Delineating Associations Between History of Migraine, Long COVID and High Risk Long COVID Comorbidities Using Informatics Methods and Tools

By

XINZHOU WANG
THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Health Informatics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Prabhu Shankar, Chair

Heejung Bang

Mark Carroll

Committee in Charge

2022

Table of Contents

Abstract

Post-acute sequelae of COVID-19 (PASC), or "long COVID", is a debilitating chronic condition many COVID-19 survivors continuously suffer from for 12 weeks after their initial Sars-Cov-2 infection. The condition is considered a "second pandemic" that also affects millions. Symptomatic and severe COVID-19 has been recognized as a risk factor for long COVID. Risk for symptomatic COVID-19 is associated with vascular comorbidities such as hypertension, cardiovascular disease (CVD), and diabetes, and migraine shares these comorbidities. Migraine is a debilitating, neurologic condition that affects 12-15% of the general population. A retrospective cross-sectional study was conducted on de-identified patient data from the University of California COVID Research Data Set (UC CORDS) using informatics methods and tools. The primary aim was to assess whether COVID-positive migraineurs were significantly at risk for long COVID. The secondary aims were to identify confounders by stratum-specific risk ratio analysis and to find out if the presence of hypertension, CVD, or diabetes in COVID-positive migraineurs is also associated with long COVID. History of migraine in COVID-19 patients was found significantly associated with long COVID (RR 2.45, 95% CI: (2.64, 2.88), p-value = <0.0001). Stratum-specific risk ratio analysis found that long COVID was significantly associated with the presence of primary hypertension (RR 2.14, 95% CI: (1.60, 2.87), p-value <0.0001), CVD (RR 1.81, 95% CI: (1.19, 2.76), p-value = 0.008), or type 2 diabetes mellitus (RR 1.85, 95% CI: (1.23, 2.79), p-value = 0.005). in specifically female COVID-positive migraineurs younger than 65 years old. This thesis also aims to be a showcase for UC CORDS for future COVID-19 researchers.

Introduction

Background of COVID-19

The coronavirus disease 2019 (COVID-19) swept across the world as one of the most infectious viral diseases in history since the first cases were observed in December 2019. As of the end of 2022, the pandemic has had more than 600 million cases worldwide, with more than 6.5 million deaths, and in the United States alone accounts for almost 100 million cases and over 1 million deaths (World Health Organization, 2021). People infected with the severe acute respiratory syndrome-associated coronavirus 2 (Sars-Cov-2) are at a higher risk of symptomatic COVID-19 illness if certain comorbidities are present. A comorbidity is a disease or condition that is present alongside at least one other disease – in this case COVID-19 – and the term implies that the two diseases are somehow connected. The most common comorbidities associated with symptomatic COVID-19 include hypertension, cardiovascular disease (CVD), and diabetes, with hypertension being the most prevalent (Booth et al., 2021; Li et al., 2020; Sanyaolu et al., 2020). A systematic review of COVID-19 autopsies found that patients with vascular disease like hypertension, heart disease, and diabetes had shorter survival time (Martín-Martín et al., 2022).

Background of "Long COVID"

While many people fully recover from acute COVID-19 within a few days or weeks, for a significant number of patients, symptoms can continue for months or reemerge after acute disease recovery as post-acute sequelae of COVID-19 (PASC), or "long COVID" (*Long COVID Information and Resources | National Institutes of Health*, n.d.). Long COVID refers to "signs and symptoms that develop during or after an infection consistent with COVID-19, continue for more than 12 weeks, and are not explained by an alternative diagnosis" and should be considered

as a diagnosis after assessing a patient's comprehensive clinical history (NICE, RCGP, and SIGN, 2022; World Health Organization, 2021b). While anybody that has been infected with the Sars-Cov-2 virus can develop long COVID (Centers for Disease Control and Prevention, 2021), long COVID disproportionately affects those who had severe COVID-19 (Carfì et al., 2020; Halpin et al., 2020; Huang et al., 2021). According to the latest Household Pulse Survey (HPS) by the U.S. Census Bureau in October 2022, 14.2% of all adults in the U.S. and 29.6% of adults who had had acute COVID-19 have experienced long COVID (Centers for Disease Control and Prevention, 2022).

Impact of long COVID.

Long COVID can trigger many symptoms and does not affect everybody the same way. The most common long COVID symptoms are neuropsychiatric, which include fatigue, headache, sleep disorder, and difficulty concentrating or brain fog. Other less reported long COVID symptoms are respiratory, inflammatory, or abdominal in nature, such as dyspnea, lung damage, myalgia, and gastrointestinal disorders (Alkodaymi et al., 2022; Canas et al., 2022). These physical and cognitive post-COVID impairments have put millions of Americans out of work, in addition to those who contracted acute COVID-19. They had such a substantial impact that long COVID qualified as a disability under the Americans with Disabilities Act as of July 2021 (Office for Civil Rights, 2021). The same HPS found that 26.2% of people with long COVID suffer from significant activity limitations (Centers for Disease Control and Prevention, 2022). While the pandemic has been slowing down since July 2022, health experts are expecting another wave at the end of 2022. As the winter approaches and 3 new Omicron variants BQ.1, BQ.1.1, and BF.7 gain more traction, COVID-19 cases and subsequently long COVID cases may experience another surge (Centers for Disease Control and Prevention, 2020; National

Geographic, 2022). Due to its widespread impact and high risk of debilitation, the identification of long COVID risk factors has been a hot topic among pandemic researchers.

While the etiology of long COVID is still under investigation, to combat long COVID, researchers have done substantial work in identifying risk factors and groups disproportionately affected by the conditions. Large studies on sociodemographic risk factors show that long COVID is more associated with the female sex and lower socioeconomic brackets; results regarding which age groups are the most at risk are still mixed at this time, ranging from younger adults to midlife adults to seniors (Office for National Statistics, 2022; Subramanian et al., 2022; Sylvester et al., 2022; Thompson et al., 2022). In addition, there is growing data that patients who were hospitalized from COVID-19 and/or had severe acute COVID-19 illness are more likely to experience long COVID symptoms. Analysis from a cohort study following inpatients at a Wuhan hospital showed that 6 months after discharge, inpatients suffered more from fatigue, sleep disorder, and anxiety or depression, and inpatients who had more severe COVID-19 illness experienced more severe dyspnea and had more abnormal results from CT scans (Carfì et al., 2020; Halpin et al., 2020; Huang et al., 2021).

Associations with migraine.

Migraine is a complex, neurovascular, episodic disorder that shares comorbidities with symptomatic COVID-19. It is a disorder that affects 12-15% of the general population, disproportionately affects women and midlife individuals, and is a major cause for disability (Lipton et al., 2007; May & Goadsby, 1999; UpToDate, 2022). Common vascular comorbidities with migraine include hypertension, coronary heart disease, stroke, and diabetes, all of which overlap with those associated with higher risk for symptomatic COVID-19 illness (Guidetti et al., 2014; Scher et al., 2005; Wang, 2010). Since these pre-existing conditions raise the risk for

symptomatic COVID-19, it would be logical to suspect that a migraineur suffering from these comorbidities would also be more at risk for symptomatic COVID-19. However, research addressing the risk for symptomatic COVID-19 in migraineurs is scarce. A case-control study found that migraineurs experienced a higher frequency of headache symptoms during acute COVID-19 than non-migraineurs (Magdy et al., 2022), but it did not address the potential effect by comorbidities. A meta-analysis of COVID-19 studies worldwide showed that higher prevalence of migraine was significantly associated with higher COVID-19 mortality, compared to other headache disorders (Shapiro et al., 2021), but the analysis did not factor in pre-existing conditions in those migraineurs. On the basis that a migraineur is more at risk for symptomatic COVID-19 due to shared vascular comorbidities, and severe COVID-19 is a risk factor for long COVID, this thesis will investigate whether a migraineur is more at risk for long COVID, in consideration of demographic risk factors and vascular comorbidities.

Problem Statement

Long COVID is a debilitating chronic condition and a "second pandemic" that affects millions worldwide. Symptomatic and severe COVID-19 has been recognized as a risk factor for long COVID. Risk for symptomatic COVID-19 is associated with vascular comorbidities such as hypertension, CVD, and diabetes, and migraine shares these comorbidities. Given that a migraineur had COVID-19 and recovered, this thesis aims to answer the primary research question of whether their migraine history increases risk for long COVID. This thesis also seeks to answer the secondary research question of whether having one of the three comorbidities is also associated with higher long COVID risk.

This thesis seeks to investigate long COVID risk in migraineurs by analyzing the large amount of de-identified patient data hosted in the University of California COVID Research

Data Set (UC CORDS), a HIPAA Limited Data Set generated from a centralized COVID-19 database with data from all UC medical centers.

Significance to the Field

The significance of this thesis is to provide evidence, or a lack thereof, of previous history of migraine as a risk factor for long COVID. This thesis aims to facilitate future research on the etiology of long COVID and on public health action for the benefit of migraineurs. It also seeks to demonstrate the depth and complexity of UC CORDS and involve more researchers into leveraging the dataset.

Research Design and Methods

Methods Summary

This thesis followed a cross sectional study design to investigate the risk of long COVID in migraineurs, based on retrospective data. Data was queried from the UC CORDS using Jupyter Notebook. Inclusion-exclusion criteria was implemented based on up-to-date COVID-19 phenotype documentation, database documentation, and standardized diagnostic codes. 2x2 contingency table for crude risk ratio was constructed, with long COVID serving as the disease outcome, and previous history of migraine serving as the exposure. Crude risk ratio of long COVID between migraineurs and non-migraineurs, 95% confidence interval (CI), and p-value were calculated. Stratum-specific risk ratios, 95% CIs, and p-values were calculated for long COVID risk factors reported in literature. The Cochran-Mantel-Haenszel (CMH) procedure was followed to identify confounders. Calculations of proportions, risk ratios, 95% confidence

intervals, p-values, and CMH Estimates were performed using Jupyter Notebook, R Studio, and Microsoft Excel. Data was visualized using Microsoft Excel.

Literature Search

Literature search was conducted in PubMed. PubMed search for COVID-19 studies was conducted using a modified version of the COVID-19 query recommended by the PubMed User Guide (PubMed, 2022). Filters for vaccines and COVID-19 testing were removed:

> ("COVID-19" OR "COVID-19"[MeSH Terms] OR "SARS-CoV-2" OR "sars-cov-2"[MeSH Terms] OR "Severe Acute Respiratory Syndrome Coronavirus 2" OR "NCOV" OR "2019 NCOV" OR (("coronavirus"[MeSH Terms]) AND 2019/11/01[PDAT] : 3000/12/31[PDAT])) AND ("Comorbid*"[all])

PubMed search for long COVID articles was conducted using the query recommended by the PubMed User Guide (PubMed, 2022):

> "post acute sequelae of Sars-CoV-2" OR ("PASC" AND ("COVID-19" OR "Sars-CoV-2")) OR "post acute sequelae of COVID" OR "COVID-19 sequelae" OR "long haul covid" OR "covid long haul*" OR "long covid" OR "long term covid" OR "chronic covid syndrome" OR "post covid syndrome" OR "post COVID-19 neurological syndrome" OR "post-acute COVID-19 syndrome" [Supplementary Concept] OR "COVID-19 post-intensive care syndrome" [Supplementary Concept]

PubMed search for migraine was conducted using the following query:

> ("Migraine" OR "Migraine Disorder*"[MeSH Terms])

PubMed filters for risk factors and comorbidities were as follows. This filter can be attached to previous queries by using the "AND" logical operator:

("Risk factor*"[all] OR "predictor*"[all] OR "comorbid*"[all])

Data Access: University of California COVID Research Data Set (UC CORDS)

Electronic health data was provided by the University of California COVID Research Data Set (UC CORDS). UC CORDS contains electronic health records (EHRs) for all patients that tested positive or negative for COVID-19, except for restricted patients. It includes Sars-Cov-2 testing results and inpatient COVID-19 treatment information (for those positive for the virus) (University of California Health, 2022).

UC CORDS is a Limited Data Set (LDS), a data set with only a limited set of identifiable patient information, such as age and dates of medical service, as defined by the Privacy Regulations issued under the Health Insurance Portability and Accountability Act (HIPAA) (Office of Human Subjects Research - Institutional Review Board, 2015). Application for read permissions for UC CORDS and read and write permissions for a personal database was accessed from the Data Sources portal on the UC Davis Health "Health Data Resources" webpage. No separate IRB approval request was required.

How to Query UC CORDS

UC CORDS data was accessed within a secure, firewalled environment via Jupyter Notebook, a server-client application that allows the editing of Python code and rich text elements via a web browser (Ingargiola, 2015). Internet access was required. Querying the UC CORDS relational database in a Python environment was made possible by a proprietary Python package similar to "psycopg2".

Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

Before proceeding to querying, the structure and standardization of UC CORDS must be understood. UC CORDS follows the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) version 5.4. (Observational Health Data Sciences and Informatics, 2022). The OMOP CDM was developed by the Observational Health Data Sciences and Informatics (OHDSI) collaborative as a global standard for observational research. The purpose of a CDM is to harmonize disparate data using its own standardized vocabularies and existing vocabularies (Blacketer, 2021). Citing the central limit theorem, the larger the sample size, the better it can predict the parameters of a population. So, to draw conclusions with statistical power, data from many patients from many sources must be gathered. Each data source, representing a unique healthcare entity, tends to have its own way of recording data, which complicates or prevents interoperability. Data standardization ultimately improves patient wellbeing by facilitating systematic and standardized research and enabling exchange of patient data. How clinical data is organized using standardized vocabularies is depicted in Figure 2.1.

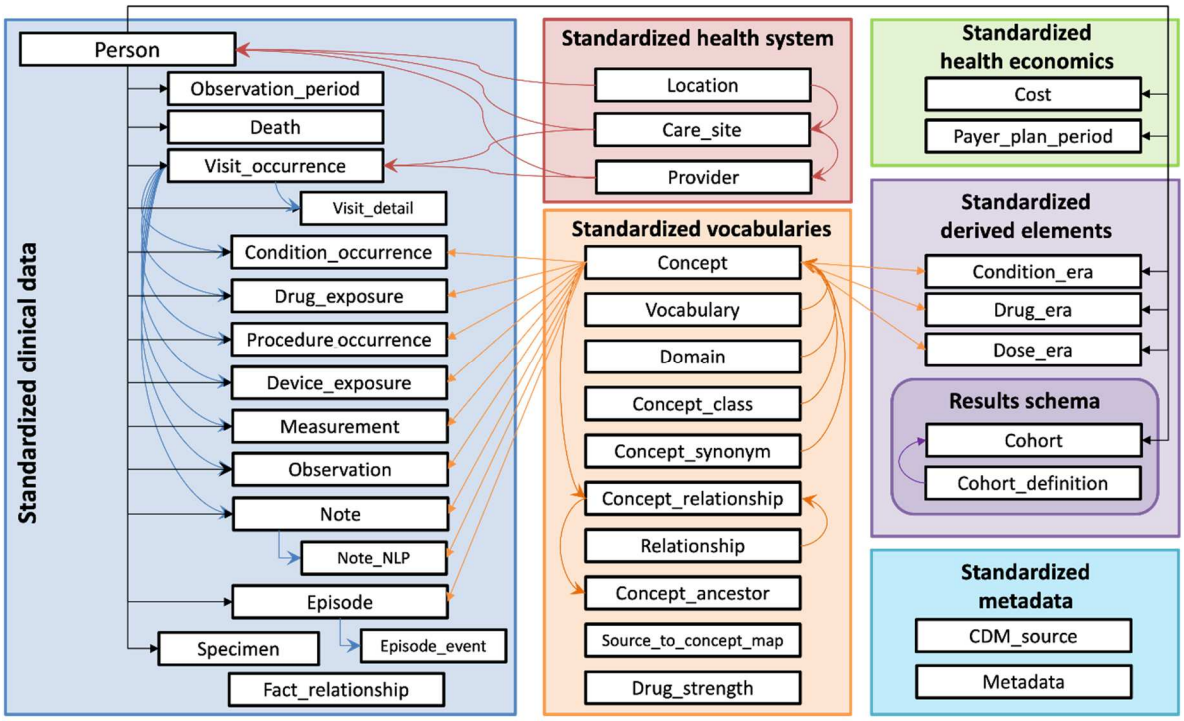

Figure 2.1: Overview of all the tables in CDM version 5.4 and their relationships. (Observational Health Data Sciences and Informatics, 2022)

Under standardized vocabularies, the most relevant elements to this thesis are the Concept and the Vocabulary. An OMOP Concept can describe any clinical event, and these Concepts are designed and always being updated to cover all clinical events (Figure 2.2). They serve as the fundamental building blocks of OMOP's clinical data tables.

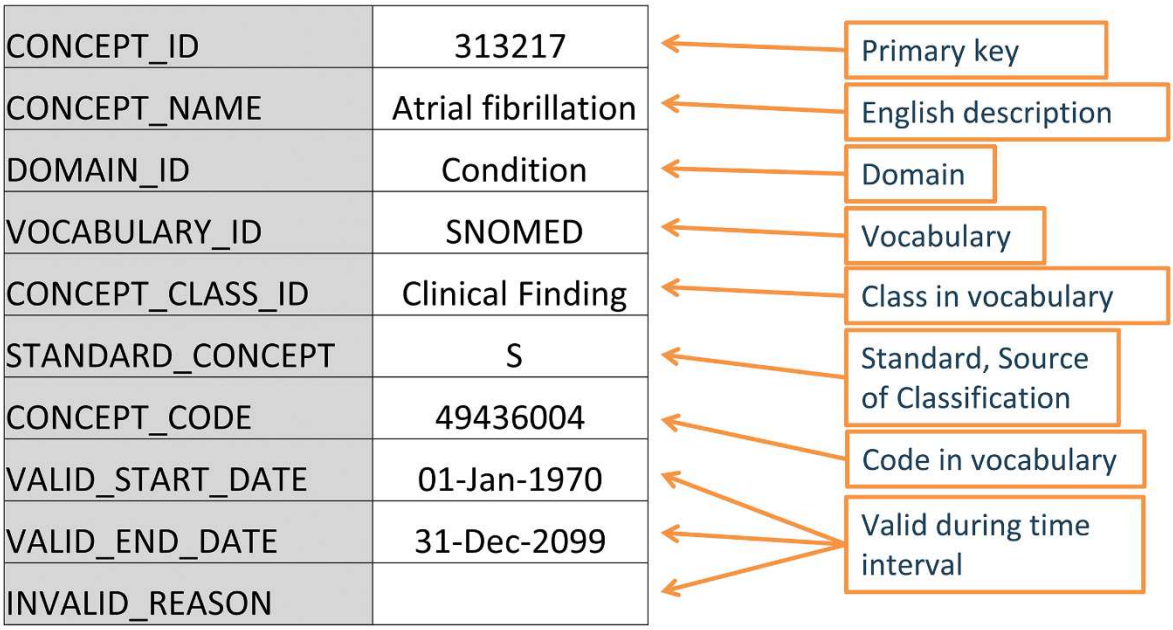

Figure 2.2: Example of a record in the CONCEPT table for the Concept “atrial fibrillation”. (Reich & Ostropolets, 2021)

A Vocabulary can describe either an internal OMOP vocabulary or an external vocabulary. Vocabularies relevant to this thesis are the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), the Logical Observation Identifiers Names and Codes (LOINC), and the Systemized Nomenclature of Medicine Clinical Terms (SNOMED-CT). Specifying the vocabulary allows precise adherence to up-to-date inclusion and exclusion criteria.

This thesis will conduct its queries within the following tables with deidentified clinical data,

- PERSON: records that uniquely identify each patient, a Person, in the source data. Data fields relevant to this thesis are person_id, year_of_birth, and gender_concept_id.

- VISIT_OCCURRENCE: the spans of time a Person continuously receives medical services within the health care system. Data fields relevant to this thesis are visit_occurrence_id, visit_start_date, and visit_end_date.
- CONDITION_OCCURRENCE: records of a Person suggesting the presence of a disease or medical condition stated as a diagnosis, a sign, or a symptom, which is either observed by a health care provider or reported by the Person. Data fields relevant to this thesis are condition_occurrence_id, condition_concept_id, and condition_source_concept_ID. Condition_concept_id is used to query SNOMED-CT codes, and condition_source_concept_id is used for ICD codes.
- MEASUREMENT: records of structured values, numerical or categorical, obtained through systematic and standardized examination or testing of a Person or their sample. Data fields relevant to this thesis are measurement_id, measurement_concept_id, measurement_date, and value_as_concept_id. Measurement_concept_id is used to query LOINC codes.

Inclusion and Exclusion Criteria

UC CORDS includes all patients that tested positive or negative for COVID-19, except for restricted patients. UC CORDS data is refreshed monthly, and patient history dates back to 1/1/2018.

For this thesis, the baseline inclusion criteria are that only "truly UC patients" – patients who have had at least 1 actual encounter in a UC medical center since 1/1/2018 – will be included. UC CORDS data was cleaned by excluding all patients with only 1 unique encounter ID. Additionally, patients with "OTHER" or "UNKNOWN" genders were eliminated.

For all disease case inclusion/exclusion criteria, Python's Regular Expression functions were used to comb through patient history. The logic is to retrieve both the list of patients that

met the inclusion criteria and the list of those for the exclusion criteria. Next, the list of exclusions, if any, is subtracted from the list of inclusions. The disease elements included in this thesis are COVID-19, long COVID, history of migraine, primary hypertension, CVD, and type 2 diabetes mellitus. All disease elements were defined by the OMOP CDM, which utilizes ICD-10-CM codes.

Representation of positive.

Before querying for positive COVID-19 test results, how the concept of a positive result is represented within a data model must be defined. Within an OMOP CDM, positive lab results are represented as “Detected”, “Positive”, “Present”, or “Reactive” (Kostka et al., 2020).

Positive for COVID-19.

Inclusion-exclusion criteria for lab-confirmed, suspected, and possible cases of COVID-19 referenced the latest version (version 4.0) of the COVID-19 phenotype documentation provided by the National COVID Cohort Collaborative (N3C) (empfff et al., 2022). A positive case is characterized as follows:

- No age or demographic restrictions.
- 1/1/2020 as the start date.
- Patient must have one or more of the COVID-19 LOINC lab codes, with a positive result.
- Patient must have one or more of the “Strong Positive” COVID-19 diagnostic codes.
- Patient must have two or more of the "Weak Positive" COVID-19 diagnostic codes, during the same encounter or on the same date, on or prior to 5/1/2020.

Positive for long COVID.

Inclusion criteria for patients with long COVID are as follows,

- No age or demographic restrictions.
- 1/1/2020 as the start date.
- Patient must have the “U09.9” ICD-10-CM diagnostic code for long COVID.

History of migraine.

Inclusion criteria for patients with a history of migraine are as follows,

- No age or demographic restrictions.
- No starting date restriction. Therefore, patient history date starts from 1/1/2018.
- Patient’s first migraine encounter date must be at least 1 day before the date of their first COVID-19 positive test result.
- Patient must have the keyword “migraine” occur at least once in ICD-10-CM code descriptions. Irrelevant conditions and non-condition concepts, such as “Cyclical vomiting syndrome unrelated to migraine”, were excluded.

Comorbidity: primary hypertension.

For hypertension, only primary hypertension is considered since secondary hypertension has causal relationships with other diseases, which may influence risk for long COVID (UpToDate, 2022b). However, a patient can develop both primary and secondary hypertension at the same time.

- Patient must have the “I10” ICD-10-CM code (ICD10Data, 2019).
- Patient was excluded if they had both the “I10” code and the “I15” code, which codes for secondary hypertension (ICD10Data, 2019).

Comorbidity: CVD.

For CVD, the diseases coronary heart disease (CHD), congestive heart failure (CHF), and stroke were queried (American Heart Association, 2017; World Health Organization, 2021b).

- For CHD, patient must have the "I25" ICD-10-CM code but must not have the following words occur in the code description: "bypass graft", "transplanted", and "dissection" (Hayes et al., 2018; ICD10Data, 2019).
- For CHF, patient must have the "I50.2", "I50.3", or "I50.4" codes for their specific mentions of congestive heart failure in their code descriptions (ICD10Data, 2019).
- For stroke, patient must have the "I63" code (ICD10Data, 2019).
- Any duplicate patients among the 3 groups were eliminated.

Comorbidity: diabetes mellitus.

For diabetes mellitus, only type 2 diabetes mellitus is considered.

- Patient must have the "E11" ICD-10-CM code (BlueCross BlueShield of Alabama, 2022).

Data Organization

Patient data was snapshotted into four categories following the 2x2 contingency table design. The data cross section was snapshot on November 13th, 2022. For the contingency table, "+" indicates exposed or with outcome, and "–" indicates unexposed or without outcome. Every patient was positive or highly likely to be positive for COVID-19. The 4 categories are as follows:

- History of migraine –, Long COVID – (M0L0)
- History of migraine –, Long COVID + (M0L1)

- History of migraine +, Long COVID – (M1L0)
- History of migraine +, Long COVID + (M1L1)

Table 3.1 depicts the 2x2 contingency table layout, with long COVID as disease outcome and history of migraine as exposure.

	Long COVID		Total
History of Migraine	+	–	
+	A	B	A+B
–	C	D	C+D
Total	A+C	B+D	A+B+C+D

Table 3.1: 2x2 contingency table data organization.

Data Analysis

Crude risk ratio.

First, crude risk ratio (RR_{crude}) for long COVID between migraineurs and non-migraineurs and its 95% CI were calculated using the following formulae:

$$RR_{crude} = \frac{\frac{A}{A+B}}{\frac{C}{C+D}}$$

Equation 3.1: Risk ratio formula.

$$95\%\ CI = \exp\left(\ln(RR_{crude}) \pm 1.96 \times \sqrt{\left(\frac{1}{A} - \frac{1}{A+B}\right) + \left(\frac{1}{C} - \frac{1}{C+D}\right)}\right)$$

Equation 3.2: 95% confidence interval for risk ratio.

The two-sided p-value was obtained using R Studio. The specific function used was the "riskratio" method within the "epitab" function, which was available from the R package "epitools".

Stratum-specific risk ratios.

Second, stratum-specific risk ratios were calculated to assess adjusted risk ratios and identify confounders using the Cochran-Mantel-Haenszel (CMH) procedure. The CMH procedure uses a dichotomous exposure variable (previous history of migraine) and a dichotomous outcome variable (long COVID) to generate an estimate of an association between the exposure and the outcome, after adjusting for confounding (LaMorte, 2016). Variables of interest in this thesis are all known or debated risk factors for long COVID: age (<65 years/≥65 years), gender (male/female), and presence of comorbidities (hypertension, CVD, and diabetes mellitus). The confounder age was transformed from continuous to binary categorical. While currently there is no consensus on which age group is the most at risk for long COVID, age can be separated into older adults (≥65 years old) and younger individuals (<65 years old) according to the CDC definition (Centers for Disease Control and Prevention, 2019) and to test the claim in UpToDate that older age was associated with increased risk of long COVID (UpToDate, 2022c).

The CMH procedure first requires a 2x2 table to be constructed for each stratum of a confounder and then a risk ratio to be calculated for each 2x2 table. The 2x2 table is designed as Table 3.1 and the risk ratio formula is the same as Equation 3.1. Next, the CMH Estimate, or the weighted average of risk ratios across all the strata for a cross sectional study, is calculated using the following formula:

$$RR_{CMH} = \frac{\sum \frac{A_i(C_i + D_i)}{n_i}}{\sum \frac{C_i(A_i + B_i)}{n_i}}$$

Equation 3.4: Cochran-Mantel-Haenszel Estimate of a risk ratio for a cross sectional study with retrospective data. i indicates a stratum; A_i, B_i, C_i, and D_i refer to the number of patients in each cell of a 2x2 table in the ith stratum; n_i represents the total number of patients in the ith stratum.

The CMH Estimate is then compared with the crude RR. If the difference between the crude RR and a CMH Estimate for a specific stratum is greater than 10%, then the variable for that stratum is considered a confounder.

The structure of strata is as follows:

1. The 1st stratum was age. Age was categorized into <65 years old and ≥65 years old groups.
2. The 2nd stratum below age was gender. Gender was categorized into males <65 years old, females <65 years old, males ≥65 years old, and females ≥65 years old groups.
3. The 3rd stratum below gender was presence of comorbidities. The 4 age-gender groups were further stratified by the 3 diseases of interest into 12 age-gender-disease groups. 3 CMH Estimates were calculated for the 3 levels of strata.

Results

I/E Criteria Results

Following the N3C COVID Phenotype Documentation, there were a total of N = 128,990 patients positive or highly likely to be positive for COVID-19. After applying the criteria that only male and female patients with more than 1 unique encounter were allowed, 21,939 patients were eliminated, making the baseline sample size N = 107,051.

Within all the COVID-19 positive patients, N = 7,057 (6.59%) had a history of migraine and N = 2,059 (1.92%) had experienced long COVID.

More data subsetting must be done for the 4 contingency table categories. N = 304 (0.28%) had both history of migraine and long COVID, N = 6,753 (6.31%) patients had history of migraine but no long COVID, N = 1,755 (1.64%) patients had long COVID but no history of

migraine, and N = 98,239 (91.77%) patients had COVID-19 but no history of migraine nor long COVID.

Within all the COVID-19 positive patients, N = 23,946 (22.37%) had primary hypertension but no secondary hypertension, N = 11,586 (10.82%) had CVD, and N = 13,668 (12.77%) had type 2 diabetes mellitus.

Demographic Statistics

Figures 3.1, 3.2, and 3.3 delineate the age distributions of all COVID-19 patients, long COVID patients, and migraine patients, respectively. The age groups used for visualization are comparable to the age groups the CDC used for its COVID Data Tracker (Centers for Disease Control and Prevention, 2020b). Detailed data for age distribution can be found in Appendix. Figure 3.4 compares the gender distributions between the 3 main groups.

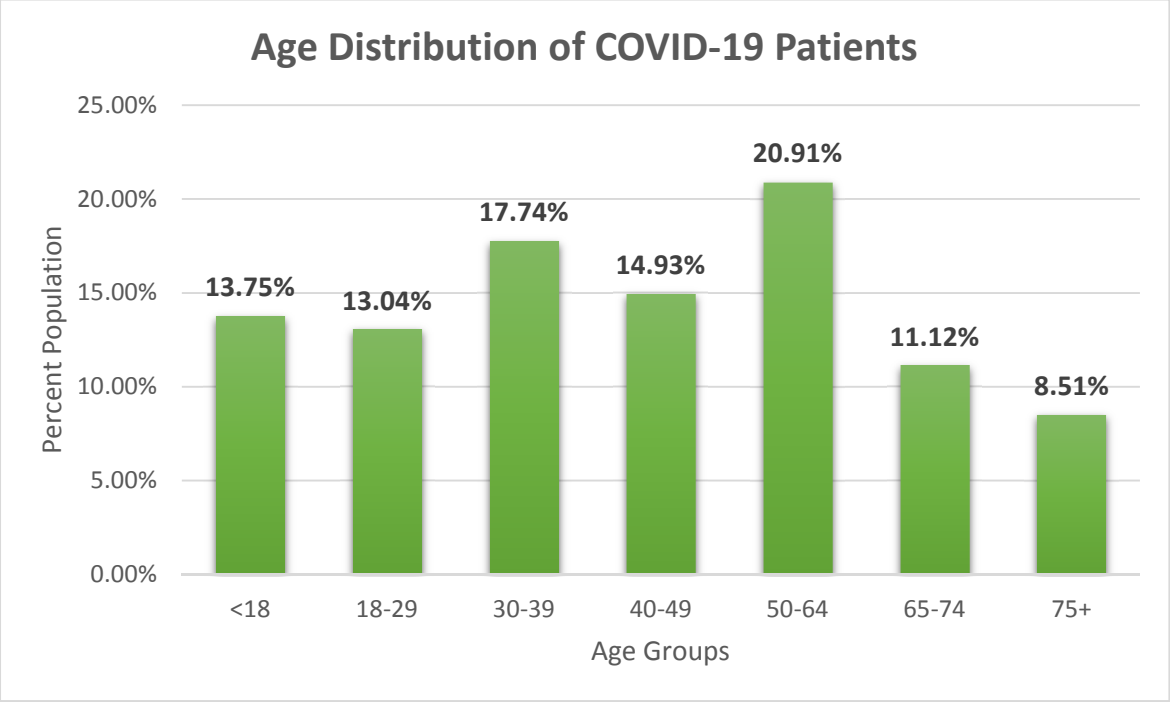

Figure 3.1: Age distribution of COVID-19 patient age groups.

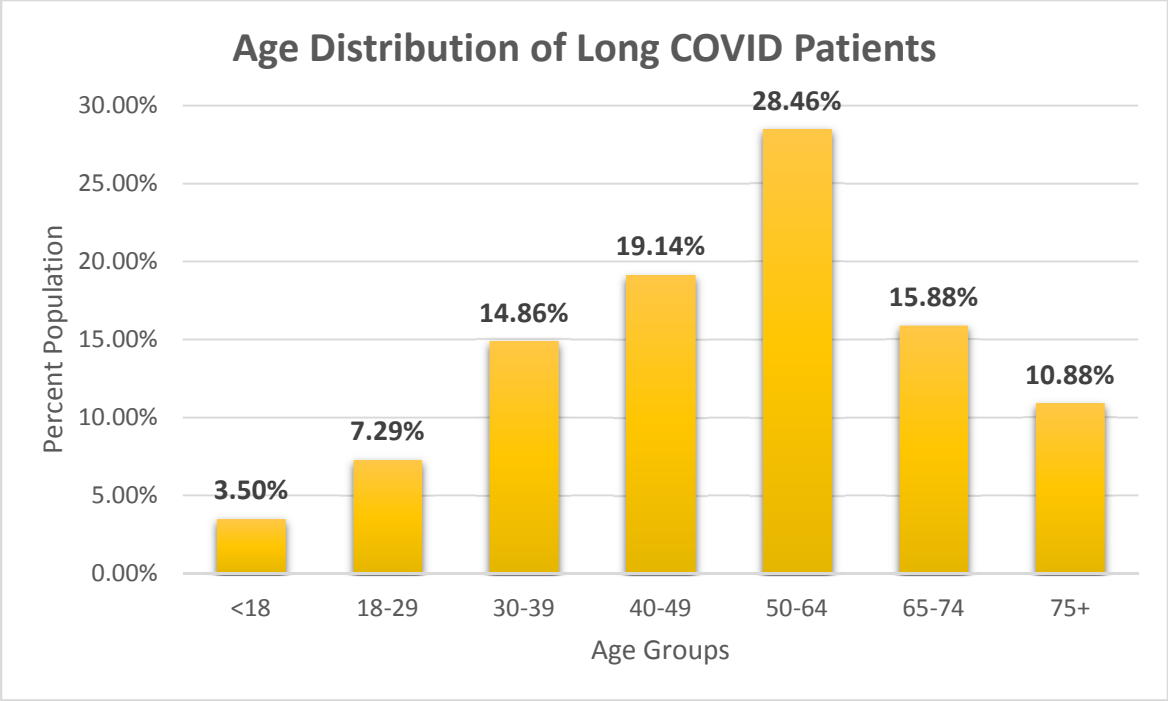

Figure 3.2: Age distribution of long COVID patient age groups.

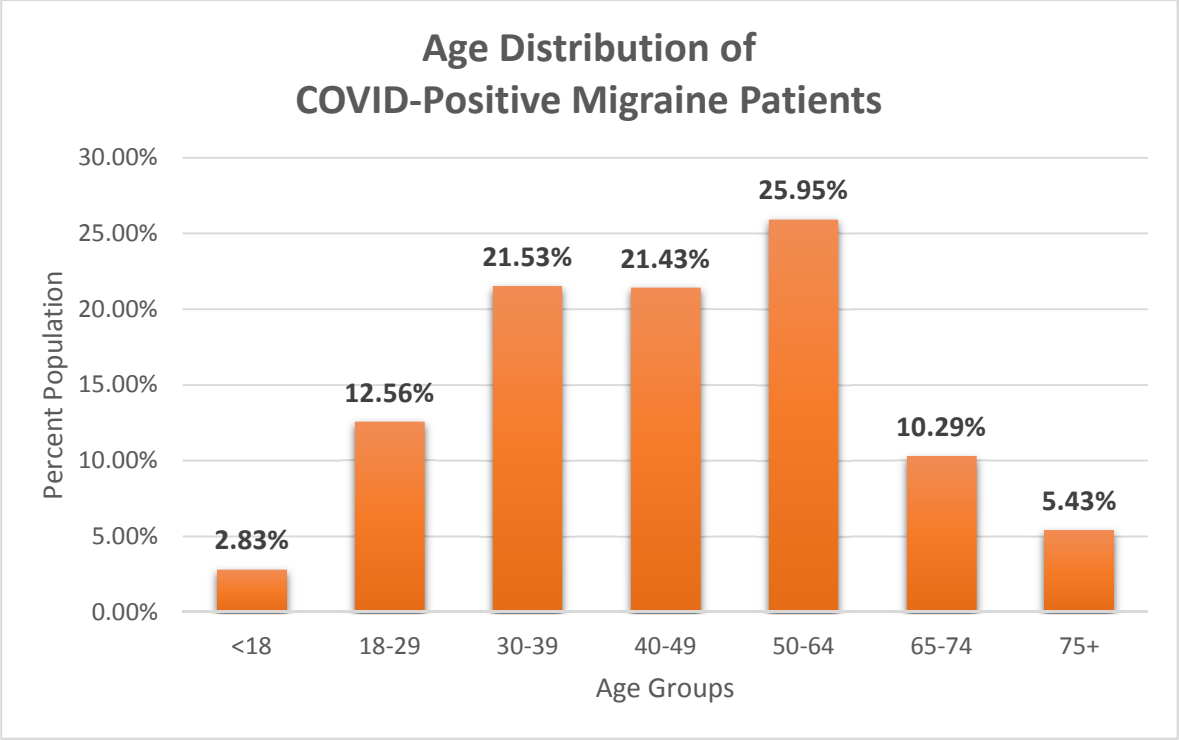

Figure 3.3: Age distribution of COVID-positive migraineur age groups.

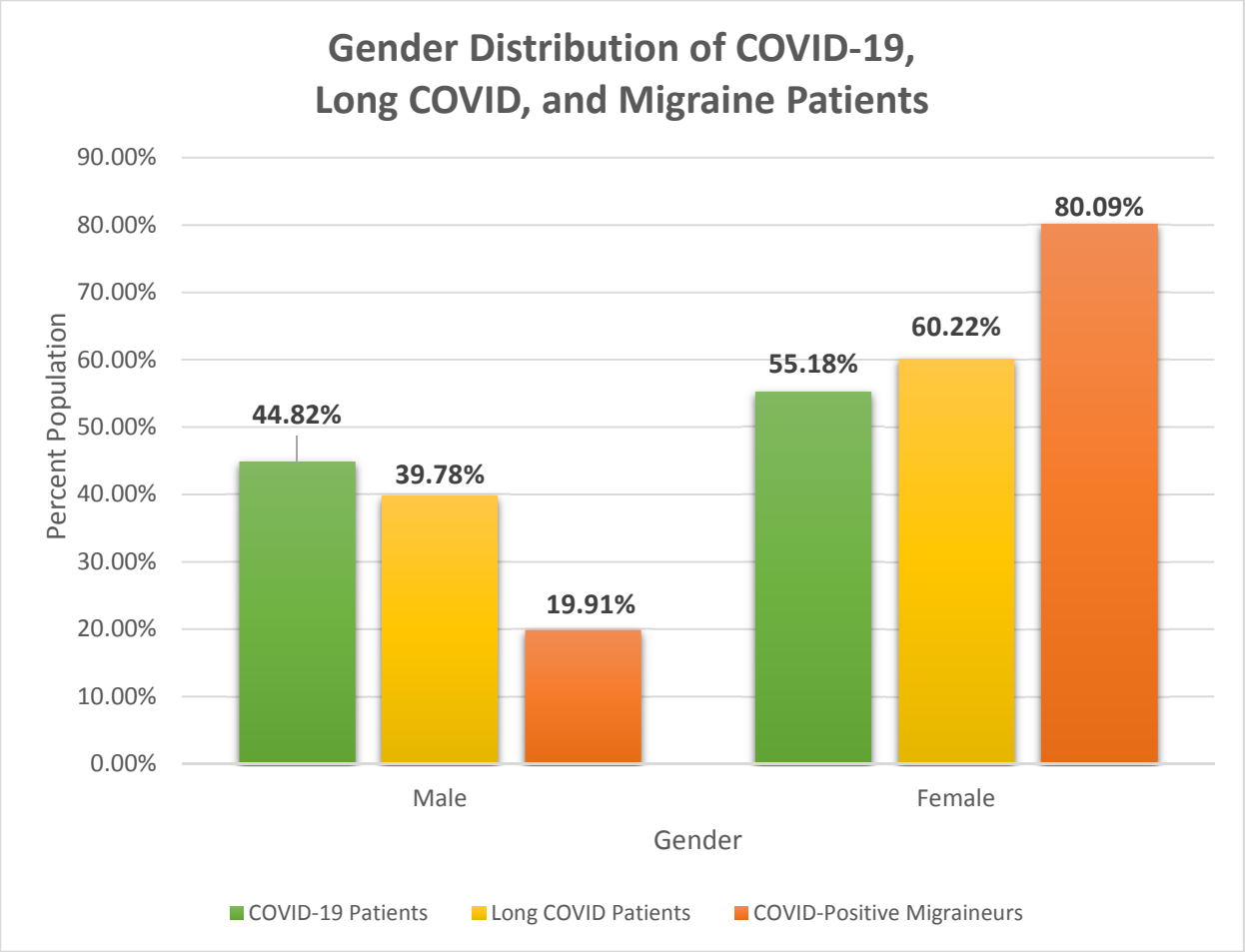

Figure 3.4: Gender distribution of COVID-19 patients, long COVID patients, and COVID-positive migraineurs.

Risk of Long COVID for Migraineurs

Table 3.3 provides all crude and stratum-specific risk ratios and their 95% CIs and p-values. Crude risk ratio of long COVID between COVID-positive migraineurs and COVID-positive non-migraineurs showed significant association (RR 2.45, 95% CI: (2.64, 2.88), p-value = <0.0001).

Crude RR: Migraineurs				
	RR	95% CI LB	95% CI UB	p-value
	2.45	2.18	2.77	<0.0001

1st Stratum: Older and Younger Migraineurs				
	RR	95% CI LB	95% CI UB	p-value
≥ 65 years	1.56	1.15	2.11	0.007
< 65 years	2.81	2.46	3.2	<0.0001
2nd Stratum: Male and Female Migraineurs				
	RR	95% CI LB	95% CI UB	p-value
Male ≥ 65 years	1.88	1.09	3.24	0.88
Female < 65 years	1.36	0.95	1.96	0.1
Male ≥ 65 years	3.01	2.25	4.02	<0.0001
Female < 65 years	2.65	2.28	3.09	<0.0001
3rd Stratum: Migraineurs with Comorbidity				
Presence of HT	RR	95% CI LB	95% CI UB	p-value
Male ≥ 65 years	1.24	0.59	2.61	0.52
Female < 65 years	1.26	0.76	2.09	0.38
Male ≥ 65 years	1.87	1.18	2.97	0.01
Female < 65 years	2.14	1.60	2.87	<0.0001
Presence of CVD	RR	95% CI LB	95% CI UB	p-value
Male ≥ 65 years	1.32	0.63	2.78	0.50
Female < 65 years	1.06	0.56	2.01	0.86
Male ≥ 65 years	1.48	0.73	2.99	0.26
Female < 65 years	1.81	1.19	2.76	0.008
Presence of DM	RR	95% CI LB	95% CI UB	p-value
Male ≥ 65 years	1.68	0.70	4.04	0.23
Female < 65 years	1.03	0.46	2.34	0.83
Male ≥ 65 years	1.66	0.82	3.34	0.16
Female < 65 years	1.85	1.23	2.79	0.005

Table 3.3: Crude and stratum-specific risk ratios of long COVID between COVID-positive migraineurs and COVID-positive non-migraineurs. RR = risk ratio, CI = confidence interval, LB = lower bound, UB = upper bound, HT = hypertension, CVD = cardiovascular disease, DM = type 2 diabetes mellitus.

Next, the 1st stratum, age-specific risk ratios, was calculated. For both ≥ 65 years and < 65 years old patients, history of migraine was significantly associated with long COVID. The CMH

Estimate for the 2 groups was 2.51, which had a 2.26% difference from the crude risk ratio. Since the percent difference was less than 10%, age was not a confounder.

The 2nd stratum was age-gender-specific risk ratios. For both male and female patients < 65 years old, having history of migraine was significantly associated with long COVID. For both male and female patients ≥ 65 years old, there is no significant risk of long COVID if they have history of migraine. The CMH Estimate for the 4 groups was 2.42, which had a 1.39% difference from the crude risk ratio. Since the percent difference was less than 10%, gender was not a confounder.

The 3rd stratum was age-gender-comorbidity risk ratios. For female patients < 65 years old only, having history of migraine and any one of the 3 comorbidities, which includes hypertension, CVD, or type 2 diabetes, was significantly associated with long COVID. All other groups had no significant association between history of migraine and long COVID. The CMH Estimate for the 12 groups was 1.66, which had a 32.42% difference from the crude risk ratio. Since the percent difference was greater than 10%, presence of any of these 3 comorbidities was a confounder.

Discussion

UC CORDS Patient Demographics Compared Against Literature

There are some notable similarities and differences in COVID-19 populations between the CDC data and UC CORDS data. CDC data was imported from the CDC COVID Data Tracker (Centers for Disease Control and Prevention, 2020b). In Figure 4.1, the UC CORDS data had less COVID-19 patients younger than 29 years old and more patients older than 65 years old. The ages 30-64 had similar proportions, with UC CORDS having 4.28% more patients than the

CDC. This shows that there were more senior COVID-19 patients who went to UC Health entities compared to health entities from across the U.S. In Figure 4.2, the gender distributions between the two datasets were very similar with a total difference of 2.96% across both male and female genders.

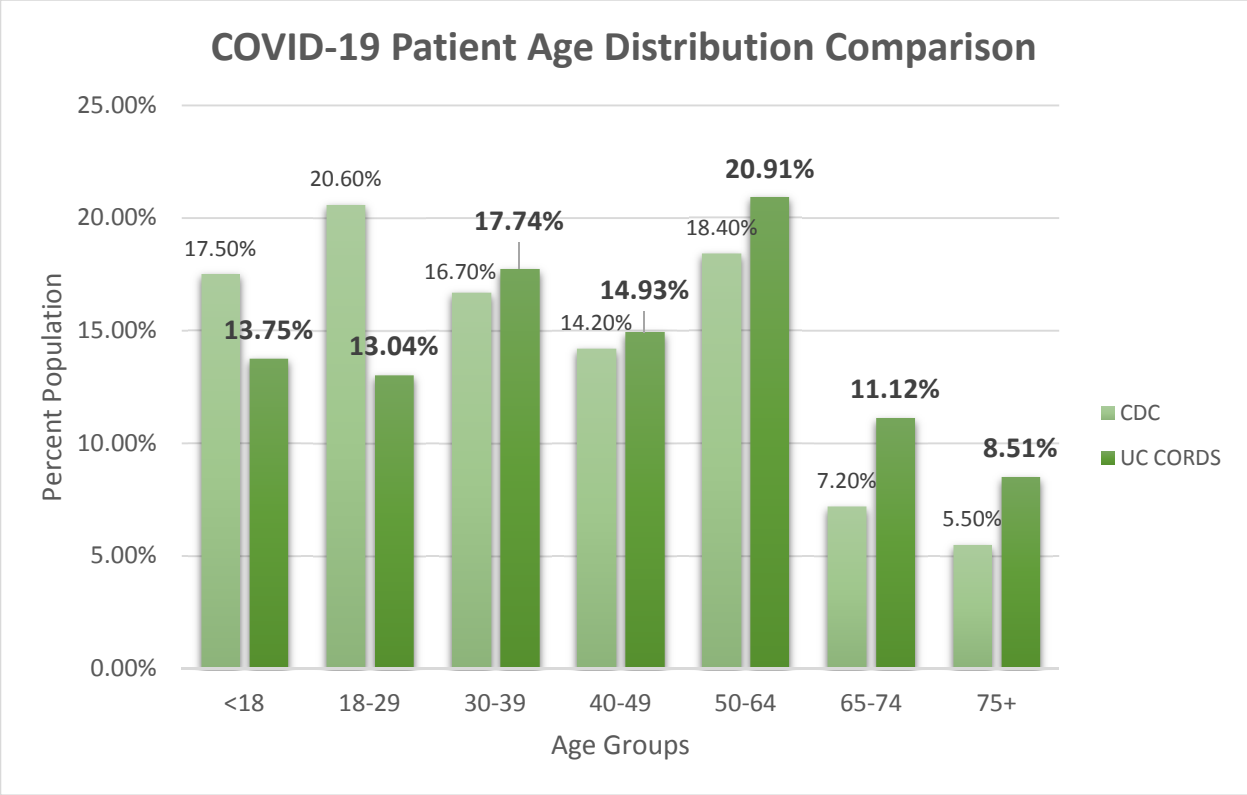

Figure 4.1: Age distribution comparison between CDC's COVID Data Tracker data (N = 90,398,392) and UC CORDS data (N = 107,051).

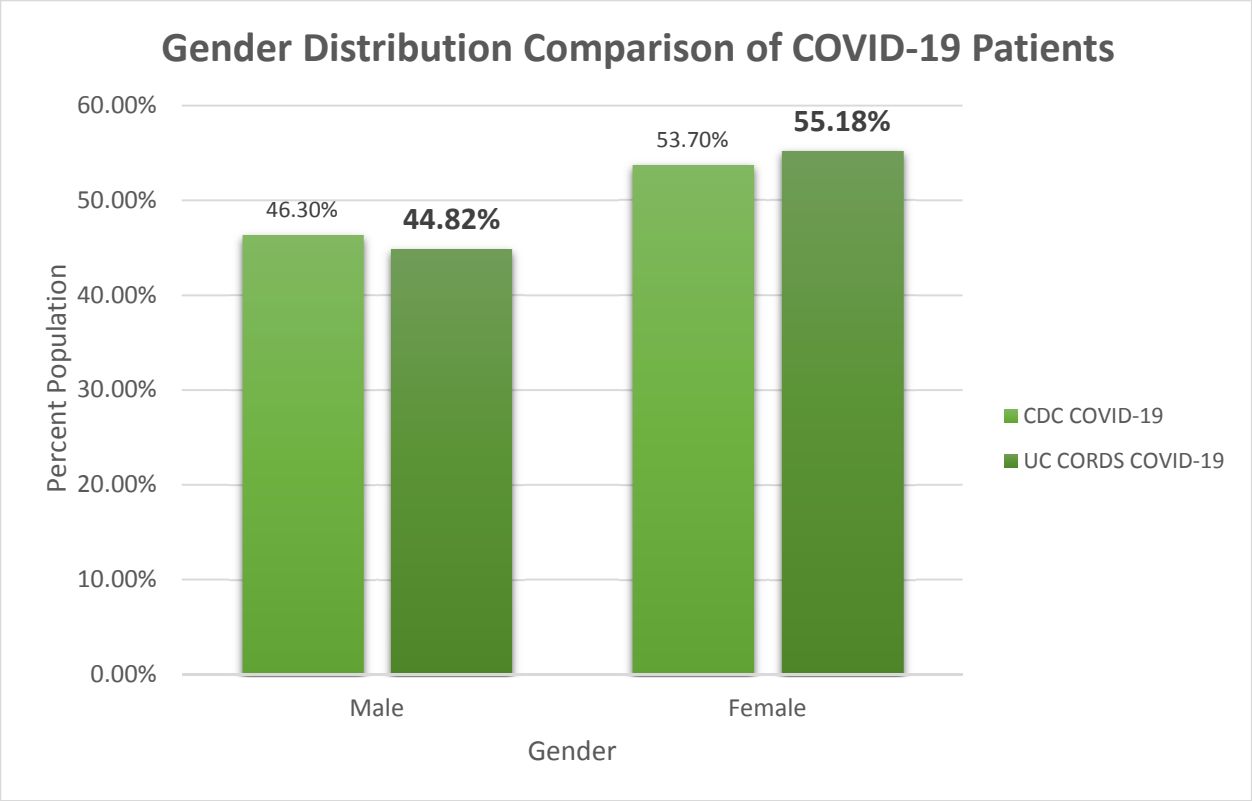

Figure 4.2: Gender distribution comparison between CDC's COVID Data Tracker data (N = 89,378,508) and UC CORDS data (N = 107,051).

In Figures 4.3 and 4.4, COVID-positive migraineurs from UC CORDS were compared with just migraineurs from the Migraine in America Symptoms and Treatment (MAST) online survey (Buse et al., 2020). Again, UC CORDS showed a higher proportion of senior patients compared to MAST data. Remarkedly, there was a 7.38% difference in the 25-34 group in favor of MAST, perhaps due to the accessibility of the online survey. 7.09% more women were documented in UC CORDS compared to the MAST survey. This could be attributed to either the higher prevalence of COVID-19 cases in women, or the greater number of women who goes to see a headache doctor due to the higher prevalence of migraine in women.

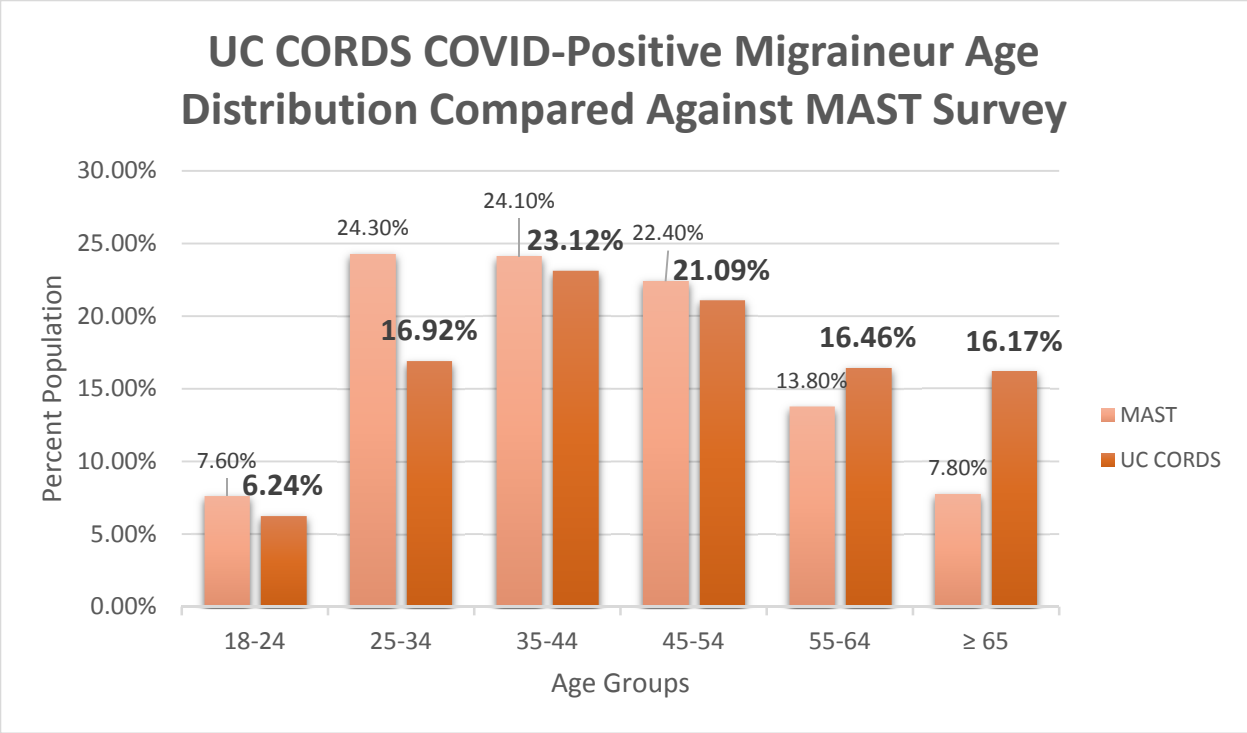

Figure 4.3: Age distribution comparison between migraineur data from the Migraine in America Symptoms and Treatment (MAST) study (N = 92,586), and COVID-positive migraineur data from UC CORDS (N = 92,329). Individuals <18 years of age were excluded.

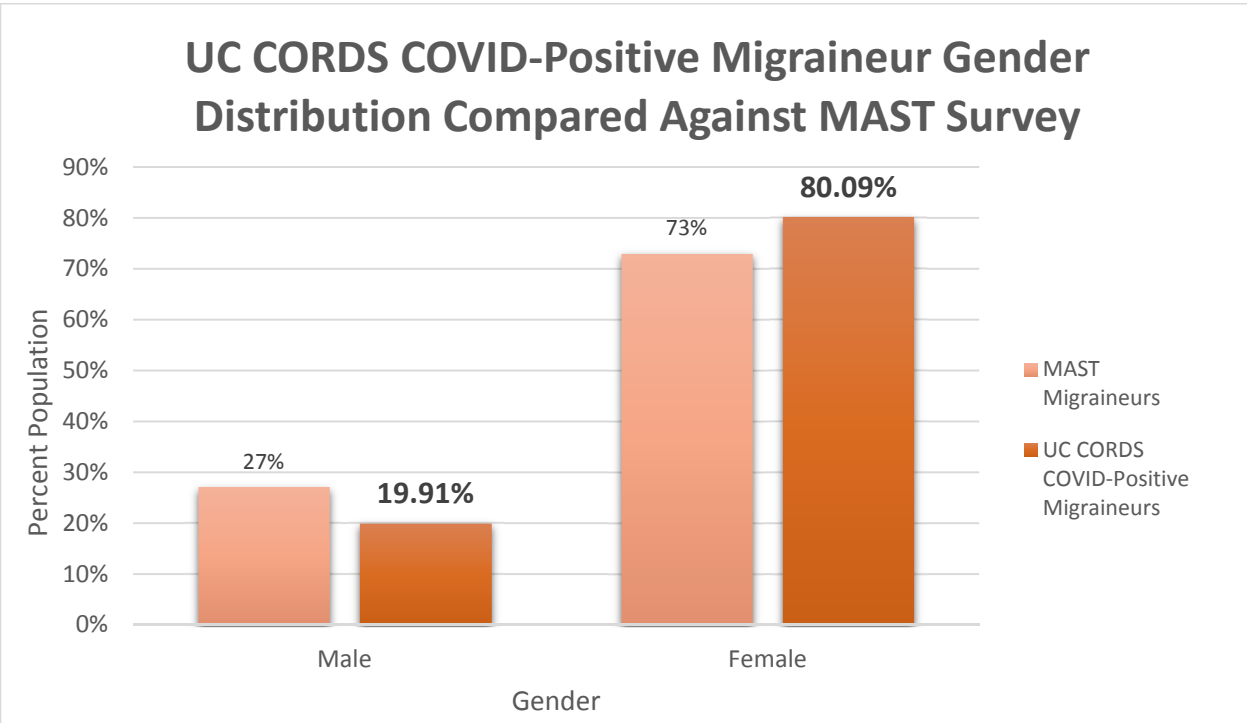

Figure 4.4: Gender distribution comparison between migraineur data from the Migraine in America Symptoms and Treatment (MAST) study (N = 92,586), and COVID-positive migraineur data from UC CORDS (N = 92,329). Individuals <18 years of age were excluded.

Long COVID data comparison between HPS and UC CORDS requires separate discussion (Figure 4.5). The HPS presents long COVID data as a percentage of adults who have ever had COVID, not as a percentage of the long COVID population. Instead, this comparison shows the difference in performance between patient surveys and EHR symptom reporting. While patient surveys can be inaccurate in recalling specific information, they are powerful in determining whether an event occurred at all. The low percentage from the UC CORDS data could be attributed by the phenomenon that many long COVID diagnoses in the EHR are secondary diagnoses. However, data and analysis included this thesis cannot validate that claim.

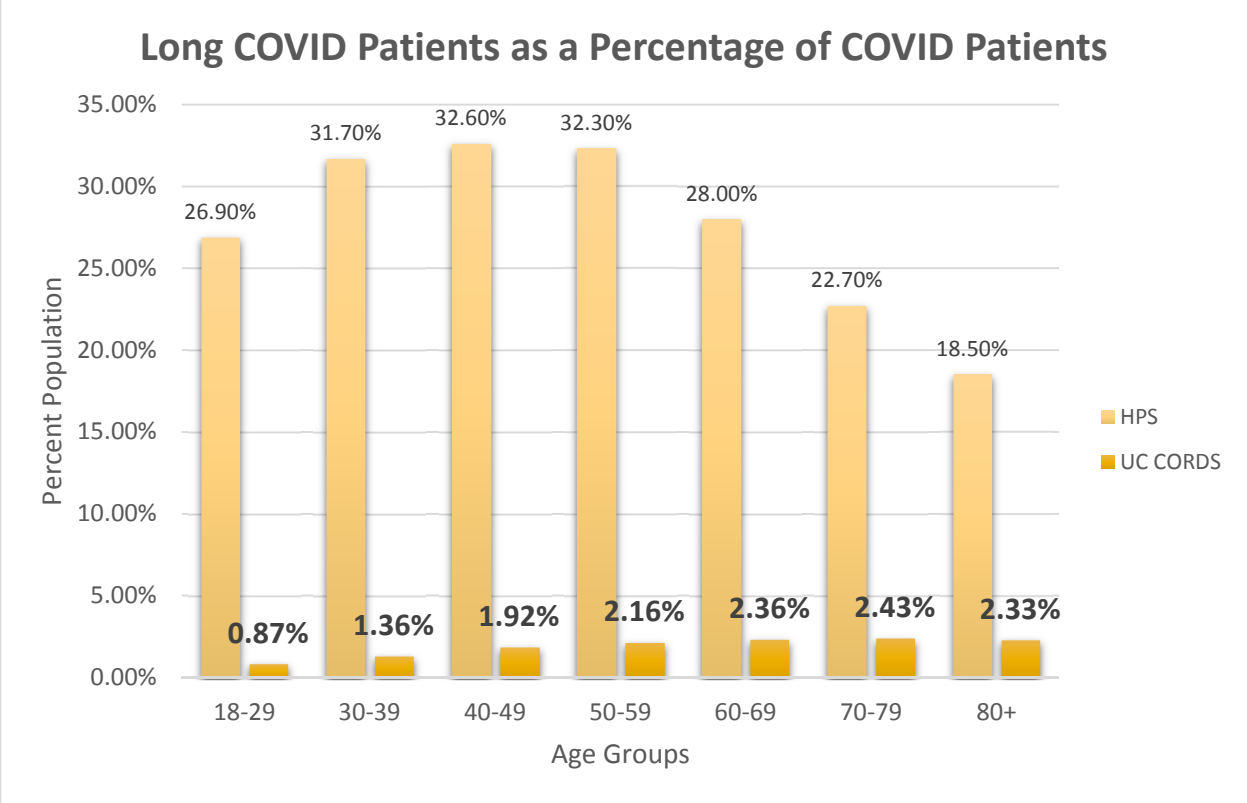

Figure 4.5: Age distribution of long COVID patients as a percentage of COVID patients, not a percentage of the total long COVID population. Compared between the Household Pulse Survey (HPS) in (N = 41,415) and UC CORDS data (N = 92,329). Individuals <18 years of age were excluded.

Risk of Long COVID for Migraineurs

There were notable but unsurprising patterns in the adjusted risk ratios regarding age. According to large migraine demographic studies, migraine affects people of working age the most, which includes those aged between 25-54 years old (Buse et al., 2020; World Health Organization, 2016). In the age-adjusted stratum, migraineurs < 65 years old had higher risk than migraineurs ≥ 65 years old (RR 2.81, p-value <0.0001 vs. RR 1.56, p-value = 0.007). This comparison held true in the gender-adjusted stratum, in which both genders of migraineurs < 65 years old had significant associations with long COVID, whereas the risk of long COVID for both genders of older migraineurs was nonsignificant. This was unsurprising because the 2 age groups were decided in consideration of risk of long COVID, not prevalence of migraine. Since most migraine patients are younger than 65 years old, the result that most of the migraineurs that get long COVID are also younger than 65 seems logical. These risk ratios also demand a deeper dive into long COVID risk for specifically migraineurs of working age.

There were surprising and unsurprising results regarding gender. Both long COVID and migraine disproportionately affect women, a phenomenon reflected by demographic statistics from governmental and scholarly sources and UC CORDS. In the 3rd stratum, female migraineurs < 65 years old was the only group that had significant risk of long COVID if they also had hypertension (RR 2.14, p-value <0.0001), CVD (RR 1.81, p-value = 0.008), or type 2 diabetes mellitus (RR 1.85, p-value = 0.005). However, in the 2nd stratum, male migraineurs < 65 years old surprisingly had higher risk of long COVID than female migraineurs < 65 years old (RR 3.01, p-value <0.0001 vs RR 2.65, p-value<0.0001). There might be more effect modification under the age and gender variables that this thesis has not discovered.

So far, only 2 other studies have specifically addressed the association of long COVID with history of migraine. Magdy et al. conducted a case-control study investigating long COVID neuropsychiatric symptoms among 204 COVID-19 survivors with migraine via patient interview (Magdy et al., 2022). The study found that the most significant long COVID neuropsychiatric symptoms in migraineurs were headache (OR 3.15, 95% CI: (1.62, 6.14), p-value < 0.001), anxiety (OR 3.267, 95% CI: (1.75, 6.11), p-value <0.001), and depression (OR 2.259, 95% CI: (1.28, 3.98), p-value = 0.004). The researchers also utilized a "Post-COVID-19 Functional Status scale" to study the impact of long COVID on migraineurs. Next, Fernández-de-las-Peñas et al. also conducted a case-control study investigating frequency and trends of any long COVID symptoms among 57 COVID-19 survivors with migraine via patient telephone interview (Fernández-de-las-Peñas et al., 2021). These researchers found that fatigue was more prevalent among migraineurs (OR 2.89, 95% CI: (1.32, 6.32), p-value = 0.008). Considering both case-control studies used patient interviews for symptom reporting, UC CORDS could serve as a great data source for more accurate characterizations of long COVID symptoms.

Data Complexity in UC CORDS

There are benefits to working with a standardized data set like UC CORDS. First, UC CORDS comprises EHRs sourced from Epic software, which is the gold standard EHR system. Health professionals must perform necessary assessments and follow up-to-date case definitions before deciding on a diagnosis and inputting information into a patient's EHR. Therefore, using recorded symptoms in EHRs instead of relying on patient interviews or surveys can increase accuracy in both the description of the symptoms and the time of the encounter. Second, there would be numerous analytic tool options available when a dataset is made for the purpose of research. For this thesis, Python was used to query UC CORDS. However, UC CORDS and the

OMOP CDM can support R and Statistical Analysis System (SAS). In addition, information (not the raw patient data) extracted from UC CORDS can be easily exported and analyzed in another software, if needed. And lastly, there is rich documentation of the OMOP CDM and UC CORDS available online at any time.

However, any data set will have its own nuances, and dealing with these nuances can be a challenge. A problem faced by every COVID researcher using UC CORDS is the visit occurrence ID complexity problem. When a patient comes in for a COVID-19 lab test, the provider will record the date and time of the visit and the test. However, in UC CORDS, a patient's encounter ID may not be constant due to refreshes from three staggered data feeds: daily COVID-related testing and hospitalization data, weekly supplemental clinical data, and monthly refreshes with all patient history. Therefore, these IDs did not always match up in the past. Data curators resolved this problem by setting these patients' encounter ID to a constant default value -1. Here comes the complexity problem: records that do not have an encounter time also use the default value -1. Unfortunately, this created three groups of patients: 1) patients with normal, non-default encounter IDs and encounter dates, 2) patients with encounter dates but has -1, and 3) patients with -1 and no encounter date.

This complexity problem can be resolved using Steve Covington's 3-part union query (Covington, 2020). Figure 4.6 depicts the logic of the query:

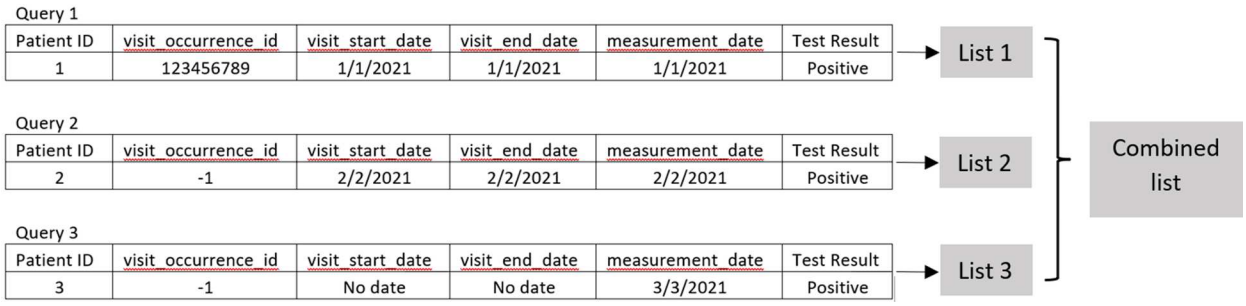

Figure 4.6: Graphical illustration of the logic of Steve Covington's 3-part union query.

Since this complexity problem created three groups of patients, this query captures exactly those groups of patients. After combining the three groups, the result is a single list of patients positive for COVID-19, without any duplicates. While nuances need to be considered, the principles upon which the data model is constructed do not change, and methods and resources available for resolving problems brought on by nuances are always growing.

Limitations and Strengths

When reviewing the UC CORDS data and the resulting analyses, several considerations and limitations should be kept in mind. Data and analyses from this study cannot be generalized to the entire Californian or U.S. population, for only patients who could access a UC medical center were included and studied. This thesis did not consider the different major variants – Wuhan, Delta, and Omicron (World Health Organization, 2022) – of the Sars-Cov-2 virus, their lethality, and their potential in developing long COVID. It also did not account for the number of recurrent COVID-19 infections, nor the vaccination status of these patients. Current literature has shown that COVID vaccines are effective at reducing long COVID risk, albeit findings about their effectiveness are still mixed (Al-Aly et al., 2022; Antonelli et al., 2021; Strain et al., 2022). The effects of these variables can be investigated in a future study since the stratum-specific risk ratio analytical method works best with only a few variables.

There are several main strengths to this thesis tied to UC CORDS. First, UC CORDS is a dataset containing millions of EHRs. EHRs ensure accurate reporting of encounter times and eliminate inconsistencies in symptom reporting. Second, the dataset contains records of patients with diverse backgrounds from different regions of California. Third, the size of UC CORDS is massive enough to produce analysis with high statistical power, which in turn enabled stratum-specific risk analysis.

Conclusion

Analysis from this thesis showed that COVID-positive migraineurs are more likely to experience long COVID than COVID-positive non-migraineurs. Stratum-specific risk ratio analysis further revealed that female COVID-positive migraineurs younger than 65 years old with comorbid primary hypertension, CVD, or type 2 diabetes mellitus are more at risk for long COVID. This thesis also aims to be a showcase of the depth and complexity of the UC CORDS and an encouragement for future researchers to leverage this standardized dataset.

References

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

Alkodaymi, M. S., Omrani, O. A., Fawzy, N. A., Shaar, B. A., Almamlouk, R., Riaz, M., Obeidat, M., Obeidat, Y., Gerberi, D., Taha, R. M., Kashour, Z., Kashour, T., Berbari, E. F., Alkattan, K., & Tleyjeh, I. M. (2022). Prevalence of post-acute COVID-19 syndrome symptoms at different follow-up periods: a systematic review and meta-analysis. *Clinical Microbiology and Infection*, *28*(5). https://doi.org/10.1016/j.cmi.2022.01.014

American Heart Association. (2017, May 31). *What is Cardiovascular Disease?* Www.heart.org; American Heart Association. https://www.heart.org/en/health-topics/consumer-healthcare/what-is-cardiovascular-disease

Antonelli, M., Penfold, R. S., Merino, J., Sudre, C. H., Molteni, E., Berry, S., Canas, L. S., Graham, M. S., Klaser, K., Modat, M., Murray, B., Kerfoot, E., Chen, L., Deng, J., Österdahl, M. F., Cheetham, N. J., Drew, D. A., Nguyen, L. H., Pujol, J. C., & Hu, C. (2021). Risk factors and disease profile of post-vaccination SARS-CoV-2 infection in UK users of the COVID Symptom Study app: a prospective, community-based, nested, case-control study. *The Lancet Infectious Diseases*, *22*(1). https://doi.org/10.1016/s1473-3099(21)00460-6

Blacketer, C. (2021). *The Book of OHDSI*. https://ohdsi.github.io/TheBookOfOhdsi/CommonDataModel.html

BlueCross BlueShield of Alabama. (2022). *Correctly Coding: Diabetes Mellitus*. https://providers.bcbsal.org/portal/web/pa/resources. https://providers.bcbsal.org/portal/documents/10226/306297/Correctly+Coding+Diabetes

+Mellitus/cf5e3336-d1b7-4abb-aa17-b03b33e35d90#:~:text=ICD%2D10%20Code%20Z79

Booth, A., Reed, A. B., Ponzo, S., Yassaee, A., Aral, M., Plans, D., Labrique, A., & Mohan, D. (2021). Population risk factors for severe disease and mortality in COVID-19: A global systematic review and meta-analysis. *PLOS ONE*, *16*(3), e0247461. https://doi.org/10.1371/journal.pone.0247461

Buse, D. C., Reed, M. L., Fanning, K. M., Bostic, R., Dodick, D. W., Schwedt, T. J., Munjal, S., Singh, P., & Lipton, R. B. (2020). Comorbid and co-occurring conditions in migraine and associated risk of increasing headache pain intensity and headache frequency: results of the migraine in America symptoms and treatment (MAST) study. *The Journal of Headache and Pain*, *21*(1). https://doi.org/10.1186/s10194-020-1084-y

Canas, L. S., Molteni, E., Deng, J., Sudre, C. H., Murray, B., Kerfoot, E., Antonelli, M., Chen, L., Rjoob, K., Pujol, J. C., Polidori, L., May, A., Osterdahl, M. F., Whiston, R., Cheetham, N. J., Bowyer, V., Spector, T. D., Hammers, A., Duncan, E. L., & Ourselin, S. (2022). *Profiling post-COVID syndrome across different variants of SARS-CoV-2*. https://doi.org/10.1101/2022.07.28.22278159

Carfì, A., Bernabei, R., & Landi, F. (2020). Persistent symptoms in patients after acute COVID-19. *JAMA*, *324*(6), 603–605. https://doi.org/10.1001/jama.2020.12603

Centers for Disease Control and Prevention. (2019, January 21). *Indicator Definitions - Older Adults | CDI | DPH | CDC*. Www.cdc.gov. https://www.cdc.gov/cdi/definitions/older-adults.html

Centers for Disease Control and Prevention. (2020a, March 28). *COVID Data Tracker*. Centers for Disease Control and Prevention. https://covid.cdc.gov/covid-data-tracker/#variant-proportions

Centers for Disease Control and Prevention. (2020b, March 28). *Demographic Trends of COVID-19 cases and deaths in the US reported to CDC*. Centers for Disease Control and Prevention. https://covid.cdc.gov/covid-data-tracker/#demographics

Centers for Disease Control and Prevention. (2021, September 16). *Long COVID or Post-COVID Conditions*. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html

Centers for Disease Control and Prevention. (2022, July 19). *Long COVID - Household Pulse Survey - COVID-19*. Www.cdc.gov. https://www.cdc.gov/nchs/covid19/pulse/long-covid.htm

Covington, S. (2020, September 18). *UCCORDS_Visit_Occurrence_IDs.sql*. Covid-Lds. https://gitlab.ri.ucdavis.edu/ucd-public/covid-lds/-/blob/dev/SQL%20Query%20Library/UC_CORDS/UC%20Davis%20--%20Visit%20Occurrence%20ID%20Complexity/UCCORDS_Visit_Occurrence_IDs.sql

empfff, Kostka, K., & Marie, K. (2022, March 11). *Latest Phenotype · National-COVID-Cohort-Collaborative/Phenotype_Data_Acquisition Wiki*. GitHub. https://github.com/National-COVID-Cohort-Collaborative/Phenotype_Data_Acquisition/wiki/Latest-Phenotype

Fernández-de-las-Peñas, C., Gómez-Mayordomo, V., García-Azorín, D., Palacios-Ceña, D., Florencio, L. L., Guerrero, A. L., Hernández-Barrera, V., & Cuadrado, M. L. (2021). Previous History of Migraine Is Associated With Fatigue, but Not Headache, as Long-Term Post-COVID Symptom After Severe Acute Respiratory SARS-CoV-2 Infection: A

Case-Control Study. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.678472

Guidetti, D., Rota, E., Morelli, N., & Immovilli, P. (2014). Migraine and Stroke: â€œVascularâ€ Comorbidity. *Frontiers in Neurology*, *5*(193). https://doi.org/10.3389/fneur.2014.00193

Halpin, S. J., McIvor, C., Whyatt, G., Adams, A., Harvey, O., McLean, L., Walshaw, C., Kemp, S., Corrado, J., Singh, R., Collins, T., O'Connor, R. J., & Sivan, M. (2020). Postdischarge symptoms and rehabilitation needs in survivors of COVID‐19 infection: A cross‐sectional evaluation. *Journal of Medical Virology*, *93*(2). https://doi.org/10.1002/jmv.26368

Hayes, S. N., Kim, E. S. H., Saw, J., Adlam, D., Arslanian-Engoren, C., Economy, K. E., Ganesh, S. K., Gulati, R., Lindsay, M. E., Mieres, J. H., Naderi, S., Shah, S., Thaler, D. E., Tweet, M. S., & Wood, M. J. (2018). Spontaneous Coronary Artery Dissection: Current State of the Science: A Scientific Statement From the American Heart Association. *Circulation*, *137*(19). https://doi.org/10.1161/cir.0000000000000564

Huang, C., Huang, L., Wang, Y., Li, X., Ren, L., Gu, X., Kang, L., Guo, L., Liu, M., Zhou, X., Luo, J., Huang, Z., Tu, S., Zhao, Y., Chen, L., Xu, D., Li, Y., Li, C., Peng, L., & Li, Y. (2021). 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *The Lancet*, *0*(0). https://doi.org/10.1016/S0140-6736(20)32656-8

ICD10Data. (2019). *The Web's Free 2019 ICD-10-CM/PCS Medical Coding Reference*. Icd10data.com. https://www.icd10data.com/

Ingargiola, A. (2015). *Jupyter/IPython Notebook Quick Start Guide — Jupyter/IPython Notebook Quick Start Guide 0.1 documentation*. Jupyter-Notebook-Beginner-Guide.readthedocs.io. https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/index.html

Kostka, K., empfff, & mim18. (2020, December 1). *Defining a "Positive" COVID Lab*. National-COVID-Cohort-Collaborative / Phenotype_Data_Acquisition. https://github.com/National-COVID-Cohort-Collaborative/Phenotype_Data_Acquisition/wiki/Defining-a-%22Positive%22-COVID-Lab

LaMorte, W. W. (2016, June 3). *The Cochran-Mantel-Haenszel Method*. Sphweb.bumc.bu.edu. https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704-EP713_Confounding-EM/BS704-EP713_Confounding-EM7.html

Li, J., Huang, D. Q., Zou, B., Yang, H., Hui, W. Z., Rui, F., Yee, N. T. S., Liu, C., Nerurkar, S. N., Kai, J. C. Y., Teng, M. L. P., Li, X., Zeng, H., Borghi, J. A., Henry, L., Cheung, R., & Nguyen, M. H. (2020). Epidemiology of COVID□19: A systematic review and meta□analysis of clinical characteristics, risk factors, and outcomes. *Journal of Medical Virology*, *93*(3). https://doi.org/10.1002/jmv.26424

Lipton, R. B., Bigal, M. E., Diamond, M., Freitag, F., Reed, M. L., & Stewart, W. F. (2007). Migraine prevalence, disease burden, and the need for preventive therapy. *Neurology*, *68*(5), 343–349. https://doi.org/10.1212/01.wnl.0000252808.97649.21

Long COVID Information and Resources | National Institutes of Health. (n.d.). NIH COVID-19 Research. Retrieved November 19, 2022, from https://covid19.nih.gov/covid-19-topics/long-covid#what-we-know-about-long-covid-1

Magdy, R., Elmazny, A., Soliman, S. H., Elsebaie, E. H., Ali, S. H., Abdel Fattah, A. M., Hassan, M., Yassien, A., Mahfouz, N. A., Elsayed, R. M., Fathy, W., Abdel-Hamid, H. M., Mohamed, J., & Hussein, M. (2022). Post-COVID-19 neuropsychiatric manifestations among COVID-19 survivors suffering from migraine: a case–control study. *The Journal of Headache and Pain*, *23*(1). https://doi.org/10.1186/s10194-022-01468-y

Martín-Martín, J., Martín-Cazorla, F., Suárez, J., Rubio, L., & las-Heras, S. M. -. (2022). Comorbidities and autopsy findings of COVID-19 deaths and their association with time to death: a systematic review and meta-analysis. *Current Medical Research and Opinion*, *38*(5), 1–22. https://doi.org/10.1080/03007995.2022.2050110

May, A., & Goadsby, P. J. (1999). The trigeminovascular system in humans: pathophysiologic implications for primary headache syndromes of the neural influences on the cerebral circulation. *Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism*, *19*(2), 115–127. https://doi.org/10.1097/00004647-199902000-00001

National Geographic. (2022, November 18). *Coronavirus in the U.S.: Where cases are growing and declining*. National Geographic. https://www.nationalgeographic.com/science/graphics/graphic-tracking-coronavirus-infections-us

NICE, RCGP, and SIGN. (2022, November 2). *MAGICapp - Making GRADE the Irresistible Choice - Guidelines and Evidence summaries*. App.magicapp.org. https://app.magicapp.org/#/guideline/EQpzKn/section/EKbyVn

Observational Health Data Sciences and Informatics. (2022). *OMOP Common Data Model*. Ohdsi.github.io. https://ohdsi.github.io/CommonDataModel/

Office for Civil Rights. (2021, July 26). *Guidance on "Long COVID" as a Disability Under the ADA, Section 504, and Section 1557*. HHS.gov. https://www.hhs.gov/civil-rights/for-providers/civil-rights-covid19/guidance-long-covid-disability/index.html#footnote10_0ac8mdc

Office for National Statistics. (2022, November 18). *Coronavirus (COVID-19) latest insights - Office for National Statistics*. Www.ons.gov.uk. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19latestinsights/infections#long-covid

Office of Human Subjects Research - Institutional Review Board. (2015, April). *HIPAA - Definition of Limited Data Set*. Www.hopkinsmedicine.org. https://www.hopkinsmedicine.org/institutional_review_board/hipaa_research/limited_data_set.html

PubMed. (2022, November 15). *Help*. PubMed. https://pubmed.ncbi.nlm.nih.gov/help/#covid19-article-filters

Reich, C., & Ostropolets, A. (2021). *The Book of OHDSI*. https://ohdsi.github.io/TheBookOfOhdsi/StandardizedVocabularies.html#concepts

Sanyaolu, A., Okorie, C., Marinkovic, A., Patidar, R., Younis, K., Desai, P., Hosein, Z., Padda, I., Mangat, J., & Altaf, M. (2020). Comorbidity and its Impact on Patients with COVID-19. *Sn Comprehensive Clinical Medicine*, *2*(8), 1–8. https://doi.org/10.1007/s42399-020-00363-4

Scher, A. I., Terwindt, G. M., Picavet, H. S. J., Verschuren, W. M. M., Ferrari, M. D., & Launer, L. J. (2005). Cardiovascular risk factors and migraine: the GEM population-based study. *Neurology*, *64*(4), 614–620. https://doi.org/10.1212/01.WNL.0000151857.43225.49

Shapiro, R. E., Gallardo, V. J., Caronna, E., & Pozo-Rosich, P. (2021). *The impact of headache disorders on COVID-19 survival: a world population-based analysis*. https://doi.org/10.1101/2021.03.10.21253280

Strain, W. D., Sherwood, O., Banerjee, A., Van der Togt, V., Hishmeh, L., & Rossman, J. (2022). The Impact of COVID Vaccination on Symptoms of Long COVID: An International Survey of People with Lived Experience of Long COVID. *Vaccines*, *10*(5), 652. https://doi.org/10.3390/vaccines10050652

Subramanian, A., Nirantharakumar, K., Hughes, S., Myles, P., Williams, T., Gokhale, K. M., Taverner, T., Chandan, J. S., Brown, K., Simms-Williams, N., Shah, A. D., Singh, M., Kidy, F., Okoth, K., Hotham, R., Bashir, N., Cockburn, N., Lee, S. I., Turner, G. M., & Gkoutos, G. V. (2022). Symptoms and risk factors for long COVID in non-hospitalized adults. *Nature Medicine*, *28*(8), 1706–1714. https://doi.org/10.1038/s41591-022-01909-w

Sylvester, S. V., Rusu, R., Chan, B., Bellows, M., O'Keefe, C., & Nicholson, S. (2022). Sex differences in sequelae from COVID-19 infection and in long COVID syndrome: a review. *Current Medical Research and Opinion*, *38*(8), 1391–1399. https://doi.org/10.1080/03007995.2022.2081454

Thompson, E. J., Williams, D. M., Walker, A. J., Mitchell, R. E., Niedzwiedz, C. L., Yang, T. C., Huggins, C. F., Kwong, A. S. F., Silverwood, R. J., Di Gessa, G., Bowyer, R. C. E., Northstone, K., Hou, B., Green, M. J., Dodgeon, B., Doores, K. J., Duncan, E. L., Williams, F. M. K., Steptoe, A., & Porteous, D. J. (2022). Long COVID burden and risk

factors in 10 UK longitudinal studies and electronic health records. *Nature Communications*, *13*(1), 3528. https://doi.org/10.1038/s41467-022-30836-0

University of California Health. (2022). *University of California COVID Research Data Set (UC CORDS)*. University of California COVID Research Data Set (UC CORDS). https://analytics.uchealth.edu/uccords/index

UpToDate. (2022a, May 4). *Pathophysiology, clinical manifestations, and diagnosis of migraine in adults*. Www.uptodate.com. https://www.uptodate.com/contents/pathophysiology-clinical-manifestations-and-diagnosis-of-migraine-in-adults?topicRef=3347&source=see_link

UpToDate. (2022b, May 25). *Overview of hypertension in adults*. Www.uptodate.com. https://www.uptodate.com/contents/overview-of-hypertension-in-adults#H9

UpToDate. (2022c, October 20). *COVID-19: Evaluation and management of adults with persistent symptoms following acute illness ("Long COVID")*. Www.uptodate.com. https://www.uptodate.com/contents/covid-19-evaluation-and-management-of-adults-with-persistent-symptoms-following-acute-illness-long-covid#H3789064069

Wang, S.-J. (2010). Comorbidities of migraine. *Frontiers in Neurology*, *4*(16). https://doi.org/10.3389/fneur.2010.00016

World Health Organization. (2016, April 8). *Headache disorders*. Www.who.int. https://www.who.int/news-room/fact-sheets/detail/headache-disorders#:~:text=and%20ill%2Dhealth.-

World Health Organization. (2021a). *WHO Coronavirus Disease (COVID-19) Dashboard*. WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int/table

World Health Organization. (2021b, June 11). *Cardiovascular Diseases (CVDs)*. Who.int; World Health Organization: WHO. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

World Health Organization. (2021c, October 6). *A clinical case definition of post COVID-19 condition by a Delphi consensus, 6 October 2021*. Www.who.int. https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1

World Health Organization. (2022, June 7). *Tracking SARS-CoV-2 variants*. Www.who.int. https://www.who.int/activities/tracking-SARS-CoV-2-variants

Appendix

Age Distribution Data

	M1L1		M1L0	
Age group	**Number of patients**	**M1L1 percent population**	**Number of patients**	**M1L0 percent population**
<18	3	0.99%	197	2.92%
18-24	12	3.95%	416	6.16%
25-29	16	5.26%	442	6.55%
30-34	28	9.21%	674	9.98%
35-39	20	6.58%	797	11.80%
40-44	43	14.14%	725	10.74%
45-49	45	14.80%	699	10.35%
50-54	32	10.53%	670	9.92%
55-59	37	12.17%	571	8.46%
60-64	24	7.89%	497	7.36%
65-69	22	7.24%	383	5.67%
70-74	10	3.29%	311	4.61%
75-79	7	2.30%	202	2.99%
80+	5	1.64%	169	2.50%
Total	304	100%	6753	100%
	M0L1		M0L0	
Age group	**Number of patients**	**M0L1 percent population**	**Number of patients**	**M0L0 percent population**
<18	69	3.93%	14453	14.71%
18-24	53	3.02%	6082	6.19%
25-29	69	3.93%	6874	7.00%
30-34	114	6.50%	8770	8.93%
35-39	144	8.21%	8448	8.60%
40-44	158	9.00%	7480	7.61%
45-49	148	8.43%	6680	6.80%
50-54	159	9.06%	6746	6.87%
55-59	169	9.63%	6780	6.90%
60-64	165	9.40%	6531	6.65%
65-69	160	9.12%	5965	6.07%
70-74	135	7.69%	4919	5.01%

75-79	93	5.30%	3702	3.77%
80+	119	6.78%	4809	4.90%
Total	1755	100%	98239	100%

Table 7.1: Age distribution of UC CORDS COVID-19 patients.

Gender Distribution Data

	M1L1		M1L0	
	Count	Proportion	Count	Proportion
Male	61	20.07%	1344	19.90%
Female	243	79.93%	5409	80.10%
Total	304	100.00%	6753	100.00%
	M0L1		M0L0	
	Count	Proportion	Count	Proportion
Male	758	43.19%	45818	46.64%
Female	997	56.81%	52421	53.36%
Total	1755	100.00%	98239	100.00%

Table 7.2: Gender distribution of UC CORDS COVID-19 patients.

Crude Risk Ratio Data

CRUDE	Long COVID		Total
History of Migraine	Yes	No	
Yes	304	6753	7057
No	1755	98239	99994
Total	2059	104992	107051

Table 7.3: 2x2 contingency table for crude risk ratio.

Age-adjusted Risk Ratio Data

AGE >= 65	Long COVID		Total
History of Migraine	Yes	No	
Yes	44	1065	1109
No	507	19395	19902
Total	551	20460	21011
AGE< 65	Long COVID		Total
History of Migraine	Yes	No	

Yes	260	5688	5948
No	1248	78844	80092
Total	1508	84532	86040

Table 7.4: 2x2 contingency tables for age-adjusted risk ratios.

Age-gender-adjusted Risk Ratio Data

MALE				FEMALE			
AGE >= 65	Long COVID		Total	AGE >= 65	Long COVID		Total
Migraine	Yes	No		Migraine	Yes	No	
Yes	13	293	306	Yes	31	772	803
No	227	9795	10022	No	280	9600	9880
Total	240	10088	10328	Total	311	10372	10683
AGE < 65	Long COVID		Total	AGE < 65	Long COVID		Total
Migraine	Yes	No		Migraine	Yes	No	
Yes	48	1051	1099	Yes	212	4637	4849
No	531	36023	36554	No	717	42821	43538
Total	579	37074	37653	Total	929	47458	48387

Table 7.5: 2x2 contingency tables for age-gender-adjusted risk ratios.

Age-gender-disease-adjusted Risk Ratio Data

Primary Hypertension							
Male				OLDER FEMALE			
AGE >= 65	Long COVID		Total	AGE >= 65	Long COVID		Total
Migraine	Yes	No		Migraine	Yes	No	
Yes	7	199	206	Yes	16	413	429
No	150	5318	5468	No	153	5013	5166
Total	157	5517	5674	Total	169	5426	5595
AGE < 65	Long COVID		Total	AGE < 65	Long COVID		Total
Migraine	Yes	No		Migraine	Yes	No	
Yes	19	357	376	Yes	62	973	1035
No	172	6199	6371	No	137	4758	4895
Total	191	6556	6747	Total	199	5731	5930
Cardiovascular Disease							
Male				Female			
AGE >= 65	Long COVID		Total	AGE >= 65	Long COVID		Total
Migraine	Yes	No		Migraine	Yes	No	
Yes	7	146	153	Yes	10	249	259
No	141	3936	4077	No	100	2650	2750

Total	148	4082	4230	Total	110	2899	3009
AGE < 65	Long COVID		Total	AGE < 65	Long COVID		Total
Migraine	Yes	No		Migraine	Yes	No	
Yes	8	136	144	Yes	29	301	330
No	95	2437	2532	No	65	1276	1341
Total	103	2573	2676	Total	94	1577	1671
Type 2 Diabetes Mellitus							
Male				Female			
AGE >= 65	Long COVID		Total	AGE >= 65	Long COVID		Total
Migraine	Yes	No		Migraine	Yes	No	
Yes	5	101	106	Yes	6	197	203
No	93	3216	3309	No	79	2684	2763
Total	98	3317	3415	Total	85	2881	2966
AGE < 65	Long COVID		Total	AGE < 65	Long COVID		Total
Migraine	Yes	No		Migraine	Yes	No	
Yes	8	156	164	Yes	28	450	478
No	108	3565	3673	No	94	2878	2972
Total	116	3721	3837	Total	122	3328	3450