NON-INTERVENTIONAL (NI) STUDY PROTOCOL

Study information

Title	A standing cohort to understand the characteristics of patients with COVID-19 and contextualize the COVID-19 complication and safety events of interests using US OPTUM EHR data
Protocol number	C4671040
Protocol version identifier	1.0
Date	29, April 2022
Research question and objectives	Research Questions: What are the demographic and clinical characteristics and healthcare utilization of all COVID-19 patients (i.e. all patients with a diagnosis of COVID-19 in the Optum database (Overall Source Cohort); and multiple subcohorts including, COVID-19 patients with characteristics similar to the Paxlovid clinical trial programs (Trial Similar Cohorts) and COVID-19 patients treated with Paxlovid (Paxlovid Subcohort))? What is background incidence of clinical manifestations of COVID-19 among COVID-19 patients including subpopulation of interests (e.g. hospitalized patients, trial similar patients)? What are the background rates of events that may be potential Paxlovid safety concerns among COVID-19 patients?

	What are the incidences of long COVID-19 or post-acute sequalae SARS CoV-2 (PASC) the characteristics of patients with long COVID? The following are the research objectives that will be addressed in this study: ***Primary Objectives:*** Objective 1: Describe the distribution of demographics, co-morbid conditions, medical history, selected biomarkers, and healthcare utilization at baseline for COVID-19 patients (including subpopulation with long COVID) Objective 2: Describe the time to clinical events (e.g. hospitalization, ICU admission, death) Objective 3: Estimate background incidence of COVID-19 manifestations/complications and safety outcome of interests following COVID-19 infection or Paxlovid exposure Objective 4: Estimate incidence of long COVID-19 or post-acute sequalae SARS CoV-2 (PASC) and assess health effects of Long COVID. Objective 5: Describe the health equity and demographic and clinical characteristics (e.g. co-morbid conditions, outcomes, treatment patterns, healthcare utilization) among patients treated with Paxlovid ***Secondary Objectives:*** Objective 1: Estimate the incidence of oxygen supplementation among COVID-19

	patients hospitalized with mechanical ventilation/ECMO ***Exploratory Objectives:*** Objective 1: Identify predictors of COVID-19 severe outcomes/manifestations and Long COVID with machine learning (ML) methods Objective 2: Predict risk of developing severe outcomes/manifestation and Long COVID at the patient level with machine learning (ML) methods
Author	Redacted Redacted Redacted Redacted Redacted Redacted Redacted Redacted Redacted Redacted

1. TABLE OF CONTENTS

2. LIST OF ABBREVIATIONS

Abbreviation	Definition
COVID-19	Coronavirus Disease 2019
CRP	C-reactive protein
DVT	Deep vein thrombosis
ECMO	Extracorporeal Membrane Oxygenation
EUA	Emergency Use Authorization
FDA	Federal Drug Administration
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
ICU	Intensive care unit
IDN	Integrated delivery network
IFN	Interferon
IL	Interleukin
ILD	Interstitial lung disease
LDH	Lactate dehydrogenase
PCR	Polymerase Chain Reaction
PT	Prothrombin time
PTT	Partial thromboplastin time
RNA	Ribonucleic acid
SARSCoV- 2	Severe Acute Respiratory Syndrome Coronavirus 2
SD	Standard deviation

TSC	Trial Similar Cohort
TNF	Tumor necrosis factor

3. RESPONSIBLE PARTIES

Principal Investigator(s) of the Protocol

Name, degree(s)	Job Title	Affiliation	Address
Redacted	RedactedRedacted RedactedRedacted RedactedRedacted Redacted	Redacted	Redacted Redacted Redacted Redacted Redacted
Redacted	RedactedRedacted Redacted	Redacted	Redacted Redacted Redacted Redacted Redacted
Redacted	RedactedRedacted RedactedRedacted Redacted	Redacted	Redacted Redacted Redacted
RedactedRedacted	Redacted RedactedRedacted RedactedRedacted	Redacted	Redacted Redacted Redacted Redacted

Co-Investigator(s) of the Protocol

Name, degree(s)	Job Title	Affiliation	Address
Redacted	RedactedRedacted Redacted	Redacted	Redacted Redacted Redacted
Redacted	RedactedRedacted RedactedRedacted	Redacted	Redacted Redacted Redacted Redacted Redacted
Redacted	RedactedRedacted Redacted	Redacted	Redacted Redacted Redacted Redacted
Redacted	RedactedRedacted Redacted	Redacted	Redacted Redacted Redacted Redacted Redacted
Redacted	RedactedRedacted Redacted	Redacted	Redacted Redacted Redacted Redacted Redacted
Redacted	RedactedRedacted RedactedRedacted	Redacted	Redacted Redacted Redacted Redacted Redacted

4. AMENDMENTS AND UPDATES

None.

.

5. MILESTONES

Milestone	Planned date
Start of data collection	1 May 2022
Interim report	30 August 2022
End of data collection	30 June 2024
Final study report	31 September 2024

6. RATIONALE AND BACKGROUND

Background: A novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV- 2), was first identified in December 2019 as the cause of a respiratory illness designated coronavirus disease 2019, or COVID-19[1]. The infection may be asymptomatic but symptomatic cases may present with cough, fever, dyspnea, headache, diarrhea, altered mental status, myalgia etc. Serious disease may manifest as acute respiratory distress syndrome, septic shock, and multi-organ dysfunction. As of January 2022, at least 324,196,758 cases have been confirmed worldwide, and at least 5,531,843 deaths have occurred[2]

A number of different classes of drugs are being evaluated or newly developed for treatment or post-exposure prophylaxis of COVID-19. As of the date of issuance of clinical trial of Paxlovid (PF-07321332/ritonavir), only 1 anti-viral drug with activity against SARSCoV-2, remdesivir, an RNA polymerase inhibitor, has received approval in hospitalized patients with COVID-19.[3] Monoclonal antibodies (containing casirivimab and imdevimab or bamlanivimab and estesevimab) received emergency use authorization for mild to moderate COVID19 in recently diagnosed high risk nonhospitalized patients. Both of these options are administered intravenously and require administration by healthcare professionals. As of October 2021, there were no approved oral alternatives to PF-07321332/ritonavir for the treatment of mild-to-moderate COVID-19 in adult patients with positive results of direct SARS-CoV-2 viral testing, and who are at high risk for progression to severe COVID-19, including hospitalization or death.[4-6]

There is an urgent and unmet need for antiviral agents that could be used for the treatment of non-hospitalized persons with COVID-19. Paxlovid, a potent and selective inhibitor of the SARS-CoV-2 3CL protease, was developed by Pfizer Inc. as an oral treatment in patients with COVID-19 in outpatient setting. The coronavirus 3CL protease is a virally encoded enzyme that is critical to the SARS-CoV-2 replication cycle, analogous to other obligatory virally encoded proteases (-e.g.-, HIV Protease, HCV Protease).[7] In December 2021, Food and Drug Administration FDA has authorized the emergency use of Paxlovid in adult and pediatric patients 12 years of age and older (weighing at least 88 pounds (40 kg)]) with

positive results of direct SARS-CoV-2 viral testing, and who are at high risk for progression to severe COVID-19, including hospitalization or death.[8]

There is limited information about the safety and effectiveness of using Paxlovid, to treat people with mild-to-moderate COVID-19 in real-world setting.[8] Understanding the epidemiology and clinical manifestations of COVID-19, as well as the background rates of potential Paxlovid related adverse events (AEs) among COVID-19 patients who are at high and standard risk of progression to severe illness in real world setting can provide critical context for interpreting potential safety events and clinical manifestations in on-going clinical trials post-marketing, with broaden patient population. This population based cohort study using US OPTUM COVID-19 EHR data will provide background rates of safety events of interest to support Paxlovid program. The Phase 3 clinical trial inclusion and exclusion criteria will be applied to a subgroup of the population to create a trial similar cohort. This study is being conducted in anticipation of the need to address possible questions about safety emanating from reports of AEs associated with the Paxlovid. The timeframes for this study will consider the health impacts as well as changes in healthcare utilization that COVID-19 may have in interpreting our understanding of potential safety events. Additional outcome events and subgroups may be incorporated in this study over time. Given the global scale of COVID 19 pandemic, there is an increasing need to better understand and characterize the post-acute sequalae SARS CoV-2 (PASC) or long COVID.[9-10] This study will also aim at describing the characteristics o19 outcome/longlong COVID including those exposed to Paxlovid in real-world setting. Such information may inform clinical trial design and disease management for the long-term health effects of the COVID patients. Moreover, this study provides insights into understanding of potential racial disparities among patients using Paxlovid. The exploratory analyses may help identify predictors of severe COVID-19 outcomes/ manifestations and long Covid and predict the risk developing the severe outcomes/ manifestations and long Covid at individual level in real world setting.

7. RESEARCH QUESTION AND OBJECTIVES

Research Question:

What are the demographic and clinical characteristics and healthcare utilization of all COVID-19 patients (Overall Source Cohort); and multiple subcohorts including COVID-19 patients with characteristics similar to the Paxlovid clinical trial programs (Trial Similar Cohorts) and COVID-19 patients treated with Paxlovid (Paxlovid Subcohort))?

What is background incidence of clinical manifestations of COVID-19 among COVID-19 patients including subpopulation of interests (e.g.,. hospitalized patients, trial similar patients)??

What are the background rates of events that may be potential Paxlovid safety concerns among COVID-19 patients?

What are the incidences of long COVID-19 or post-acute sequalae SARS CoV-2 (PASC) the characteristics of patients with long COVID?

The following are the research objectives that will be addressed in this study:

Primary Objectives:

Objective 1: Describe the distribution of demographics, co-morbid conditions, medical history, selected biomarkers, and healthcare utilization at baseline for COVID-19 patients(including subpopulation with long COVID)

Objective 2: Describe the time to clinical events (e.g. hospitalization, ICU admission, death)

Objective 3: Estimate background incidence of COVID-19 manifestations/complications and safety outcome of interests following COVID-19 infection or Paxlovid exposure

Objective 4: Estimate incidence of long COVID-19 or post-acute sequalae SARS CoV-2 (PASC) and assess health effects of Long COVID.

Objective 5: Describe the health equity and demographic and clinical characteristics (e.g. co-morbid conditions, outcomes, treatment patterns, healthcare utilization) among patients treated with Paxlovid

Secondary Objectives:

Objective 1: Estimate the incidence of oxygen supplementation among COVID-19 patients hospitalized with mechanical ventilation/ECMO

Exploratory Objectives:

Objective 1: Identify predictors of COVID-19 severe outcomes/manifestations and Long COVID with machine learning (ML) methods

Objective 2: Predict risk of developing severe outcomes/manifestation and Long COVID with machine learning (ML) methods at the patient level

8. RESEARCH METHODS

8.1. Study design

This is a population-based database cohort study with retrospective and prospective data collection utilizing electronic healthcare data in the US. This study will include adults (≥18 years of age) and pediatric patients (<18 years of age) with COVID-19 diagnosis (i.e. positive PCR or antigen results of direct SARS-CoV-2 viral testing or have confirmed COVID 19 diagnosis codes). All Paxlovid users will be included in the study. Within the population-based general cohort (COVID-19 Cohort), a subcohort will be created based on

similar eligibility criteria as those used in select Paxlovid clinical trials (e.g. C4671005, C4671002) to attempt to emulate the characteristics of select clinical trial program populations (Paxlovid Trial Similar COVID-19 Subcohorts).

8.2. Setting

This study will be conducted using Optum's COVID-19 Electronic Health Record Data during the period of January 01, 2020, to June 30, 2024. Details about the database are found in Section 8.4. The COVID-era cohorts will be refreshed annually to include additional individuals, follow-up time, and events. These refreshes and cohorts will accordingly be described in a Statistical Analysis Plan (SAP) or protocol amendment as appropriate.

8.2.1. Study population and cohorts

The study population will consist of patients who tested positive or were diagnosed for COVID-19 (i.e. positive PCR or antigen results of direct SARS-CoV-2 viral testing or have confirmed COVID-19 diagnosis codes). All Paxlovid users will be included in the study.

Within the study population, three source cohorts of general population with COVID-19 will be formed (Figure 1):

- 1a: Overall COVID-19 patients.
- 1b: Hospitalized patients (i.e. hospitalized within 30 days after index date) for COVID-19
- 1c: Non-hospitalized patients (i.e. not hospitalized within 30 days after index date) with COVID-19

Within source cohort of 1a, 1b, and 1c, the following trial similar subcohorts, will be formed:

Patients at high risk of progression to severe illness (Figure 1)

- 2a: Subgroup of the general population cohort 1a who meet relevant inclusion/exclusion criteria of the Paxlovid clinical trial (C4671005).
- 2b: Subgroup of the hospitalized patient cohort 1b who meet relevant inclusion/exclusion criteria of the Paxlovid clinical trial (C4671005).
- 2c: Subgroup of the non-hospitalized patient cohort 1c who meet relevant inclusion/exclusion criteria of the Paxlovid clinical trial (C4671005).

Patients at standard risk of progression to severe illness (Figure 1)

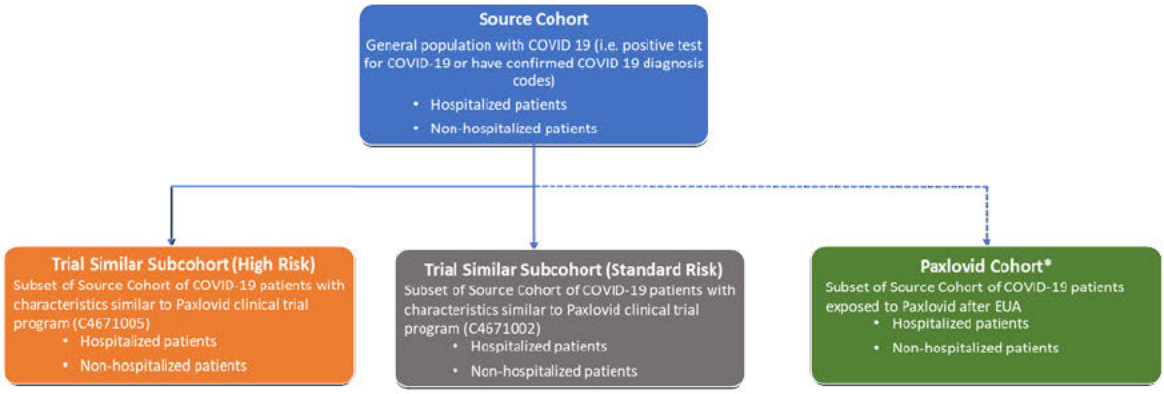
- 3a: Subgroup of the general population cohort 1a who meet relevant inclusion/exclusion criteria of the Paxlovid clinical trial (C4671002).
- 3b: Subgroup of the hospitalized patient cohort 1b who meet relevant inclusion/exclusion criteria of the Paxlovid clinical trial (C4671002).

- 3c: Subgroup of the non-hospitalized patient cohort 1b who meet relevant inclusion/exclusion criteria of the Paxlovid clinical trial (C4671002).

A sub-cohort of patients who were exposed to Paxlovid in real world settings will be created. If a patient had an exposure to paxlovid without COVID PCR/antigen testing/diagnosis data, this patient can be captured in the Paxlovid cohort.

Additional special populations (e.g. immunocompromised patients, pediatric population, severe renal impairment, etc.) may be added as a subset.

Figure 1. Flowchart outlining the study cohort and subcohorts



Note: Any addition of a special population (e.g. immunocompremised patients, pediatric population, severe renal cohort etc.) as well as patients with multiple infections of COVID 19 or patients received COVID 19 vaccines can be a subset of either source cohort or high/standard risk cohort

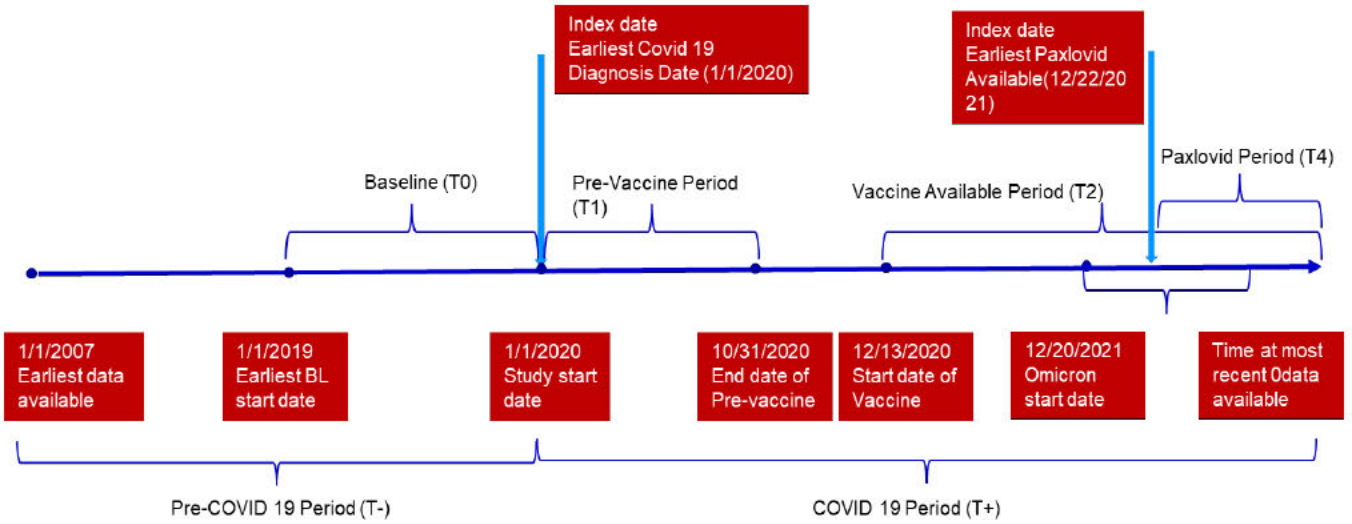
* For Paxlovid cohort, if a patient had an exposure to paxlovid without COVID PCR testing/diagnosis data, this patient can be captured in the Paxlovid cohort.

8.2.2. Time periods

8.2.2.1. Study period

The study will include an overall study period and four sub-periods for each source cohort and subcohort (Figure 2):

Figure 2. Flowchart outlining the study periods



- Study period: COVID-19-era (T+) will include Jan 1, 2020 – most recent data available at the time of analytic dataset creation. The study period will be further stratified as follows:
  - Study period i (T1): Active COVID-19-era (Pre-Vaccine Period) will include Jan 1, 2020 – Oct 31, 2020
  - Study period ii (T2): Active COVID-19-era (Vaccine Available Period) will include Dec 13, 2020 – most recent data available at the time of analytic dataset creation.
  - Study period iii (T3): Omicron Period (Approximate time period when Omicron is dominated) will include Dec 20, 2021 – (TBD).
  - Study period iv (T4): Paxlovid Period (Paxlovid Available after EUA) will include Dec 22, 2021 (i.e. EUA approval) to most recent data available at the time of analytic dataset creation.
  - Note: The period(s) for future variant(s) may be added.

8.2.2.2. Baseline period

The 365 days before index date (see Section 8.2.2.3) will be the baseline period (T0, Figure 2).

8.2.2.3. Index date

The index date for the COVID-19 Source and the Trial Similar Subcohort will be defined as the date of first date tested positive or diagnosed for COVID-19 (i.e. positive PCR or antigen results of direct SARS-CoV-2 viral testing or have confirmed COVID-19 diagnosis codes) during the study period.

The index date for the Paxlovid cohort will be defined as the date of first prescription of Paxlovid during the study period.

8.2.2.4. Follow-up

The follow-up period starts at index date and participants will be followed until the date of the first occurring of the following events:

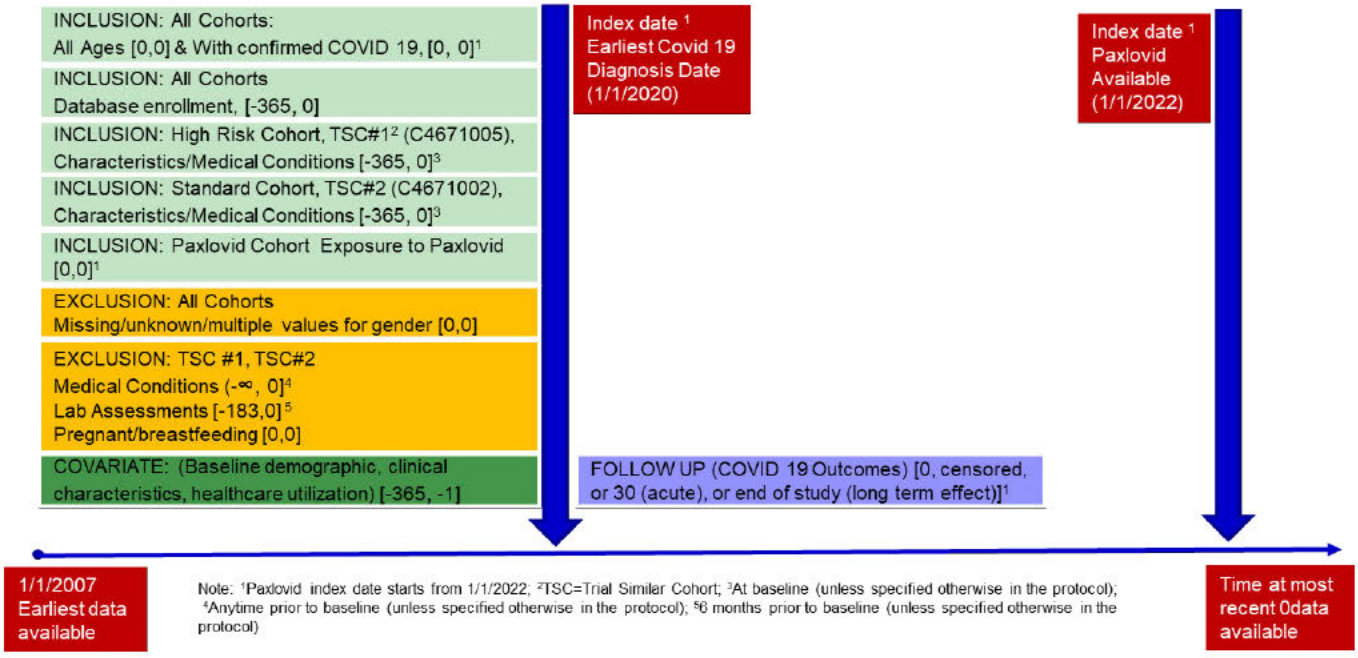
- Death
- At the end of their database disenrollment from the database
- At the end of study period (Section 8.2.2.1)
- Occurrence of the outcome of interest

Follow-up time for hospitalized patients will meet the criterion in Section 8.2.1 (i.e. hospitalized within 30 days of COVID 19 diagnosis). Follow-up windows for long COVID will be defined in SAP.

8.2.3. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria for the source cohort, trial similar cohorts (TSC), and Paxlovid cohort are described respectively in this section. Figure 3 shows the inclusion and exclusion assessment windows.

Figure 3. Inclusion & Exclusion Assessment Windows (Days)



8.2.3.1. COVID-19 Source Cohort

The inclusion and exclusion criteria to ascertain the source cohort (Section 8.2.1) are listed as follows:

Inclusion Criteria

1. Tested positive or diagnosed for COVID-19 (i.e. positive PCR or antigen results of direct SARS-CoV-2 viral testing or have confirmed COVID-19 diagnosis codes: SARS-CoV-2L severe acute respiratory syndrome coronavirus 2; PCR: polymerase chain reaction) during the study period.
2. At least 12 months of ≥1 healthcare activity prior to index date (Section 8.2.2.3)
3. Integrated delivery network information to allow linkage of outpatient and inpatient records

Additionally, the source cohort is further stratified into two subpopulations: hospitalized - those who were hospitalized within 30 days after the index date (Section 8.2.1); and non-hospitalized - those without hospitalization within 30 days after the index date (Section 8.2.1).

Exclusion criteria

1. Missing, unknown, or multiple values for gender

8.2.3.2. Trial Similar Cohort (TSC)

In addition to inclusion criteria in section 8.2.3.1, patients in the Paxlovid trial similar cohorts must meet the following inclusion and exclusion criteria. Although not an exact match, the listed criteria provide a subcohort with patient characteristics that are as similar as possible to the patients in the Paxlovid clinical trial programs.

(1) High Risk Cohort (TSC to C4671005)

Inclusion Criteria

1. Has at least 1 characteristic or underlying medical condition associated with an increased risk of developing severe illness from COVID-19 at baseline (unless specified otherwise) including:

    a. ≥60 years of age at the index date
    b. Body mass index (BMI) >25 kg/m^2
    c. Current smoker (cigarettes smoking at baseline)
    d. Immunosuppressive disease (e.g., bone marrow or organ transplantation or primary immune deficiencies) OR prolonged use of immune-weakening medications:

- Has received treatment with biologics (eg, infliximab, ustekinumab), immunomodulators (eg, methotrexate, 6MP, azathioprine) or cancer chemotherapy within 90 days prior to index date.
- HIV infection

e. Chronic lung disease (i.e. Asthma, chronic obstructive pulmonary disease [COPD], interstitial lung disease [ILD])
f. Known diagnosis of hypertension
g. Cardiovascular disease (CVD), defined as history of any of the following: myocardial infarction, stroke, transient ischemic attack (TIA), heart failure (HF), angina with prescribed nitroglycerin, coronary artery bypass graft surgery (CABG), percutaneous coronary intervention (PCI), carotid endarterectomy, and aortic bypass
h. Type 1 or Type 2 diabetes mellitus
i. Chronic kidney disease (CKD) (excluding dialysis or end stage of renal disease)
j. Sickle cell disease
k. Neurodevelopmental disorders or other conditions that confer medical complexity
l. Any cancer diagnoses, other than localized skin cancer (i.e. NMSC) at baseline, within one year prior to index date

Exclusion Criteria

Participants are excluded from the study if any of the following criteria apply:

Medical Conditions (any time prior to index date, unless specified otherwise)

1. History of hospitalization within 30 days prior to Index date
2. Known medical history of active liver disease (other than nonalcoholic hepatic steatosis), including chronic or active hepatitis B or C infection, primary biliary cirrhosis, or acute liver failure.
3. Receiving dialysis or end stage of renal disease

Other Exclusions:

4. Females who are pregnant or breastfeeding within 9 months prior to index date.

(2) Standard Risk Cohort

Inclusion Criteria

See source cohort inclusion criteria in Section 8.2.3.1

Exclusion Criteria

Has at least 1 characteristic or underlying medical condition associated with an increased risk of developing severe illness from COVID-19, including:

- ≥65 years of age at the index date
- BMI≥30 kg/m^2
- "High Risk Cohort – Inclusion Criteria 1.c – 1.l " (Section 8.2.3.2)
- "High Risk Cohort – Exclusion Criteria #1 - #4" (Section 8.2.3.2)

8.2.3.3. Paxlovid Cohort

Inclusion/exclusion criteria:

Patients who met the inclusion and exclusion criteria in source cohort and exposed to Paxlovid after EUA approval on December 22, 2021. If a patient had an exposure to paxlovid without COVID PCR testing/diagnosis data, this patient can be captured in the Paxlovid cohort.

8.3. Outcomes

Outcome events will be the pre-specified potential safety concerns, including those that were part of Paxlovid or identified after the EUA post marketing[8]. The list of events of interest is provided below. This list of outcome events was chosen based on clinician's expert opinion, biologic plausibility, and feasibility of identification in EHR or claims data. Other potential safety concerns that arise during the post marketing, clinical trials, or based on emerging literature related to Paxlovid safety will be included in subsequent refreshes of the cohorts, or as rapid queries in response to regulatory authority requests.

Potential Safety Outcomes of Interests*

1) Renal impairment

2) Hepatic impairment

3) Other potential safety events of interest (e.g., Diarrhea, Dysgeusia, Hypersensitivity reaction/allergic reactions, and other additional potential safety events of interest may be added over time)

* The potential safety outcomes of interests listed above can be further evaluated by the covariates including (demongraphics, drug-drug interactions etc.)

COVID-19 complications/interventions (within 30 days of COVID 19 diagnosis):

- Pneumonia

- Kidney failure
- Thrombotic event
- Acute respiratory distress syndrome (ARDS)
- Heart failure
- Sepsis/septic shock
- Hospitalization
- All-cause mortality
- In hospital death
- Multi-organ dysfunction/failure
- Mechanical ventilation/ECMO (Extracorporeal membrane oxygenation)
- Vasopressor use

Time to clinical events:

- Time from index date (Section 8.2.2.3) to admission
- Time from index date (Section 8.2.2.3) to the events above

COVID-19 long term health effects:

The following conditions may be considered for COVID-19 long term effects. This is not an exclusive list. Specific definition and follow up windows of long COVID (e.g., greater than 4 weeks, 12 weeks, or 6 months after a COVID diagnosis)[9-10] will be further defined in SAP.

- Abdominal pain
- Acute kidney injury
- Alopecia
- Body aches
- Chest pain
- Cognitive impairment
- Confusion/delirium
- Nasal congestion
- Cough
- Anxiety disorders
- Depressive disorder (includes major depressive disorder, dysthymic disorder)
- Post-traumatic stress disorder
- Substance use disorders
- Suicidal behavior and suicide ideation
- Diarrhea
- Nausea
- Palpitations or fast-beating or pounding heart
- Fatigue
- Hair loss
- Headache

- Intermittent fever
- Arthralgia or joint pain
- Anosmia or loss of smell
- Ageusia or loss of taste
- Lung function abnormalities
- Myalgia or muscle pain
- Myocarditis
- Pericarditis
- Post-COVID syndrome (i.e., post-acute sequalae SARS CoV-2 (PASC) defined by ICD-10-CM code U09.9)
- Rash
- Dyspnea or shortness of breath
- Insomnia
- Others to be added

COVID-19 Re-infection:

Re-infection of COVID-19 within certain time windows post Paxlovid prescription date. This will be further defined in SAP.

8.4. Covariates

Covariates will include baseline demographic, clinical characteristics, medications, and health utilization, as well as those needed for the study cohort inclusion and exclusion criteria. These are not intended to be the exclusive lists. Adjustments of the covariates may be made and additional variables may be added in SAP as needed over time. The code lists of conditions and medications (if applicable) will be provided in SAP.

Demographics:

- Age
- Gender
- Race
- Ethnicity
- Region
- Insurance

Clinical Characteristics:

- Asthma
- Autoimmune diseases
    - Multiple Sclerosis; Sjogren's syndrome; Guillain-Barre; idiopathic thrombocytopenia purpura; autoimmune thyroiditis; giant cell arteritis;

psoriasis; systemic or cutaneous lupus erythematosus; autoimmune arthritis/rheumatoid arthritis, glomerulonephritis

- BMI
- Cancer
- Cerebrovascular disease or stroke (including transient ischemic attack)
- CKD
- COPD
- ILD
- Coronary heart disease
- Cardiomyopathies
- Deyo-Charlson Comorbidity Index Score
- Hepatitis B (HBV), Hepatitis C (HCV), HIV infection
- Hyperlipidemia
- Hypertension
- Liver disease
- Sickle cell disease
- Smoking
- Diabetes mellitus
- Pregnancy status
- C-Reactive Protein (CRP; within 10 days prior to the index date)
- Lactate dehydrogenase (LDH; within 10 days prior to the index date)
- Ferritin (within 10 days prior to the index date)
- D-dimer (within 10 days prior to the index date)
- Note: A focus of these lab tests would be for inpatients.

Medications that are contraindicated with Paxlovid (FDA EUA label):

- Alpha1-adrenoreceptor antagonist: alfuzosin
- Analgesics: pethidine, propoxyphene
- Antianginal: ranolazine
- Antiarrhythmic: amiodarone, dronedarone, flecainide, propafenone, quinidine
- Anti-gout: colchicine
- Antipsychotics: lurasidone, pimozide, clozapine
- Ergot derivatives: dihydroergotamine, ergotamine, methylergonovine
- HMG-CoA reductase inhibitors: lovastatin, simvastatin
- PDE5 inhibitor: sildenafil (Revatio®) when used for pulmonary arterial hypertension (PAH)
- Sedative/hypnotics: triazolam, oral midazolam
- Anticancer drugs: apalutamide
- Anticonvulsant: carbamazepine, phenobarbital, phenytoin
- Antimycobacterials: rifampin
- Herbal products: St. John's Wort (hypericum perforatum)

Medications for treatment of COVID-19/Vaccination for COVID-19[11]:

Anti-SARS-CoV-2 Antibody Products

Bamlanivimab/Etesevimab

Casirivimab/Imdevimab

Sotrovimab

Tixagevimab/Cilgavimab

Bebtelovimab

Convalescent plasma

Antiviral therapy

Remdesivir

Molnupiravir

Chloroquine /hydroxychloroquine

Azithromycin

Lopinavir/ritonavir

Ivermectin

Nitazoxanide

Ribavirin

Inteferon-beta

Immunomodulators

Systemic Corticosteroids

Dexamethasone

Methylprednisolone

Fluvoxamine

Interleukin inhibitors

Tocilizumab

Sarilumab

Antithrombotic Therapy

Heparin

Covid-19 vaccines

Health Utilization:

- ER visit
- Office visits
- Telemedicine visits
- Urgent care visits
- Number of hospitalizations
- Length of hospital stay

Paxlovid User Characteristics (in addition to Clinical Characteristics, Medications in "Covariates" in the section 8.4).

- Demographics
  - Age
  - Gender
  - Race
  - Ethnicity
  - Region
  - Insurance
- Comorbidities
  - Renal impairment
  - Hepatic impairment
  - Top 50 Comorbidities in the database
- Medications
  - Top 50 prescriptions in the database
- Paxlovid use
  - Days since index date (COVID-19 diagnosis)
  - Provider

- Health Utilization:
  - ER visit
  - Office visits
  - Telemedicine visits
  - Urgent care visits
  - Number of hospitalizations
  - Length of hospital stay

8.5. Data sources

Given the urgent need to clinically understand the novel virus of COVID-19, Optum developed a low latency data pipeline that enables minimal data lag, while preserving as much clinical data as possible. The data are sourced from Optum's longitudinal electronic health record (EHR) repository, which is derived from dozens of healthcare provider organizations in the US, including more than 700 hospitals and 7,000 clinics. The data are certified as deidentified by an independent statistical expert following the Health Insurance Portability and Accountability Act of 1996 (HIPAA) statistical deidentification rules and managed according to Optum® customer data use agreements. The COVID-19 dataset incorporates a wide swathe of raw clinical data, including new, unmapped COVID19-specific clinical data points from both inpatient and ambulatory electronic medical records, practice management systems, and numerous other internal systems. Information is processed from across the continuum of care, including acute inpatient stays and outpatient visits. The COVID-19 data capture point-of-care diagnostics specific to the COVID-19 patient during

initial presentation, acute illness, and convalescence, with over 500 mapped labs and bedside observations, including COVID-19-specific testing.

The Optum COVID-19 EHR dataset elements included patient-level information: demographics, mortality (captured from Social Security Administration Death Master File, electronic medical records and Centers for Medicare and Medicaid Services), as well as clinical interventions, such as medications prescribed and administered, laboratory results, and vital signs and other biometric measures. The data are comprised of multiple tables that can be linked by a common patient identifier (an anonymous, randomized string of characters). The COVID-19 patient data included patients in the EHR database who had documented clinical care from January 2007 through to the most current monthly data release with a documented exposure to, or had been tested for, SARS-CoV-2 (positive or negative result), and/or had a diagnosis of COVID-19, or acute respiratory illness, after February 1, 2020. Thus, not all patients within the Optum® COVID-19 EHR dataset had received a diagnosis of COVID-19. Patients with COVID-19 were identified via a diagnosis code for SARS-CoV-2, a positive test for SARS-CoV-2 active infection (antigen and/or polymerase chain reaction), and/or a positive antibody test. The Optum® COVID-19 EHR dataset included medical records from 2007, allowing for the utilization of patients' medical history in the analysis.

Additionally, supplemental claims/EHR data (Optum Claims and Optum Market Clarity) with longer latency but likely larger size of Paxlovid users may be used as additional analyses of Paxlovid (if needed) in this study.

8.6. Study size

This is a descriptive study with no minimum sample size. This study will be updated with new data annually (or as needed) and made readily accessible to address research questions. All eligible patients during the relevant study periods will be included. If there is a need to add on post-hoc comparative studies analyses, which are not part of the scope of this current study, the feasibility of robust comparisons given accrued sample size will be assessed as described in the SAP.

8.7. Data management

All study data exist as structured data by the time of study. Analyses will be conducted using SAS (version 9.4 or later, SAS Institute, Cary, NC, USA). As the database will be regularly updated, date and version of the database will be specified in the report of the study, and the intermediate datasets will be archived if the report will be submitted to a regulatory agency or will be published.

The study will be conducted by Quantitative Epidemiology and Analytic Team, Global Medical Epidemiology, Pfizer. The analytic datasets and programs will be stored according to Pfizer's procedures.

8.8. Data analysis

Baseline characteristics will be summarized for Source Cohort, Trial Similar Subcohorts, and Paxlovid cohort. Means with standard deviations, medians with interquartile will be provided for continuous variables. Numbers and percentages will be provided for dichotomous variables or categorical variables.

For dichotomous endpoints such as COVID-19 complications and outcomes, the crude cumulative incidence and incidence rate of each endpoint will be estimated for the Source Cohort and the Trial Similar Subcohorts. For incidence calculation, only the first of each event occurring in the risk window will be included in the numerator. How to handle prevalence cases prior to the index date for incidence calculation will be further described in SAP depending on the nature of outcomes (e.g. acute vs. chronic evets). Cumulative incidence and its 95% CI will be estimated and illustrated using 1-Kaplan-Meier curves. Incidence rate will be calculated as the number of patients who experience the event divided by the total person time at risk, along with the 95% CI. The follow-up will start on the index date and will end on the first occurrence of an endpoint, disenrollment, study end date, or death, whichever comes first. Different endpoints will not censor each other's follow-up. Subpopulation analyses, as needed, will be detailed in SAP that may include those immunocompremised patients, pediatric population, severe renal patients, Paxlovid users with contraindication medications, patients with multiple infections of COVID 19, or patients received COVID 19 vaccines etc.

For exploratory analysis, machine learning models (e.g. LASSO Cox regression, Random Survival Forest (RSF), and XGBoosting (XGB) models) will be used to identify important predictors associated with progression to COVID-19 severe outcomes/manifestations as well as long COIVD. An ensemble method will be used to rank predictors across models and find the strongest predictors for each outcome. Patient-level prediction for the risk of developing severe outcomes/manifestation and long COVID will be performed using the top ensemble predictors. C-index and dynamic AUCs will be used to evaluate model performance.
All data analysis will be executed using statistical software SAS version 9.4 or later.

Detailed methodology for statistical analyses in this study will be documented in a statistical analysis plan (SAP), which will be dated, filed, and maintained. The SAP may modify the plans outlined in the protocol; any major modifications of primary endpoint definitions or their analyses would be reflected in a protocol amendment.

8.9. Quality control

Analyses are programmed according to the specifications in the protocol, and the SAP, and documented in a programming plan as appropriate. Final deliverables will be reviewed and verified by a second, independent programmer who may also perform double programming. All quality checks will be documented in the programming plan.

8.10. Limitations of the research methods

COVID-19 is a new disease with evolving diagnostic mechanisms and care. A COVID-19 diagnosis code was established on 01 April 2020, months after the initial identification of cases and circulation in the US. The validity of the diagnosis codes remains to be determined. The spectrum of clinical manifestations of SARS-CoV-2 is still not fully described. Many infected persons are thought to have mild to unrecognizable symptoms and may therefore not present for care, testing or diagnosis. These infections will likely not be identifiable in the datasets. This may overestimate the incidence of severe disease in relation to overall SARS-CoV-2 infection but should be generalizable in relation to hospitalized patients in relevant analyses.

Similarly, patients with chronic diseases or in inpatient settings, the elderly, or other risk groups may be more likely to be under ongoing care or thought to be at higher risk and therefore may be more likely to seek care or be referred for testing. Differences among at-risk subpopulations may arise from access to care/health care seeking behaviors rather than true differences in risk. Generalizability may be further limited as SARS-CoV-2 testing patterns and treatment are evolving over time.

Other limitations are similar to those from observational studies in general, including the inability to obtain detailed clinical information on disease severity and activity due to the nature of secondary databases. In addition, EHR data only offer the prescribed drug information which does not indicate that the medication was consumed or that it was taken as prescribed. Misclassification in diagnosis data is unavoidable in RWD. Some conditions to be used to assess long-term health effects (e.g. Dysgeusia, Diarrhea, Myalgia etc.) may not be well captured in the database if the patient is not seen by a health care provider to report these types of events. We use the data from the integrated delivery network (IDN) that manages end-to-end healthcare facilities of one or more healthcare centers in a defined geographical location and aims at providing full spectrum service of inpatient and outpatient for a patient to help reduce the concern of the discontinuity in patient service. We also plan to use the linked data to Claims as sensitivity (or additional) analyses when needed to expand the analyses by adding additional patients or patient records beyond EHR via close or open claims. Lastly, we would also like to note that there is a lack of complete similarity between the Trial Similar Subcohort and the clinical trial patients.

8.11. Other aspects

Not applicable.

9. PROTECTION OF HUMAN SUBJECTS

9.1. Patient information

This study involves data that exist in anonymized structured format and contain no patient personal information. In these data sources, individual patient data are not retrieved or validated, and it is not possible to link (i.e., identify a potential association between) a

particular product and medical event for any individual. Thus, the minimum criteria for reporting an adverse event (AE) (i.e., identifiable patient, identifiable reporter, a suspect product, and event) cannot be met.

9.2. Patient consent

As this study involves anonymized structured data, which according to applicable legal requirements do not contain data subject to privacy laws, obtaining informed consent from patients by Pfizer is not required.

9.3. Institutional review board (IRB)/Independent ethics committee (IEC)

Independent Ethics Committee (IEC)/ Institutional Review Board (IRB) review is not required.

9.4. Ethical conduct of the study

The study will be conducted in accordance with legal and regulatory requirements, as well as with scientific purpose, value and rigor and follow generally accepted research practices described in Guidelines for Good Pharmacoepidemiology Practices (GPP) issued by the International Society for Pharmacoepidemiology (ISPE).

10. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

This study involves a combination of existing structured data and unstructured data, which will be converted to structured form during the implementation of the protocol solely by a computer using automated/algorithmic methods, such as natural language processing.

In these data sources, individual patient data are not retrieved or validated, and it is not possible to link (i.e., identify a potential association between) a particular product and medical event for any individual. Thus, the minimum criteria for reporting an adverse event (AE) (i.e., identifiable patient, identifiable reporter, a suspect product, and event) cannot be met.

11. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

Study reports will be generated including an interim report based on updated data, and a final study report.

In the event of any prohibition or restriction imposed (e.g., clinical hold) by an applicable competent authority in any area of the world, or if the party responsible for collecting data from the participant is aware of any new information which might influence the evaluation of the benefits and risks of a Pfizer product, Pfizer should be informed immediately.

12. REFERENCES

1. Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, et al. Remdesivir for the Treatment of Covid-19 - Final Report. N Engl J Med. 2020

2. John Hopkins University. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Available from: https://coronavirus.jhu.edu/map.html. Accessed on: Jan, 15, 2022.

3. FDA. https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-covid-19. Accessed February 6, 2020.

4. US Food and Drug Administration. Fact Sheet for Health Care Providers. Emergency Use Authorization (EUA) of Bamlanivimab and Etesevimab. Report Number: 04 June 2021. 2021. Available from: https://www.fda.gov/media/145802/download. Accessed on: 01 July 2021.

5. US Food and Drug Administration. Fact sheet for health care providers emergency use authorization (EUA) of Sotrovima. Report Number: 04 June 2021. Available from: https://www.fda.gov/media/149534/download. Accessed on: 01 July 2021.

6. US Food and Drug Administration. US Food and Drug Administration (2021). Fact sheet for health care providers emergency use authorization (EUA) of Regen-Covtim (casirivimab and imdevimab). Report Number: 04 June 2021. Available from: https://www.fda.gov/media/145611/download. Accessed on: 01 July 2021

7. Anand K, Ziebuhr J, Wadhwani P, et al. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science. 2003;300(5626):1763-7.

8. US Food and Drug Administration. Fact Sheet for Patients, Parents, and Caregivers. Emergency Use Authorization (EUA) of Paxlovid https://www.fda.gov/media/155051/download. Accessed on: 15 Jan 2022

9. Aakriti G, Mahesh VM, Claire M. et. al., Post-acute COVID-19 syndrome, *Nature Medicine* | VOL 27 | April 2021 | 601–615 | www.nature.com/naturemedicine

10. Ziyad AA, Yan X, Benjamin B, High-dimensional characterization of post-acute sequelae of COVID-19, *Nature* | Vol 594 | 10 June 2021 |

11. NIH (National Institute of Health) COVID-19 Treatment Guidelines, Table of Contents | COVID-19 Treatment Guidelines (nih.gov), last update: April 1st. Accessed on: April 24, 2022

ANNEX 1. LIST OF STAND ALONE DOCUMENTS

None.

ANNEX 2. ENCEPP CHECKLIST FOR STUDY PROTOCOLS

Not applicable.

ANNEX 3. ADDITIONAL INFORMATION