19 vaccination status and risk of incident PCC among individuals who test positive for SARS-CoV-2.

STUDY AIMS

PHASE I Assess the 6-month healthcare utilization associated with COVID-19

Aim 1a.) We will describe the trajectory of weekly counts (i.e. counts over time) of utilization (combined and stratified by outpatient, inpatient, ED, virtual settings) at each participating site, and by age, sex,

and race/ethnicity, over the study period for those in the study population with a COVID-19 positive laboratory test and a COVID-test-negative matched population, starting from the time of COVID-19 test. We will use the encounter distributions to verify the approach to define the index date. Next, we will focus on utilization across settings for the 6 months following index dates, as well as corresponding (by calendar month and day) pre-periods in 2019.

Aim 1b.) Use an interrupted time series (ITS) approach to estimate the difference in changes in utilization, by healthcare setting (i.e. stratified by outpatient, inpatient, ED, virtual; and outpatient vs. inpatient), in the pre-period and post-period for those with a COVID-19 positive test compared to a COVID-test-negative matched population.

Aim 1c.) Use ITS approach to estimate the difference in changes in utilization (all settings combined) in the pre- vs. post-period for those with a COVID-19 positive test vs. a COVID-test-negative matched population, stratified by age, sex, race/ethnicity and clinical characteristics of interest.

Aim 1d.) Use ITS approach to estimate the difference in changes in utilization (all settings combined) in the pre- vs. post-period for those with a COVID-19 positive test vs. a COVID-test-negative matched population, stratified by select PCC types.

a.) Conduct analyses for 3 and 6-month follow-up periods. We will examine longer time periods if data are available.

PHASE II

Aim 2) Assess the association between prior COVID-19 vaccination status and risk of incident PCC among individuals with SARS-CoV-2 infection

a.) Conduct retrospective cohort analyses, stratified by calendar period (pre- and post-Omicron variant), age group (<18, 18-64 and ≥65 years) and severity of SARS-CoV-2 infection (hospitalization yes/no).

APPROACH

Aim 1

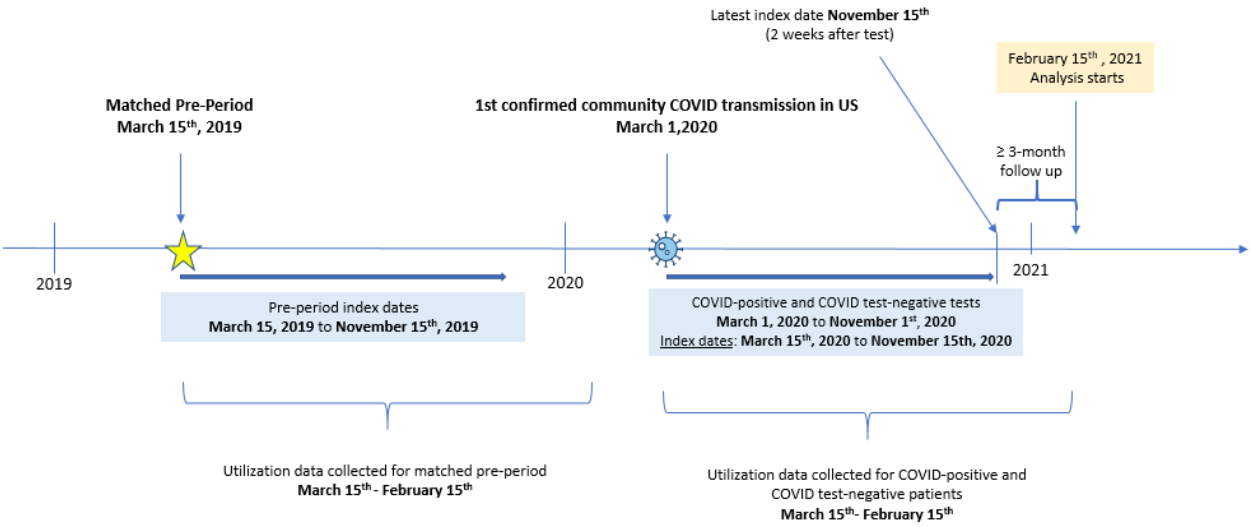
Study Design: Retrospective cohort study.

Environment: VSD multi-site study.

Study Period: We will include utilization, comorbidity, clinical, and other data on COVID-19 cases (defined by positive SARS-CoV-2 laboratory test) and the COVID-test-negative matched population from March 1, 2020, to February 15, 2021, and from March 15, 2019 to February 15, 2020. For complete assessment of utilization, we will also require continuous membership (allowing for a 31-day gap) from 12 months prior to the index date and 3 months (and ultimately 6 months or more when more time accumulates, which will extend the end of follow-up to 5/15/2021 or later) after the index date. The COVID-test-negative matched population will be defined as those that have a negative COVID-19

diagnostic test from March 1, 2020, through 3 months (and ultimately 6 months) following their index date.

Figure 1. Timeline for Cohort Construction



The study timeline is as follows: an index date is defined as 14 days following the SARS-CoV-2 test date during March 1st - Nov 1st 2020. For the analysis, a matching pre-period was used to estimate a difference-in-difference effect size among positive SARS-CoV-2 cases compared with the test-negative cohort.

Index date: The index date will be defined as the date 2 weeks following the date of positive SARS-CoV-2 laboratory test. This allows a 2-week washout window immediately post-positive lab test. For example, if a patient is diagnosed with COVID-19 by positive lab test on May 1st, the index date for this patient will be May 15th. The COVID-19-test-negative matched population may be matched to COVID-19 positive patients on VSD site, age, sex and race/ethnicity, and date of test, based on data availability.

Study Cohort: The cohort will include patients of all ages who have a positive laboratory test for SARS-CoV-2 from March 1, 2020 to November 1, 2020 and those who test negative to SARS-CoV-2, matched by test date to a COVID positive case. We will restrict the study to patients with positive COVID-19 tests until November 1, 2020, to allow at least 3-months of follow-up. This date assumes that study analyses will begin no earlier than February 15, 2021.

The COVID-19-test-negative population will include patients matched to COVID-19-positive patients using the matching criteria described above.

Covariate Data to Collect: Although our interrupted time series approach removes the need to adjust for comorbidities (pre-post comparison), it is possible that the COVID-19-positive population differs from the COVID-19-negative population on prevalence of certain factors that are also associated with utilization. Therefore, we plan to collect comorbidity data in the 12 months prior to the index date, including demographic characteristics (age, sex, race/ethnicity, Medicaid), prior healthcare utilization (outpatient, inpatient, ED settings 12 months prior), clinical comorbidities (cancer, chronic kidney

disease, cardiac disease, organ transplant, sickle cell disease, diabetes I and II, down syndrome, asthma, cardiovascular disease, cystic fibrosis, hypertension, liver disease, fibrosis, thalassemia, immunodeficiency, and dementia [comorbidities included according to CDC COVID-19 high risk category definition, see Excel file]; and obesity based on body mass index [BMI]), and other potential covariates of interest for stratified subgroup analyses. In addition, for analysis of 6-month follow-up period, we will collect COVID-19 vaccine information among the COVID-19-positive and COVID-19-negative population since the follow-up period and vaccination period will possibly overlap.

STUDY AIMS

Aim 1a: ***We will describe the trajectory of weekly counts (i.e. counts over time) of utilization (combined and stratified by outpatient, inpatient, ED, virtual settings) at each participating site, and by age, sex, and race/ethnicity, over the study period for those in the study population with a COVID-19 positive laboratory test and a COVID-test-negative matched population, starting from the date of COVID-19 test. We will use the encounter distributions to verify the approach to define index date. Next, we will focus on utilization across settings for the 3 months following index dates, as well as corresponding (by calendar month and day) pre-periods in 2019.***

For this Aim we will first identify all patients with a positive SARS-CoV-2 laboratory test from March 1, 2020 to November 1st, 2020. We will then match these COVID-positive patients to those with test-negative-COVID-19 results on VSD site, age, sex, race/ethnicity, and test date based on data availability. We will describe the trajectory of weekly counts (i.e. counts over time) of utilization (combined and stratified by outpatient, inpatient, ED, virtual settings) at each participating site, and by age, sex, and race/ethnicity, over the study period for those in the study population with a COVID-19 positive laboratory test and a COVID-test-negative matched population.

We propose an index date defined as the date 2 weeks following the date of positive or negative SARS-CoV-2 laboratory test. However, with these first descriptive analyses, we plan to verify this approach with utilization data and therefore plan to capture utilization starting at the time of the laboratory test, rather than only 14 days after.

Next, we will capture utilization data from the identified COVID-19 positive and COVID-19 negative patients during 3 months after a corresponding index date in 2019. We will also estimate and describe the trajectory of weekly counts (i.e. counts over time) of utilization in 2019 (combined and stratified by outpatient, inpatient, ED, virtual settings) at each participating site, and by age, sex, and race/ethnicity.

Based on the index date, which we will initially define as the date 2 weeks following the date of positive or negative SARS-CoV-2 laboratory test, we will define a 3-month post-index time window and a 3-month pre-index time window for each patient. The 3 months in the pre-period will correspond to those in the post-period: for example, if a patient tests positive for COVID-19 on May 1, 2020, the post-index date period will cover May 15th – August 15th. The pre-index date period will be May 15th – August 15th of 2019. This logic will also be applied to the COVID-19-negative patients, anchored on their index date.

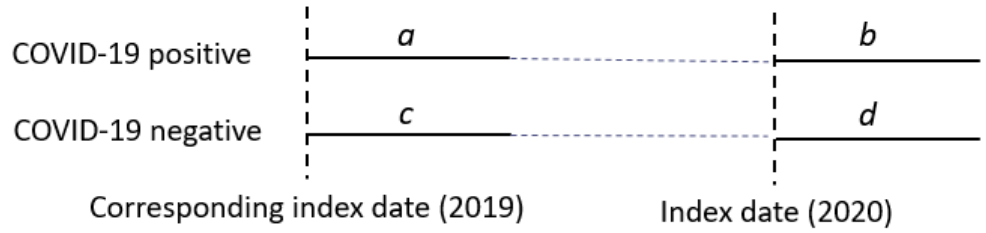
During the 3-month pre- and post-index date periods of both COVID-19-positive and COVID-19-negative patients, we will explore the utilization trajectory and possible secular clusters after COVID-19 infection. We will describe the clinical and demographic characteristics of COVID-19 positive and COVID-19 negative patients.

Aim 1b.) ***Use an interrupted time series (ITS) approach to estimate the difference in changes in utilization, by healthcare setting (i.e. stratified by outpatient, inpatient, ED, virtual; and outpatient vs. inpatient), in the pre-period and post-period for those with a COVID-19 positive test compared to a matched COVID-test-negative population.***

For this aim we will conduct interrupted time series analyses, as illustrated below. Specifically, we plan to use this design because inherent adjustment for secular confounding, including various restrictions on healthcare utilization, other community lockdown measures, and changing rates of SARS-CoV-2 infection over time are built into the design. First, we will start with a Difference-in-Difference (DiD) approach. It is the simplest approach in the ITS family and will be an informative first step, comparing the average encounter rate during the assessment period[13]. The illustration of interrupted time series design and parameter estimates can be found in **Figure 2** and **Table 1.** ITS may model a linear trajectory or even a smoothing spline curve for utilization rate; we will therefore plot encounter rates over time and explore secular trends prior to determining the final model.

Once the final model is identified, we will conduct stratified analyses by individual setting (outpatient, inpatient, ED, virtual), and by inpatient vs. outpatient setting.

Figure 2. Difference-in-Difference Approach as a Special Case of the Interrupted Time Series Design with a Control Group



The difference-in-difference study design involves estimating the difference in changes in healthcare utilization in the post- vs. pre-period for those with a COVID-19 positive test (parameters b & a, respectively) vs. a COVID-test-negative matched population (d & c, respectively).

Table 1. Parameter Estimates for the Difference-in-Difference Approach

Parameter estimate	2019	2020	Rate Ratio (RR) or Ratio of Rate Ratio (RRR) calculations
Rate among COVID-19 positive	*a*	*b*	$RR_1 = b/a$
Rate among COVID-19 negative	*c*	*d*	$RR_2 = d/c$
Ratio of Rate Ratio (RRR)	-	-	$RRR = bc/ad$

Aim 1c.) ***Use ITS approach to estimate the difference in changes in utilization (all settings combined) in the pre- vs. post-period for those with a COVID-19 positive test vs. a COVID-test-negative matched***

population, stratified by age, sex, race/ethnicity and clinical characteristics of interest (i.e. comorbidities, prior influenza and/or pneumococcal vaccination, etc.).

This aim will provide high-level insights into populations that may be at higher risk for PCC. We will conduct similar ITS analyses as Aim 1b stratified by race/ethnicity, age group, sex, and other limited variables, depending on statistical power. We will have a flag for care setting if additional analyses are needed by setting.

Aim 1d.) ***Use ITS approach to estimate the difference in changes in utilization (all settings combined) in the pre- vs. post-period for those with a COVID-19 positive test vs. a COVID-test-negative matched population, stratified by PCC type (i.e. neurologic, respiratory, gastrointestinal, cardiac, psychiatric, other).***

We will conduct analyses for 3 and 6-month follow-up periods. We will examine longer time periods if data are available.

Similar to **Aim 1c)**, this aim will provide high-level insights into the burden and frequency of each type of PCC outcome. We will first use a list of 44 PCC outcomes that includes conditions such as arrhythmias, autonomic disfunction, depression, etc. We may group these conditions into broader categories based on the affected organ system such as neurologic, respiratory, gastrointestinal, cardiac, psychiatric, and other. To define utilization by type, we will identify all visits that have at least 1 ICD-10 code that falls within each type. For example, for autonomic dysfunction, we will define all visits (from all settings) that have at least 1 code related to autonomic dysfunction in the pre- and post-periods for both COVID-19 positive and negative matched controls. We will describe the clinical and demographic characteristics of those with each type of COVID-19 PCC and will then conduct similar ITS analyses as Aim 1b, stratified by PCC type. We will have a flag for care setting if additional analyses are required by setting. For analysis of 6-month follow-up period, we will conduct a sensitivity analysis by excluding the follow up time after receiving the COVID-19 vaccine.

Potential Limitations to Phase I

There are potential limitations to this study. First, we make assumptions that utilization occurring in all settings in the 3- and 6-months post COVID-19 disease are potentially related to, or attributable to COVID-19, as we are not requiring a COVID diagnosis code at every encounter. We do not believe that coding is occurring reliably across settings and across time. To address this concern, we use ITS and matched cohort designs to assess the incremental utilization associated with COVID-19. Second, we assume that comorbidity status in the 12 months prior to index date will be similar to the distribution in the post-period, although status may change during this time period. Third, the length of PCC may differ by system/symptoms. Further, PCC of a particular type may be longer in time than what our analyses allow given limitations on follow-up time. Fourth, there might be a delay from time of symptom onset to appointment time, so we may overestimate the duration of symptoms, as reflected by healthcare utilization. Fifth, some patients with COVID-19 may not be diagnosed, resulting in misclassification. Sixth, while the CDC includes smoking and pregnancy as potential risk conditions for COVID-19, we do not include these in our study.

PHASE II

For Phase II, we aim to assess the association between prior COVID-19 vaccination status and risk of incident Post-Covid Conditions (PCC) among individuals with SARS-CoV-2 infection. A small number of population studies have attempted to explore this association previously,[14-16] however most have not included a full 6-months of follow-up,[15] were not able to distinguish between hospitalized and non-hospitalized patients, or contained low numbers of vaccinated individuals.[16] Of note, the study objective of the proposed study is not to assess the impact of vaccination on patients who are already experiencing PCC outcomes (i.e., we are assessing the effect of vaccination on incident PCC).

Methods

Study population

To assess the association between prior COVID-19 vaccination status and PCC over 6 months, we plan to select participants of all ages with at least one positive SARS-CoV-2 test (PCR or antigen) administered across all care settings from March 1st, 2021, to February 28th, 2022, and 1-year continuous health plan membership (allowing for a 31-day administrative gap) prior to and 30 days after the SARS-CoV-2 positive test date. This study period was selected because a large proportion of the US population had received a COVID-19 vaccination and a sufficiently large proportion of all SARS-CoV-2 infections to date had occurred over this period, including during a period when the Omicron variant pre-dominated transmission in the US. **Figure 3** displays the infection rates in California as an example. Further, the study period covers time periods during which the main variants of concern were circulating in the US. For the main analyses, we plan to define SARS-CoV-2 positive patients into two exposure groups:

1. **Vaccinated patients** at the time of documented SARS-CoV-2 positive test (i.e., breakthrough infection), defined as patients with at least one dose of COVID-19 vaccine documented in the EHR. Vaccination records starting from December 14, 2020, will be used to define COVID-19 vaccination status. In analysis is stratified by dose, the most recent dose at least 14 days prior to the documented SARS-CoV-2 infection will be defined as follows:
    a. **Dose 1**: patients with documented SARS-CoV-2 infection at least 14 days following receipt of the first dose of COVID-19 vaccine.
    b. **Dose 2**: patients with documented SARS-CoV-2 infection at least 14 days following receipt of a second dose of COVID-19 vaccine.
    c. **Dose 3 or more:** patients with documented SARS-CoV-2 infection at least 14 days following a third dose or greater of COVID-19 vaccine.

    For participants with multiple doses documented within 7 days apart, the latter vaccination records will be treated as erroneous data entry (duplicates), and the earliest dose will be included in the analysis.

2. **Unvaccinated patients**, defined as patients without a documented COVID-19 vaccination at least 14 days prior to the date of SARS-CoV-2 positive test.

Exclusion criteria

- SARS-CoV-2 positive tests among persons that received a COVID-19 vaccine within 14 days prior to or within 30 days after their SARS-CoV-2 positive test date. The rationale for this exclusion criteria is as follows: i) for patients that received vaccinations within 14 days prior to the date of SARS-CoV-2 infection, insufficient time may have elapsed to mount a full immune response and to consider that individual "vaccinated" prior to infection; ii) since follow-up for PCC outcomes begins 30 days after the SARS-CoV-2 positive test date and we want to define exposure status prior to the SARS-CoV-2 positive test date, we exclude individuals who received a COVID-19 vaccine within 30 days after their SARS-CoV-2 positive test date.
- Patients with missing information on age or sex, since these are important potential confounders required for adjustment.
- Persons with documented receipt of COVID-19 vaccination not routinely administered in the United States, or any dose of Janssen (Johnson & Johnson) COVID-19 vaccine. This exclusion criterion was due to the potential heterogeneous immunological effects of other vaccines that will be difficult to account for in the analysis. In addition, mRNA vaccines are recommended over Janssen vaccine due to safety concerns. The different number of doses in the primary series for mRNA versus Janssen vaccine complicates the interpretation of number of vaccine doses.[17] Furthermore, following exploratory analysis among KPSC members, the proportion of study participants receiving these types of vaccines is thought to be very small.

Figure 3. 7-day average COVID-19 infections per 100 000 people and dominating circulating SARS-CoV-2 variants in California*

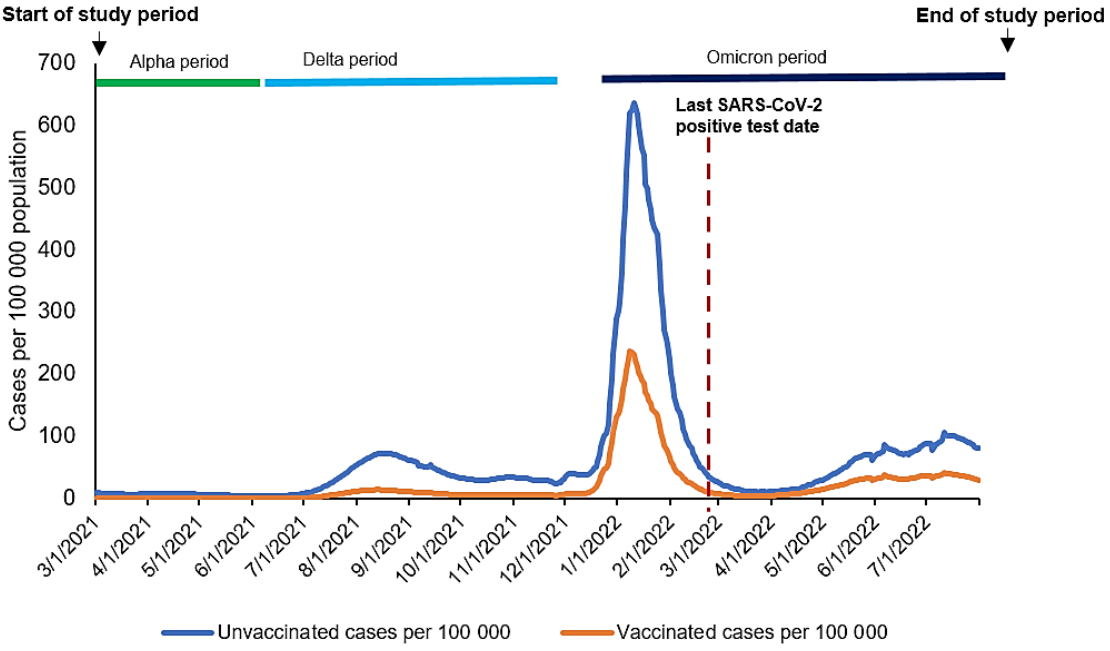


Figure shows the distribution of SARS-CoV-2 cases over time in California, by vaccination status (unvaccinated/vaccinated). The absolute number of cases per 100,000 persons is consistently higher among unvaccinated individuals compared with vaccinated individuals. There is a large spike in cases during the start of the Omicron period, around January 2022.

*Data from California Department of Public Health: https://data.chhs.ca.gov/dataset/covid-19-time-series-metrics-by-county-and-state (Accessed 9/26/2022)

Index date

The index date will be defined as the date of positive SARS-CoV-2 laboratory test from March 1st, 2021, to February 28th, 2022 (**Figure 3**). A person may be included in the analysis more than once if a re-infection occurs over the study period, defined as a SARS-CoV-2 positive test (PCR or antigen) occurring at least 90 days from the previous SARS-CoV-2 positive test date.

Outcomes

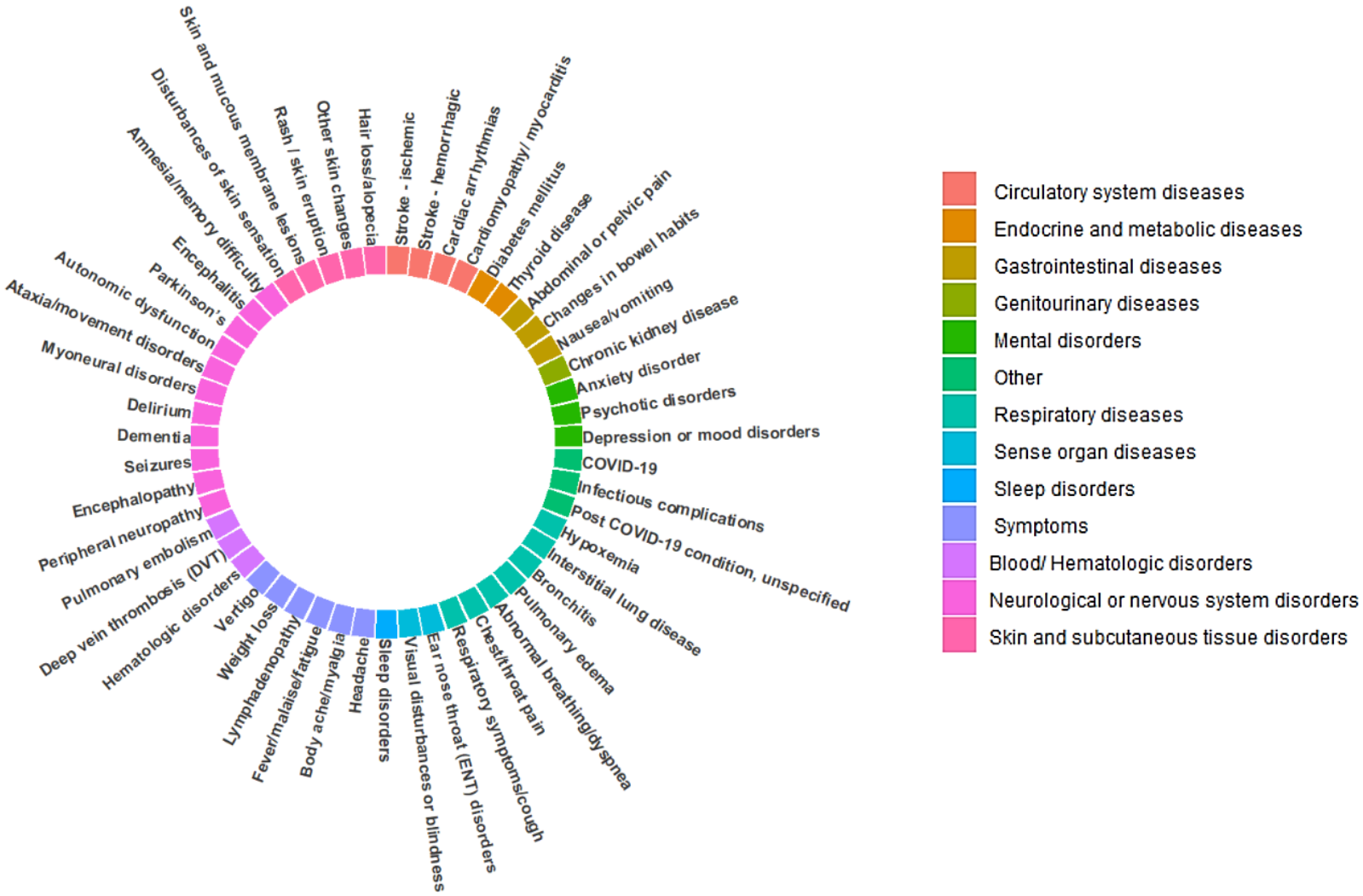
Study participants will be followed up for 6 months from the date of documented SARS-CoV-2 infection occurring within the study period. The primary outcomes will be defined as any new-onset PCC outcomes occurring ≥30 days after the SARS-CoV-2 positive test. PCC outcomes will include a list of 51 outcomes broadly categorized into 13 disease categories (**Figure 4; Appendix A**). When defining new-onset PCC, different lookback periods will be applied to define pre-existing illness for each PCC outcome separately (**Appendix A**). If an individual has EHR documentation of a pre-specified outcome in the defined lookback period, this outcome will not contribute to the analysis over follow-up (i.e., restricting the study to new-onset PCC events only). Individual PCC sub-conditions will be defined according to a pre-specified list of diagnosis codes (International Classification of Diseases, 10th Revision [ICD-10] codes). In an effort to standardize our approach with others, we performed a rapid review of the literature to identify large-scale studies that investigated PCC using EHR data between vaccinated and unvaccinated test-positive groups (**Appendix B**).[18-20] In summary, the PCC category and ICD-10 codes used across studies were similar to those used for Phase I of the current study,[21] which themselves were developed in consultation with CDC and were derived from a combination of self-reported survey- and EHR- based studies that displayed evidence of an association with COVID-19 after the acute stage of illness.

Where disease categories display statistically significant associations with vaccination status, individual sub-conditions will be included in the analysis as secondary outcomes. For example, if 'circulatory system disorders' is significant, all sub-conditions within this category (i.e., arrythmias, myocarditis/pericarditis etc.) will be assessed as secondary outcomes. Where power permits, severe PCC categories will also be assessed as secondary outcomes, defined as an inpatient encounter with a PCC diagnosis code documented as the primary diagnosis for the encounter.

Follow-up

Participants will contribute person-time to the analyses from 30 days following the date of the SARS-CoV-2 positive test (i.e. the index date). This allows a 30-day washout window immediately post-positive lab test. For example, if a patient tests positive for SARS-CoV-2 on May 1st, 2021, follow-up for this patient will start on May 31st, 2021. Follow-up will continue for 5 months after this date, corresponding to 6 months following the date of the SARS-CoV-2 positive test (**Figure 3**). Otherwise, where applicable, patients will be censored at the date of receipt of an additional dose of COVID-19 vaccine (or first dose, if previously unvaccinated), termination of health plan membership, or death (whichever occurs first). For patients with an additional SARS-CoV-2 positive test occurring over the study period and at least 90 days following the initial SARS-CoV-2 positive test, follow-up will restart from 30 days following the additional positive SARS-CoV-2 test.

Figure 4. Post-COVID Conditions (PCC) outcomes and corresponding categories



The 51 post-COVID conditions are distributed widely across 13 body systems.

Statistical analysis

Characteristics of the study population will be described using mean and standard deviation for continuous variables and frequency and percentage for categorical variables. Differences between vaccinated and unvaccinated groups will be compared using an independent *t* test or χ^2 test. Absolute standardized difference will be calculated to assess the balance of covariates.

Poisson regression analysis with robust variance (GEE models) will be used to assess the association of PCC categories with vaccination status (exposure). Relative risk (RRs) estimates and corresponding 95% confidence intervals will be calculated. Models will also be adjusted for VSD site, date of SARS-CoV-2 positive test (+/-calendar week), age, sex, severity of infection (hospital admission within 7 days of SARS-CoV-2 positive test), prior SARS-CoV-2 infection (defined as a documented SARS-CoV-2 positive test at least 90 days prior to the index date), healthcare utilization in the year prior SARS-CoV-2 positive test date (number of inpatient visits, number of ED visits, number of outpatient and virtual visits), time since most recent dose of vaccine (<90, 90-180, >180 days), number of vaccine doses received (1-3+), influenza vaccination in the 2 years prior, Medicaid status, race/ethnicity (Hispanic, Black, Asian, White and Other/Unknown), and Charlson comorbidities in the year prior to the index date.

We will perform pre-specified subgroup analysis on the primary PCC outcomes by age group (<18, 18-64, ≥65 years), since prior evidence has suggested that PCCs differ by age, particularly among children compared with adult populations.[22] We will also stratify the analysis by calendar study period in intervals approximately corresponding to the predominance of selected variants of concern in the US population (i.e. March 1st 2021 – November 30th 2021 [pre-Omicron]; December 1st 2021 – February 28th 2022 [Omicron]), as displayed in **Figure 3**. Additionally, we will stratify the analysis by most recent vaccine dose received (1,2, 3+) and by severity of SARS-CoV-2 infection (hospitalization within 7 days of index date). Statistical significance will be considered as two-sided p-values <0.05. Bonferroni correction will be used to correct for multiple testing.

Sensitivity analysis: Matched cohort

As a sensitivity analysis, vaccinated (all doses) and unvaccinated patients will be matched (1:N, depending on data availability) on VSD site, date of SARS-CoV-2 positive test (+/-calendar week), age, sex, and severity of infection (hospital admission within 7 days of SARS-CoV-2 positive test). Additional variables will be included as adjustment variables in the matched models, including prior SARS-CoV-2 infection, number of vaccine doses received, time since most recent dose, race/ethnicity, comorbidity status, Medicaid status, influenza vaccination, and healthcare utilization in the year prior.

Potential limitations: Phase II

There are a number of potential limitations that are specific to Phase II of the proposed study. -First, although the proposed analysis is designed to assess the relative risk of PCC outcomes between vaccinated and unvaccinated test-positive cohorts, the study is not able to assess the absolute incidence of PCC outcomes. Although the PCC outcome codes captured in the 6 months after the SARS-CoV-2 positive test date may be temporally associated with infection, they might not necessarily be

attributable to COVID-19. Second the study population may not be representative of the general US population. Similarly, SARS-CoV-2 testing availability varied over time, and during periods of low testing capacity, may have been directly proportional to need or vulnerability such as age or other potential confounders. However, since the analyses adjust for calendar date of SARS-CoV-2 positive test and age, these factors are accounted for when comparisons are made between vaccinated and unvaccinated individuals. Third, PCC outcome captured in EHR may not be complete. Indeed, previous large-scale studies have used natural language processing to extract additional PCC outcomes from the EHR of adults with COVID-19.[23] However, this is unlikely to impact the main findings of the current study, unless the contribution of NLP to EHR-derived outcomes differs between the unvaccinated and vaccinated population. EHR may also be incomplete for SARS-CoV-2 infection status, particularly for milder cases, or those occurring during later study periods when at-home rapid diagnostic tests were widely available. Finally, the planned analyses will not account for COVID-19 treatment, which may be associated with initial disease severity (particularly in inpatient care settings) and affect subsequent PCC outcomes. However, effective antiviral drugs for the treatment of mild-to-moderate COVID-19 were not widely prescribed during the study period, and hospital admission is a sufficient proxy measure of inpatient treatment.[24]

Data Management: Phase I & II

Data files: CONSTANT, ENROLL, INPT, OUTPT, VACCINE from VSD DDF or cycle files, ancillary COVID-19 DxID and COVID-19 lab data, and ancillary HTWT files. Individual-level data from each VSD site will be required to complete the analysis for the proposed study. For example, for phase II: i) Poisson regression analysis with robust variance requires individual-level data; ii) The proposed sensitivity analysis of 1:1 matching requires individual-level data; and iii) The analysis requires the index date, timing of vaccination, timing of outcomes and timing of censoring events, all of which are individual-level attributes. Given the large number of exposure (vaccination) patterns, covariates and outcomes, aggregated data is not feasible. However, where possible, covariates will be based on derived variables instead of extracting raw data.

Chart review (Phase I): We will likely conduct preliminary chart abstractions at KPSC (approximately 100 reviews total) as part of this study, primarily to determine whether providers attribute the neurologic/respiratory/gastrointestinal/cardiac/psychiatric visits to COVID-19. If needed, we can expand chart review to additional sites.

Data management: The VSD team at KPSC will be responsible for data management activities, including data extraction and consolidation between sites, study documentation and archival. All electronic documents, KPSC data sets, and files relevant to the project will be stored on KPSC computers, which have restricted access. Currently, we propose this study as a multi-site study. SAS programs will be developed at KPSC and sent to participating sites for approval prior to data extraction. Site data managers will be responsible for developing an ancillary HTWT files.

Human Subjects Protection

Human subjects: Privacy and confidentiality will be strictly protected according to VSD standard procedures. There will be minimal risks to patient privacy and confidentiality. All information will be stored on secure KPSC computers and at participating sites. This study will be covered under KPSC's

umbrella VSD IRB approval, which includes a waiver for the requirement to obtain HIPAA authorizations. Participating sites may seek IRB approval as needed.

Timeline for Phase I

December 2020	Discuss concept on December VSD Project Call; submit full proposal to CDC and sites for comment and invite sites to participate
March 2021	Revise proposal based on reviewer comments; circulate DDM programs for review and approval
March 2021	Creation of ancillary HTWT file at participating sites
March-April 2021	Data extraction and cleaning, limited chart review
April-May 2021	Data analyses (3 month follow up)
June 2021	Data refresh for 6-month follow up
January-March 2022	Manuscript preparation, updated data analysis
June 2022	Submit manuscript to CDC for clearance; submit manuscript for publication
July 2022	Publication: Phase I

Estimated timeline for Phase II

September 2022	Discuss concept for Phase II on September VSD Project Call
November 2022	Submit full proposal to CDC and sites for comment and invite sites to participate; Revise proposal based on reviewer comments
December 2022	Circulate DDM programs for review and approval
December 2022- January 2023	Data extraction and cleaning
February-April 2023	Data analysis Phase II
April – May 2023	Draft manuscript
May 2023	Send manuscript to co-authors for review
June 2023	Co-author review of manuscript and revisions
July 2023	Submit manuscript to CDC for clearance; submit manuscript for publication

REFERENCES

1. Carfi A, Bernabei R, Landi F, Gemelli Against C-P-ACSG. Persistent Symptoms in Patients After Acute COVID-19. *JAMA.* 2020;324(6):603-605.
2. Carvalho-Schneider C, Laurent E, Lemaignen A, et al. Follow-up of adults with noncritical COVID-19 two months after symptom onset. *Clin Microbiol Infect.* 2020.
3. Datta SD, Talwar A, Lee JT. A Proposed Framework and Timeline of the Spectrum of Disease Due to SARS-CoV-2 Infection: Illness Beyond Acute Infection and Public Health Implications. *JAMA.* 2020;324(22):2251-2252.

4. Del Rio C, Collins LF, Malani P. Long-term Health Consequences of COVID-19. *JAMA.* 2020;324(3):1723-1724.
5. Halpin SJ, McIvor C, Whyatt G, et al. Postdischarge symptoms and rehabilitation needs in survivors of COVID-19 infection: A cross-sectional evaluation. *J Med Virol.* 2021;93(2):1013-1022.
6. Huang C, Huang L, Wang Y, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet.* 2021;397(10270):220-232.
7. Toy S, Reddy R, Hernandez D. Doctors Begin to Crack Covid's Mysterious LongTerm Effects; Severe fatigue, memory lapses, heart problems affect patients who weren't that badly hit initially; 'It's been so long'. *Wall Street Jounal* Nov 1, 2020.
8. Tenforde MW, Kim SS, Lindsell CJ, et al. Symptom Duration and Risk Factors for Delayed Return to Usual Health Among Outpatients with COVID-19 in a Multistate Health Care Systems Network - United States, March-June 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69(30):993-998.
9. Datta SD, Talwar A, Lee JT. A Proposed Framework and Timeline of the Spectrum of Disease Due to SARS-CoV-2 Infection: Illness Beyond Acute Infection and Public Health Implications. *JAMA.* 2020.
10. National Center for Immunization and Respiratory Diseases (U.S.),Division of Viral Diseases. Late Sequelae of COVID-19. **https://stacks.cdc.gov/view/cdc/97200**. Accessed January 22, 2021.
11. Taquet M, Luciano S, Geddes JR, Harrison PJ. Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort studies of 62 354 COVID-19 cases in the USA. *Lancet Psychiatry.* 2020.
12. Hechter RC, Qian L, Yan S, Luo Y, Krishnarajah G, Tseng HF. Impact of the change of copay policy in Medicare Part D on zoster vaccine uptake among Medicare beneficiaries in a managed care organization. *BMC Health Serv Res.* 2017;17(1):503.
13. Lopez Bernal J, Cummins S, Gasparrini A. Difference in difference, controlled interrupted time series and synthetic controls. *Int J Epidemiol.* 2019;48(6):2062-2063.
14. Antonelli M, Pujol JC, Spector TD, Ourselin S, Steves CJ. Risk of long COVID associated with delta versus omicron variants of SARS-CoV-2. *Lancet.* 2022;399(10343):2263-2264.
15. Notarte KI, Catahay JA, Velasco JV, et al. Impact of COVID-19 vaccination on the risk of developing long-COVID and on existing long-COVID symptoms: A systematic review. *E Clinic Med.* 2022;53:101624.
16. Zisis SN, Durieux JC, Mouchati C, Perez JA, McComsey GA-O. The Protective Effect of Coronavirus Disease 2019 (COVID-19) Vaccination on Postacute Sequelae of COVID-19: A Multicenter Study From a Large National Health Research Network. (2328-8957 (Print)).
17. *Interim Clinical Considerations for Use of COVID-19 Vaccines: Appendices, References, and Previous Updates* **https://www.cdc.gov/vaccines/covid-19/clinical-considerations/interim-considerations-us-appendix.html#appendix-a**. Accessed Nov 10, 2022.
18. Al-Aly Z, Bowe B, Xie YA. Long COVID after breakthrough SARS-CoV-2 infection. *Nat Med.* 2022;28:1461–1467.
19. Taquet M, Dercon Q, Harrison PJ. Six-month sequelae of post-vaccination SARS-CoV-2 infection: A retrospective cohort study of 10,024 breakthrough infections. *Brain Behav Immun.* 2022;103:154-162.
20. Zisis SN, Durieux JC, Mouchati C, Perez JA, McComsey GA. The Protective Effect of Coronavirus Disease 2019 (COVID-19) Vaccination on Postacute Sequelae of COVID-19: A Multicenter Study From a Large National Health Research Network. *Open Forum Infectious Diseases.* 2022;9(7):ofac228.

21. Tartof SY, Malden DE, Liu IA, et al. Health Care Utilization in the 6 Months Following SARS-CoV-2 Infection. *Jama Netw Open.* 2022;5(8).
22. Fischer HA-O, Qian L, Skarbinski J, et al. Development and validation of a prediction algorithm to identify birth in countries with high tuberculosis incidence in two large California health systems. *PLoS One.* 2022;17(8)(1932-6203).
23. Pfaff ER, Girvin AT, Bennett TD, et al. Identifying who has long COVID in the USA: a machine learning approach using N3C data. *Lancet Dig Health.* 2022;4:E532-E541.
24. Gold JA KJ, Magid J, et al. Dispensing of Oral Antiviral Drugs for Treatment of COVID-19 by Zip Code–Level Social Vulnerability — United States, December 23, 2021–May 21, 2022. *MMWR Morb Mortal Wkly Rep.* 2022;71:825-829.

Appendix A. Pre-selected disease categories and ICD-10 codes used to define Post COVID Conditions

Endocrine and metabolic diseases

Disease category	ICD-10 code	Look-back period (prior to index date)
Diabetes mellitus (Type I and II)	E10, E11	12 mo
Thyroid disease	E03, E06	12 mo

Circulatory system diseases

Disease category	ICD-10 code	Look-back period (prior to index date)
Stroke – ischemic or unspecified	I63, I64, I69, G45, G46	12 mo
Stroke - hemorrhagic	I60, I61, I62	12 mo
Cardiac arrhythmias/ Postural orthostatic tachycardia syndrome (POTS)	R00, I47, I48, I49	12 mo
Cardiomyopathy, myocarditis, endocarditis	I30, I40, I51.4, B33, I51.81	12 mo

Blood/ Hematologic disorders

Disease category	ICD-10 code	Look-back period (prior to index date)
Hematologic disorders	D72.81, D69	12 mo
Deep vein thrombosis (DVT)	I82	12 mo
Pulmonary embolism	I26	12 mo

Respiratory diseases

Disease category	ICD-10 code	Look-back period (prior to index date)
Interstitial lung disease	J84	12 mo
Bronchitis	J20, J40, J41, J42	30 d
Pulmonary edema	J81	12 mo
Abnormal breathing/dyspnea	R06	30 d
Chest/throat pain	R07	30 d
Respiratory symptoms/cough	R05	30 d
Hypoxemia	R09.02	12 mo

Mental disorders

Disease category	ICD-10 code	Look-back period (prior to index date)
Anxiety disorder	F40, F41, F42, F43, F44, F45, F48, R45	12 mo
Psychotic disorders	F20, F21, F22, F23, F24, F25, F28, F29	12 mo
Depression or mood disorders	F30, F31, F32, F33, F34, F38, F39	12 mo

Neurological or nervous system disorders

Disease category	ICD-10 code	Look-back period (prior to index date)
Peripheral neuropathy	G50, G51, G52, G53, G54, G55, G56, G57, G58, G59, G61, G62, G64, G65	12 mo
Encephalopathy	R40.0, R44	12 mo
Seizures	G40, G41	12 mo
Dementia	F01, F02, F03, G31	12 mo
Delirium	F05	12 mo
Myoneural disorders	G72, M60	12 mo
Ataxia/movement disorders	G26, R26, R27	12 mo
Autonomic dysfunction	I95.1, G90, R55	12 mo
Parkinson's / extrapyramidal syndromes	G21, G24, G25	12 mo
Encephalitis	A85, A86, G04, G05, R29	12 mo
Amnesia/memory difficulty	R41	12 mo

Genitourinary diseases

Disease category	ICD-10 code	Look-back period (prior to index date)
Chronic kidney disease	N18, N19	12 mo

Gastrointestinal diseases

Disease category	ICD-10 code	Look-back period (prior to index date)
Abdominal or pelvic pain	R10	30 d
Changes in bowel habits	K58, K59, A08, A09, R19.4, R19.7	30 d
Nausea/vomiting	R11	30 d

Symptoms

Disease category	ICD-10 code	Look-back period (prior to index date)
Headache	G43, G44, R51	30 d
Body ache/myalgia	M02, M25, M79	30 d
Fever/malaise/fatigue	R50, R61, R53, G93.3	30 d
Lymphadenopathy	R59	12 mo
Weight loss	R63, R64	12 mo
Vertigo	A88, H81, R42	12 mo

Skin and subcutaneous tissue disorders

Disease category	ICD-10 code	Look-back period (prior to index date)
Disturbances of skin sensation	R20	30 d
Skin and mucous membrane lesions	B09	30 d
Rash and other nonspecific skin eruption	R21	30 d
Other skin changes	R23	30 d
Hair loss/alopecia	L63, L65	12 mo
Disturbances of skin sensation	R20	30 d

Sense organ diseases

Disease category	ICD-10 code	Look-back period (prior to index date)
Ear nose throat (ENT) disorders	H90, H91, H92, H93, J31, R43, R13	12 mo
Visual disturbances or blindness	H53, H54	12 mo

Sleep disorders

Disease category	ICD-10 code	Look-back period (prior to index date)
Sleep disorders	G47, F51	12 mo

Other

Disease category	ICD-10 code	Look-back period (prior to index date)
COVID-19	U07.1, J12.82, B97.29, B34.2	12 mo
Infectious complications	M35.81, B94	12 mo
Post COVID-19 condition, unspecified	U09.9	12 mo

Appendix B. Large-scale studies investigating Post COVID Conditions (PCC) using electronic health record (EHR) data among vaccinated and unvaccinated test-positive groups

Reference & Region	Study period	Study population	Study design	Follow-up	PCC Outcome definition
Taquet (2022) US	Jan – Aug 2021	18,958 persons (age NR) with confirmed COVID-19 or positive COVID-19 test (RT-PCR) from across the country. 9479 unvaccinated;2,996 partially vaccinated; 6,957 fully vaccinated. Vaccination status recorded 2 weeks prior to SARS-CoV-2 infection	Nested case-control (propensity score matching)	FU for symptom diagnosis in 6 months post- infection	Any diagnosis: abnormal bleeding, anxiety/depression, chest/throat pain, cognitive symptoms, fatigue, headache, myalgia, other pain, and death
Al-Aly (2022) US	Jan - Oct 2021	33,940 participants with break-through infection compared with 113,474 contemporary controls with SARS-CoV-2 infection and no prior history of vaccination.	Case-control	FU for 6 months post-infection	Pre-specified outcomes from diagnoses, medications and laboratory results: cardiovascular system disorders, coagulation and hematologic disorders, fatigue, gastrointestinal disorders, kidney disorders, mental health disorders, metabolic disorders, musculoskeletal disorders, neurologic disorders, pulmonary disorders. Death and at least one post-acute sequelae analyzed separately.
Zisis (2022) US	Sept 2020- Dec 2021	25 225 vaccinated patients diagnosed with COVID-19 after at least a week of administration of the complete vaccine, and approx. 1.5M controls (no differences in demographics or comorbidity status)	Case-control	≥4 weeks from initial infection. Assessment of symptoms for 3-month FU post-diagnosis.	New, continuing, or recurrent symptoms using EHR: Hypertension, diabetes, thyroid disease, heart disease, malignant neoplasm, thrombosis, rheumatoid arthritis, mental disorders, respiratory symptoms, headache, fatigue, body ache, diarrhea or constipation.

EHR = Electronic Health Records; FU = Follow-up; NR = Not reported

References

Al-Aly, Z., Bowe, B. & Xie, Y. Long COVID after breakthrough SARS-CoV-2 infection. Nat Med 28, 1461–1467 (2022). https://doi.org/10.1038/s41591-022-01840-0

Taquet M, Dercon Q, Harrison PJ. Six-month sequelae of post-vaccination SARS-CoV-2 infection: A retrospective cohort study of 10,024 breakthrough infections. Brain Behav Immun. 2022 Jul;103:154-162. doi: 10.1016/j.bbi.2022.04.013. Epub 2022 Apr 18. PMID: 35447302; PMCID: PMC9013695.

Zisis S, Durieux J, Mouchati C, Perez J, McComsey G. The protective effect of COVID-19 vaccination on post-acute sequelae of COVID-19 (PASC): A multicenter study from a large national health research network. Open Forum Infect Dis. 2022;fac228.