RISK FACTORS FOR PROLONGED SYMPTOMS OF COVID-19 IN A YOUNG ADULT POPULATION

by

MICHELLE BENTIVEGNA, MPH

(Under the Direction of MARK EBELL)

ABSTRACT

Problem: Studies on long COVID-19 primarily focus on older ages. Our goal is to better understand the risk factors for long COVID-19 and predict which young adults are at increased risk for developing this condition. **Methods:** We performed a meta-analysis to identify risk factors for long COVID-19 that have been identified in literature including all adult ages. A survey was then administered to University of Georgia students in a cross-sectional study to identify risk factors associated with the development of long COVID in a young adult population. We then used our knowledge of associated risk factors from the first two aims to develop a clinical prediction rule (CPR) to predict young adults at risk for long COVID-19. **Results:** The meta-analysis identified 46 studies that reported risk factors for long COVID-19. We found statistically significant associations of age 60 to 70 years old, female sex, white race, having any comorbidity, being hospitalized, and having moderate or critical COVID-19 infection with an increased risk of long COVID-19. A total of 1,983 students completed our survey with 230 (11.6%) identified as having long COVID-19. Use of the lasso technique for variable selection yielded a logistic regression model with 7 risk factors. This model included testing positive before December 1, 2021 as well as six symptoms: body or muscle aches (aOR: 1.22),

loss of taste or smell (aOR: 2.16), shortness of breath (aOR: 1.97), diarrhea (aOR: 1.68), brain fog (aOR:1.45), and new anxiety or depression (aOR: 1.32). Our CPR included risk factors female sex, having any comorbidity, and number of symptoms during acute infection and had an area under the cure (AUC) of 0.702. The model classified patients into low, moderate, and high risk groups with about 6 percent of the low risk and 75 percent of the high risk group having long COVID. **Conclusion:** Long COVID-19 is an outcome of COVID-19 that is a burden for all age groups, including young adults. We were able to successfully create a simple risk score that can identify young adults at high risk for long COVID-19.

INDEX WORDS: COVID-19, long COVID, risk factors, clinical prediction rules, college students

RISK FACTORS FOR PROLONGED SYMPTOMS OF COVID-19 IN A YOUNG ADULT POPULATION

by

MICHELLE BENTIVEGNA, MPH

BACHELOR OF SCIENCE IN BIOLOGY, VALDOSTA STATE UNIVERSITY, 2017

MASTER OF PUBLIC HEALTH, UNIVERSITY OF GEORGIA, 2019

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

RISK FACTORS OF PROLONGED COVID-19 IN A YOUNG ADULT POPULATION

by

MICHELLE BENTIVEGNA

Major Professor:	MARK EBELL
Committee:	JESSICA KNIGHT
	YE SHEN
	MICHAEL MOORE

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2023

ACKNOWLEDGEMENTS

Undertaking a PhD is no simple task that has been truly life-changing for me. It would not have been possible without the support and guidance that I received from so many people.

I would first like to express my gratitude to my advisor and committee chair, Dr. Mark Ebell, for his constant guidance throughout my time at UGA. From convincing me to even pursue a PhD to helping me through all stages of this degree, I am truly thankful. Under your guidance I have been able to develop into a better researcher, writer, and presenter. Thank you for many years of guidance, wisdom, and patience.

To the rest of my committee, I would like to thank you for your support throughout my candidacy. I have learned so much from all of you, from your courses I was a part of to your thoughtful guidance and feedback on each aspect of my dissertation.

I have also been supported by my fellow PhD students, particularly Cassie Hulme, throughout my PhD. I couldn't have survived that last six years without you Cassie. From the early days of spending hours doing homework in computer labs, stressing about exams, reviewing thousands of abstracts for many systematic reviews, and often questioning all of our life choices, you have been there throughout my entire graduate school career. There is nobody else I would have wanted to be a "package deal" with. I am so grateful for the friendship we have developed over the years. I can't wait to support you throughout the rest of your PhD journey and see all of the great things you accomplish.

Finally, I would not have been able to achieve this goal without the support of my family. To my fiancé Daniel, thank you for believing in me, holding me accountable, and pushing me to

complete this goal. While also writing your own dissertation and achieving your goals, you still took the time to support me and help me achieve my own goals. To my mom, thank you for being my biggest supporter throughout my educational career. I couldn't have dreamed of getting a PhD without your constant love and support. Thank you for believing in me. Now you can officially say you are the mom to two doctors! Lastly, I would like to thank my brother, Anthony. I look up to you in so many ways and your example pushed me to pursue a PhD. I am so grateful to have had your advice, support, and guidance throughout my PhD career and life.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Statement of the Problem

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), also known as COVID-19, is a novel coronavirus that originated in China. Cases of this virus quickly spread to other countries and eventually globally. By March 2020, COVID-19 was declared a pandemic by the World Health Organization (WHO).[1] COVID-19 has caused an estimated 759 million cases and about 6 million deaths as of March 2023.[2] Presentation, symptoms, and severity vary by individual and can range from mild to severe. Most individuals are infected and recover from the virus within two weeks.[3] However, there is evidence that survivors of COVID-19 can experience long lasting medical and psychological consequences.

Individuals experiencing symptoms of COVID-19 for four weeks or longer are considered to have "long COVID-19" according to CDC guidance.[4] It is also referred to in the literature as "post-acute sequelae of COVID" or PASC or "long hauler syndrome". Prolonged symptoms can vary by individual and length of time since COVID-19 diagnosis. Studies generally indicate respiratory symptoms, fatigue, pain, anxiety or depression, and cognitive or psychological changes as commonly reported prolonged symptoms.[5–7] Due to the wide array of symptoms and no formal definition, it is difficult for healthcare providers to diagnose individuals with long COVID-19. Healthcare providers have a difficult task to decide if the wide array of symptoms a patient is experiencing are the result of prior COVID-19 infection or have some other explanation.

Similarly, little is known as to why certain individuals experience prolonged symptoms of COVID-19. Generally, incidence of health outcomes is tied to age and underlying medical conditions. However long COVID-19 has been observed to span multiple age groups with varying degrees of health.[8] Ongoing research has just begun to identify risk factors for long COVID. However, these are just initial qualitative analyses and have yet to draw formal conclusions as to what is causing this outcome. Despite current vaccination efforts, understanding the long-term effects of the virus is urgent, as there are many multi-organ system impacts we do not yet know how to treat.

Evidence has shown that young adults are also experiencing prolonged symptoms of COVID-19.[9] Initially in the COVID-19 pandemic young adults were often not experiencing symptoms but still becoming infected. Those individuals who remained asymptomatic would often not know that they had the virus and then easily spread it to others. This often makes it difficult to quantify the true number of cases and the impact of COVID-19 in young adults. This has yielded limited research focused on this age group surrounding outcomes of COVID-19. As time from the onset of the pandemic and research surrounding the topic has grown, indications of prolonged symptoms in young adults have emerged. Like older adults, they too are experiencing similar prolonged symptoms for weeks post COVID-19 infection. Like all other age groups little is known as to what is causing long COVID. Therefore, research to understand the epidemiology, risk factors, diagnosis and prognosis of long COVID in young adults is important to fully understand COVID-19 and its long term effects on the population.

Aims of this Dissertation

Our primary goal for these three aims is to predict which young adults are at increased risk for experiencing long COVID-19. Our three aims are:

1. Identify risk factors that can be used to predict which adults are at increased risk for long COVID-19.
2. Identify college students with prolonged symptoms and their risk factors.
3. Predict which young adults are at increased risk for developing long COVID-19.

Each aim has its own set of objectives and methods that are described in detail in Chapter 3. Table 1.1 presents a brief description of each aim.

Table 1.1 Objective, Data Source, and Methods for each aim.

Aim	Objective	Data Source	Methods
1	Use a systematic review of the literature to determine the risk factors that put individuals with COVID-19 at an increased risk for experiencing long COVID-19	Pubmed, WHO COVID-19 Database, MedRXIV	Systematic review, random effects meta-analysis
3	Identify college students with prolonged symptoms of COVID-19 and their risk factors	Qualtrics survey administered to enrolled UGA students	Logistic regression, Lasso
2	Develop a clinical prediction rule (CPR) for predicting young adults who are risk for developing long COVID-19	National COVID Cohort Collaborative (N3C)	Logistic regression, Lasso

Aim 1 is focused on identifying the risk factors or predictors of long COVID-19. We will conduct a systematic review of the literature to identify all studies reporting common risk factors. Pooled estimates of the risk factors association with long COVID will be produced and evaluated. We hypothesize there may be some association of age, initial disease severity, and

comorbidities with the outcome. We will use Aim 2 to identify college students that are experiencing long COVID-19 and identify their risk factors for developing this outcome. We will administer a survey to all enrolled University of Georgia students (UGA) and ask those with prior COVID-19 infection since August 2021 to complete. This survey will hope to capture students with and without prolonged COVID-19 symptoms and compare risk factors that they have to identify those associated with long COVID in this population. Risk factors identified in the first two aims will then be included in a model in Aim 3 to develop a clinical prediction rule (CPR) for predicting long COVID in young adults. We will use a publicly available dataset to identify individuals with our outcome and then determine the presence or absence of the risk factors. We will use logistic regression to develop a CPR, presented as a point score that can be used to classify individuals with COVID-19 as low, moderate, or high risk of developing long COVID.

Dissertation Outline

Chapter 1 of this dissertation has provided a brief introduction to long COVID-19 and the impact in the young adult population. Chapter 2 will describe the literature surrounding COVID-19, long COVID-19, and the experience of COVID-19 in young adults in detail. Chapter 3 will describe the methods used in each aim of this dissertation. Chapter 4 is a brief introduction to the results. Chapters 5, 6, and 7 will be manuscripts prepared for publication representing each of the three aims. Chapter 8 will be a summarization of the findings of each aim and conclusions.

CHAPTER TWO

LITERATURE REVIEW

COVID-19 Overview

Biology of Coronavirus

Coronavirus (CoV) is derived from the word 'corona' meaning crown.[10] The name of this virus comes from its structure, first seen under an electron microscope in the 1960s.[11] The crown-like structure is caused by spike proteins on the surface of the virus that mediate virus entry into cells and induce a host of immune responses.[12] CoVs are single stranded RNA viruses in the Coronaviridae family.[13] These viruses are an important cause of respiratory outbreaks and endemic respiratory infections, as well as numerous other diseases in animal species.[13] After initial discovery, prior to the emergence of SARS-CoV-1 in 2002, CoV generally caused mild to moderate disease in humans. SARS-CoV-1, MERS, and SARS-CoV-2 were all discovered later and cause more severe disease in humans.[14]

Coronavirus Outbreaks

In November of 2002 the first cases of severe acute respiratory distress syndrome (SARS) began to emerge in China.[15] Three months later, about 300 cases had been reported.[15] As individuals continued to travel, the number of SARS cases began to grow resulting in roughly 8,000 cases and more than 700 deaths across 27 countries.[15] By April of 2003, the WHO determined that the epidemic was caused by a CoV, one of the first major outbreaks of severe disease caused by a CoV in humans.[15] A number of infection control measures helped to stop the transmission of

SARS and prevented further human infection. However, 10 years later another novel CoV emerged. Middle East respiratory syndrome coronavirus (MERS-CoV) was first reported in Saudi Arabia in 2012.[16] Since the first case, there have been approximately 2000 laboratory confirmed cases of MERS resulting in 790 deaths.[17] As indicated by the name, a number of the outbreaks occurred across the Middle East, eventually spreading to 27 countries.[17] A number of these clusters occurred in the Republic of Korea. Like SARS, infection prevention measures and guidance by the WHO have limited cases.[17] While both viruses caused initial 'media panic', human-to-human transmission of these viruses were quickly limited and these outbreaks did not have greater impacts.[17]

COVID-19 Pandemic

Based on previous experiences with CoV through the SARS and MERS outbreaks, it was only a matter of time before another virus caused more significant human illness. In December 2019, the first cases of an unknown pneumonia were first reported to the World Health Organization (WHO).[18] By January 2020, the cause of these cases was determined to be severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Early in the outbreak, a number of infected persons were reported to be linked to a seafood market in Wuhan, China.[18] Later, new cases continued to emerge that had no connection to the market which suggested significant person-to-person transmission of the virus.[19] In January 2020 organizations such as the WHO and Centers for Disease Control (CDC) began to put into action protocols created during the MERS outbreak to help mitigate impacts of SARS-CoV-2.[18] The virus continued to spread through international travel, with cases appearing first in Thailand and Japan and then quickly spreading to Europe, with Italy becoming the newest epicenter.[18] Within the United States, the CDC began screening

passengers on direct and connecting flights from China at major international airports. Case numbers of SARS-CoV-2 quickly began to rise worldwide and by March of 2020, the WHO declared Coronavirus Disease 2019 (COVID-19) a pandemic.

Following classification as a pandemic, SARS-CoV-2 continued to have major impacts on the world. As of February 2023, there have been over 750 million confirmed cases of COVID-19 resulting in more than 6 million deaths.[2] Many organizations and governments proposed mitigation efforts since the onset of the pandemic to help slow transmission. However, SARS-CoV-2 has proved to be highly transmissible and has changed over time as a result of mutations creating a number of new variants. To date, the WHO has identified five variants of concern (VOC). These are classified as a VOC because of increase in transmissibility, increase in virulence, or decrease in effectiveness of public health and social measures. A brief overview of each VOC is included in Table 2.1.[20]

Table 2.1. Description of COVID-19 Variants of Concern[20]

Variant	**Country Identified**	**Date First Identified**
Alpha	United Kingdom	September 2020
Beta	South Africa	May 2020
Gamma	Brazil	November 2020
Delta	India	October 2020
Omicron	Multiple Countries	November 2021

Epidemiology of COVID-19

SARS-CoV-2 Mechanism

Sequencing of the novel coronavirus has revealed more than 80% and 50% similarity to SARS and MERS respectively.[21] The mechanism and structure of SARS-CoV-2 are what make the virus so pathogenic and highly transmissible. The spike proteins on the surface of the virus recognize and bind to angiotensin converting enzyme 2 (ACE2) receptors. These receptors are widespread in the human body, appearing in the lungs, heart, kidneys, intestines, esophagus, liver, and blood vessels.[21] This allows for the virus to have an impact across organ systems causing various symptoms as well as potential widespread complications. SARS-CoV-2 has continued to thrive as the virus mutates to create new variants. Genetic sequencing as well as monitoring by researchers and government agencies have identified impacts SARS-CoV-2 variants may have in relation to disease severity, vaccine efficacy, and disease transmissibility.[22]

Transmission

Individuals are most commonly infected with SARS-CoV-2 through exposure to respiratory fluids carrying the virus.[23] Exposure most commonly occurs through inhalation of respiratory droplets and aerosols, deposition of virus carrying droplets on the mucous membranes of the mouth, nose, or eyes, and the touching of these mucous membranes by hands that have come in contact with virus containing fluids or by touching surfaces with virus on them.[23]

Studies on the transmission dynamics of COVID-19 have estimated the reproductive number using various models. The average basic reproductive number for the ancestral strain of SARS-CoV-2 is estimated to range from 2.24 to 3.58.[24] In similar studies the mean incubation period was estimated to be 5.2 days.[3] However, this period can extend to 14 days with the

average time of symptom onset occurring around 5 days. Estimates of transmission dynamics can vary based on COVID-19 variant as well. Research is ongoing to quantify the reproductive number of each VOC. In a systematic review, including five studies estimating the basic reproductive number of the Delta variant, they estimated an average basic reproductive number of 5.08 (95% CI: 3.2-8.0).[25] This is noticeably higher than the ancestral strain of COVID-19 and aligns with the characteristics of Delta infections having a higher viral load and longer duration of shedding causing higher transmissibility.[25] While research is ongoing surrounding formal estimates of the same values for the Omicron variant, it is observed to be more transmissible but less severe, in terms of outcomes, when compared to Delta.[26] Estimates from a study of Omicron cases in Denmark indicate that the effective reproductive number of Omicron is 3.19 times that of Delta.[27] Similar estimates were made in studies performed in South Korea as well.[28] These estimates are difficult to compare to the ancestral strain of COVID due to the implementation of vaccines and varying rates of vaccine coverage.

Asymptomatic and presymptomatic infections can make it difficult to quantify transmission of COVID-19. Many individuals who never develop or have not yet developed symptoms are unlikely to be tested which can lead to increased transmission of COVID-19. Several studies have attempted to quantify asymptomatic transmission with a number suspecting that about one-third of COVID-19 infections are asymptomatic.[29] To account for unidentified cases and differences in case definitions, estimates of infection fatality ratios (IFR) have been made in various studies. Early in the pandemic using data of cases in China, an overall IFR of 0.66 percent was estimated for COVID-19.[30] The IFR was also seen to increase with increasing age (Table 2.2). In the same study, the proportion of individuals requiring hospitalization were estimated and also increased with age. [30] Similar estimates have been made, using the same

methods, in the United Kingdom resulting in an IFR of 0.9 percent early in the pandemic.[31] While overall estimates of IFR differ, similar patterns in relation to age have been seen across studies. It is important to note that these estimates are made from data early in the pandemic. Current estimates of IFR are likely now lower due to improvements in treatment and changes in virulence of newer variants, especially Omicron.

Table 2.2. Summary of infection fatality ratio and individuals requiring hospitalization based on data from early COVID-19 cases in China[30]

Age (Years)	Proportion of infected individuals requiring hospitalization	Infection Fatality Ratio (%)
0-9	0.00	0.002
10-19	0.0408	0.006
20-29	1.04	0.03
30-39	3.43	0.08
40-49	4.25	0.15
50-59	8.16	0.60
60-69	11.8	2.2
70-79	16.6	5.1
80+	18.4	9.3

Diagnosis and Treatment

Symptoms

For persons who do experience symptoms related to COVID-19, they can vary from mild to severe. A summary of the prevalence of commonly reported symptoms in several cohort studies is presented in Table 2.3. Among these five cohort studies the most commonly reported symptom was fever followed by dyspnea and sputum production. While many of these symptoms are similar to those of influenza, the progression of COVID-19 can result in a higher case fatality rate. Typically many individuals with COVID-19 have a symptomatic period and then recover. However, some patients progress into an inflammatory response, known as the cytokine storm,

resulting in inflammatory fluid in the lungs, declining oxygen levels, hospitalization, and even death.[32]

Table 2.3. Cohort studies reporting common COVID-19 symptoms

	Guan, 2020[33]	Chen, 2020[34]	Shi, 2020[35]	Huang, 2020[36]	Yang, 2020[37]
Patients (n)	1081	99	21	41	52
Fever (%)	44	83	86	98	89
Dyspnea (%)	19	31	43	54	64
Sputum (%)	34	-	14	27	-
Rhinorrhea (%)	5	4	24	-	6
Sore throat (%)	14	5	-	-	-
Headache (%)	14	8	10	5	6
Diarrhea (%)	4	2	5	2	-
Nausea/Vomiting (%)	5	1	10	-	4
Myalgia (%)	15	11	-	-	12

Risk Factors for Severe Disease

Those with increased age and underlying medical conditions are at an increased risk for developing severe COVID-19 illness.[19] For individuals with progressive COVID-19 infection, a number of biomarkers could be indicative of more severe disease and a higher mortality risk. While research is still ongoing for predicting severe outcomes of COVID-19, there is evidence that biomarkers including C-reactive protein (CRP), Interleukin 6 (IL-6), Interleukin 10 (IL-10), lymphocytes, thrombocytes, and D-dimer are associated with poor outcomes and more severe COVID-19.[38,39] Table 2.4 presents the results of a systematic review of 22 studies reporting risk factors for severe COVID-19 disease.[40] Summary estimates of the relative risks (RR), corresponding confidence intervals, and the number of studies and patients included in the estimate are presented. Risk factors most strongly associated with severe disease included dyspnea, elevated procalcitonin, chronic heart disease, and chronic kidney disease.

Table 2.4 Association of risk factors with severe COVID-19 disease[40]

Variables	Studies	Total patients	RR (95% CI)
Male sex	11	4582	1.30 (1.11-1.53)
Coronary heart disease	9	4364	3.69 (1.75-7.77)
Chronic kidney disease	3	3516	3.02 (0.63-14.6)
Diabetes mellitus	11	4582	2.57 (1.59-4.17)
Chronic obstructive pulmonary disease	7	1033	2.47 (1.34-4.52)
Hypertension	11	4582	2.29 (1.61-3.26)
Dyspnea	7	1473	6.28 (3.10-12.7)
Fever	10	4504	1.13 (1.05-1.22)
Cough	9	1775	1.11 (0.87-1.42)
Procalcitonin > 0.05 ng/mL	3	448	4.06 (0.65-25.3)
C-reactive protein > 5 to 0 mg/L	3	448	1.68 (1.47-1.93)

COVID-19 Testing

To identify individuals with current and prior COVID-19 infection, several diagnostic tests have been developed. Nucleic acid amplification tests (NAAT), including real time reverse transcription polymerase chain reaction (RT-PCR), are used for detection of viral RNA.[41] Alternatively, rapid diagnostic tests (Ag-RDT) can be used to identify current COVID-19 infection by detecting viral antigens.[41] NAAT tests are considered the gold standard test with sensitivity and specificity estimated at 95.2% and 98.9% based on a systematic review of 11 studies measuring accuracy of these tests.[42] Ag-RDT are not as sensitive, with sensitivity and specificity near 70% and 100%, but have the advantage of rapid turn-around, low cost, and availability for home testing.[43,44] NAAT tests can be used as confirmatory test for asymptomatic individuals or those with known exposures with negative Ag-RDT.[41] There can also be differences in diagnostic accuracy with Ag-RDT over time. Sensitivity of the test can be higher

in the first week after symptom onset (78.3%, 95% CI: 71.1%-84.1%) than in the second week of symptoms (51.0%, 95% CI: 40.8%-61.0%).[42]

Samples for both RT-PCR and Ag-RDT were primarily collected via nasal swabs. However, multiple samples are sometimes required and the tests can be invasive. Collection can cause patients to sneeze or cough and potentially increase transmission to healthcare workers. Early in the pandemic, there have also been issues with having adequate testing supplies including swabs and transport media. A non-invasive alternative is saliva sampling. Studies indicate that use of saliva samples for RT-PCR test yield good diagnostic abilities with sensitivities and specificities of 88% (95% CI 82%–92%) and 92% (95% CI 75%–98%), respectively.[45] Use of saliva testing will more easily allow patients to self-collect specimens and allow for more efficient and wide spread COVID-19 testing.

For research and surveillance purposes, a serology test can be used to identify individuals who have had a previous COVID-19 infection. These tests work by identifying antibodies produced in response to infection. This test is performed by taking a blood or saliva sample to measure for various levels of antibodies in response to COVID-19 infection. IgG antibodies are most commonly measured with these tests as they persist the longest and may reflect longer term immunity, however they take a longer time to appear. Similarly IgM antibodies can also be measured because they quickly occur after infection.[46] A systematic review from the National Academies indicates that IgM can be detected a median of 5 days after symptom onset as compared to IgG which appears a median of 14 days after onset.[47] These antibodies can take weeks to develop and therefore, serology tests should not be used for diagnosis of active COVID-19 infection.

Treatment

Various treatments for COVID-19 have been studied throughout the pandemic in an attempt to treat this novel virus. Currently the U.S. Food and Drug Administration (FDA) has approved one antiviral for COVID-19 treatment, remdesivir.[48] Through various clinical trials, remdesivir has been shown to shorten recovery time for adults hospitalized with COVID-19 by about 5 days.[49] The greatest benefit was seen in those patients only requiring low flow oxygen. The same study has also shown a reduction in 14-day mortality in the remdesivir group (7.1% vs 11.9%), however this was not statistically significant.

Due to the urgency of the pandemic, a number of other treatments have been granted emergency use authorizations (EUA) by the FDA. EUAs allow for unapproved drugs or unapproved uses of approved drugs when needed. In December of 2021, the FDA issued an EUA for a novel oral antiviral, ritonavir-boosted nirmatelvir (Paxlovid). Evidence from initial trials indicate that the antiviral reduced the proportion of individuals with COVID-19 related hospitalization or death when compared to placebo (0.8% vs 6.3%, $p < .001$). All-cause mortality was also less likely in the treatment group (0.0% vs 1.1%, $p = .001$).[50] This is the first COVID-19 treatment of its kind in the form of a pill that can be taken orally. Paxlovid is recommended for treatment of mild to moderate COVID-19 in adults and children 12 years and older who are at risk for severe disease and is most effective within five days of symptom onset.[51]

Treatment with systemic corticosteroids was found to be beneficial for persons with severe COVID-19 . Treatment with corticosteroids has shown to reduce mortality in patients requiring mechanical ventilation as a result of COVID-19. In a clinical trial randomizing over 2,000 patients to dexamethasone, 28 day mortality was significantly lower in the dexamethasone group compared to usual care (rate ratio, 0.83; 95% confidence interval (0.75-0.93) $P<0.001$).[52] This

degree of benefit was associated with disease severity and showed a 12 percent reduction in mortality for those on mechanical ventilation. [52] However, outcomes were not improved with a trend toward worse outcomes in non-hospitalized patients not requiring supplemental oxygen. This indicates that steroids could cause worse outcomes if administered early in treatment as they suppress the immune system, but can be important should a patient become more severe and require mechanical ventilation. Similar results were identified in a meta-analysis including 1700 patients that had received dexamethasone, hydrocortisone, or methylprednisolone.[53]

Use of monoclonal antibodies for treatment of individuals at risk for severe disease in the outpatient setting have shown some benefit. Monoclonal antibodies are laboratory made molecules that act as antibodies to help fight COVID-19 infection. They are able to recognize and respond to the virus more effectively and help prevent severe outcomes and limit symptoms.[54] Monoclonal antibodies have shown to reduce viral load and decrease the severity of COVID-19 symptoms.[55,56] Monoclonal antibodies that target interleukin-6 (IL-6) inhibit inflammation in order to reduce viral load while others that target the spike proteins on the CoV surface work by inhibiting replication. In a trial comparing use of bamlanivimab and etesevimab, evidence of significant reduction in viral load when the treatments were used in combination were observed.[57] Treatment with these monoclonal antibodies was suggested for individuals at increased risk for severe infection including adults older than 65 years, those with suppressed immune systems, and with underlying medical conditions.[54] However these treatments, specifically bamlanivimab and etesevimab, have been shown to be ineffective against the Omicron variant and have ceased use as of January 2022. The FDA later approved an EUA for bebtelovimab which has shown to be effective against the Omicron variant.

Prevention

Public Health Measures

Due to the novelty of SARS-CoV-2, lack of any population level immunity, and the lack of effective treatments early in the pandemic, a number of public health measures were put in place worldwide to help slow the spread of the virus. Shutdowns of schools, businesses, and non-essential entities were used to limit interactions between individuals and prevent further spread. A study analyzing incidence rates before and after implementation of various distancing measures, including closures, restrictions on gatherings, and lock-down orders, found an overall 13% reduction in COVID-19 incidence due to their implementation.[58] While evidence shows that shut-downs are effective it became apparent that shutdowns could not feasibly continue. Public health measures including remaining six feet apart from others, wearing masks in indoor areas, frequent testing, improving ventilation systems, and monitoring of symptoms were developed. A study done by the CDC compared cases and deaths of COVID-19 before and after implementation of masks mandates. In counties where masks were mandated, there was a 0.5 percent decrease in cases 1-20 days following implementation.[59] This reduction in cases was continually observed incrementally every 20 days post implementation through day 100. [59] Similar results of the benefits of the use of face masks to reduce COVID-19 incidence were also found in a study of grade schools in Arkansas.[60] In the United States the Centers for Disease Control (CDC) recommends face masks for anyone over the age of 2 who is not fully vaccinated in indoor public places.[61] Their recommendations also include social distancing measures, avoiding crowded and poorly ventilated areas, frequent hand washing, covering coughs and sneezes, and cleaning high use surfaces.[61]

In some countries, intermittent lockdowns occurred based on case rates. Many countries also instituted quarantine and isolation guidance to prevent further transmission of the virus. Many included a quarantine period for those who came into contact with a COVID-19 positive individual and an isolation period for those who are COVID-19 positive themselves. These recommendations are different between countries and also differ between US states. The CDC currently recommends an isolation period of 5 days following a positive COVID-19 test or since symptoms began, followed by 5 days of mask-wearing in public.[62] Those who have come in close contact with a COVID-19 positive individual should quarantine for at least 5 days post contact unless they are fully vaccinated.[62]

COVID-19 Vaccines

In the year following the start of the SARS-CoV-2 pandemic, a number of vaccines were developed, evaluated in large randomized trials, and became available for administration. Prior knowledge from previous vaccine development aided in the rapid development of COVID-19 vaccines. As of January 2022 there are 140 vaccines in clinical development worldwide.[63] These vaccines all range from one to three doses with some administered via injection and others orally. In the United States three vaccines were approved initially under EUA and then formally by the FDA They are given in one or two doses with booster doses recommended 6 months later.[64] The vaccines developed by Pfizer-BioNTech and Moderna are mRNA vaccines. These vaccines consist of a short strand of mRNA that encodes the SARS-CoV-2 spike protein wrapped in a lipid coat, and after being absorbed by a human cell leads to the production of spike proteins. An immune response then occurs to produce antibodies to help fight COVID-19 infection.[65] Alternatively, the Johnson and Johnson vaccine uses a viral vector mechanism. This

vaccine uses a vector virus to produce spike proteins and trigger the same immune response.[66] Both two dose mRNA vaccines have been shown to have an efficacy in preventing symptomatic disease of 95% and 94% respectively,[67,68] while the single dose viral vector vaccine has about a 67% efficacy.[69] Overall, vaccination for COVID-19 continues to be the preferred method to reduce incidence and stop further transmission.

Concerns of breakthrough cases for vaccinated individuals are still of concern, especially related to COVID-19 variants. In a presentation by the CDC, they estimated that the risk of infection among vaccinated individuals compared to unvaccinated is 21 vs 177 per 100,000.[70] Vaccine efficacy for the Delta variant is estimated to remain high ranging from 87 to 90 percent.[70] However, with a surge in Delta variant cases, it did appear efficacy was reduced for those who are immunocompromised or elderly. Similar studies have been conducted to assess vaccine efficacy in the face of the Omicron variant as well. In a study of over 800,000 individuals with the Omicron variant, it was observed that the initial two dose series of the Pfizer vaccine provided limited protection against symptomatic COVID-19 (Efficacy: 65.6%).[71] However, implementation of a booster shot following the initial two doses provided increased protection against the Omicron variant.

Long COVID-19

Prolonged symptoms and side effects of COVID-19 can be referred to as long COVID-19, post COVID-19 condition, or post-acute sequelae of SARS-CoV-2 (PASC). There are many definitions of long COVID-19 that generally differ based on the minimum length of time symptoms have occurred. The World Health Organization (WHO) defines post COVID-19 condition as occurring in those with confirmed SARS CoV-2 infection, 3 months post onset with

symptoms that have lasted for at least 2 months and cannot be explained by another diagnosis.[72] The CDC defines it as a wide range of new, returning, or ongoing health problems people can experience four or more weeks after being infected with the virus that causes COVID-19.[4]

Individuals classified as having long COVID-19 can experience the disease in many different ways. The most common is by experiencing either new or ongoing symptoms that can last weeks or months post infection. Table 2.5 summarizes several cohort studies that aim to identify individuals with long COVID and their symptoms. Across all three studies a range of about 14 to 53 percent of patients reported at least one persistent symptom following their diagnosis. All studies included those with mild to severe disease, including some who had been hospitalized and some not. The most common symptoms reported differ in each study. However, in each study patients commonly reported dyspnea, loss of taste or smell, and weakness. A systematic review, that included these studies, measuring prolonged symptoms 6 months post COVID-19 diagnosis indicated that over half of the COVID-19 survivors experienced long COVID with those patients most commonly reporting functional mobility impairments, pulmonary abnormalities, and mental health disorders.[73]

It has been assumed that the prevalence of long COVID is tied to initial disease severity. Though most studies measuring long COVID have included primarily patients who had been previously hospitalized, there is evidence that long COVID can occur in those with mild to moderate disease as well. A study that compared over 4,000 healthcare workers, with and without mild COVID-19 over 8 months, persistent symptoms including anosmia, fatigue, dyspnea, and sleeping disorders, among other symptoms, were reported.[74] However, symptoms related to neurologic or psychiatric illness following COVID-19 infection were more commonly seen in those who were previously hospitalized or admitted to the intensive care unit (ICU).[75]

While long COVID can occur in individuals with milder initial disease severity, the type or severity of these prolonged symptoms could potentially be linked to disease severity.

Table 2.5. Summary of studies reporting patients with long COVID-19

	Carvalho-Schneider, 2020[76]	Carfi, 2020[5]	Logue, 2021[77]
Country	France	Italy	United States
Age	49.0 (Mean)	62 (Median)	48 (Mean)
Length of follow up (days)	60	60	180
At least 1 persistent symptom N (%)	86 (66.0)	159 (32.6)	17 (26.6)
Persistent symptoms reported (%)	Loss of taste or smell (23.0)	Fatigue (53.0)	Fatigue (13.6)
	Dyspnea (30.0)	Dyspnea (43.0)	Loss of taste or smell (13.6)
	Weakness (40.0)	Joint Pain (28.0)	Brain fog (2.3)
		Weakness (16.0)	

Research is ongoing to identify to what extent these persistent symptoms are specific to long COVID-19. A retrospective cohort study, including over 200,000 COVID-19 survivors looked at nine core long COVID symptoms to try and identify their uniqueness to COVID-19.[78] These included breathing difficulties, fatigue, chest/throat pain, headache, abdominal symptoms, myalgia, other pain, cognitive symptoms, and anxiety/depression. In a comparison with outcomes of influenza, all nine features were reported more frequently after COVID-19 than influenza with an overall hazard ratio of 1.65. [78] This indicates that these symptoms are more common after having COVID-19, however about 43 percent of patients with prior influenza diagnosis did also experience these same symptoms. [78]

COVID-19 in Young Adults

Older adults and those with other health conditions are at the greatest risk for severe illness from COVID-19. The demographic that is less commonly studied are young adults of ages roughly 18 to 30 years old. This age group is generally considered young, healthy, and not at an increased risk for severe outcomes of COVID-19. However, there is evidence of higher than expected numbers of cases as well as increased transmission in young adults. In a study done prior to widespread vaccination, it was shown that adults 20-34 and 35-49 were the only age groups that maintained a reproductive number (R) above one.[79] It was also estimated at that time that 75 of every 100 COVID infections were occurring among adults in these age groups.[79] Though there appears to be high transmission amongst younger adults, outcomes are generally considered to be more mild when compared to older adults. This can be seen as previously mentioned in the estimation of IFR and proportion needing hospitalization in Table 2.2. Generally those of younger age are less likely to experience severe outcomes or death due to COVID-19.

The increased transmission among young adults has been generally attributed to their behaviors and beliefs throughout the pandemic. In a study by the CDC, interviews to assess attitudes and perceptions of the pandemic amongst young adults and their surrounding community were done in Winnebago County, Wisconsin. [80] As a result of the interviews they concluded that young adults share a belief that the severity of disease outcomes in younger persons is not great enough to change behaviors such as social interactions or gatherings and the following of public health measures.[80] This pattern of increased transmission was observed following students return to universities after their shutdown. In August 2020, COVID-19 incidence was generally lower around the United States compared to earlier in the summer.

However, an increase in incidence was noticed in adults 18-22 years which was then followed by an increase in adults 60 years and older.[81] In a comparison of counties with large universities with in-person and remote education in the Fall 2020 semester, those counties with in-person universities experienced a 56% increase in COVID-19 incidence compared to a 17.9% decrease in those counties with remote education.[81] These trends continued to occur as students traveled away from the university and returned. Transmission is also estimated to be higher than reported due to a number of young adults having asymptomatic COVID-19. In a study at the University of Georgia, it was estimated that approximately two thirds of infected students had asymptomatic COVID-19 infection.[9]

For those young adults that do experience symptoms of COVID-19, there is some evidence that some are experiencing symptoms for a prolonged period of time. In a study of about 430 students at the University of Georgia, 11 percent with previously confirmed COVID-19 reported prolonged symptoms for a median of 132 days.[9] Another study at a small university indicated that 51 percent of college students experienced COVID-19 symptoms for greater than 28 days.[82] While there is limited research on young adults and long COVID, there is evidence that this age group is experiencing long term effects of the virus. Common prolonged symptoms reported by young adults include chest pain, fatigue, fever, loss of taste and smell, and headaches.[82] While these symptoms are aligned with those of acute COVID-19, further research will be important to differentiate the symptoms related to acute and long COVID-19.

Gaps in the Literature

Since SARS-CoV-2 is a newly identified human pathogen, much is unknown about COVID-19 and its outcomes. Due to the rate that research is being published on the topic, there is a constant

flood of information that is influencing decisions in relation to policies, treatment and prevention surrounding COVID-19. However, this virus has not been around long enough for us to have a full grasp on its impacts, especially in relation to its long term effects.

Much is unknown about long COVID including risk factors for those likely to develop this outcome. A number of studies have aimed to identify these risk factors but there is a great degree of heterogeneity with regards to patient selection, symptoms recorded, and duration of follow-up. Also, there are currently several clinical prediction rules (CPR) for COVID-19 diagnosis and disease severity. However, there are none to date for determining who is at increased risk for long COVID. Finally, research surrounding young adults and their outcomes of COVID-19 are limited.

We will address these gaps in the literature through the three aims of this dissertation. A systematic review will first be conducted to determine common predictors for long COVID-19. Then we will evaluate long COVID and its risk factors in a college student population at the University of Georgia. Lastly, we will develop our own CPR using knowledge of common predictors from the systematic review and developing it using electronic health record data.

CHAPTER 3

METHODS

AIM 1: Risk factors of Long COVID-19: A Systematic Review and Meta-Analysis

Background

Individuals with previous mild to severe COVID-19 can experience prolonged symptoms referred to as long COVID. Currently in the United States, the CDC classifies long COVID as a range of new, returning, or ongoing health problems people can experience four or more weeks after being infected with COVID-19.[4] While experiences with this outcome differ by person, many report respiratory, neurologic, and psychiatric symptoms.[6,7,83]

The risk factors for developing long COVID are largely unknown. A number of studies have attempted to identify these risk factors and report that increased age, higher BMI, and male sex all appear to be associated with long COVID.[84] Though associations with increasing age are reported, there is also some evidence that younger individuals are also experiencing prolonged symptoms as a result of COVID-19 infection.[9]

Current systematic reviews on long COVID primarily focus on signs and symptoms that patients are experiencing. Some do report risk factor information that is included in studies, but do so in a qualitative way due to differences in studies and limited data at the time of publication.[85,86] As long COVID becomes more prevalent, the number of publications on the topic has also grown. A comprehensive list of risk factors and the magnitude of their association with long COVID is needed to truly determine what causes patients to develop prolonged symptoms, and how to identify persons at increased risk.

Objective

Our objective is to use a systematic review of the literature to determine the risk factors that put individuals with COVID-19 at an increased risk for experiencing long COVID-19 and to quantify the strength of those associations.

Methods

Inclusion and Exclusion Criteria

Studies will be considered for inclusion in the systematic review if they are original research conducted primarily in an adult population. Participants in the studies must also have laboratory confirmed COVID-19 and have symptoms recorded for at least four weeks post COVID-19 diagnosis. Studies must also report at least one potential risk factor for long COVID which can include demographics, such as age and sex, BMI, vital signs, common symptoms, comorbidities, such as diabetes, heart disease, COPD, or asthma, and results of laboratory tests. To complete a meta-analysis of the data, included studies must also report sufficient data to calculate relative risks or odds ratios for each risk factor.

A study would not be considered for inclusion if it included only a specialized population such as only pregnant women or only immunocompromised patients. Also, those studies that only study a child population would be excluded. No limitations will be set for country or language of the publications.

Search Strategy

We will use PubMed, the WHO COVID-19 Database, and the MedRXIV preprint server to identify all literature that highlights risk factors of long COVID-19. This systematic review will be conducted following the PRISMA guidelines.[87] A flow diagram is presented in Figure 3.1 to illustrate the search process following these guidelines. Each step of the search strategy and analysis will be performed by two researchers working in parallel. In the case of disagreement, a final decision will be made by consensus or with help of a third researcher when necessary.

Search strategies for each of the included databases were informed by work of previous researchers.[88–90] Additionally, search strategies of related systematic reviews that used the same databases were reviewed. Multiple terms for long COVID-19 such as "long covid" and "post-covid sequelae" as well as terms for prognostic studies such as "risk factor" and "prognosis" were used to develop a search strategy. The search strategy for each database are included in Appendix A, Table A1.

Data Abstraction

All abstracts of articles identified through the search process will be reviewed by two researchers in parallel. For any abstract that appears to be related or of interest the full article will be obtained and reviewed by two researchers. The full articles will be screened by each researcher using the inclusion criteria. Study characteristics including patient population, setting, and length of follow up will also be included in this table. Risk factor data including the number of individuals with and without the risk factors and how many observed the outcome of interest will also be abstracted. For continuous risk factors, data including mean, standard deviation, median, and interquartile range will be abstracted depending on how the risk factor is reported.

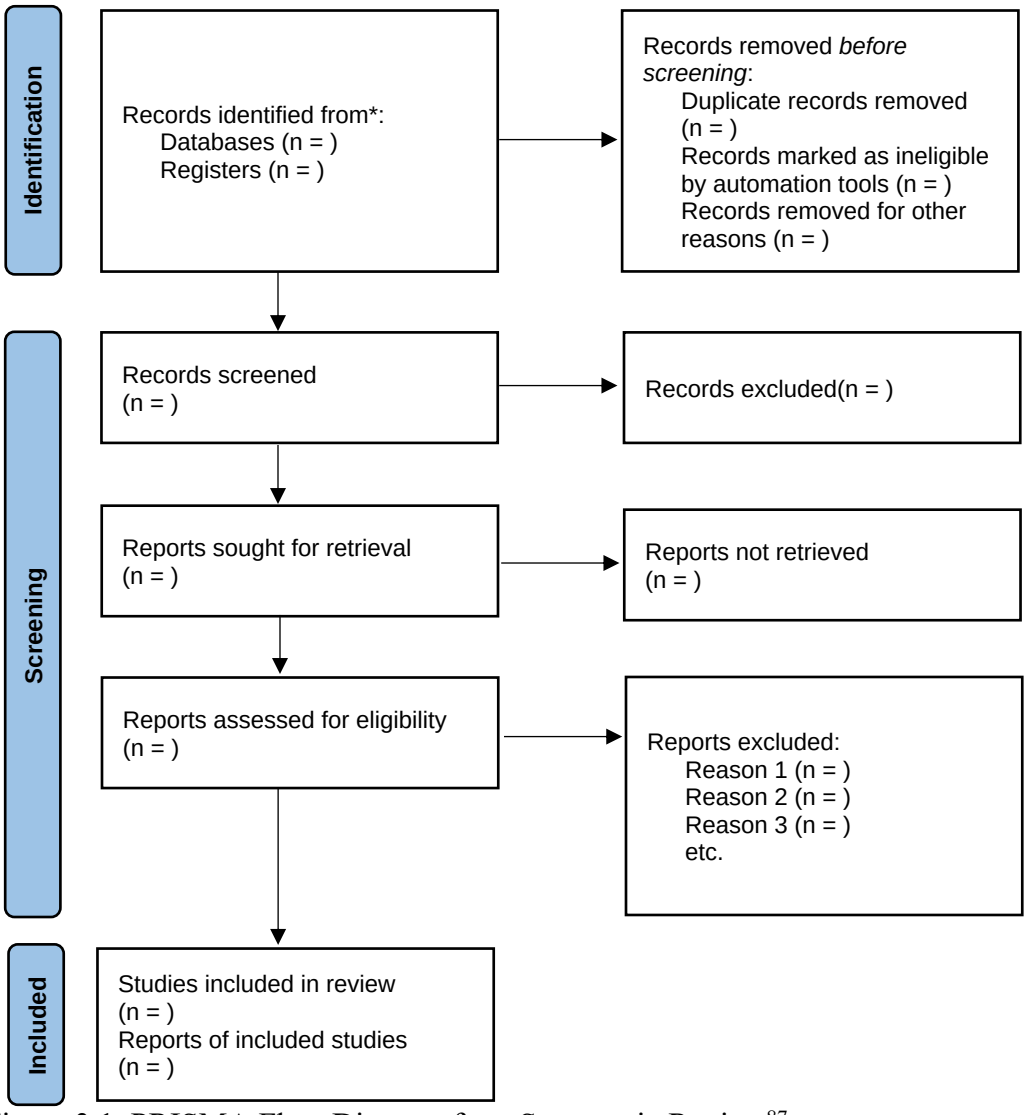

Figure 3.1. PRISMA Flow Diagram for a Systematic Review[87]

Data Cleaning and Preparation

To make risk factors reported in each study comparable, similar risk factors will be grouped together when it is considered clinically reasonable. For example, lab tests such as lymphocyte count <0.8 and <1.0 would be grouped together. In the case where risk factors are reported with different units, results will be converted to a common set of units.

For continuous risk factors that are not reported in mean and standard deviation, these values will be estimated using median and interquartile range. The mean will be estimated by using methods proposed by Wan et al.[91] The lower (q1) and upper (q3) bounds will be added to the median (m) and divided by a constant of 3. The standard deviation will be estimated by subtracting q1 from q3 and dividing by $\eta(n)$, which will be determined using the sample size and Table 2 from the Wan publication. These values will be calculated using the below equation (Equation 3.1) and R statistical software.

Equation 3.1. Estimation of median and interquartile range

$$\eta(n) = 2E(Z_{(3Q+1)})$$

Quality Assessment

To assess the study quality of each of our included studies, we will use the Quality in Prognostic Studies (QUIPS) tool adapted for our systematic review.[92] This quality tool was created to assess how well a prognostic study's methods limited potential biases. QUIPS is divided into five domains: 1) Study Participation, 2) Study Attrition, 3) Prognostic Factor Measurement, 4) Outcome Measurement, 5) Study Confounding, 6) Statistical Analysis and Reporting. These

categories were created by grouping many commonly reported sources of biases seen in prognostic studies into a manageable 6-category tool.

All sections are judged as high, moderate, or low risk of bias. Definitions of each level of bias are prespecified for each domain. Aligned with our inclusion criteria, a study will be considered low risk of bias if both inpatient and outpatient participants are included, while those studies that include only a specialized population will be considered high risk of bias. In terms of study attrition, those studies with complete participant ascertainment or less than 10 percent loss to follow up will be considered low risk of bias. Review of the literature and guidance from a physician will help us to define study bias in terms of prognostic factor measurement. Low bias studies will include those that use clearly defined and typical cutoff values. To assume limited bias in terms of outcome measurement, the outcome in each study must have a clear and reproducible definition. Lastly for a study to be considered low risk of bias, use of a multivariate analysis is necessary to show assessment of confounding. The full adapted tool is included in Appendix A, Table A4. Quality of each included study will be assessed by two researchers in parallel. Disagreements will be resolved by consensus or with the aid of a third researcher when necessary.

Analysis

Random Effects Meta-Analysis

R version 4.1.2 and the meta package will be used to perform a random effects meta-analysis of data gathered from included studies. We chose to use a random effects meta-analysis to account for both within and between study variability.[93] With the likely differences between studies in

terms of outcome definition, prognostic factor assessment, and study population as well as a desire to make inferences beyond the observed studies, a random effects analysis is appropriate.

Both categorical and continuous risk factor data will be uploaded into R. Pooled estimates of the relative risks (categorical) and unstandardized mean difference (continuous) for each risk factor will be generated. For continuous risk factors, median and interquartile ranges will be converted into mean and standard deviation estimates using methods outlines by Wan et al.[91] For each risk factor, a forest plot will also be created with the number of studies and patients for each summary estimate noted.

Assessment of Heterogeneity

To compare results of each included study and consider the consistency of their findings we will use a combination of graphical representations and statistical tests. We will visually observe generated forest plots to assess any noticeable variance between included studies. These plots will also report the I^2 statistic. This statistic indicates what proportion of the observed variance would remain if we could eliminate sampling error. There is always some level of heterogeneity present due to the inherent differences between studies making a statistical test for heterogeneity not of great value. Also I^2 tends to underestimate between study variability if there are less than ten studies included. [94] Therefore we will use the values to help guide our assessment of heterogeneity between studies but will not report on the specific values, especially if the number of studies included is limited.

Assessment of Publication Bias

In the case of publication bias, also known as non-reporting bias, studies with statistically significant results tend to be reported over those without. [94] This can lead to bias in the results of a meta-analysis due to a systematic bias in the reporting of results. A funnel plot can be created to visualize this type of bias. This plot graphs a measure of variance versus a measure of effect. Larger studies tend to have less variance and cluster at the top of the plot. Alternatively, smaller studies have more variance and tend to spread out at the bottom of the plot.[95] A symmetrical plot, in the shape of a funnel, indicates no publication bias. An asymmetrical or skewed plot indicates possible publication bias. This bias can also be tested using Egger's test which performs a linear regression of the intervention effect estimates on their standard errors using inverse variance weighting.[96] This essentially tests if the association between the effect of the intervention and the measure of study size is more than we could expect by chance.[94] A significant result from this test indicates potential publication bias. However, when there are too few studies included in meta-analysis the power of Egger's test tends to be too low. Therefore, this test should only be performed when ten or more studies are included in analysis.

To minimize publication bias in this meta-analysis we chose to include papers from a preprint server in hopes of identifying all related literature, regardless of result. We will use the 'metabin' function in the R meta package to generate a funnel plot to visualize publication bias. A plot for each risk factor with ten or more studies will be generated. We will then visually analyze the plot for symmetry. The funnel plots will also be statistically evaluated using Egger's test using the 'metabias' function in R, permitting enough studies are included.

AIM 2: Risk Factors of Long COVID-19 in College Students

Background

On college campuses, initial transmission of COVID-19 was minimal due to campus closures and the implementation of virtual learning during the spring and summer of 2020. By Fall of 2020, some schools began to allow students to come back to campus. This resulted in a large increase in COVID-19 transmission on many campuses. This is likely due to student gatherings and their beliefs and behaviors related to social distancing measures upon initial return.[80] On campus, universities also provided many resources for students to combat and prevent COVID-19 illness. By the Spring of 2021, most universities provided free and accessible COVID-19 testing, access to vaccinations, and implementation of social distancing measures in all facilities. Despite these mitigations efforts, students continue to experience COVID-19 illness and some adverse events as a result.

Additional research has shown that college students are reporting prolonged symptoms as a result of COVID-19 infection. Prolonged symptoms commonly include chest pain, fatigue, fever, loss of taste and smell, and headaches.[82] Two studies in college students have demonstrated long COVID existing in a range of 11 to 50 percent of students. [9,82] This wide range of prevalence is likely due to differing definition of long COVID. Research on this topic surrounding just young adults, more specifically college students, is very limited. This age group is not thought to be of increased risk for adverse outcomes such as severe disease, death, or prolonged symptoms. Further research is needed to understand why college students are experiencing prolonged symptoms and who is most at risk. Lastly, focusing on students on the same college campus will allow us to compare outcomes and experiences of students with COVID-19 living in similar environments and with access to the same resources.

Objective

To identify risk factors for prolonged symptoms of COVID-19 in college students by comparing those with long-COVID-19 to a group without prolonged symptoms. We also aim to estimate the prevalence of prolonged symptoms of COVID-19 among college students.

Methods

Setting and Population

The University of Georgia (UGA) is the nation's first land grant institution with 40,118 students enrolled as of 2022.[97] Of those students, 30,166 are undergraduate, 8,304 are graduate, and 1,648 are professional students. Table 3.1 includes demographics of enrolled students at UGA as of the Fall 2021 semester.[97] Overall the university is about 59 percent female and has an average student age of 22 years old.

Throughout the COVID-19 pandemic, COVID-19 testing for both symptomatic and asymptomatic students has been provided on campus. Enrolled students with COVID-19 symptoms can make appointments at the University Health Center (UHC) and receive a PCR COVID-19 test via nasal swab. For those students who are asymptomatic, walk-in testing was provided via saliva sample at UHC as well as pop-up testing sites throughout campus. Promotions for asymptomatic testing before and after scheduled campus breaks were also frequently made to promote students getting tested before and after travel. Public testing sites as well as private were also available throughout the Athens, Georgia community. This population is ideal for our study as it captures a population of young adults that have easy access to COVID-

19 testing regardless of symptom status. This will allow for us to hopefully capture students with prior COVID-19 diagnoses and assess their risk of long COVID.

Table 3.1 Characteristics of enrolled UGA students Fall 2021 (Total Students= 40,118) [97]

Characteristic	Frequency (%)
Sex	
Male	16,480 (41.1)
Female	23,578 (58.8)
Not Reported	60 (0.15)
Race	
American Indian	43 (0.1)
Asian	4,489 (11.2)
Black/African-American	3,258 (8.1)
Hawaiian/Pacific Islander	25 (0.06)
Hispanic	2,527 (6.3)
Multiracial	1,634 (4.1)
White	26,503 (66.1)
Not Reported	1,639 (4.1)
Age	
17 and younger	200 (0.5)
18-20	17,862 (44.5)
21-24	14,770 (36.8)
25-29	3,738 (9.3)
30-34	1,651 (4.1)
35-39	760 (1.9)
40 and older	1,137 (2.8)

Study Design

We plan to conduct a cross-sectional study in students at UGA with previous COVID-19 infection to identify predictors of long COVID-19. Our outcome, long COVID-19, will be defined as a patient experiencing at least one symptom that they did not have prior to COVID-19 diagnosis and lasted for four weeks or longer. This is the same definition for long COVID given by the CDC.[4] To be included in the study, participants must be between 18 and 30 years old and have a previous COVID-19 diagnosis, either via real time reverse transcription chain reaction

(RT-PCR) or rapid diagnostic test (Ag-RDT). To minimize potential for recall bias, due to the length of time since disease diagnosis, we are limiting patient inclusion to those with a positive COVID-19 test since August 1, 2021. This date was chosen to align with more recent case surges as a result of COVID-19 variants. As of August 2021 more than 98 percent of cases in the United States were a result of the Delta variant.[98] This was then followed by another surge of cases caused by the Omicron variant peaking in January 2022.

Data Collection

We will generate a survey to gather information on students with potential long COVID. Qualtrics, a secure subscription based service provided by UGA free of charge, will be used to generate and administer our survey. Protocols, consent forms, and our survey will be submitted to the UGA institutional review board (IRB) for approval before the start of data collection. A participant consent letter as well as the survey are included in Appendix B. Aside from our outcome data, our survey will aim to capture risk factors participants may have for long COVID, including demographics, preexisting medical conditions, vaccination status, symptoms, and the severity of their illness.

For determination of predictors of long COVID we will use the UGA student email listserv. This listserv includes the email addresses of all currently enrolled UGA students. Included in the email will be a description of our study, inclusion criteria, consent information, and a link to the survey. Students will answer questions to confirm that they meet our inclusion criteria and that they agree to complete the survey. Only those that meet inclusion criteria will be able to complete the rest of the survey. Those who do not will be excluded from any further analysis. Participants who are included will be evaluated for our outcome. Final numbers of total

survey responses and those included in the study will be provided in a flow diagram as well as characteristics of study participants and survey response data.

Analysis

Data will be stored in a Google Drive account in a password protected file. This file will only be accessible by myself, the dissertation committee, and any other approved data analyzers. A codebook will be created and data cleaned by myself. All analysis will be done using R statistical software.

Analysis for predictors of long COVID-19

Exploratory data analysis will first be done, which will include examination of the data set and any needed data cleaning. Then an item analysis will be done for each variable individually to assess for missing data, lack of variability, and outliers. Covariates that will be included in the model will be derived from our survey responses. The outcome of the final model will be the presence or absence of long COVID-19. All dichotomous and categorical variables will be plotted using a stem and leaf plot to assess variability. Continuous variables will be plotted using histograms. Correlation analysis will then be done to address any potential collinearity. Those variables with high correlation coefficients will be considered for combination or exclusion. All variables to be included in analysis, following correlation analysis, will be moved to a final data set before generating our model. Bivariate analysis will be done to assess the relationship of long COVID-19 with each potential predictor identified in our survey. Logistic regression will be used to asses associations of each predictor with our outcome.

Finally, we will build a multivariate logistic regression model using the least absolute shrinkage and selection operator technique (Lasso).[99] This is a type of penalized regression that uses a flexible shrinkage approach that is effective when the number of events per predictor could be less than 10.[100] We chose lasso for our variable selection method as it tends to be more sensitive and more precise than traditional stepwise selection methods. Lasso constrains the sums of the absolute values of the regression coefficients and is able to shrink the predictor coefficients to zero to exclude them from the final model.[100] The amount of shrinkage that is applied to the Lasso regression is determined by a tuning parameter lambda. This resulting value thus determines variable selection for the final model. We will use the glmnet package in R to perform the regression. As a result we will report predictors included in the final model with their corresponding beta coefficient and odds ratio.

AIM 3: Development and Validation of a Clinical Prediction Rule for Predicting Long COVID-19 in Young Adults

Background

While treating patients, clinicians typically make multiple decisions regarding the diagnosis, prognosis, and treatment plan for disease. Clinical prediction rules (CPR) are tools derived from original research that attempt to reduce the uncertainty of medical decision making by standardizing the collection and interpretation of clinical data.[101] They can be used to support decision-making around diagnosis, the need for hospital referral, and treatment.

The first stage of creating a CPR is the development stage. This first stage typically includes three steps: 1) identification of predictors, 2) assessment of the presence or absence of these predictors in patients with and without the outcome of interest, and 3) statistical analysis.[101] Researchers use extensive literature review as well as knowledge of disease to compile a list of predictors. These predictors can include symptoms, signs, medical history, and laboratory tests. Following identification of predictors, a patient population is selected and assessed for the presence or absence of the predictors and the outcome of interest. This group of patients is often referred to as the development or derivation population as it is the population that will be used to develop the CPR to predict the outcome. To make the final CPR generalizable to other populations, use of a diverse population is necessary. Generally, the development population should resemble the clinical population researchers want to apply the model to. Once data collection is complete, statistical analysis is used, primarily logistic regression models, to identify which predictors have the strongest association with the outcome and which can be removed from the final model. Using the regression coefficients of the final fitted model, point

values are assigned to each parameter to create a CPR that is easy to implement in clinical practice.

Several clinical prediction rules have been developed to determine the risk of long COVID-19, however many use machine learning approaches with models that have a high data collection burden and are complex.[102,103] These types of models are useful at a large scale, for example in predicting impacts of long COVID-19 on the healthcare system or gaining more knowledge on the outcome while definitions are unclear. However, these complex prediction models cannot be easily memorized or implemented in a healthcare setting for a physician to determine a treatment plan for an individual patient at risk for long COVID-19. Simple prediction rules have also been developed, including predictors such as age, sex, and number of symptoms during acute infection, but these do not necessarily take into account the patients pre-existing medical conditions.[104] Lastly, all existing CPR for long COVID-19 include all age groups. These rules have generally found an association with middle to older age and long COVID-19. While it is commonly true that older individuals experience more complications related to COVID-19, young adults are also experiencing these complications. Young adults in particular are reporting prolonged symptoms of COVID-19 that are lasting for as long as one year after infection.[9]

Objective

We aim to develop a CPR for predicting which young adults are at increased risk for developing long COVID-19.

Methods

Data Source

To develop our CPR, we will use data from the National COVID Cohort Collaborative (N3C).[105] This registry dataset includes an enclave of data from participating institutions electronic health records (EHR). To date, there are about 18 million COVID-19 positive patients across 77 healthcare sites included. N3C includes individuals both with lab confirmed COVID-19 and those with related symptoms. Data on the medical history, comorbid diagnoses, demographics, immunizations, and laboratory results are also included. Patient records are also linked across sources to remove duplicates and link patient visits. N3C includes individuals with confirmed COVID-19 as well as controls without. For our study, only young adults between the ages of 18 and 30 years old with lab-confirmed COVID-19 will be included.

Identification of risk factors and outcome

Our outcome of interest is long COVID-19. There is currently no standard for defining long COVID-19 and International Classification of Diseases, Tenth Revision (ICD-10) codes in the EHR are not widely used. We chose to use the definition set by the Centers for Disease Control (CDC) for long COVID-19. This definition classifies the outcome as a range of new, returning, or ongoing health problems people can experience four or more weeks after being infected with COVID-19.[4] From N3C, individuals with lab-confirmed COVID-19 visiting a provider more than 4 weeks after their diagnosis and reporting at least one related symptom will be classified as having long COVID-19. These symptoms include abdominal pain, new anxiety or depression, brain fog, chest pain, chills or sweats, cough, fatigue, fever, headache, nausea or vomiting,

congestion, shortness of breath, sore throat, or loss of taste or smell as listed by CDC for common symptoms of long COVID-19. [4]

Risk factors that will be included in our CPR will be identified through our research in both aims 1 and 2 of this dissertation. Risk factors for the model include demographics, symptoms, disease severity, and comorbidities. A variable for a date of positive test, classified by those with a positive test either before or after December 1, 2021, will be included to account for differences among more severe COVID-19 variants. This date coincides with the Omicron variant that was first identified in November 2021 and became the predominant strain in the United States by December 2021. [20] Risk factors for COVID-19 vaccination status will be left out of this analysis due to missingness of this information. N3C reports that many of the healthcare records included in their database are not linked to vaccination records causing a large amount of missing data for these variables. We also chose not to include variables related to race in our multivariate analysis. Race in clinical prediction models, if there is no biological association, can represent underlying health disparities rather than a medical need for this adjustment in the model.[106]

Analysis

Development of CPR

Bivariate analysis will be performed using logistic regression for all variables to asses association with the outcome. Cut points for continuous variables will be selected by inspection of histograms. Our cohort will be randomly split into derivation and validation datasets using a 80:20 ratio. The derivation dataset will be used to develop our model and the validation dataset to test the ability of the model to predict individuals with long COVID-19.

Typically in medical research, stepwise selection is used to build a prediction model. These methods are based on inclusion or exclusion of risk factor into a model based on p-values, Akaike Information Criterion (AIC), or Bayesian Information Criterion (BIC).[99] These methods can be fairly subjective and lead to biased estimates and overfit models.[99] To minimize overfitting and extreme predictions, we will be using the least absolute shrinkage and selection operator technique (lasso) to develop a logistic regression model for predicting long COVID-19. This is a type of penalized regression that uses a flexible shrinkage approach that is effective when the number of events per risk factor could be less than 10.[100] Lasso is also preferred over other penalized regression methods as it is able to produce a simpler model with fewer risk factor.[100] This is optimal for developing a CPR that can be easily implemented in clinical practice. Lasso constrains the sums of the absolute values of the regression coefficients and is able to shrink the risk factor coefficients to zero to exclude them from the final model.[100] The amount of shrinkage that is applied to the Lasso regression is determined by a tuning parameter lambda. This resulting value thus determines variable selection for the final model. We will use the glmnet package in R to perform the regression. As a result we will report risk factors included in the final model with their corresponding beta coefficient.

Discrimination

A receiver operating characteristic (ROC) curve is a plot of the true positive rate (sensitivity) against the false positive rate (1-specificity) for a range of possible cutoffs to define an abnormal test. We will apply our clinical prediction rule to our data and calculate the sensitivity and specificity to build an ROC curve. This curve will also generate a value for the area under the curve (AUC), also known as the c-statistic. The c-statistic indicates the probability that our test

will correctly classify a patient with the outcome as having the outcome and a patient who does not have the outcome as not having the outcome. The c-statistic can range from 0.5, for a test that does not discriminate between patients with and without disease at all, to 1 for a test that perfectly classifies them. We will use the AUC to measure the discrimination of our CPR. This will be done in both the derivation and validation groups and results compared. Guidelines for interpreting the AUROCC indicate 1 is perfect discrimination, 0.99-0.9 is excellent, 0.89 to 0.8 is good, 0.79 to 0.70 is fair, 0.51 to 0.69 is poor, and 0.5 is worthless.[107]

Presentation of CPR

Our goal is to generate a CPR that is easy to implement in a healthcare setting. We want to present a CPR that allows a provider to classify a patient into a risk group for developing long COVID-19. To achieve this, we plan to present our final CPR as a point score. A point score assigns a numerical value to each risk factor in the model and allows a provider to add the total points to easily classify patient risk of developing the outcome. We plan to achieve this by using the regression coefficients of our final model and rounding them to the nearest number resulting in a weighted score.[108] Using the regression coefficients for development of the point score will result in a narrow range of points and a linear distribution of the final scores.[109]

Classification Accuracy

We then want to use our resulting CPR to classify both cohorts into low, moderate, and high risk groups. We will apply the CPR to the derivation group and calculate individual point scores. We will stratify scores into risk groups based on visual inspection of the point score distribution to

create groups that would be most useful for clinical decision making. The CPR will then be applied to the validation group and results classified into risk groups.

External Validation

Our CPR will also externally validated using a dataset of college students at the University of Georgia gathered in Aim 2. UGA students between the ages of 18 and 30 years old and who had a positive COVID-19 test after August 1, 2021 will be included in the dataset. Students will be emailed a survey that gathered information on their prior COVID-19 infection including dates of infection, symptom information, and comorbidities. The CPR will be applied to this dataset and the AUROCC will be calculated. The points for each participant will be calculated and the distribution of the points into low, moderate, and high risk groups will be observed.

CHAPTER 4

RESULTS

Introduction

The results of each aim of this dissertation are presented as individual manuscripts in chapters 5, 6, and 7. These chapters correspond to aims 1, 2, and 3 respectively. Each manuscript contains an introduction, methods, results, discussions, and related tables and figures. A single reference list is included at the end of this dissertation. There may be some repetition between information presented in chapters 1 through 3. Following the results of all three aims, chapter 8 will summarize all results and discuss future directions for research.

CHAPTER 5

Risk Factors of Long COVID-19: A Systematic Review and Meta-Analysis

Introduction

Individuals with previous mild to severe COVID-19 can experience prolonged symptoms referred to as long COVID-19. Currently in the United States, the CDC classifies long COVID as a range of new, returning, or ongoing health problems people can experience four or more weeks after being infected with COVID-19.[4] While experiences with this outcome differ by person, many report respiratory, neurologic, and psychiatric symptoms.[6,7,83]

The risk factors for developing long COVID are largely unknown. A number of studies have attempted to identify these risk factors and report that increased age, higher BMI, and female sex all appear to be associated with an increased risk of long COVID.[84] Though an association with increasing age has been reported, there is some evidence that younger individuals are also experiencing prolonged symptoms as a result of COVID-19 infection.[9]

Current systematic reviews on long COVID primarily focus on signs and symptoms that patients are experiencing. Some do report risk factor information that is included in studies, but do so in a qualitative way due to differences in studies and limited data at the time of publication.[85,86] As long COVID becomes more prevalent, the number of publications on the topic has also grown. In this study we will aim to identify risk factors that put individuals at an increased risk for development of long COVID-19.

Methods

Inclusion Criteria

To be included in the meta-analysis studies had to report the association between at least one risk factor and a prolonged symptom or health problem after COVID-19 infection. These risk factors can include demographics, such as age and sex, BMI, vital signs, common symptoms, comorbidities, such as diabetes, heart disease, COPD, or asthma, and results of laboratory tests. The outcome of long COVID-19 is defined as new, returning, or ongoing health problems occurring four or more weeks after being infected with COVID-19. Therefore, all studies must report symptoms for at least four weeks following initial diagnosis. Included studies must also have a patient population of primarily adults with laboratory confirmed COVID-19. Studies were only included if sufficient data was reported for calculation of relative risk. This includes reporting the number of patients with and without the risk factor for the outcomes of long COVID-19 and those without long COVID-19.

Studies were excluded if they included primarily children. They were also excluded if they focused on a specialized population such as individuals with cancer, HIV, or pregnant women. Studies that included less than 50 patients were also excluded. There were no limitations set to the language of the publications and pre-print studies were also included.

Search Strategy

A search of both Medline and the WHO COVID-19 database were done using a combination of key terms for COVID-19, long COVID-19, and risk factors.[110,111] The full search strategy for these databases is included in Appendix A, Table A1. Additionally, the MedRxIV preprint server

was searched to identify additional preprint studies using similar terms. Use of preprint studies allow for the inclusion of non-significant results that might not otherwise be published.

Data Abstraction

All abstracts were reviewed in parallel by two investigators. For any abstract of interest, the full article was obtained and also reviewed by two investigators. Information including study characteristics, study quality information, and risk factor data were abstracted from all studies meeting inclusion criteria. Risk factor data for both continuous and categorical data were abstracted and analyzed. All abstraction was done in parallel and all discrepancies were discussed and resolved by consensus.

Data Preparation

To make risk factors reported in each study comparable, similar risk factors were grouped together when it was considered clinically reasonable. For example, lab tests such as lymphocyte count <0.8 and <1.0 would be grouped together. In the case where risk factors are reported with different units, results were converted to a common set of units. A full list of the original and grouped risk factor names are included in Appendix A, Table A2 and A3.

For continuous risk factors that are not reported in mean and standard deviation, these values were estimated using median and interquartile range. The mean was estimated by using methods proposed by Wan et al.[91] The lower (q1) and upper (q3) bounds were added to the median (m) and divided by a constant of 3. The standard deviation was estimated by subtracting q1 from q3 and dividing by $\eta(n)$, which was determined using the sample size and Table 2 from the Wan publication.

Quality Assessment

The Quality in Prognostic Studies (QUIPS) tool was used to assess the quality of included studies. The tool was adapted for our use and definitions of low, moderate, and high risk of bias were prespecified for each of the domains. The full tool is included in Appendix A, Table A4. All included studies were assessed for quality by two investigators with all discrepancies being discussed and resolved by consensus. Overall risk of bias was determined for all included studies. Studies with only one or two categories marked with moderate risk of bias were considered moderate risk of bias overall. If there were three or more categories with moderate or at least one category with high risk, then the study was considered to have high risk of bias overall. Those with low risk across all categories were considered low risk of bias.

Analytic Strategy

R version 4.1.2 and the meta package was used to perform a random effects meta-analysis of data gathered from included studies. Both categorical and continuous risk factor data were uploaded into R. Pooled estimates of the relative risks (categorical) and unstandardized mean difference (continuous) for each risk factor were generated. For continuous risk factors, median and interquartile ranges were converted into mean and standard deviation estimates using methods outlines by Wan et al.[91] For each risk factor, a forest plot was also created with the number of studies and patients for each summary estimate noted. These plots will also report the I^2 statistic. This statistic indicates what proportion of the observed variance would remain if we could eliminate sampling error. I^2 tends to underestimate between study variability if there are less than ten studies included. [94] Therefore we will use the values to help guide our assessment of

heterogeneity between studies but will not report on the specific values, especially if the number of studies included is limited.

Assessment of Publication Bias

We used the 'funnel' function in the R meta package to generate a funnel plot to visualize publication bias. A plot for each risk factor with ten or more studies was generated. Plots were then visually analyzed for symmetry. We also statistically evaluated for publication bias using Egger's test. This essentially tests if the association between the effect of the intervention and the measure of study size is more than we could expect by chance.[94] A significant result from this test indicates potential publication bias. However, when there are too few studies included in meta-analysis the power of Egger's test tends to be too low. Therefore, this test was only be performed when ten or more studies are included in analysis.

Results

Search results

Figure 5.1 includes a summary of our search process for relevant studies. A total of 1,157 studies were identified and 224 were reviewed in full for inclusion. For our meta-analysis, 46 studies met all inclusion criteria and were used for analysis.

Characteristics of the included studies are shown in Table 5.1. Studies were performed worldwide including 9 set in the United States, 5 in Germany, and 5 in Italy. Dates of positive COVID-19 tests for participants ranged from March 2020 to December 2021. Participants in these studies were followed up for at least 4 weeks after diagnosis and up to 1 year. The mean or median age ranged from 29 to 67 years and 19 of the 46 studies included primarily men.

Study quality

Study quality of all included studies was evaluated using the QUIPS tool. Table 5.2 includes these results. In terms of study participation 30 studies had moderate risk of bias as they were performed only in inpatient or only outpatient settings. Twenty-two studies had moderate risk of bias for study attrition due to a loss to follow-up greater than 10 percent. Prognostic factor measurement, outcome measurement, and statistical analysis were clearly defined and indicated a low risk of bias across all studies. Twelve studies did not include a multivariate analysis and were given a high risk of bias for study confounding. Overall, 7 studies had a low risk of bias across all QUIPS categories. We classified 27 studies has having an overall moderate risk of bias. This was determined by having one to two categories of moderate risk of bias. Twelve studies had high risk of bias overall due to either have three or more categories of moderate risk or at least one category of high risk of bias.

Categorical risk factors

Summary estimates of the relative risks and their corresponding confidence intervals for each categorical risk factor reported in 3 or more studies is shown in Table 5.3. Among demographic risk factors, age 60 to 70 years, male sex, white race, other race, and BMI of 18.5 to 25 kg/m^2 were significantly associated with our outcome. Among these risk factors, relative risks for age 60 to 70 years and white race were greater than 1 indicating they were associated with the development of long COVID. Male sex, other race, and BMI of 18.5 to 25 kg/m^2 showed relative risks less than 1 indicating individuals with this risk factor were more likely to not develop long COVID-19. Twelve comorbidities were also significantly associated. These

included liver disease, arthritis, autoimmune conditions, and mental health conditions. Hearing loss, back pain, hospitalization, allergic rhinitis, moderate severity, and critical severity were also associated with long COVID-19. Among these statistically significant risk factors, liver disease, arthritis, and autoimmune conditions were the most strongly associated with our outcome. COVID-19 severity in the included studies was defined the WHO definition for these risk categories.[112] Many common COVID-19 symptoms including cough, loss of taste and smell, and fatigue were not significantly associated with our outcome near the null value. While our review included studies of primarily adults, several studies were included as they had less the 10 percent children. These included children aged 0 to 19 years. Our results show that age less than 20 had a strong association with our outcome but was not statistically significant. All forest plots for risk factors significantly associated with long COVID and their I^2 value are included in Appendix A. Overall there was moderate heterogeneity amongst categorical risk factors. Five risk factors had a high I^2 value over 80 percent while the rest were primarily below 50 percent.

Continuous risk factors

Table 5.4 includes risk factors reported as continuous variables, the unstandardized mean difference between patients with and without the risk factor, and their corresponding confidence intervals. Risk factors that had a statistically significant association with long COVID included increased LDH, length of hospital stay, and the number of symptoms at onset. Platelet count was also strongly associated with our outcome, however this association was not significant. Both age and BMI were reported in 30 and 12 studies respectively. Older age showed an association with long COVID while lower BMI indicated an association with the outcome, but results indicate a non-significant association with the outcome. There was a range of heterogeneity

amongst continuous risk factors. Several risk factors including number of symptoms and LDH had a high I^2 value at 94 percent. Other risk factors, including length of hospital stay, showed no heterogeneity with an I^2 value of 0 percent.

Additional sensitivity analysis of both categorical and continuous risk factors are included in Appendix A (Tables A4-A7). Risk factors were grouped by the setting of the study either being classified as inpatient, outpatient, or inpatient and outpatient. Also studies were grouped by long COVID-19 definition, noted by the minimum length of time a patient must have symptoms after COVID-19 infection in order to be considered having long COVID.

Assessment of publication bias

Publication bias for risk factors reported in 10 or more studies was assessed via forest plots and Egger's test. The results of Egger's test are included in Table 5.5. Egger's test indicates publication bias with p-values less than 0.05. Based on that standard there was evidence of publication bias for the risk factors diabetes, hypertension, kidney disease, and male sex. Funnel plots for these risk factors are included in Appendix A. Studies reporting diabetes, hypertension, and kidney disease were overwhelmingly larger studies reporting relative risks near the null value. There were not many smaller studies included and they did not report many extreme associations. Male sex indicated a slight publication bias based on Egger's test.

This risk factor was reported in both larger and smaller studies and did include a range of associations. However, there were several studies reporting more extreme associations, towards the outcome of long COVID.

Discussion

The objective of our meta-analysis was to identify the risk factors associated with long COVID-19. We were able to identify several associations that were both strongly and significantly associated with our outcome of long COVID-19. Though we did find some associations with long COVID-19 there are several limitations to consider with the studies included in this meta-analysis.

First, some of our risk factors that had the strongest associations and were significant were only reported in a few studies. For example, increased LDH had the strongest association with our outcome (RR=20.33), but was only reported in three studies. Similarly symptoms that were strongly associated including hearing loss and back pain were only reported in four and three studies respectively. It is important to consider the number of studies these risk factors were reported in, as those reported across more studies are substantiated by more evidence. Among symptom risk factors we would expect more common symptoms of COVID-19 to be more strongly associated such as cough, shortness of breath, or fatigue. However these risk factors were reported in more studies and their overall risk estimate remained around the null value. It is also important to consider risk factors that we found minimal or no significant association with the outcome. The framework of a meta-analysis only allows the summation of risk factors with sufficient data. Just because no association between a risk factor and the outcome were found in our study does not mean it is not associated with the outcome. There just may not be enough evidence reported across literature to determine an association of the two.

Due to the novelty of long COVID-19 the definitions for the outcome varied greatly across studies. Some studies defined long COVID by prolonged symptoms of varying lengths while others specified specific symptoms that must be present for a certain period. We used a

more general definition by setting an inclusion criteria of a minimum follow-up of 4 weeks. This followed the CDC definition of long COVID while also allowing us to capture studies with different definitions for the outcome. However, the difference in definitions and how long COVID was defined could lend to some of the heterogeneity amongst studies. In future research aimed at identifying risk factors of long COVID, a more clear definition of the outcome will help make studies more comparable.

Lastly, there was evidence of some issues with both study quality and publication bias. Twelve of the included studies were considered to be high risk of bias. This included generally small studies, all with less than 500 participants. Whereas, included larger studies were considered to be moderate or low risk of bias overall. These studies generally had bias issues with not including multivariate models and loss to follow-up due to the design of the study. Related are the issues with publication bias. Funnel plots for those risk factors displaying evidence of publication bias showed mostly larger studies with estimates pooled near the null value. Some with significant associations and some not. Missing were the medium to smaller sized studies that showed more extreme associations. Overall these results indicate that more quality studies of varying size are needed in this subject area. A comparison of studies ranging in size and setting will allow for future research to make more concrete conclusions about the risk factors for long COVID.

Conclusion

We found that age 60-70 years, female sex, white race, a number of comorbidities and symptoms, hospitalization, increased LDH, length of hospital stay, and number of symptoms were all significantly associated with development of long COVID-19. However to draw causal

conclusions about these associations more research is needed. There is a need for an understanding of what long COVID truly is and how to define it. Once a clear definition is formed, this will make classifying patients with the outcome more simple and allow for the true identification of the risk factors of long COVID. Studies both large and small are also needed in order to help identify the true association of these risk factors with this outcome. We hope future research in this area will lead to the development of clinic tools that will help identify and treat prolonged symptoms of COVID-19 effectively.

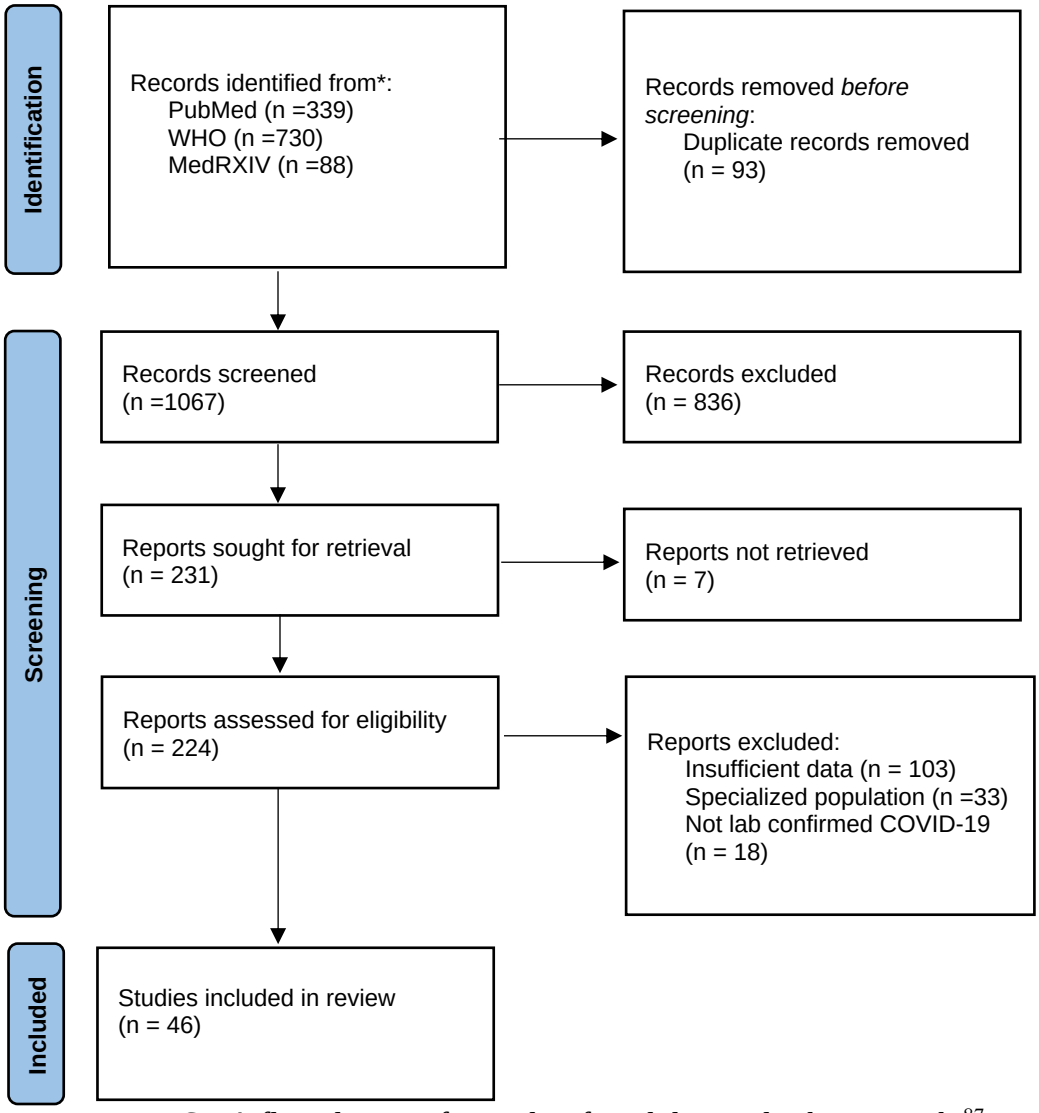

Figure 5.1: PRISMA flow diagram for studies found during database search [87]

Table 5.1. Study characteristics of all included studies

Author, Year	Country	N	Study Design	Setting	Date of Positive COVID-19 Test	Follow-up Period	Mean or Median Age	Percent Male
Sudre, 2021[104]	United Kingdom, United States, Sweden	4182	Cohort	Outpatient	March 25, 2020 - June 30, 2020	4-12 weeks	42 (32-53)	28.5
Augustin, 2021[113]	Germany	958	Cohort	Outpatient	April 6, 2020	4-7 months	43 (31-54)	46.5
Peghin, 2021[114]	Italy	599	Cohort	Inpatient and Outpatient	March 1, 2020 - May 30, 2020	6 months	53 (15.8)	46.6
Naik, 2021[115]	India	1234	Cohort	Inpatient	September 2020	45-181 days	41.6 (14.2)	69.4
Asadi-Pooya, 2021[116]	Iran	4681	Cohort	Inpatient	February 19, 2020 - November 20, 2020	3-12 months	52 (15)	52.9
Subramania, 2022[117]	United Kingdom	384137	Cohort	Outpatient	January 31, 2020 - April 15, 2021	88-153 days	44.1 (17.0)	44.8
Bai, 2022[118]	Italy	377	Cohort	Inpatient	March 1, 2020 - November 1, 2020	1-3 months	57 (49-68)	63.7
Bliddal, 2021[119]	Denmark	185	Cohort	Outpatient	Before August 12, 2020	4-12 weeks	LC: 47.4 (15.2) No LC: 45.4 (15.5)	36.8
Cervia, 2022[120]	Switzerland	134	Cohort	Outpatient	April - August 2020	6-12 months	LC: 52 (34-65) No LC: 34 (27-50)	64.2
Vimercati, 2021[121]	Italy	352	Cohort	Outpatient	March 8, 2020 - March 15, 2021	12 months	45.4 (13.2)	42

Kim, 2022 (a)[122]	South Korea	241	Cohort	Inpatient and Outpatient	February 18, 2020 - March 14, 2020	6-12 months	37 (26-51)	32
Becker, 2021[123]	Switzerland	90	Cohort	Inpatient	March - June 2020	1 month	60.1 (15.1)	62
Nune, 2021[124]	United Kingdom	89	Cohort	Inpatient	February 15, 2020 - July 31, 2020	3-9 months	67.5 (?)	61
Desgranges, 2022[125]	Switzerland	418	Cohort	Outpatient	February 26, 2020 - April 27, 2020	3-10 months	LC: 42 (33-54) No LC: 41 (30-54)	37.5
Hossain, 2021[126]	Bangladesh	2693	Cohort	Outpatient	June-November 2020	4-12 weeks	LC: 38.86 (11.3) No LC: 38.7 (11.4)	71.8
Lui, 2021[127]	Hong Kong	204	Cohort	Inpatient	July 21, 2020 - December 21, 2020	3 months	55 (44.3-63.0)	46.6
Förster, 2022[128]	Germany	1459	Cohort	Inpatient and Outpatient	March 1, 2020 - September 30, 2020	Mean: 219 (32.6)	53 (37-62)	43.5
El Otmani, 2022[129]	Morocco	118	Case-Control	Outpatient	February 2021 - April 2021	4-12 weeks	29 (21-54)	29
Helmsdal, 2022[130]	Faroe Islands	170	Cohort	Inpatient and Outpatient	March - April 2020	19-23 months	40.0 (19.4)	45.3
Messin, 2021[131]	France	74	Cohort	Inpatient	March 2020	4-7 months	52.3 (18)	40.5
Silverberg, 2022[132]	United States	372	Cohort	Outpatient	May 13, 2020 - July 6, 2020	9-11 months	42 (31-54)	52.1
Sadat Larijani, 2022[133]	Iran	254	Cohort	Inpatient and Outpatient	March-October 2021	4-24 weeks	41 (35-49)	53.9

Faycal, 2021[134]	France	429	Cohort	Outpatient	March 10, 2020 - May 18, 2020	30-60 days	41.6 (30-51.5)	27.5
Chudzik, 2022[135]	Poland	488	Cohort	Outpatient	2020-2022	3 months	45.7 (12.5)	37
Buonsenso, 2022[136]	Italy	155	Cohort	Inpatient and Outpatient	April 2020	12 months	46.5 (7.3)	49.7
de Oliveira, 2022[137]	Brazil	439	Cohort	Inpatient	July 1, 2020 - March 31, 2021	90-201 days	58 (47-67)	50.3
Yellumahanthi, 2022[138]	United States	53	Cohort	Outpatient	March 13, 2020 - March 12, 2021	3 months	NR	45.3
Hill, 2022[139]	United States	8325	Case-Control	Inpatient and Outpatient	March 1, 2020 - December 1, 2021	NR	LC: 52.3 (15.5) No LC: 46 (17.8)	37.2
Yoo, 2022[140]	United States	1038	Cohort	Inpatient and Outpatient	April 2020-February 2021	30-90 days	60 (37-83)	50.4
Chan Sui Ko, 2022[141]	France	316	Cohort	Inpatient	February 2, 2020 - December 28, 2020	4 months	64.1 (14.3)	59.2
Loosen, 2022[142]	Germany	50402	Cohort	Outpatient	March 2020 - March 2021	90-183 days	48.8 (19.3)	45.5
Spinicci, 2022[143]	Italy	428	Cohort	Outpatient	June 2020-June 2021	4-12 weeks	64 (54-76)	59
Garcia-Abellan, 2022[144]	Spain	72	Cohort	Inpatient	March 10 2020 - June 30, 2020	1-12 months	LC: 59.5 (53-71) No LC: 60 (52-71)	61.1
Caze, 2022[145]	Brazil	412	Cohort	Inpatient and Outpatient	September 2020 - April 2021	30-250 days	LC: 40 (32-51) No LC: 35 (28-47)	44.7
Khalaf, 2022[146]	Egypt	538	Cross-sectional	Inpatient and Outpatient	March-May 2020	3-7 months	41.2 (14.8)	54.1

Knight, 2022[147]	United States	435	Cohort	Inpatient and Outpatient	March 13, 2020 - August 17, 2020	1 week-9 months	54 (18-99)	39.8
Kim, 2022 (b)[148]	South Korea	170	Cohort	Inpatient and Outpatient	February 17, 2020 - March 24, 2020	6-12 months	51 (37-60)	40
Sneller, 2022[149]	United States	189	Cohort	Inpatient and Outpatient	NR	105-210 days	50 (38-58)	45
Zulu, 2022[150]	Zambia	302	Cohort	Outpatient	March-June 2020	2 months	32 (1-85)	42.4
Wu, 2022[151]	United States	308	Cohort	Inpatient and Outpatient	March 10 2020 - March 31, 2021	12 weeks	46	42.7
Zuschlag, 2022[152]	Germany	162	Cohort	Inpatient	January 1, 2020 - December 31, 2020	12 months	LC: 62.2 (16.9) No LC: 66.6 (17.2)	54.3
Varghese, 2022[153]	Germany	116	Cohort	Inpatient	NR	44-82 days	41 (31-55)	85
Senjam, 2021[154]	India	773	Cross sectional	Inpatient and Outpatient	January 1, 2021 - April 30, 2021	4 weeks-6 months	34 (27-44)	56.4
Zayet, 2021[155]	France	354	Cohort	Inpatient and Outpatient	March 1, 2020 - May 31, 2020	9-12 months	49.6 (18.7)	37
Bell, 2021[156]	United States	303	Cohort	Outpatient	Before February 24, 2021	3-12 months	43.6 (16.6)	30
Ong, 2021[157]	Singapore	183	Cohort	Inpatient	January 30, 2020 - August 14, 2020	30-360 days	44 (33-56)	75.4

Table 5.2. Quality assessment of included studies using the QUIPS tool

Name, Year	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting	Overall
Sudre, 2021	M	L	L	L	L	L	M
Augustin, 2021	M	M	L	L	L	L	M
Peghin, 2021	L	M	L	L	L	L	M
Naik, 2021	M	L	L	L	L	L	M
Asadi-Pooya, 2021	M	L	L	L	L	L	M
Subramania, 2022	M	L	L	L	L	L	M
Bai, 2022	M	M	L	L	L	L	M
Bliddal, 2021	M	M	L	L	L	L	M
Cervia, 2022	M	M	L	L	L	L	M
Vimercati, 2021	M	L	L	L	H	L	H
Kim, 2022	L	M	L	L	L	L	M
Becker, 2021	M	L	L	L	L	L	M
Nune, 2021	M	M	L	L	L	L	M
Desgranges, 2022	M	L	L	L	L	L	M
Hossain, 2021	M	L	L	L	L	L	M
Lui, 2021	M	L	L	L	L	L	M
Förster, 2022	L	M	L	L	L	L	M
El Otmani, 2022	M	L	L	L	H	L	H
Helmsdal, 2022	L	L	L	L	L	L	L
Messin, 2021	M	M	L	L	H	L	H
Silverberg, 2022	M	L	L	L	L	L	M
Larijani, 2022	L	L	L	L	L	L	L
Faycal, 2021	M	L	L	L	L	L	M
Chudzik, 2022	M	L	L	L	L	L	M
Buonsenso, 2022	L	L	L	L	L	L	L

de Oliveira, 2022	M	M	L	L	H	L	H
Yellumahanthi, 2022	M	M	L	L	H	L	H
Hill, 2022	L	L	L	L	L	L	L
Yoo, 2022	L	M	L	L	L	L	M
Chan Sui Ko, 2022	M	M	L	L	L	L	M
Loosen, 2022	M	L	L	L	L	L	M
Spinicci, 2022	M	L	L	L	L	L	M
Garcia-Abellan, 2022	M	M	L	L	L	L	M
Caze, 2022	L	M	L	L	L	L	M
Khalaf, 2022	L	L	L	L	L	L	L
Knight, 2022	L	L	L	L	H	L	H
Kim, 2022	L	M	L	L	H	L	H
Sneller, 2022	L	L	L	L	L	L	L
Zulu, 2022	M	M	L	L	H	L	H
Wu, 2022	L	L	L	L	L	L	L
Zuschlag, 2022	M	M	L	L	H	L	H
Varghese, 2022	M	M	L	L	H	L	H
Senjam, 2021	L	M	L	L	L	L	M
Zayet, 2021	L	L	L	L	H	L	H
Bell, 2021	M	M	L	L	H	L	H
Ong, 2021	M	M	L	L	L	L	M

Table 5.3. Association between categorical variables and long COVID-19

Risk Factor	Studies	Patients	RR (95% CI)
Demographics			
Age (Years)			
Age < 20	3	484	3.18 (0.66-15.43)
Age 15-34	9	390773	0.90 (0.56-1.45)
Age 18-49	4	3424	0.84 (0.68-1.05)
Age 30-40	8	418379	1.03 (0.92-1.15)
Age 40-50	7	417950	0.99 (0.96-1.01)
Age 50-60	6	30800	1.03 (0.89-1.19)
Age 50-69	4	2995	1.52 (0.68-3.37)
Age 60-70*	8	418473	1.33 (1.01-1.76)
Age > 50	3	736	1.01 (0.62-1.63)
Age ≥ 65	9	420594	1.25 (0.86-1.82)
Sex (Male)*	42	536941	0.82 (0.76-0.87)
Race/Ethnicity			
Black	6	465868	1.14 (0.72-1.79)
White*	6	465868	1.07 (1.06-1.08)
Asian	4	464522	0.91 (0.73-1.14)
Hispanic	5	51976	1.04 (0.85-1.27)
Other*	5	465605	0.88 (0.80-0.96)
BMI (kg/m^2)			
18.5-25*	4	415207	0.97 (0.95-0.99)
25-30	4	415207	1.01 (0.99-1.03)
≥ 30	4	1578	1.12 (0.67-1.89)
Comorbidities			
Liver disease*	3	2228	5.55 (2.22-13.84)
Arthritis*	4	2798	1.75 (1.48-2.07)
Autoimmune condition*	6	416112	1.60 (1.32-1.95)
Mental health problem*	5	466468	1.59 (1.38-1.84)
Chronic lung disease*	17	525064	1.54 (1.21-1.97)
Asthma*	11	468776	1.50 (1.27-1.78)
Kidney disease*	14	60246	1.43 (1.12-1.83)
Obesity*	11	105610	1.28 (1.11-1.49)
Diabetes*	23	113971	1.24 (1.07-1.43)
Hypertension*	22	110825	1.22 (1.10-1.36)
Former smoker*	4	415054	1.15 (1.13-1.16)
Any Comorbidity*	13	9202	1.15 (1.05-1.26)
Cancer	9	53391	1.15 (0.91-1.46)

Heart disease	21	112521	1.11 (0.88-1.39)
Neurologic disease	3	5634	1.05 (0.92-1.19)
Smoker	20	421271	0.97 (0.84-1.12)
Drinks alcohol	4	1598	0.90 (0.80-1.01)
HIV	3	780	0.29 (0.04-2.06)
Symptoms			
Hearing loss*	4	414980	1.57 (1.16-2.13)
Back pain*	3	415953	1.44 (1.15-1.81)
GI problems	3	420146	1.44 (0.99-2.10)
Skin condition	3	1055	1.29 (0.75-2.24)
Hospitalized*	10	8036	1.28 (1.08-1.52)
Allergic rhinitis*	5	1312	1.22 (1.02-1.46)
Sore throat	5	1660	1.18 (0.89-1.56)
Dyspnea	8	2710	1.09 (0.81-1.47)
Loss of taste	6	2195	1.02 (0.73-1.42)
Loss of smell	8	2880	1.01 (0.80-1.27)
Headache	9	416886	0.98 (0.64-1.49)
Respiratory problems	4	6076	0.97 (0.81-1.17)
Fatigue	8	2682	0.96 (0.68-1.36)
Cough	9	3063	0.92 (0.75-1.13)
Chest pain	4	1351	0.84 (0.44-1.61)
Fever	9	3063	0.83 (0.46-1.48)
Myalgia	8	2575	0.79 (0.44-1.40)
Rhinorrhea	5	1406	0.71 (0.25-1.98)
Abdominal Pain	4	1233	0.67 (0.25-1.82)
Nausea/Vomiting	6	2098	0.54 (0.22-1.30)
Severity			
Asymptomatic	6	2370	0.27 (0.16-0.45)
Mild	10	7018	0.93 (0.85-1.02)
Moderate*	10	7018	1.39 (1.17-1.65)
Severe	9	5784	1.37 (0.91-2.06)
Critical*	5	1597	3.97 (2.14-7.35)

*indicates risk factors significantly associated with long COVID

Table 5.4. Association between continuous variables and long COVID-19

Risk Factor	Studies	Patients	WMD (95% CI)
Demographics			
Age	30	69237	1.93 (0.43-3.43)
BMI	12	6231	0.81 (0.34-1.27)
Labs			
LDH (U/L)*	3	852	20.33 (2.90-37.77)
Platelet ($10^3/mm^3$)	3	932	6.26 (-4.24-16.76)
IL-6 (pg/mL)	3	322	1.28 (-0.74-3.30)
D-dimer (mg/L)	3	888	0.06 (-0.07-0.18)
Leukocytes ($10^9/L$)	6	1582	0.01 (-0.30-0.33)
Creatinine (mg/dl)	4	1286	0.0 (-0.03-0.03)
CRP (mg/L)	8	1970	0.0 (-0.02-0.02)
Hepatic transaminases, ALT	3	932	-0.45 (-1.64, 0.73)
Hemoglobin (g/dl)	4	1286	-0.31 (-0.68-0.06)
Lymphocytes ($10^9/L$)	5	1238	-0.07 (-0.20-0.05)
Symptoms			
Length of hospital stay (days)*	5	5659	1.66 (1.28-2.04)
Number of symptoms	6	3706	1.15 (0.20-2.11)

WMD= Weighted Mean Difference

*indicates risk factors significantly associated with long COVID

Table. 5.5. Assessment of publication bias using Egger's test for risk factors reported in 10 or more studies

Risk Factor	Intercept (95% CI)	p-value
Categorical Risk Factors		
Any comorbidity	0.83 (-0.42-2.09)	0.22
Asthma	0.40 (-0.67-1.47)	0.48
Chronic lung disease	-1.75 (-3.63-0.13)	0.09
Diabetes*	-1.08 (-1.95- -0.2)	0.03
Heart disease	0.83 (-0.42-2.09)	0.22
Hospitalized	0.59 (-0.8-1.97)	0.43
Hypertension*	-1.40 (-2.04- -0.76)	0.00
Kidney disease*	-1.40 (-2.04- -0.76)	0.00
Obesity	-0.80 (-2.13-0.54)	0.27
Severity (Mild)	0.67 (-1.5-2.84)	0.56
Severity (Moderate)	-0.65 (-2.54-1.24)	0.52
Sex (Male)*	1.22 (0.08-2.37)	0.04
Smoker	-0.34 (-1.11-0.44)	0.41
Continuous Risk Factors		
Age	-1.63 (-3.26-0)	0.06
BMI	1.42 (0.19-2.65)	0.05

*Risk factors indicating potential publication bias

CHAPTER 6

Risk Factors of Long COVID-19 in College Students

Introduction

On college campuses, initial transmission of COVID-19 was minimal due to campus closures and the implementation of virtual learning during the spring and summer of 2020. By Fall of 2020, some schools began to allow students to come back to campus. This resulted in a large increase in COVID-19 transmission on many campuses. This was likely due to student gatherings and their beliefs and behaviors related to social distancing measures upon initial return.[80] On campus, universities also provided many resources for students to prevent COVID-19 illness. By the Spring of 2021, most universities provided free and accessible COVID-19 testing, access to vaccinations, and implementation of social distancing measures in all facilities. Despite these mitigation efforts, students continue to experience COVID-19 illness and some adverse events as a result.

Research has shown that, as in other populations, some college students are reporting prolonged symptoms as a result of COVID-19 infection.[9,82] Prolonged symptoms commonly include chest pain, fatigue, fever, loss of taste and smell, and headaches, and the syndrome is called “long COVID” or “post-acute sequelae of COVID-19” or PASC.[82] Two studies in college students have demonstrated long COVID existing in a range of 11 to 50 percent of students. [9,82] This wide range of prevalence is likely due to differing definition of long COVID. Research on this topic regarding just young adults, is very limited. This age group is not thought to be at increased risk for adverse outcomes such as severe disease, death, or prolonged symptoms.

Further research is needed to understand why college students are experiencing prolonged symptoms and who is most at risk. Lastly, focusing on students on the same college campus will allow us to compare outcomes and experiences of students with COVID-19 living in similar environments and with access to the same resources. As a result of this research, our goal is to identify risk factors for long COVID among college students at the University of Georgia (UGA).

Methods

Setting and Population

UGA is the nation's first land grant institution with 40,118 students enrolled as of 2022.[97] Throughout the COVID-19 pandemic, COVID-19 testing for both symptomatic and asymptomatic students has been provided on campus. Enrolled students with COVID-19 symptoms could make appointments at the University Health Center (UHC) and receive a PCR COVID-19 test via nasal swab. For those students who are asymptomatic, walk-in testing was provided via saliva sample at UHC as well as pop-up testing sites throughout campus. Promotions for asymptomatic testing before and after scheduled campus breaks were also frequently made to promote students getting tested before and after travel. Public testing sites as well as private were also available throughout the Athens, Georgia community. This population is ideal for our study as it captures a population of young adults that have easy access to COVID-19 testing regardless of symptom status. This will allow for us to capture students with prior COVID-19 diagnoses and assess their risk of long COVID.

Study Design

We conducted a cross-sectional study in students at UGA with previously confirmed COVID-19 infection. This study was approved by the University of Georgia Internal Review Board (IRB). Our outcome, long COVID-19, was defined as a patient experiencing at least one symptom that they did not have prior to COVID-19 diagnosis and lasted for four weeks or longer. This is the same definition for long COVID given by the CDC.[4] To be included in the study, participants had to be between 18 and 30 years old and have a previous COVID-19 diagnosis, either via real time reverse transcription chain reaction (RT-PCR) or rapid diagnostic test (Ag-RDT). To minimize potential for recall bias, due to the length of time since disease diagnosis, we limited patient inclusion to those with a positive COVID-19 test since August 1, 2021. This date was also chosen to align with more recent case surges as a result of COVID-19 variants. By August 2021 more than 98 percent of cases in the United States were a result of the Delta variant.[98] This was then followed by another surge of cases caused by the Omicron variant peaking in January 2022.

Data Collection

To capture information on student risk factors and experience with COVID-19, we generated a survey using Qualtrics, a secure subscription-based service provided by UGA free of charge. Protocols, consent forms, and our survey were approved by the UGA IRB before the start of data collection. Participant consent letters and the survey are included in Appendix B. Aside from our outcome data, our survey aimed to capture risk factors participants may have for long COVID, including demographics, preexisting medical conditions, vaccination status, symptoms, and the severity of their illness. We were also interested in the impacts of COVID-19 variant on

development of prolonged symptoms. We chose to include a variable distinguishing the date of positive COVID-19 test before and after December 1, 2021. This date was chosen to align with the date that the Omicron variant became the dominant strain in the United States. [20]

We contacted students by obtaining a list of email addresses for all enrolled UGA students as of the Fall 2022 semester. Included in the email was a description of our study, inclusion criteria, consent information, and a link to the survey. Students answered questions to confirm that they met our inclusion criteria and that they agree to complete the survey. Only those that met inclusion criteria were able to complete the rest of the survey. Those who did not were excluded from any further analysis.

Analysis

All analysis was done using R statistical software. Exploratory data analysis was done first, which included examination of the data set and any needed data cleaning. Then an item analysis was done for each variable individually to assess for missing data and outliers. Covariates that are included in the model were derived from our survey responses. The outcome of the final model is the presence or absence of long COVID-19. Correlation analysis was done to address any potential collinearity. Those variables with high correlation coefficients were considered for combination or exclusion. All variables included in the analysis, following correlation analysis, were moved to a final data set before generating our model. Continuous variables are presented as median and range and categorical variables are presented as a frequency and percentage. Our dataset had minimal missing data as students were not able to proceed in the survey without at least selecting 'prefer not to answer'. This allowed for us to do a complete-case analysis of the observations obtained. Bivariate and multivariate logistic regression was used to analyze risk

factors associated with long COVID-19. P-values less than 0.05 were deemed statistically significant.

Finally, we built a multivariate logistic regression model using the least absolute shrinkage and selection operator technique (Lasso). We chose lasso for our variable selection method as it tends to be more sensitive and more precise than traditional stepwise selection methods. Lasso aims to identify variables and regression coefficients that limit prediction error. This is done by imposing a constraint on model parameters that shrinks regression coefficients towards zero to exclude from the final model. Lasso is generally preferred over other penalized regression methods as it is able to produce a simpler model with fewer predictors.[100]

Results

Study population

A flow diagram of student participation and the number with and without long COVID-19 is included in Figure 6.1. The initial student participation email was sent on September 12, 2022. One reminder email was sent to students a week later and the survey remained open until October 20, 2022. Emailed invitations to participate were sent to 40,056 students. Overall, 3,103 students began the survey and 2,936 consented to participation. Among those who consented to participation, 953 did not meet inclusion either due to age or not having a positive COVID-19 test since August 1, 2021. In total, 1,983 UGA students completed the survey. After reviewing the responses of the surveys 230 (11.6%) students had long COVID while 1,753 did not.

Participant Characteristics

Characteristics of participating students including demographics, year in school, vaccination status, and comorbidities are included in Table 6.1. The overall UGA student population is about 59 percent female, 66 percent white race, with an average age of 22 years old, as of the Fall 2021 semester. [97] The average age of our study population was 21 years old with 73.8 percent being female and 72.7 percent being of white race. Graduate or professional students comprised 28.5 percent of the study population which is similar to the overall UGA population of 25 percent. UGA also reports having 26 percent of students reported as seniors, however in our study we had only 19 percent, classified as year 4 and year 5 students. The distribution of the remaining students was similar to the overall UGA student population. Half of our study population reported having three or more COVID-19 vaccines. Students most commonly reported the comorbidities of anxiety, depression, and asthma.

Symptom duration and severity

COVID-19 symptom information is included in Table 6.2. Students with long COVID-19 reported experiencing symptoms for a median of 34.5 days compared to those without long COVID reporting symptoms for a median of 8 days. Among UGA students, 36.5 percent reported experiencing symptoms for 30 to 60 days, 7.8 percent for 60 to 90 days, 8.9 percent for 90 to 180 days, and 2.2 percent for more than 180 days.

The most commonly reported symptoms across all students included runny nose or congestion (76.4%), sore throat (71.1%), and muscle or body aches (62.9%). Participants ranked the severity of their symptoms on a scale of 0 to 10, with 0 as not having the symptom and 10 being very severe. Fatigue was reported as the most severe symptom overall with a median of 8

and 6 among each group respectively. Overall, students with long COVID-19 reported more severe symptoms than those without long COVID (Median: 5.33 vs 4.67, p=<0.01). This was determined by averaging the three most severe symptoms as reported by students.

Bivariate and Multivariate Analysis

The results of the bivariate analysis are included in Table 6.3. Variables that were experienced in less than 1 percent of our study population were excluded from further analysis. This was done to remove variables that were generally not of interest or related to our study population and to simplify our final model. There were no significant independent associations for most demographic risk factors and long COVID-19. Having three or more doses of a COVID-19 vaccine was significantly associated with not having long COVID (Odds ratio [OR] 0.55, 95% Confidence Interval [CI]: 0.34-0.92). Testing positive for COVID-19 prior to December 1, 2021 was also significantly association with the outcome (OR 5.17 95% CI: 3.76-7.08). Anxiety or other mental health conditions were also significantly associated with the development of long COVID (OR 1.40, 95% CI: 1.03-1.88). Having no comorbidities was associated with a lower likelihood of reporting long COVID (OR 0.63, 95% CI: 0.48-0.84). Twelve different symptoms were significantly associated with an increased likelihood of reporting long COVID including loss of taste or smell, new anxiety/depression, shortness of breath, and abdominal pain.

Our final logistic regression model using the Lasso technique is included in Table 6.4. Seven variables remained in the final model, which included primarily symptom risk factors. The Lasso technique did initially include being a year 3 student (adjusted odds ratio [aOR] 0.98) and sore throat (aOR 0.99) in the model. However, we chose to remove these variables from the final model as they both indicated a reduced risk of long COVID and were very close to the null

value, likely not aiding in predicting individuals with the outcome. The remaining variables were all indicative of long COVID. These included symptoms muscle or body aches, loss of taste or smell, shortness of breath, diarrhea, brain fog, and new anxiety or depression, and testing positive for COVID-19 before December 1, 2021. Date positive (aOR 4.06) and loss of taste or smell (aOR 2.16) had the strongest association with reporting long COVID.

Discussion

Overall, 11.6 percent of our study population was identified as having long COVID-19. Date positive before December 2021 as well as six different symptoms experienced during infection were included in our final logistic regression model. These include the symptoms muscle or body aches, loss of taste or smell, shortness of breath, diarrhea, brain fog or difficulty concentrating, and new anxiety or depression. The presence of all of these risk factors were associated with an increased risk of developing long COVID. This logistic regression could help to identify individuals at risk for long COVID-19.

Comparison to current literature

There are some similarities and differences in our final model when compared to results of similar literature that are important to highlight. The largest difference between our study and similar studies identifying long COVID-19 risk factors are the age of the study population. While our study focused on young adults (median: 21 years old), many similar studies include all adults with the average age around 50 to 60 years old.[84,114,120,140] As shown in this area of research, there is a known association of increased age with long COVID-19, particularly ages 50 to 59 years.[84,120,158] Older individuals likely experience more medical complications and are generally

not in as good of health as young adults. Due to this difference in age our study likely did not find some of the same associations with the outcome that these studies that focused on all adults did.

The difference in the average age of our study and other comparable studies can be seen in the proportion of participants with long COVID-19. In our study we found that about 12 percent of the study population was experiencing long COVID. This is slightly lower than other studies, with proportions ranging from about 13 percent to 63 percent, that included all adult ages. [84,114,120,140] However, in another study done in a student population at UGA, 11.4 percent of the study population reported prolonged symptoms.[9] While the previous UGA study was not focused on prolonged symptoms it did find a proportion of long COVID that was comparable to our results. When comparing these results to studies with older participants we would expect a lower percent of long COVID due to younger individuals having generally better health. However, the proportion of young adults at UGA with long COVID does indicate that prolonged symptoms of COVID-19 are a problem among these young adults.

The results of our study also showed similarities and difference in terms of risk factors that were associated with long COVID. Overall, most studies found associations with some demographic and symptom associated risk factors. Age, female sex, and BMI were common demographic risk factors found to be associated with long COVID in other studies.[84,114,119,120,140] We did not have any associations with demographic risk factors. These associations could be potentially confounded by age. Older adults with higher BMI might be likely to develop comorbidities and other health conditions that put them at risk for long COVID-19 that younger adults have not yet developed or been diagnosed with. While we did not find significant associations with these risk factors we did find associations with a number of symptom

associated risk factors. Several studies found associations with the number of symptoms during infection.[84,114] While this risk factor was not included in our final model, we did find an association with long COVID-19 and number of symptoms during infection. Many studies also reported their most commonly reported symptoms which included brain fog, shortness of breath, and muscle or body aches.[119,129,152] These were commonly reported among our student population and were found to have a statistically significant association with our outcome. Like comparable studies, we did not find many associations with any reported comorbidities. We assumed this was due to the age of our study population. However, when compared to similar studies many also did not find significant associations with any comorbidities.[84,114,119,120]

Strengths and limitations

Our study did have some limitations. First, all data included for these analyses were self-reported data. With the idea of prolonged symptoms of COVID-19 in the news, individuals might be more likely to report prolonged symptoms. Also, this type of data may not be as accurate as prospectively collected data or data from electronic health records (EHR). In future research this would be better managed by using EHR data from physician visits to accurately capture patient symptoms after COVID-19 infection.

Second, due to the design of this study there is likely an inherent issue with recall bias. Our survey asked participants to recall dates of COVID-19 tests, dates symptoms ended, and estimate how long and severe each symptom was. The longer since the participant had COVID-19, the harder this information is to remember. Those individuals with more severe COVID-19 or who had unusual symptoms might be more likely to recall this information rather than those with more mild and simple cases. Similar to related literature we also found an increased risk for

long COVID-19 in women when compared to men. Related research has not found a biological reason for this association but it is speculated that women are generally more likely to visit a provider for their prolonged symptoms and are generally more in-tune with their health. [118] If this is true then women are more likely than men to recall their prolonged illness and may lend to an inflated association of gender with long COVID-19.

Lastly, there were also likely issues with selection bias making our study population not representative of 18 to 30 year old adults as a whole. This sample included just UGA students who were living in similar conditions with the same access to COVID-19 testing and treatment. It is likely these results are more representative of college students rather than the age group as a whole. Also, students who either had a more severe course of COVID-19 or noticeable prolonged symptoms may have been more likely to respond to our survey than those students who had minimal to no complications related to their COVID-19 illness. We tried to minimize these biases by not stating in recruitment letters that our research was focused on prolonged symptoms, however those students who were more impacted by COVID-19 were still more likely to respond to our survey. It is important to make this consideration before generalizing these results to larger populations.

Our study was able to begin to identify risk factors that might indicate increased risk of long COVID in young adults. We identified symptoms associated with the outcome that have also been reported in similar literature, lending to the evidence of their association. We also identified a significant proportion of our study population with long COVID-19. This indicates a potential problem with this outcome in this age group that could lead to future healthcare implications and considerations.

Conclusions

Overall, our results indicate that long COVID-19 is a fairly common problem among college students. About 11.6 percent of our study population experienced prolonged symptoms after COVID-19 diagnosis. We found the strongest associations for loss of taste or smell and date of positive test with a higher likelihood of developing long COVID-19. These risk factors could help to identify college students at risk. Long COVID is an outcome of COVID-19 that could put future strain on the healthcare system. Identifying the causes of this phenomenon is vital for future treatment planning and minimization of impacts in this age group and beyond. Larger prospective studies focused on this age group will be important to identifying those at risk for development of long COVID-19 and allow for generalization of future findings to larger populations.

TABLE AND FIGURES

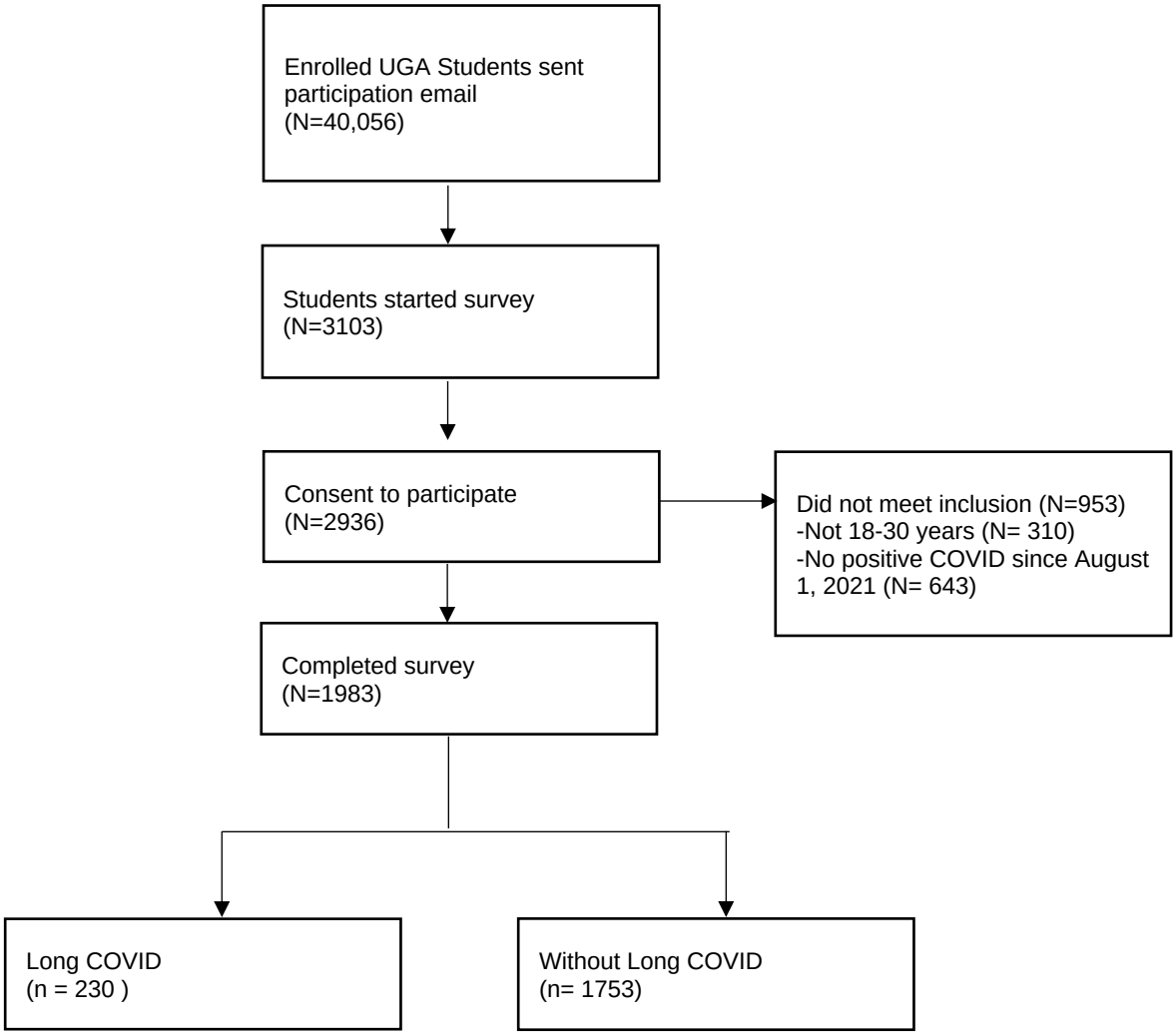

Figure 6.1. Flow diagram of student participation

Table 6.1. Demographics of study participants

	Total (N=1983)	Long COVID (N=230)	No Long COVID (N=1753)
Age (years)			
Median [Min,Max]	21.0 [18.0,30.0]	21.0 [18.0,30.0]	21.0 [18.0,30.0]
Gender			
Male	466 (23.5%)	54 (23.5%)	412 (23.5%)
Female	1464 (73.8%)	167 (72.6%)	1297 (74.0%)
Non-binary/third gender	41 (2.1%)	8 (3.5%)	33 (1.9%)
Prefer not to answer	8 (0.4%)	1 (0.4%)	7 (0.4%)
Race			
Alaska Native/American Indian/Native Hawaiian	6 (0.3%)	1 (0.4%)	5 (0.3%)
Asian/Pacific Islander	341 (17.2%)	36 (15.7%)	305 (17.4%)
Black/African American	126 (6.4%)	9 (3.9%)	117 (6.7%)
White/Caucasian	1442 (72.7%)	175 (76.1%)	1267 (72.3%)
Other	49 (2.5%)	7 (3.0%)	42 (2.4%)
Prefer not to answer	15 (0.8%)	2 (0.9%)	13 (0.7%)
Ethnicity			
Hispanic	148 (7.5%)	23 (10.0%)	125 (7.1%)
Non-Hispanic	1799 (90.7%)	206 (89.6%)	1593 (90.9%)
Prefer not to answer	32 (1.6%)	1 (0.4%)	31 (1.8%)
Year in School			
Year 1	405 (20.4%)	56 (24.3%)	349 (19.9%)
Year 2	306 (15.4%)	37 (16.1%)	269 (15.3%)
Year 3	319 (16.1%)	26 (11.3%)	293 (16.7%)
Year 4	309 (15.6%)	30 (13.0%)	279 (15.9%)
Year 5	67 (3.4%)	10 (4.3%)	57 (3.3%)
Graduate/Professional	565 (28.5%)	69 (30.0%)	496 (28.3%)
Other	5 (0.3%)	2 (0.9%)	3 (0.2%)

	Total (N=1983)	Long COVID (N=230)	No Long COVID (N=1753)
Prefer not to answer	3 (0.2%)	0 (0%)	3 (0.2%)
BMI (kg/m^2)			
Mean (SD)	23.5 (5.00)	23.8 (5.57)	23.5 (4.92)
Vaccination Status			
Not vaccinated	131 (6.6%)	23 (10.0%)	108 (6.2%)
1 dose	78 (3.9%)	14 (6.1%)	64 (3.7%)
2 doses	689 (34.7%)	85 (37.0%)	604 (34.5%)
3 or more doses	999 (50.4%)	105 (45.7%)	894 (51.0%)
Prefer not to answer	15 (0.8%)	3 (1.3%)	12 (0.7%)
Date of Positive Test			
Before 12/1/21	245 (12.4%)	81 (53.0%)	164 (9.4%)
Number of comorbidities			
Median [Min, Max]	1.00 [1.00,5.00]	1.00 [1.00-4.00]	1.00 [1.00, 5,00]
Comorbidities			
Diabetes	8 (0.4%)	2 (0.9%)	6 (0.3%)
Heart disease	4 (0.2%)	0 (0%)	4 (0.2%)
Heart attack	1 (0.1%)	1 (0.4%)	0 (0%)
Congestive heart failure	1 (0.1%)	0 (0%)	1 (0.1%)
Hypertension	15 (0.8%)	2 (0.9%)	13 (0.7%)
High cholesterol	28 (1.4%)	5 (2.2%)	23 (1.3%)
History of stroke	0 (0%)	0 (0%)	0 (0%)
Autoimmune disorder	34 (1.7%)	7 (3.0%)	27 (1.5%)
HIV	1 (0.1%)	0 (0%)	1 (0.1%)
Hepatitis C	0 (0%)	0 (0%)	0 (0%)
Asthma	208 (10.5%)	29 (12.6%)	179 (10.2%)
Chronic lung disease	1 (0.1%)	1 (0.4%)	0 (0%)
Cancer	1 (0.1%)	0 (0%)	1 (0.1%)
Depression	282 (14.2%)	42 (18.3%)	240 (13.7%)
Pregnant	2 (0.1%)	0 (0%)	2 (0.1%)

	Total (N=1983)	Long COVID (N=230)	No Long COVID (N=1753)
Obesity	62 (3.1%)	11 (4.8%)	51 (2.9%)
Anxiety or other mental health conditions	475 (24.0%)	70 (30.4%)	405 (23.1%)
ADHD	12 (0.6%)	3 (1.3%)	9 (0.5%)
Chronic migraine	10 (0.5%)	2 (0.9%)	8 (0.5%)
Seizure disorder	4 (0.2%)	1 (0.4%)	3 (0.2%)
Other	68 (3.4%)	7 (3.0%)	61 (3.5%)
No comorbidities	1146 (57.8%)	115 (50.0%)	1031 (58.8%)
Prefer not to answer	17 (0.9%)	1 (0.4%)	16 (0.9%)

Table 6.2. Description of COVID-19 symptoms during acute stage of infection

	Total (N=1983)	Long COVID (N=230)	No Long COVID (N=1753)
Symptom duration (days)			
Median [Min,Max]	9.00 [0,397]	34.5 [29.0,397]	8.00 [0,28.0]
Number of symptoms			
Median [Min,Max]	7.00 [1.00, 16.0]	8.00 [1.00, 16.0]	6.00 [1.00,16.0]
Symptoms			
Fever	1152 (58.1%)	167 (72.6%)	985 (56.2%)
Chills/Sweats	1043 (52.6%)	139 (60.4%)	904 (51.6%)
Muscle or Body Aches	1247 (62.9%)	179 (77.8%)	1068 (60.9%)
Loss of Taste or Smell	464 (23.4%)	113 (49.1%)	351 (20.0%)
Runny Nose/Congestion	1516 (76.4%)	184 (80.0%)	1332 (76.0%)
Sore Throat	1409 (71.1%)	159 (69.1%)	1250 (71.3%)
New or Worsening Cough	1034 (52.1%)	131 (57.0%)	903 (51.5%)
Shortness of breath or Difficulty Breathing	534 (26.9%)	108 (47.0%)	426 (24.3%)
Chest Pain	249 (12.6%)	49 (21.3%)	200 (11.4%)
Fatigue or Tiredness	1503 (75.8%)	192 (83.5%)	1311 (74.8%)
Nausea or Vomiting	220 (11.1%)	35 (15.2%)	185 (10.6%)
Headache	1159 (58.4%)	158 (68.7%)	1001 (57.1%)
Abdominal Pain	90 (4.5%)	21 (9.1%)	69 (3.9%)
Diarrhea	229 (11.5%)	46 (20.0%)	183 (10.4%)
Brain Fog/Difficulty Concentrating	692 (34.9%)	120 (52.2%)	572 (32.6%)
New Anxiety/Depression	119 (6.0%)	35 (15.2%)	84 (4.8%)
Other	20 (1.0%)	5 (2.2%)	15 (0.9%)
Prefer not to answer	1 (0.1%)	0 (0%)	1 (0.1%)
No symptoms	50 (2.5%)	0 (0%)	50 (2.9%)
Sneezing	4 (0.2%)	0 (0%)	4 (0.2%)
Skin condition	8 (0.4%)	1 (0.4%)	7 (0.4%)

	Total (N=1983)	Long COVID (N=230)	No Long COVID (N=1753)
Loss of appetite	2 (0.1%)	1 (0.4%)	1 (0.1%)
Dizziness/Fainting	6 (0.3%)	1 (0.4%)	5 (0.3%)
Symptom severity, Median [Min, Max]			
Overall average	4.67 [0,10.0]	5.33 [0,10.0]	4.67 [0,10.0]
Fever	3.00 [0,10.0]	4.00 [0,10.0]	2.00 [0,10.0]
Chills/Sweats	1.00 [0,10.0]	3.00 [0,10.0]	1.00 [0,10.0]
Muscle or Body Aches	3.00 [0,10.0]	6.00 [0,10.0]	3.00 [0,10.0]
Loss of Taste or Smell	0 [0,10.0]	0 [0,10.0]	0 [0,10.0]
Runny Nose/Congestion	5.00 [0,10.0]	5.00 [0,10.0]	5.00 [0,10.0]
Sore Throat	4.00 [0,10.0]	5.00 [0,10.0]	4.00 [0,10.0]
New or Worsening Cough	1.00 [0,10.0]	4.00 [0,10.0]	1.00 [0,10.0]
Shortness of breath or Difficulty Breathing	0 [0,10.0]	0 [0,10.0]	0 [0,10.0]
Chest Pain	0 [0,10.0]	0 [0,10.0]	0 [0,10.0]
Fatigue or Tiredness	6.00 [0,10.0]	8.00 [0,10.0]	6.00 [0,10.0]
Nausea or Vomiting	0 [0,10.0]	0 [0,10.0]	0 [0,10.0]
Headache	3.00 [0,10.0]	5.00 [0,10.0]	2.00 [0,10.0]
Abdominal Pain	0 [0,10.0]	0 [0,8.00]	0 [0,10.0]
Diarrhea	0 [0,10.0]	0 [0,10.0]	0 [0,10.0]
Brain Fog/Difficulty Concentrating	0 [0,10.0]	2.50 [0,10.0]	0 [0,10.0]
New Anxiety/Depression	0 [0,10.0]	0 [0,10.0]	0 [0,10.0]
Other	0 [0,10.0]	0 [0,10.0]	0 [0,10.0]
Skin condition	0 [0,10.0]	0 [0,10.0]	0 [0,10.0]
Loss of appetite	0 [0,10.0]	0 [0,9.00]	0 [0,10.0]
Dizziness/Fainting	0 [0,10.0]	0 [0,5.00]	0 [0,10.0]

Table 6.3. Bivariate Analysis

	OR (95% CI)	p-value
Age (years)	0.99 (0.95-1.04)	0.81
Gender		
Male	1.02 (0.73-1.40)	0.73
Female	1.00 [Reference]	
Non-binary/third gender	1.88 (0.80-3.95)	0.80
Race		
Alaska Native/American Indian/Native Hawaiian	1.69 (0.09-10.90)	0.64
Asian/Pacific Islander	1.00 [Reference]	
Black/African American	0.65 (0.29-1.34)	0.27
White/Caucasian	1.17 (0.81-1.73)	0.42
Other	1.41 (0.55-3.21)	0.44
Ethnicity		
Hispanic	1.42 (0.87-2.23)	0.14
Non-Hispanic	1.00 [Reference]	
Prefer not to answer	0.25 (0.01-1.17)	0.09
Year in School		
Year 1	1.15 (0.79-1.68)	0.46
Year 2	0.99 (0.64-1.50)	0.96
Year 3	0.64 (0.40-1.02)	0.06
Year 4	0.77 (0.48-1.20)	0.27
Year 5	1.26 (0.58-2.48)	0.52
Graduate/Professional	1.00 [Reference]	
BMI	1.01 (0.98-1.04)	0.34
Vaccination Status		
Not vaccinated	1.00 [Reference]	
1 dose	1.03 (0.48-2.12)	0.94
2 doses	0.66 (0.40-1.11)	0.11

	OR (95% CI)	p-value
3 or more doses	0.55 (0.34-0.92)	0.02
Date of Positive Test		
Before 12/1/21	5.17 (3.76-7.08)	<0.01
Number of comorbidities	1.19 (0.96-1.46)	0.09
Comorbidities		
High cholesterol	1.61 (0.54-3.96)	0.34
Autoimmune disorder	1.94 (0.77-4.27)	0.12
Asthma	1.22 (0.79-1.83)	0.35
Depression	1.36 (0.93-1.93)	0.10
Obesity	1.62 (0.79-3.04)	0.16
Anxiety or other mental health conditions	1.40 (1.03-1.88)	0.03
Other	0.84 (0.35-1.74)	0.67
None	0.63 (0.48-0.84)	<0.01
Number of symptoms	1.20 (1.14-1.26)	<0.01
Symptoms		
Overall symptom severity	1.86 (1.64-2.14)	<0.01
Fever	1.96 (1.45-2.67)	<0.01
Chills/Sweats	1.37 (1.03-1.81)	0.03
Muscle or Body Aches	2.11 (1.54-2.96)	<0.01
Loss of Taste or Smell	3.74 (2.82-4.98)	<0.01
Runny Nose/Congestion	1.14 (0.82-1.63)	0.44
Sore Throat	0.83 (0.62-1.12)	0.22
New or Worsening Cough	1.26 (0.95-1.68)	0.23
Shortness of Breath or Difficulty Breathing	2.67 (2.02-3.54)	<0.01
Chest Pain	2.05 (1.43-2.88)	<0.01
Fatigue or Tiredness	1.55 (1.09-2.26)	0.02
Nausea or Vomiting	1.48 (0.99-2.17)	0.05

	OR (95% CI)	p-value
Headache	1.56 (1.17-2.10)	<0.01
Abdominal Pain	2.39 (1.41-3.92)	<0.01
Diarrhea	2.09(1.45-2.97)	<0.01
Brain Fog/Difficulty Concentrating	2.18 (1.65-2.87)	<0.01
New Anxiety/Depression	3.48 (2.26-5.26)	<0.01

Table 6.4 Final multivariate logistic regression model using lasso technique for risk factors of long COVID-19

Variable	B coefficient	aOR
Muscle or body aches	0.20	1.22
Loss of taste or smell	0.77	2.16
Shortness of breath	0.68	1.97
Diarrhea	0.52	1.68
Brain fog/Difficulty concentrating	0.37	1.45
New anxiety/Depression	0.28	1.32
Date Positive: Before 12/1/21	1.40	4.06

CHAPTER 7

Development and Validation of a Clinical Prediction Rule for Predicting Long COVID-19 in Young Adults

Introduction

While treating patients, clinicians typically make multiple decisions regarding the diagnosis, prognosis and treatment plan for disease. Clinical prediction rules (CPR) are tools derived from original research that attempt to reduce the uncertainty of medical decision making by standardizing the collection and interpretation of clinical data.[101] They can be used to support decision-making around diagnosis, the need for hospital referral, and treatment. This is particularly important for evaluating patients for long term effects of COVID-19.

Several clinical prediction rules have been developed to determine the risk of long COVID-19, however many use machine learning approaches with models that have a high data collection burden and are complex.[102,103] These types of models are useful at a large scale, for example in predicting impacts of long COVID-19 on the healthcare system or gaining more knowledge on the outcome while definitions are unclear. However, these complex prediction models cannot be easily memorized or implemented in a healthcare setting for a physician to determine a treatment plan for an individual patient at risk for long COVID-19. Simple prediction rules have also been developed, including predictors such as age, sex, and number of symptoms during acute infection, but these do not necessarily take into account the patients pre-existing medical conditions.[104] Lastly, all existing CPR for long COVID-19 include all age groups. These rules have generally found an association with middle to older age and long

COVID-19. While it is commonly true that older individuals experience more complications related to COVID-19, young adults are also experiencing these complications. Young adults in particular are reporting prolonged symptoms of COVID-19 that are lasting for as long as one year after infection.[9]

We have assembled a dataset of young adults that have been diagnosed with COVID-19 and have information on the progression of their illness following initial infection. Using this dataset, our primary goal is to develop a CPR that is simple and easy for a physician to implement to predict the likelihood of a young adult patient presenting with acute COVID-19 to develop long COVID-19.

Methods

Study Population

To develop our CPR, we used data from the National COVID Cohort Collaborative (N3C).[105] This registry dataset includes an enclave of data from participating institutions electronic health records (EHR). To date, there are about 18 million COVID-19 positive patients across 77 healthcare sites included. N3C includes individuals both with lab confirmed COVID-19 and those with related symptoms. Data on the medical history, comorbid diagnoses, demographics, immunizations, and laboratory results are also included. Patient records are also linked across sources to remove duplicates and link patient visits. N3C includes individuals with confirmed COVID-19 as well as controls without. For our study, only young adults between the ages of 18 and 30 years old with lab-confirmed COVID-19 were included.

Identification of risk factors and outcome

Our outcome of interest is long COVID-19. There is currently no standard for defining long COVID-19 and International Classification of Diseases, Tenth Revision (ICD-10) codes in the EHR are not widely used. We chose to use the definition set by the Centers for Disease Control (CDC) for long COVID-19. This definition classifies the outcome as a range of new, returning, or ongoing health problems people can experience four or more weeks after being infected with COVID-19.[4] From N3C, individuals with lab-confirmed COVID-19 visiting a provider more than 4 weeks after their diagnosis and reporting at least one related symptom were classified as having long COVID-19. These symptoms included abdominal pain, new anxiety or depression, brain fog, chest pain, chills or sweats, cough, fatigue, fever, headache, nausea or vomiting, congestion, shortness of breath, sore throat, or loss of taste or smell as listed by CDC for common symptoms of long COVID-19. [4]

Risk factors included in our CPR were identified through our research in both aims 1 and 2 of this dissertation. Risk factors for the model include demographics, symptoms, disease severity, and comorbidities. A variable for a date of positive test, classified by those with a positive test either before or after December 1, 2021, was included to account for differences between Omicron and the more severe previous COVID-19 variants. This date coincides with the Omicron variant that was first identified in November 2021 and became the predominant strain in the United States by December 2021. [20] Risk factors for COVID-19 vaccination status were left out of this analysis due to missingness of this information. N3C reports that many of the healthcare records included in their database are not linked to vaccination records causing a large amount of missing data for these variables. We also chose not to include variables related to race in our multivariate analysis. Race in clinical prediction models, if there is no biological

association, can represent underlying health disparities rather than a medical need for this adjustment in the model.[106]

Development of CPR

Bivariate analysis was performed using logistic regression for all variables to assess the association with the outcome. A p-value less than 0.05 was considered statistically significant. Cut points for continuous variables were selected by inspection of histograms. To produce a simpler CPR, all comorbidities and symptoms were combined into a single variable respectively. The risk factor 'number of symptoms during acute infection' is the total number of previously listed symptoms a patient has from the date of COVID-19 diagnosis until 14 days after.

Our cohort was randomly split into derivation and validation datasets using a 80:20 ratio. The training dataset was used to develop our prediction model and the test dataset to test the ability of that model to predict individuals with long COVID-19. To minimize overfitting and extreme predictions, we used the least absolute shrinkage and selection operator technique (Lasso) for variable selection to develop a logistic regression model for predicting long COVID-19. [100] Lasso is also preferred over other penalized regression methods as it is able to produce a simpler model with fewer predictors.[100] This is optimal for developing a CPR that can be easily implemented in clinical practice. The model was evaluated in the derivation group to identify models with the highest area under the receiver operating characteristics curve (AUROCC). Guidelines for interpreting the AUROCC indicate 1 is perfect discrimination, 0.99-0.9 is excellent, 0.89 to 0.8 is good, 0.79 to 0.70 is fair, 0.51 to 0.69 is poor, and 0.5 is worthless.[107]

Presentation of CPR

In order to generate a simple CPR that can be implemented in a healthcare setting and will classify patients into risk groups, we presented our final CPR as a point score. A point score assigns a numerical value to each risk factor in the model and allows a provider to add the total points to easily classify patient risk of developing the outcome. We used the regression coefficients of our final model by dividing each beta-coefficient by the smallest beta-coefficient and rounding to the nearest number resulting in a weighted score.[108] Using the regression coefficients for development of the point score will result in a narrow range of points and a linear distribution of the final scores.[109]

We stratified scores into risk groups based on visual inspection of the point score distribution to create groups that would be most useful for clinical decision making. The primary goal was to minimize the moderate risk group and to primarily classify patients as low or high risk of developing long COVID-19. The CPR was then applied to the validation group and results classified into risk groups. All analyses were performed in R version 4.2.2.

External Validation

Our CPR was also externally validated using a dataset of college students at the University of Georgia. UGA students between the ages of 18 and 30 years old and who had a positive COVID-19 test after August 1, 2021 were included in the dataset. Student were emailed a survey that gathered information on their prior COVID-19 infection including dates of infection, symptom information, and comorbidities. The CPR was applied to this dataset and the AUROCC was calculated. The points for each participant were calculated and the distribution of the points into low, moderate, and high risk groups were observed.

Results

Study population

After identifying all COVID-19 patients in the N3C registry between the ages of 18 and 30 years old, 1,268,174 patients were included in our analysis. The characteristics of these patients are included in Table 7.1. A total of 126,656 patients (10.0%) had a diagnosis of long COVID-19. Overall, the study population had an average age of 24 years old and was primarily female (58.7%). Patients included in the study were primarily of white race (63.3%) and were Non-Hispanic (86.3%). About 2 percent of patients included in our analysis were hospitalized and about 70 percent had a positive COVID-19 test before December 1, 2021. About one-third of the study population reporting having at least one comorbidity. During acute infection (from diagnosis date until day 14) 10.9 percent of the patients were recorded has having 1-2 symptoms while 2.7 percent had 3 or more symptoms.

Bivariate analysis

Older age (odds ratio [OR]: 1.009 95% confidence interval [CI]: 1.007-1.010) and female sex (OR: 1.91 95% CI:1.89-1.94) were associated with an increased likelihood of long COVID-19. Having a COVID-19 associated hospitalization was also associated with an increased risk of long COVID-19 (OR: 1.37, 95% CI: 1.32-1.42). Similarly, having a positive COVID-19 test before December 1, 2021 (before Omicron) was associated with an increased risk of long COVID (OR: 1.27, 95% CI: 1,25-1.28). All comorbidities except HIV infection were associated with long COVID-19. This was the same for all reported symptoms. All symptoms included in analysis indicated an increased risk of developing long COVID-19.

Multivariate Analysis

Our final logistic regression model is included in Table 7.2. Use of the lasso technique for variable selection yielded a model with four risk factors for long COVID in our population of young adults. Female sex, having any comorbidity, testing positive for COVID-19 prior to December 1, 2021, and the number of symptoms present were all chosen by the lasso technique and indicated an increased likelihood for the development of long COVID-19. For calculation of the point score we chose to remove date positive from the final model. This was done as date positive before December 1, 2021 would not be useful in a clinical setting as all patients presenting would have COVID-19 after this date. This is shown in Table 7.2 where we did not calculate a point value for this variable. The resulting model had a fair discrimination based on the area under the curve (AUC) of 0.702. Point scores derived from the coefficients of the resulting model are also included in Table 7.2.

Classification Accuracy

The derivation cohort included 1,012,012 patients with 100,899 (10.0%) having a diagnosis of long COVID-19. A total of 253,467 patients with 25,749 (10.1%) having long COVID-19 were included in the validation cohort. The performance of the risk score in the derivation and validation groups is summarized in Table 7.3. The model classified about 64 percent of patients in the low-risk group with approximately a 6 percent risk of long COVID-19 in both the derivation and validation cohorts. The point range for the moderate risk group was chosen in order to minimize the likelihood of long COVID-19 in this group and maximizing the number of people classified to the low and high risk groups. About 6 percent of patients were classified into

the high risk group with about 75 percent having long COVID-19 in both the derivation and validation cohorts.

External Validation

The model was also applied to an external population of 1,983 college students. Overall, this population had 11.6 percent of students with long COVID-19. The CPR did not perform as well in this smaller data set with an AUC of 0.612. In the college student population 20 percent of students were classified into the low risk with 7.6 percent having long COVID-19. The high risk group included 36 percent of this data set with 14.7 percent having long COVID-19.

Discussion

We developed and validated a simple CPR for detecting long COVID-19 in young adults. The final rule included female sex, having any comorbidity, and the number of symptoms present during infection. The model had good accuracy with an AUC of 0.723. Furthermore, it was able to classify the majority of patients as low risk for long-COVID, and was able to identify a much smaller group at very high risk of long COVID.

Other related publications have found similar results to our clinical prediction rule. Several have found an increased risk of developing long COVID-19 in women.[102,104] Other studies, like ours, saw that women were more likely to have physician visits following COVID-19 diagnosis which could lend to this association. Currently, there is no biological explanation for an increased risk of long COVID-19 in women. We also included having any comorbidity in our CPR. Having a comorbidity has been shown in other studies to increase the risk of a poor prognosis with COVID-19.[159] The CDC provides a list of comorbidities that are associated with

long COVID-19 in the literature.[160] We chose to include a single 'any comorbidity' variable to simplify the model.

The number of symptoms reported during acute infection was also included in our final model. Bivariate analysis showed that all included symptoms were associated with the development of long COVID-19. To make the model simpler we included symptoms as a single variable. This can also act as a proxy for severity of disease. Generally the more symptoms a patient is experiencing, the more severe their illness is.[104] Lastly, the date of COVID-19 infection was chosen to be in or model however not included in the point score. The date of positive infection before and after December 1, 2021 was chosen to differentiate between earlier variants and the Omicron variant of COVID-19. Our model found an association with earlier variants and development of long COVID-19. This echoes research stating that the while the Omicron variant is more infectious, it generally causes less severe outcomes.[161] The decreased severity of Omicron may be inherent to the variant, or may also be associated with immunity due to prior infection with COVID-19 and the availability of vaccinations.

Strengths and Limitations

Our study had several strengths. The first was the access to a large set of EHR data that captures a diverse group of individuals with COVID-19 across the United States. This allowed us to analyze data on a large array of patients and to create a more representative sample of young adults with COVID-19. This large dataset also linked patient records across medical visits so we were able to follow the progression of their disease and determine their risk of prolonged symptoms. Second, even though our CPR is simple, it still had good overall accuracy in identifying patients with and without long COVID. While patients included in our study were

seen in a variety of settings, including outpatient and inpatient, this CPR is simple enough to be used in any outpatient setting as it does not require any laboratory testing.

There are also some limitations to our study. The data used in this analysis were obtained retrospectively from the EHR. This did not allow us to verify diagnoses, limited some of the data we had access to, and lent to issues with missing data. For follow-up visits we were only able to capture chief complaints listed by ICD-10 code in the EHR. Therefore, we likely did not capture all symptoms patients were experiencing and did not know if symptoms following COVID-19 were directly related to that diagnosis. Similarly, because patients were seen in a variety of healthcare settings, likely including drive up testing sights and telehealth visits, common vital signs were missing from a large number of these records. These are tests that are commonly captured in the outpatient setting and inclusion in the CPR could likely improve accuracy. Vaccine data was also largely missing from this dataset. The EHR and vaccine data are commonly not linked to each other causing much of this data to be missing. In further analysis we would either have to employ more complex imputation methods to fill in this missing data or choose healthcare sites where vaccine information can be linked to patient records.

Another limitation was the absence of a widely used reporting method for long COVID-19. The CDC implemented ICD-10 code U09.9 on October 1, 2021 to capture patients with post COVID-19 condition in the EHR. However, due to uncertain definitions of this outcome, this diagnosis has not been widely adopted by physicians. When we tried to use this code to define our outcome for analysis, only 0.4 percent of patients in our dataset were classified as having long COVID-19. Once definitions of long COVID-19 are more clear, physicians will able to more easily diagnose patients and will likely adopt the use of this ICD-10 code. Lastly, there could also be some issue with selection bias in this data set. N3C is a large dataset that receives

data from a wide array of health care systems across the United States however, many are large research based institutions. If an individual does not live in an area of one of these healthcare sites, perhaps in more rural areas, they are less likely to be included into this dataset. Ensuring that both small and large healthcare systems are included in this dataset would capture a more representative sample of the United States population with COVID-19.

Lastly, we were able to externally validate our CPR in a small dataset of college students. However, the model did not perform as well with an AUC of 0.612. This indicates that further work is needed to improve our model internally before generalizing to other populations. Issues with model performance in other datasets are likely due to some issues with missing data in the N3C dataset that made our data not truly comparable to other populations. Further work to employ imputation methods or generate a CPR with prospective data will help improve development of a CPR for long COVID-19 while also making these results more generalizable to other populations.

Conclusion

As a result of these analyses, we were able to generate a CPR that had a good accuracy internally in predicting young adults with long COVID-19. Our model included risk factors such as female sex, having, any comorbidity, and the number of symptoms during acute infection. These risk factors all indicated an increased risk of long COVID and could be used in a model to help physicians identify young adults at risk for long COVID-19.

TABLES AND FIGURES

Table 7.1 Characteristics of study population and bivariate analysis

	Total (N=1,268,174)	**Long COVID (N=126,656)**	**No Long COVID (N=1,141,518)**	**OR (95% CI)**
Age (years)				
Mean (SD)	24.18 (3.73)	24.29 (3.77)	24.16 (3.73)	1.009 (1.007-1.01)
Age Group				
18-21 years	372433 (29.4)	36302 (28.7)	336131 (29.4)	REF
22-25 years	384680 (30.3)	37172 (29.3)	347508 (30.4)	0.95 (0.93-0.96)
26-30 years	511061 (40.3)	53182 (42.0)	457879 (40.1)	1.08 (1.07-1.09)
Gender				
Male	521464 (41.1)	35464 (28.0)	486000 (42.6)	REF
Female	744688 (58.7)	91145 (72.0)	653543 (57.3)	1.91 (1.89-1.94)
Non-binary/Third Gender	1942 (0.2)	44 (0.0)	1898 (0.2)	0.18 (0.13-0.25)
Race				
American Indian/Alaska Native	6098 (0.5)	1027 (0.8)	5071 (0.4)	1.88 (1.76-2.01)
Native Hawaiian/Other Pacific Islander	2197 (0.2)	277 (0.2)	1920 (0.2)	1.28 (1.13-1.46)
Asian	29344 (2.3)	2879 (2.3)	26465 (2.3)	REF
Black/African American	189547 (14.9)	24681 (19.5)	164866 (14.4)	1.43 (1.41-1.45)
White/Caucasian	803364 (63.3)	78940 (62.3)	724424 (63.5)	0.96 (0.95-0.97)
Other	51818 (4.1)	1182 (0.9)	50636 (4.4)	0.20 (0.19-0.21)
Ethnicity				
Hispanic	173785 (13.7)	19648 (15.5)	154137 (13.5)	1.17 (1.15-1.19)
BMI (kg/m^2)				
Mean (SD)	29.78 (8.53)	31.70 (10.08)	29.76 (8.52)	
Hospitalization				
Hospitalized	27604 (2.2)	3506 (2.8)	24098 (2.1)	1.37(1.32-1.42)

	Total (N=1,268,174)	Long COVID (N=126,656)	No Long COVID (N=1,141,518)	OR (95% CI)
Date Positive				
Before 12/01/21	866557 (68.3)	92210 (72.8)	774347 (67.8)	1.27 (1.25-1.28)
Comorbidities				
Any Comorbidity	447073 (35.3)	79270 (62.6)	367803 (32.2)	3.51 (3.47-3.55)
Diabetes	28037 (2.2)	5463 (4.3)	22574 (2.0)	2.21 (2.14-2.27)
Heart Disease	1268 (0.1)	319 (0.3)	949 (0.1)	3.18 (2.80-3.60)
Heart Attack	1567 (0.1)	336 (0.3)	1231 (0.1)	2.50 (2.22-2.82)
Congestive Heart Failure	5350 (0.4)	1286 (1.0)	4064 (0.4)	2.78 (2.61-2.96)
Hypertension	55336 (4.4)	11236 (8.9)	44100 (3.9)	2.45 (2.40-2.50)
Autoimmune Disorder	8309 (0.7)	1870 (1.5)	6439 (0.6)	2.65 (2.52-2.79)
HIV	2781 (0.2)	661 (0.5)	2120 (0.2)	2.84 (2.60-3.10)
Chronic Lung Disease	87169 (6.9)	18882 (14.9)	68287 (6.0)	2.76 (2.71-2.80)
Cancer	7502 (0.6)	1724 (1.4)	5778 (0.5)	2.68 (2.54-2.83)
Depression	120755 (9.5)	28278 (22.3)	92477 (8.1)	3.26 (3.22-3.31)
Pregnant	138022 (10.9)	23010 (18.2)	115012 (10.1)	1.98 (1.95-2.01)
Obesity	233180 (18.4)	49297 (38.9)	183883 (16.1)	3.32 (3.28-3.36)
Substance Abuse	89628 (7.1)	16469 (13.0)	73159 (6.4)	2.20 (2.16-2.24)
Tobacco Smoker	33146 (2.6)	7705 (6.1)	25441 (2.2)	2.85 (2.78-2.93)
Number of Acute Symptoms				
1-2	138073 (10.9)	24427 (19.3)	113646 (10.0)	2.16 (2.13-2.19)
3 or more	34416 (2.7)	6509 (5.1)	27907 (2.4)	2.19 (2.10-2.22)
Symptoms				
Abdominal Pain	7507 (0.6)	5337 (4.2)	2170 (0.2)	4.22 (3.86-4.62)
New Anxiety/Depression	20668 (1.6)	14815 (11.7)	5853 (0.5)	5.92 (5.52-6.34)
Body/Muscle Aches	1654 (0.1)	503 (0.4)	1151 (0.1)	3.20 (2.82-3.62)

	Total (N=1,268,174)	Long COVID (N=126,656)	No Long COVID (N=1,141,518)	OR (95% CI)
Brain Fog/Difficulty Concentrating	658 (0.1)	567 (0.4)	91 (0.0)	9.65 (7.88-11.68)
Chest Pain	19 (0.0)	11 (0.0)	8 (0.0)	11.12 (3.40-26.57)
Chills/Sweats	672 (0.1)	120 (0.1)	552 (0.0)	2.15 (1.71-2.66)
Cough	20889 (1.6)	3913 (3.1)	16976 (1.5)	3.35 (3.12-3.59)
Fatigue	5392 (0.4)	3706 (2.9)	1686 (0.1)	7.91 (7.34-8.51)
Fever	9610 (0.8)	1572 (1.2)	8038 (0.7)	2.82 (2.57-3.08)
Headache	6393 (0.5)	2970 (2.3)	3423 (0.3)	3.85 (3.53-4.19)
Nausea/Vomiting	3911 (0.3)	2267 (1.8)	1644 (0.1)	4.94 (4.51-5.39)
Runny Nose/Congestion	6461 (0.5)	3099 (2.4)	3362 (0.3)	3.53 (3.21-3.89)
Shortness of Breath/Difficulty Breathing	9396 (0.7)	2739 (2.2)	6657 (0.6)	11.67 (10.95-12.42)
Sore Throat	19946 (1.6)	8780 (6.9)	11166 (1.0)	2.29 (2.10-2.49)
Loss of Taste or Smell	1963 (0.2)	225 (0.2)	1738 (0.2)	3.41 (3.26-3.58)

Table 7.2. Multivariate Logistic Regression via Lasso Technique

Variable	β Coefficient	β/lowest β	Points
Result of Lasso Technique (AUC=0.702)			
Female	0.44	1.00	1
Any Comorbidity	1.11	2.52	3
Date Positive: before 12/1/21	-	-	-
Number of Acute Symptoms			
1 to 2 symptoms	0.57	1.30	1
3 or more symptoms	0.66	1.50	2

Table 7.3. Classification accuracy of prediction rule in the derivation cohort, validation cohort, and external college student population

	Derivation Cohort	Validation Cohort	College Student Population
Risk Group (Points)	**Long COVID, n/Total (%)**	**Long COVID, n/Total (%)**	**Long COVID, n/Total (%)**
Overall	100,899/1,012,012 (10.0)	227,718/253,467 (10.1)	230/1,983 (11.6)
Low (0-2)	36,867/648,634 (5.7)	9,437/162,151 (5.8)	31/408 (7.6)
Moderate (3-4)	50,197/305,989(16.4)	12,767/77,026 (16.6)	88/752 (11.7)
High (5-7)	43,554/57,389 (75.9)	10,745/14,290 (75.2)	106/722 (14.7)

CHAPTER 8

CONCLUSIONS

This concluding chapter serves as a summary of the problems addressed by this dissertation, the results of each of the three aims, a comparison of their results, and to discuss future directions for research.

The Problem

SARS-CoV-2 as of March 2023 has caused an estimated 759 million cases and 6 million deaths worldwide.[2] Presentation, symptoms, and severity of COVID-19 infection vary by individual and can lead to a wide range of severe outcomes. One of these outcomes includes long COVID-19, also known a post-acute sequelae of COVID-19 or PASC. Due to the wide array of symptoms it is difficult for healthcare providers to diagnose individuals with long COVID-19. Healthcare providers have a difficult task to decide if the wide array of symptoms a patient is experiencing are the result of prior COVID-19 infection or have some other explanation.

Currently, we do not know what all of the true causes of long COVID-19 are or the groups that are most at risk for this outcome. Generally, these more severe health outcomes are associated with age and underling medical conditions. However, long COVID-19 has been observed to span multiple age groups with varying degrees of health.[8] Early evidence indicates that young adults are also experiencing prolonged symptoms of COVID-19.[9] As this age group is generally not of concern for severe health outcomes, there is limited research surrounding young adults and long COVID. Little is known about what is causing these prolonged symptoms in this

age group and beyond. Research understanding the risk factors and prognosis of long COVID in young adults is important to fully understand the long term impacts COVID-19 will have on this population.

Aim 1: Risk Factors of Long COVID-19: A Systematic Review and Meta-Analysis

In our first aim, we analyzed current literature outlining risk factors for long COVID-19. Our search yielded 46 studies included for meta-analysis. Positive COVID-19 tests in the included studies ranged from March 2020 to December 2021 with the average age of participants in the studies ranging from 29 to 67 years old.

Our meta-analysis identified significant associations between age 60 to 70 years (RR: 1.33 95% CI: 1.01-1.76), female sex (RR: 0.82 95% CI: 0.76-0.87), and white race (RR: 1.07 95% CI 1.06-1.08) and the development of long COVID-19. A patient having any comorbidity (RR:1.15 95% CI:1.05-1.26) was also significantly associated with long COVID. Long COVID-19 was also more likely in patients who were hospitalized (RR: 1.28 95% CI:1.08-1.52) as well as in those with moderate (RR:1.39 95%CI: 1.17-1.65) or critical COVID-19 (RR:3.97 (2.14-7.35). A longer hospital stay was also significantly associated with developing long COVID-19 (RR:1.66 95%CI: 1.28-2.04). Lastly, increased LDH was significantly associated with the development of long COVID (RR:20.33 95%CI: 2.90-37.77).

Aim 2: Risk Factors of Long COVID-19 in College Students

The second aim of this dissertation used a cross-sectional design to identify risk factors of long COVID-19 amongst 18 to 30 year old students at the University of Georgia. An email survey was sent to all enrolled students as of the Fall 2021 semester (n=40,056) and 1,983 students met

inclusion criteria and fully completed the survey. After identifying students with symptoms lasting 4 weeks or longer after infection, 230 (11.6%) were classified as having long COVID.

Univariate analysis indicated a significant association with the number of symptoms during acute infection (OR: 1.20 95% CI: 1.14-1.26), having 3 or more doses of a COVID-19 vaccine (OR: 0.55 95% CI: 0.34-0.92), having no comorbidities (OR:0.63 95%CI: 0.48-0.84), as well as a number of symptoms including shortness of breath, brain fog, and loss of taste or smell. Having more COVID-19 vaccine doses and no comorbidities was associated with a reduced risk of developing long COVID while more symptoms during the acute infection as well as specific symptoms were associated with an increased risk. Significant variables were then included into a logistic regression model using the lasso technique for variable selection. This yielded a model with seven risk factors. This included testing positive before December 1, 2021 (aOR: 4.06) and six symptom risk factors. These symptoms included body or muscle aches (aOR: 1.22), loss of taste or smell (aOR: 2.16), shortness of breath (aOR: 1.97), diarrhea (aOR: 1.68), brain fog (aOR:1.45), and new anxiety or depression (aOR: 1.32).

Aim 3: Development and Validation of a Clinical Prediction Rule for Predicting Long COVID-19 in Young Adults

We used the National COVID Cohort Collaborative (N3C) database to develop a clinical prediction rule (CRP) for identifying young adults at increased risk for long COVID-19. Our dataset included 1,268,174 patients with 126,656 (10.0%) having long COVID. Univariate analysis found statistically significant associations with older age, female sex, being hospitalized, having a positive COVID-19 test before December 1, 2021, having any comorbidity, and the number of symptoms during acute infection with an increased risk of long COVID.

The dataset was split into a derivation and validation cohort with each cohort having 10 percent of patients with long COVID-19. Use of the lasso technique for variable selection yielded a model with the risk factors female sex, having any comorbidity, and the number of symptoms during acute infection. The resulting model had good discrimination with an area under the curve (AUC) of 0.702. The resulting logistic regression model was turned into a point score and applied to both the derivation and validation cohorts. The model classified 64 percent of patients into low risk groups with about 6 percent of the patients having long COVID-19 in the derivation and validation cohorts. About 6 percent of patients were classified into the high risk group with 75 percent having long COVID-19 in both the derivation and validation cohorts.

Comparison of Results

Together, the three aims of this dissertation were able to identify risk factors that could help to identify young adults at risk for long COVID-19. While aim one included all adult age groups, there were similar results found in the second and third aims that included only young adults. Both aims two and three indicated a long COVID prevalence of about 10 to 11 percent amongst young adults. This is slightly higher than the prevalence estimated by the CDC of 7.5 percent in all adults.[158] In the first and third aims we were able to identify demographic risk factors significantly associated with long COVID-19 including older age, female sex, and white race. These aims also found significant associations with being hospitalized and having any comorbidity. All three aims indicated an increased risk of long COVID-19 in patients with more symptoms during acute infection, though this was statistically significant in only aims two and three. All three aims found minimal to no associations with any laboratory test, though this is likely due to the limited reporting of this data in the included studies and data sets. Aim two of

this dissertation included self-reported data of a single college student population, aim three included EHR data for young adults across the United States, and studies included in aim one included both self-reported and EHR data. Differences in the findings between these aims are partially due to the difference in the type of data used in these studies.

Future Directions

Further research needs to be conducted to truly determine the risk factors that cause long COVID-19 in young adults as well as to develop a more clear definition of this condition. This research needs to be done in the form of large prospective studies that follow young adult patients from COVID-19 infection forward. This will allow for the direct measurement of symptoms the patient is experiencing during infection as well as how those symptoms progress and evolve over time. These larger prospective studies will also allow for the measurement of common vital signs and laboratory tests during and after acute COVID-19 infection. Frequent measurement of these values will allow for researchers to observe any changes in these tests and possibly help increase the accuracy of identifying individuals at risk for long COVID-19. With data from these large prospective studies further work could be done to externally validate CPRs to help predict individuals at risk for long COVID-19. In particular use of our CPR in external populations with the addition of vital signs and laboratory tests could be a highly accurate tool used in long COVID-19 research. Lastly, not enough time has passed since the start of the COVID-19 pandemic to truly understand the long term impacts of this virus. Research surrounding severe outcomes and long COVID-19 need to continue into the future to truly identify the impact of this virus on patients in the long term.

REFERENCES

1. CDC Museum COVID-19 Timeline | David J. Sencer CDC Museum | CDC. https://www.cdc.gov/museum/timeline/covid19.html. Accessed February 1, 2022.
2. WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data. https://covid19.who.int/. Accessed February 1, 2022.
3. Zhai P, Ding Y, Wu X, Long J, Zhong Y, Li Y. The epidemiology, diagnosis and treatment of COVID-19. *Int J Antimicrob Agents*. 2020;55(5):105955. doi:10.1016/J.IJANTIMICAG.2020.105955
4. Centers for Disease Control and Prevention. Post-COVID Conditions | CDC. Saving Lives, Protecting People. https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects.html. Published 2021. Accessed August 1, 2021.
5. Carfì A, Bernabei R, Landi F. Persistent symptoms in patients after acute COVID-19. *JAMA - J Am Med Assoc*. 2020;324(6):603-605. doi:10.1001/JAMA.2020.12603
6. Chopra V, Flanders SA, O'Malley M, Malani AN, Prescott HC. Sixty-Day Outcomes Among Patients Hospitalized With COVID-19. *https://doi.org/107326/M20-5661*. 2020;174(4):576-578. doi:10.7326/M20-5661
7. Huang C, Huang L, Wang Y, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *www.thelancet.com*. 2021;397. doi:10.1016/S0140-6736(20)32656-8
8. Taquet M, Dercon Q, Luciano S, Geddes JR, Husain M, Harrison PJ. Incidence, co-occurrence, and evolution of long-COVID features: A 6-month retrospective cohort study of 273,618 survivors of COVID-19. *PLOS Med*. 2021;18(9):e1003773. doi:10.1371/JOURNAL.PMED.1003773
9. Ebell MH, Forgacs D, Shen Y, et al. Prevalence of previous infection with SARS-CoV-2 and persistent symptoms at a large university. *medRxiv*. July 2021:2021.07.08.21260201. doi:10.1101/2021.07.08.21260201
10. Umakanthan S, Sahu P, Ranade A V., et al. Origin, transmission, diagnosis and management of coronavirus disease 2019 (COVID-19). *Postgrad Med J*. 2020;96(1142):753-758. doi:10.1136/postgradmedj-2020-138234
11. Weiss SR, Navas-Martin S. Coronavirus Pathogenesis and the Emerging Pathogen Severe Acute Respiratory Syndrome Coronavirus. *Microbiol Mol Biol Rev*. 2005;69(4):635-664. doi:10.1128/MMBR.69.4.635-664.2005
12. Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. doi:10.1146/annurev-virology-110615-042301
13. Cascella M, Rajnik M, Cuomo A, Dulebohn SC, Di Napoli R. Features, Evaluation, and Treatment of Coronavirus (COVID-19). *StatPearls*. January 2022. https://www.ncbi.nlm.nih.gov/books/NBK554776/. Accessed February 1, 2022.
14. Sheervalilou R, Shirvaliloo M, Dadashzadeh N, et al. COVID-19 under spotlight: A close look at the origin, transmission, diagnosis, and treatment of the 2019-nCoV disease. *J Cell Physiol*. 2020;235(12):8873-8924. doi:10.1002/JCP.29735
15. De Wit E, Van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights

into emerging coronaviruses. *Nat Rev Microbiol 2016 148*. 2016;14(8):523-534. doi:10.1038/nrmicro.2016.81
16. About Middle East Respiratory Syndrome (MERS) | CDC. https://www.cdc.gov/coronavirus/mers/about/index.html. Accessed February 1, 2022.
17. Al-Omari A, Rabaan AA, Salih S, Al-Tawfiq JA, Memish ZA. MERS coronavirus outbreak: Implications for emerging viral infections. 2018. doi:10.1016/j.diagmicrobio.2018.10.011
18. CDC Museum COVID-19 Timeline | David J. Sencer CDC Museum | CDC. https://www.cdc.gov/museum/timeline/covid19.html. Accessed November 13, 2021.
19. CDC Weekly C. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) — China, 2020. *China CDC Wkly*. 2020;2(8):113-122. doi:10.46234/CCDCW2020.032
20. Tracking SARS-CoV-2 variants. https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/. Accessed March 9, 2022.
21. Kumar V, Doshi KU, Khan WH, Rathore AS. COVID-19 pandemic: mechanism, diagnosis, and treatment. *J Chem Technol Biotechnol*. 2021;96(2):299-308. doi:10.1002/JCTB.6641
22. Vasireddy D, Vanaparthy R, Mohan G, Varun Malayala S, Atluri P. Review of COVID-19 Variants and COVID-19 Vaccine Efficacy: What the Clinician Should Know? *J Clin Med Res*. 2021;13(6):317-325. doi:10.14740/jocmr4518
23. Scientific Brief: SARS-CoV-2 Transmission | CDC. https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/sars-cov-2-transmission.html. Accessed February 1, 2022.
24. Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int J Antimicrob Agents*. 2020;55(3):105924. doi:10.1016/J.IJANTIMICAG.2020.105924
25. Liu Y, Rocklöv J. The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. doi:10.1093/jtm/taab124
26. Lewnard JA, Hong VX, Patel MM, Kahn R, Lipsitch M, Tartof SY. Title: Clinical outcomes among patients infected with Omicron (B.1.1.529) SARS-CoV-2 variant in southern California. doi:10.1101/2022.01.11.22269045
27. Ito K, Piantham C, Nishiura H. Relative instantaneous reproduction number of Omicron SARS-CoV-2 variant with respect to the Delta variant in Denmark. *J Med Virol*. 2021. doi:10.1002/JMV.27560
28. Kim D, Jo J, Lim J-S, Ryu S. Serial interval and basic reproduction number of SARS-CoV-2 Omicron variant in South Korea. *medRxiv*. December 2021:2021.12.25.21268301. doi:10.1101/2021.12.25.21268301
29. Sah P, Fitzpatrick MC, Zimmer CF, et al. Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis. *Proc Natl Acad Sci U S A*. 2021;118(34). doi:10.1073/PNAS.2109229118/-/DCSUPPLEMENTAL
30. Verity R, Okell LC, Dorigatti I, et al. Articles Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis*. 2020;20:669-677. doi:10.1016/S1473-3099(20)30243-7
31. Ferguson NM, Laydon D, Nedjati-Gilani G, et al. of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. doi:10.25561/77482

32. Maggi L, Russo RC, Ragab D, et al. The COVID-19 Cytokine Storm; What We Know So Far. *Front Immunol | www.frontiersin.org*. 2020;1:1446. doi:10.3389/fimmu.2020.01446
33. Guan W, Ni Z, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med*. 2020;382(18):1708-1720. doi:10.1056/NEJMOA2002032/SUPPL_FILE/NEJMOA2002032_DISCLOSURES.PDF
34. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020;395(10223):507-513. doi:10.1016/S0140-6736(20)30211-7
35. Shi H, Han X, Jiang N, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect Dis*. 2020;20(4):425-434. doi:10.1016/S1473-3099(20)30086-4/ATTACHMENT/D9FABEDA-1DE4-4419-A7A0-3E2DFCDACBC4/MMC2.PDF
36. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395(10223):497-506. doi:10.1016/S0140-6736(20)30183-5/ATTACHMENT/D5332CA1-83D8-4C4C-BC57-00A390BF0396/MMC1.PDF
37. Yang X, Yu Y, Xu J, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med*. 2020;8(5):475-481. doi:10.1016/S2213-2600(20)30079-5
38. Malik P, Patel U, Mehta D. Biomarkers and outcomes of COVID-19 hospitalisations: systematic review and meta-analysis. *BMJ Evidence-Based Med*. 2021;26:107-108. doi:10.1136/bmjebm-2020-111536
39. Keddie S, Ziff O, Chou MKL, et al. Laboratory biomarkers associated with COVID-19 severity and management. *Clin Immunol*. 2020;221:108614. doi:10.1016/J.CLIM.2020.108614
40. Bentivegna M, Hulme C, Ebell MH. Primary Care Relevant Risk Factors for Adverse Outcomes in Patients With COVID-19 Infection: A Systematic Review. *J Am Board Fam Med*. 2021;34(Supplement):S113-S126. doi:10.3122/JABFM.2021.S1.200429
41. Recommendations for national SARS-CoV-2 testing strategies and diagnostic capacities. https://www.who.int/publications/i/item/WHO-2019-nCoV-lab-testing-2021.1-eng. Accessed February 1, 2022.
42. Dinnes J, Deeks JJ, Berhane S, et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Syst Rev*. 2021;2021(3). doi:10.1002/14651858.CD013705.PUB2/MEDIA/CDSR/CD013705/IMAGE_N/NCD013705-TST-050.PNG
43. Wang YH, Wu CC, Bai CH, et al. Evaluation of the diagnostic accuracy of COVID-19 antigen tests: A systematic review and meta-analysis. *J Chin Med Assoc*. 2021;84(11):1028-1037. doi:10.1097/JCMA.0000000000000626
44. Jegerlehner S, Suter-Riniker F, Jent P, Bittel P, Nagler M. Diagnostic accuracy of a SARS-CoV-2 rapid antigen test in real-life clinical settings. *Int J Infect Dis*. 2021;109:118-122. doi:10.1016/J.IJID.2021.07.010
45. Atieh MA, Guirguis M, Alsabeeha NHM, Cannon RD. The diagnostic accuracy of saliva testing for SARS-CoV-2: A systematic review and meta-analysis. *Oral Dis*. 2021;00:1-15. doi:10.1111/ODI.13934
46. Deeks JJ, Dinnes J, Takwoingi Y, et al. Antibody tests for identification of current and past infection with SARS-CoV-2. *Cochrane Database Syst Rev*. 2020;2020(6).

doi:10.1002/14651858.CD013652
47. National Academies of Sciences E and M. Rapid Expert Consultation on SARS-CoV-2 Viral Shedding and Antibody Response for the COVID-19 Pandemic (April 8, 2020). *Rapid Expert Consult SARS-CoV-2 Viral Shedding Antib Response COVID-19 Pandemic (April 8, 2020)*. April 2020. doi:10.17226/25774
48. Know Your Treatment Options for COVID-19 | FDA. https://www.fda.gov/consumers/consumer-updates/know-your-treatment-options-covid-19. Accessed December 1, 2021.
49. Beigel JH, Tomashek KM, Dodd LE, et al. Remdesivir for the Treatment of Covid-19 — Final Report. *N Engl J Med*. 2020;383(19):1813-1826. doi:10.1056/NEJMOA2007764/SUPPL_FILE/NEJMOA2007764_DATA-SHARING.PDF
50. EPIC-HR: Study of Oral PF-07321332/Ritonavir Compared With Placebo in Nonhospitalized High Risk Adults With COVID-19 - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT04960202. Accessed March 9, 2022.
51. Coronavirus (COVID-19) Update: FDA Authorizes First Oral Antiviral for Treatment of COVID-19 | FDA. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-first-oral-antiviral-treatment-covid-19. Accessed March 9, 2022.
52. P H, WS L, JR E, et al. Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med*. 2021;384(8):693-704. doi:10.1056/NEJMOA2021436
53. Sterne JAC, Murthy S, Diaz J V., et al. Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19: A Meta-analysis. *JAMA*. 2020;324(13):1330-1341. doi:10.1001/JAMA.2020.17023
54. Lloyd EC, Gandhi TN, Petty LA. Monoclonal Antibodies for COVID-19. *JAMA*. 2021;325(10):1015-1015. doi:10.1001/JAMA.2021.1225
55. Chen P, Nirula A, Heller B, et al. SARS-CoV-2 Neutralizing Antibody LY-CoV555 in Outpatients with Covid-19. *N Engl J Med*. 2021;384(3):229-237. doi:10.1056/NEJMOA2029849/SUPPL_FILE/NEJMOA2029849_DATA-SHARING.PDF
56. Weinreich DM, Sivapalasingam S, Norton T, et al. REGN-COV2, a Neutralizing Antibody Cocktail, in Outpatients with Covid-19. *N Engl J Med*. 2021;384(3):238-251. doi:10.1056/NEJMOA2035002/SUPPL_FILE/NEJMOA2035002_DATA-SHARING.PDF
57. Gottlieb RL, Nirula A, Chen P, et al. Effect of Bamlanivimab as Monotherapy or in Combination With Etesevimab on Viral Load in Patients With Mild to Moderate COVID-19: A Randomized Clinical Trial. *JAMA*. 2021;325(7):632-644. doi:10.1001/JAMA.2021.0202
58. Islam N, Sharp SJ, Chowell G, et al. Physical distancing interventions and incidence of coronavirus disease 2019: natural experiment in 149 countries. *BMJ*. 2020;370. doi:10.1136/BMJ.M2743
59. Association of State-Issued Mask Mandates and Allowing On-Premises Restaurant Dining with County-Level COVID-19 Case and Death Growth Rates — United States, March 1–December 31, 2020 | MMWR. https://www.cdc.gov/mmwr/volumes/70/wr/mm7010e3.htm. Accessed March 9, 2022.
60. Donovan C V., Rose C, Lewis KN, et al. SARS-CoV-2 Incidence in K–12 School

Districts with Mask-Required Versus Mask-Optional Policies — Arkansas, August–October 2021. *MMWR Morb Mortal Wkly Rep*. 2022;71(10):384-389. doi:10.15585/MMWR.MM7110E1
61. How to Protect Yourself & Others | CDC. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html. Accessed November 15, 2021.
62. COVID-19 Quarantine and Isolation | CDC. https://www.cdc.gov/coronavirus/2019-ncov/your-health/quarantine-isolation.html. Accessed November 15, 2021.
63. COVID-19 vaccine tracker and landscape. https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines. Accessed December 1, 2021.
64. Different COVID-19 Vaccines | CDC. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines.html. Accessed December 1, 2021.
65. Understanding mRNA COVID-19 Vaccines | CDC. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines/mrna.html. Accessed February 1, 2022.
66. Understanding Viral Vector COVID-19 Vaccines | CDC. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines/viralvector.html. Accessed February 1, 2022.
67. Chagla Z. The BNT162b2 (BioNTech/Pfizer) vaccine had 95% efficacy against COVID-19 ≥7 days after the 2nd dose. *https://doi.org/107326/ACPJ202102160-015*. 2021;174(2):JC15. doi:10.7326/ACPJ202102160-015
68. Baden LR, El Sahly HM, Essink B, et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N Engl J Med*. 2021;384(5):403-416. doi:10.1056/NEJMOA2035389/SUPPL_FILE/NEJMOA2035389_DATA-SHARING.PDF
69. Sadoff J, Gray G, Vandebosch A, et al. Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19. *N Engl J Med*. 2021;384(23):2187-2201. doi:10.1056/NEJMOA2101544/SUPPL_FILE/NEJMOA2101544_DATA-SHARING.PDF
70. McMorrow M. Improving communications around vaccine breakthrough and vaccine effectiveness. https://www.documentcloud.org/documents/21026654-57c98604-3b54-44f0-8b44-b148d8f75165. Accessed March 9, 2022.
71. Andrews N, Stowe J, Kirsebom F, et al. Covid-19 Vaccine Effectiveness against the Omicron (B.1.1.529) Variant. *N Engl J Med*. March 2022. doi:10.1056/NEJMOA2119451
72. A clinical case definition of post COVID-19 condition by a Delphi consensus, 6 October 2021. https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1. Accessed November 8, 2021.
73. Groff D, Sun A, Ssentongo AE, et al. Short-term and Long-term Rates of Postacute Sequelae of SARS-CoV-2 Infection: A Systematic Review. *JAMA Netw open*. 2021;4(10). doi:10.1001/JAMANETWORKOPEN.2021.28568
74. Havervall S, Rosell A, Phillipson M, et al. Symptoms and Functional Impairment Assessed 8 Months After Mild COVID-19 Among Health Care Workers. *JAMA*. 2021;325(19):2015-2016. doi:10.1001/JAMA.2021.5612
75. Taquet M, Geddes JR, Husain M, Luciano S, Harrison PJ. 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. *The lancet Psychiatry*. 2021;8(5):416-427. doi:10.1016/S2215-0366(21)00084-5

76. Carvalho-Schneider C, Laurent E, Lemaignen A, et al. Follow-up of adults with noncritical COVID-19 two months after symptom onset. *Clin Microbiol Infect*. 2021;27:258. doi:10.1016/j.cmi.2020.09.052
77. Logue JK, Franko NM, McCulloch DJ, et al. Sequelae in Adults at 6 Months After COVID-19 Infection. *JAMA Netw Open*. 2021;4(2):e210830-e210830. doi:10.1001/JAMANETWORKOPEN.2021.0830
78. Taquet M, Dercon Q, Luciano S, Geddes JR, Husain M, Harrison PJ. Incidence, co-occurrence, and evolution of long-COVID features: A 6-month retrospective cohort study of 273,618 survivors of COVID-19. *PLOS Med*. 2021;18(9):e1003773. doi:10.1371/JOURNAL.PMED.1003773
79. Monod M, Blenkinsop A, Xi X, et al. Age groups that sustain resurging COVID-19 epidemics in the United States. *Science*. 2021;371(6536). doi:10.1126/SCIENCE.ABE8372
80. Wilson RF, Sharma AJ, Schluechtermann S, et al. Morbidity and Mortality Weekly Report Factors Influencing Risk for COVID-19 Exposure Among Young Adults Aged 18-23 Years-Winnebago County, Wisconsin, March-July 2020. 2020;16(41). doi:10.1023/A:1025023600517
81. Leidner AJ, Barry V, Bowen VB, et al. Morbidity and Mortality Weekly Report Opening of Large Institutions of Higher Education and County-Level COVID-19 Incidence-United States. https://www.cdc.gov/. Accessed February 1, 2022.
82. Walsh-Messinger J, Manis H, Vrabec A, et al. The Kids Are Not Alright: A Preliminary Report of Post-COVID Syndrome in University Students. *medRxiv*. November 2020:2020.11.24.20238261. doi:10.1101/2020.11.24.20238261
83. Carfì A, Bernabei R, Landi F. Persistent Symptoms in Patients After Acute COVID-19. *JAMA*. 2020;324(6):603. doi:10.1001/JAMA.2020.12603
84. Sudre CH, Murray B, Varsavsky T, et al. Attributes and predictors of long COVID. *Nat Med 2021 274*. 2021;27(4):626-631. doi:10.1038/s41591-021-01292-y
85. Akbarialiabad H, Mohammad ·, Taghrir H, et al. Long COVID, a comprehensive systematic scoping review. *Infection*. 2021;49:1163-1186. doi:10.1007/s15010-021-01666-x
86. Michelen M, Manoharan L, Elkheir N, et al. Characterising long COVID: a living systematic review. *BMJ Glob Heal*. 2021;6:5427. doi:10.1136/bmjgh-2021-005427
87. Page MJ, Mckenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. doi:10.1136/bmj.n71
88. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Int J Surg*. 2010;8(5):336-341. doi:10.1016/J.IJSU.2010.02.007
89. McKeever L, Nguyen V, Peterson SJ, Gomez-Perez S, Braunschweig C. Demystifying the Search Button. *J Parenter Enter Nutr*. 2015;39(6):622-635. doi:10.1177/0148607115593791
90. Wilczynski NL, Haynes B, Team H. Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. 2004. http://www.biomedcentral.com/1741-7015/2/23. Accessed March 10, 2022.
91. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*. 2014;14(1):1-13. doi:10.1186/1471-2288-14-135/TABLES/3

92. Grooten WJA, Tseli E, Äng BO, et al. Elaborating on the assessment of the risk of bias in prognostic studies in pain rehabilitation using QUIPS-aspects of interrater agreement. *Diagnostic Progn Res*. 2019;3(1). doi:10.1186/S41512-019-0050-0
93. Erez A, Bloom MC, Wells MT. USING RANDOM RATHER THAN FIXED EFFECTS MODELS IN META-ANALYSIS: IMPLICATIONS FOR SITUATIONAL SPECIFICITY AND VALIDITY GENERALIZATION. *Pers Psychol*. 1996;49(2):275-306. doi:10.1111/J.1744-6570.1996.TB01801.X
94. Higgins J, Thomas J, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions | Cochrane Training. https://training.cochrane.org/handbook/current. Accessed March 10, 2022.
95. Page MJ, Sterne JAC, Higgins JPT, Egger M. Investigating and dealing with publication bias and other reporting biases in meta-analyses of health research: A review. *Res Synth Methods*. 2021;12(2):248-259. doi:10.1002/JRSM.1468
96. Egger M, Smith GD, Schneider M, Minder C. Papers Bias in meta-analysis detected by a simple, graphical test.
97. UGA Fact Book | Office of Institutional Research. https://oir.uga.edu/factbook/. Accessed March 24, 2022.
98. Yu F, Lau L-T, Fok M, Lau JY-N, Zhang K. COVID-19 Delta variants—Current status and implications as of August 2021. *Precis Clin Med*. 2021;4(4):287-292. doi:10.1093/PCMEDI/PBAB024
99. Steyerberg EW. Clinical Prediction Models. 2009. doi:10.1007/978-0-387-77244-8
100. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events. *BMJ*. 2015;351. doi:10.1136/BMJ.H3868
101. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' Guides to the Medical Literature: XXII: How to Use Articles About Clinical Decision Rules. *JAMA*. 2000;284(1):79-84. doi:10.1001/JAMA.284.1.79
102. Pfaff ER, Girvin AT, Bennett TD, et al. Identifying who has long COVID in the USA: a machine learning approach using N3C data. *Lancet Digit Heal*. 2022;4(7):e532-e541. doi:10.1016/S2589-7500(22)00048-6
103. Socia D, Larie D, Feuerwerker S, An G, Cockrell C. Prediction of Long COVID Based on Severity of Initial COVID-19 Infection: Differences in predictive feature sets between hospitalized versus non-hospitalized index infections. *medRxiv Prepr Serv Heal Sci*. January 2023. doi:10.1101/2023.01.16.23284634
104. Sudre CH, Murray B, Varsavsky T, et al. Attributes and predictors of long COVID. *Nat Med 2021 274*. 2021;27(4):626-631. doi:10.1038/s41591-021-01292-y
105. N3C Data Overview | National Center for Advancing Translational Sciences. https://ncats.nih.gov/n3c/about/data-overview. Accessed March 10, 2022.
106. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms. *N Engl J Med*. 2020;383(9):874-882. doi:10.1056/NEJMMS2004740/SUPPL_FILE/NEJMMS2004740_DISCLOSURES.PDF
107. Cantor SB, Kattan MW. Determining the area under the ROC curve for a binary diagnostic test. *Med Decis Mak*. 2000;20(4):468-470. doi:10.1177/0272989X0002000410
108. Van Der Helm-Van Mil AHM, Cessie S Le, Van Dongen H, Breedveld FC, Toes REM, Huizinga TWJ. A prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis: How to guide individual treatment decisions. *Arthritis Rheum*. 2007;56(2):433-440. doi:10.1002/ART.22380

109. Ebell M. AHRQ white paper: Use of clinical decision rules for point-of-care decision support. *Med Decis Mak*. 2010;30(6):712-721. doi:10.1177/0272989X10386232
110. medRxiv.org - the preprint server for Health Sciences. https://www.medrxiv.org/. Accessed January 20, 2023.
111. PubMed. https://pubmed.ncbi.nlm.nih.gov/. Accessed January 20, 2023.
112. Marshall JC, Murthy S, Diaz J, et al. A minimal common outcome measure set for COVID-19 clinical research. *Lancet Infect Dis*. 2020;20(8):e192-e197. doi:10.1016/S1473-3099(20)30483-7
113. Augustin M, Schommers P, Stecher M, et al. Post-COVID syndrome in non-hospitalised patients with COVID-19: a longitudinal prospective cohort study. 2021. doi:10.1016/j.lanepe.2021.100122
114. Peghin M, Palese A, Venturini M, et al. Post-COVID-19 symptoms 6 months after acute infection among hospitalized and non-hospitalized patients. 2021. doi:10.1016/j.cmi.2021.05.033
115. Naik S, Nath Haldar S, Soneja M, et al. Post COVID-19 sequelae: A prospective observational study from Northern India. *Drug Discov Ther*. 2021;15(5):254-260. doi:10.5582/ddt.2021.01093
116. Akbar Asadi-Pooya A, Akbari A, Emami A, et al. Risk Factors Associated with Long COVID Syndrome: A Retrospective Study. *Iran J Med Sci*. 2021;46(6). doi:10.30476/ijms.2021.92080.2326
117. Subramanian A, Nirantharakumar K, Hughes S, et al. Symptoms and risk factors for long COVID in non-hospitalized adults. 1909. doi:10.1038/s41591-022-01909-w
118. Bai F, Tomasoni D, Falcinella C, et al. Female gender is associated with long COVID syndrome: a prospective cohort study. 2021. doi:10.1016/j.cmi.2021.11.002
119. Bliddal S, Banasik K, Birger Pedersen O, et al. Acute and persistent symptoms in non-hospitalized PCR-confirmed COVID-19 patients. *Sci Reports* |. 123AD;11:13153. doi:10.1038/s41598-021-92045-x
120. Cervia C, Zurbuchen Y, Taeschler P, et al. Immunoglobulin signature predicts risk of post-acute COVID-19 syndrome. doi:10.1038/s41467-021-27797-1
121. Vimercati L, De Maria L, Quarato M, et al. Clinical Medicine Association between Long COVID and Overweight/Obesity. *J Clin Med*. 2021;10:4143. doi:10.3390/jcm10184143
122. Kim Y, Kim S-W, Chang H-H, Tae Kwon K, Bae S, Hwang S. COVID-19 viral load not associated with disease severity: findings from a retrospective cohort study. 2021. doi:10.1186/s12879-022-07062-6
123. Becker C BK. View of Long COVID 1 year after hospitalisation for COVID-19: a prospective bicentric cohort study. https://smw.ch/index.php/smw/article/view/3094/5152. Accessed January 12, 2023.
124. Nune A, Durkowski V, Titman A, et al. Incidence and risk factors of long COVID in the UK: A single-centre observational study. *J R Coll Physicians Edinb*. 2021;51(4):338-343. doi:10.4997/JRCPE.2021.405
125. Desgranges F, Tadini E, Munting A, et al. Post-COVID-19 Syndrome in Outpatients: a Cohort Study. doi:10.1007/s11606-021-07242-1
126. Anwar Hossain M, Amran Hossain KM, Saunders K, et al. Prevalence of Long COVID symptoms in Bangladesh: a prospective Inception Cohort Study of COVID-19 survivors. *BMJ Glob Heal*. 2021;6:6838. doi:10.1136/bmjgh-2021-006838
127. Tak Wai Lui D, Ho Lee C, Sun Chow W, et al. Long COVID in Patients With Mild to

Moderate Disease: Do Thyroid Function and Autoimmunity Play a Role? 2021. doi:10.1016/j.eprac.2021.06.016
128. Förster C, Colombo MG, Wetzel A-J, Martus P, Joos S. M E D I C I N E Persisting Symptoms After COVID-19 Prevalence and Risk Factors in a Population-Based Cohort. doi:10.3238/arztebl.m2022.0147
129. Otmani H El, Nabili S, Berrada M, et al. Prevalence, characteristics and risk factors in a Moroccan cohort of Long-Covid-19. *Neurol Sci*. 1:3. doi:10.1007/s10072-022-06138-0
130. Helmsdal G, Hanusson KD, Kristiansen MF, et al. Long COVID in the Long Run-23-Month Follow-up Study of Persistent Symptoms. doi:10.1093/ofid/ofac270
131. Messin L, Puyraveau M, Benabdallah Y, et al. COVEVOL: Natural Evolution at 6 Months of COVID-19. 2021. doi:10.3390/v13112151
132. Silverbergid JI, Zyskindid I, Naiditch H, et al. Predictors of chronic COVID-19 symptoms in a community-based cohort of adults. 2022. doi:10.1371/journal.pone.0271310
133. Larijani MS, Ashrafian F, Amiri FB, et al. Characterization of long COVID-19 manifestations and its associated factors: A prospective cohort study from Iran. 2022. doi:10.1016/j.micpath.2022.105618
134. Faycal A, Ndoadoumgue AL, Sellem B, et al. Prevalence and factors associated with symptom persistence: A prospective study of 429 mild COVID-19 outpatients. 2021. doi:10.1016/j.idnow.2021.11.003
135. Chudzik M, Lewek J, Kapusta J, Banach M, Jankowski P, Bielecka-Dabrowa A. Clinical Medicine Predictors of Long COVID in Patients without Comorbidities: Data from the Polish Long-COVID Cardiovascular (PoLoCOV-CVD) Study. *J Clin Med*. 2022;2022:4980. doi:10.3390/jcm11174980
136. Buonsenso D, Rosaria Gualano M, Francesca Rossi M, et al. Post-Acute COVID-19 Sequelae in a Working Population at One Year Follow-Up: A Wide Range of Impacts from an Italian Sample. *Int J Environ Res Public Heal*. 2022. doi:10.3390/ijerph191711093
137. Ferreira De Oliveira J, Eliane De Ávila R, Ramos De Oliveira N, et al. Persistent symptoms, quality of life, and risk factors in long COVID: a cross-sectional study of hospitalized patients in Brazil. *Int J Infect Dis*. 2022;122:1044-1051. doi:10.1016/j.ijid.2022.07.063
138. Yellumahanthi DK, Barnett B, Barnett S, Yellumahanthi S. COVID-19 Infection: Its Lingering Symptoms in Adults. *Cureus*. 2022;14(5). doi:10.7759/CUREUS.24736
139. Hill {elaine, Mehta H, Sharma S, et al. Risk Factors Associated with Post-Acute Sequelae of SARS-CoV-2 in an EHR Cohort: A National COVID Cohort Collaborative (N3C) Analysis as part of the NIH RECOVER program. doi:10.1101/2022.08.15.22278603
140. Yoo SM, Liu TC, Motwani Y, et al. Factors Associated with Post-Acute Sequelae of SARS-CoV-2 (PASC) After Diagnosis of Symptomatic COVID-19 in the Inpatient and Outpatient Setting in a Diverse Cohort. doi:10.1007/s11606-022-07523-3
141. Chan Sui Ko A, Candellier A, Mercier M, et al. Number of initial symptoms is more related to long COVID-19 than acute severity of infection: a prospective cohort of hospitalized patients. *Int J Infect Dis*. 2022;118:220. doi:10.1016/J.IJID.2022.03.006
142. Loosen SH, Jensen BEO, Tanislav C, Luedde T, Roderburg C, Kostev K. Obesity and lipid metabolism disorders determine the risk for development of long COVID syndrome: a cross-sectional study from 50,402 COVID-19 patients. *Infection*. 2022;50(5):1165. doi:10.1007/S15010-022-01784-0

143. Spinicci M, Graziani L, Tilli M, et al. Infection with SARS-CoV-2 Variants Is Associated with Different Long COVID Phenotypes. *Viruses*. 2022;14(11):2367. doi:10.3390/V14112367
144. García-Abellán J, Fernández M, Padilla S, et al. Immunologic phenotype of patients with long-COVID syndrome of 1-year duration. *Front Immunol*. 2022;13. doi:10.3389/FIMMU.2022.920627/FULL
145. Caze AB, Cerqueira-Silva T, Bomfim A. Prevalence and risk factors for long COVID after mild disease: a longitudinal study with a symptomatic control group. doi:10.1101/2022.09.15.22279958
146. Khalaf M, Alboraie M, Abdel-Gawad M, et al. Prevalence and Predictors of Persistent Symptoms After Clearance of SARS-CoV-2 Infection: A Multicenter Study from Egypt. 2022. doi:10.2147/IDR.S355064
147. Knight DRT, Munipalli B, Logvinov II, et al. Perception, Prevalence, and Prediction of Severe Infection and Post-acute Sequelae of COVID-19. *Am J Med Sci*. 2022;363(4):295-304. doi:10.1016/J.AMJMS.2022.01.002
148. Kim Y, Kim SW, Chang HH, Kwon KT, Hwang S, Bae S. One Year Follow-Up of COVID-19 Related Symptoms and Patient Quality of Life: A Prospective Cohort Study. *Yonsei Med J*. 2022;63(6):499. doi:10.3349/YMJ.2022.63.6.499
149. Sneller MC, Liang ; C Jason, Marques AR, et al. A Longitudinal Study of COVID-19 Sequelae and Immunity: Baseline Findings Background: A substantial proportion of persons who de. 2022. doi:10.7326/M21-4905
150. Exnobert Zulu J, Banda D, Zajac Hines J, et al. Two-month follow-up of persons with SARS-CoV-2 infection-Zambia. *Pan Afr Med J*. 2020;(26):41. doi:10.11604/pamj.2022.41.26.30721
151. Wu Q, Ailshire J, Crimmins E. Long COVID and Symptom Trajectory in a Representative Sample of Americans. *Res Sq*. March 2022. doi:10.21203/RS.3.RS-1440503/V1
152. Zuschlag D, Grandt D, Custodis F, Christian Braun ·, Winfried Häuser ·. A retrospective cohort single-center study. doi:10.1007/s00482-022-00626-0
153. Varghese J, Sandmann S, Vollenberg R, et al. Follow Up of COVID-19 Features in Recovered Adults without Comorbidities-Persistent Symptoms and Lab-Abnormalities. 2020. doi:10.21203/rs.3.rs-116030/v1
154. Senjam SS, Balhara YPS, Kumar P, et al. Assessment of Post COVID-19 Health Problems and its Determinants in North India: A descriptive cross section study. *medRxiv*. October 2021:2021.10.03.21264490. doi:10.1101/2021.10.03.21264490
155. Zayet S, Zahra H, Royer P-Y, et al. microorganisms Post-COVID-19 Syndrome: Nine Months after SARS-CoV-2 Infection in a Cohort of 354 Patients: Data from the First Wave of COVID-19 in Nord Franche-Comté Hospital, France. 2021;9:1719. doi:10.3390/microorganisms9081719
156. Bell Id ML, Catalfamo Id CJ, Farland L V, et al. Post-acute sequelae of COVID-19 in a non-hospitalized cohort: Results from the Arizona CoVHORT. 2021. doi:10.1371/journal.pone.0254347
157. Wei Xiang Ong S, Fong S-W, Edward Young B, et al. Open Forum Infectious Diseases Persistent Symptoms and Association With Inflammatory Cytokine Signatures in Recovered Coronavirus Disease 2019 Patients. doi:10.1093/ofid/ofab156
158. Nearly One in Five American Adults Who Have Had COVID-19 Still Have "Long COVID." https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2022/20220622.htm.

Accessed February 6, 2023.
159. Ebell MH, Hamadani R, Kieber-Emmons A. Development and Validation of Simple Risk Scores to Predict Hospitalization in Outpatients with COVID-19 Including the Omicron Variant. *J Am Board Fam Med*. 2022;35(6):1058-1064. doi:10.3122/JABFM.2022.220056R1
160. Underlying Medical Conditions Associated with Higher Risk for Severe COVID-19: Information for Healthcare Professionals | CDC. https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/underlyingconditions.html. Accessed March 2, 2023.
161. Strasser ZH, Greifer N, Hadavand A, Murphy SN, Estiri H. Estimates of SARS-CoV-2 Omicron BA.2 Subvariant Severity in New England. *JAMA Netw Open*. 2022;5(10):e2238354-e2238354. doi:10.1001/JAMANETWORKOPEN.2022.38354

APPENDIX A

Table A1. Search strategy

Database	Search Terms	Database Link
PubMed	(“betacoronavirus”[mh] OR “coronavirus”[tiab] OR “corona-virus”[tiab] OR “COVID-19”[tiab] OR “COVID19”[tiab] OR “2019-nCoV”[tiab] OR “nCoV”[tiab] OR “SARS-CoV-2”[tiab] OR “SARSCOV2”[tiab] OR “2019-nCov”[tiab] OR “2019 coronavirus”[tiab] OR "novel coronavirus"[tiab]) AND ("post-acute COVID-19 syndrome" [Supplementary Concept] OR "long-COVID"[tiab] OR “long COVID”[tiab] OR “long-haul COVID”[tiab] OR “persistent COVID-19”[tiab] OR “post-covid syndrome” [tiab] OR “post-covid sequelae” OR “long-term impact” [tiab] OR “long-term effect” [tiab] OR “prolonged symptom”) AND (“risk factors”[tiab] OR “risk factor”[tiab] OR “predictor”[tiab] OR “predictors”[tiab])	https://pubmed.ncbi.nlm.nih.gov
WHO COVID-19 Database	(tw:(post-acute COVID-19 syndrome)) OR (tw:(long-COVID)) OR (tw:(long COVID)) OR (tw:(long-haul COVID)) OR (tw:(persistent COVID-19)) OR (tw:(post-covid syndrome)) OR (tw:(post-covid sequelae)) OR (tw:(long-term impact)) OR (tw:(long-term effect)) OR (tw:(prolonged symptom)) AND (tw:(risk factors)) OR (tw:(risk factor)) OR (tw:(predictor)) OR (tw:(predictors))	https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/
MedRXIV	“long covid”	https://www.medrxiv.org

Table A2. Original and grouped categorical risk factors

Author, Year	RiskFactorOriginal	RiskFactorNew
de Oliveria, 2022	Abdominal pain	Abdominal pain
Messin, 2021	Abdominal pain	Abdominal pain
Sadat Larijani, 2022	Abdominal pain	Abdominal pain
Wu, 2022	Abdominal pain	Abdominal pain
Helmsdal, 2022	Age 0-17	Age < 20
Zulu, 2022	Age 0-19	Age < 20
Zuschlag, 2022	Age < 20	Age < 20
El Otmani, 2022	Age < 35	Age < 35
Faycal, 2021	Age <= 30	Age <= 30
Loosen, 2022	Age <= 30	Age <= 30
Buonsenso, 2022	Age <=46	Age <=46
El Otmani, 2022	Age > 35	Age > 35
Faycal, 2021	Age > 40	Age > 40
Buonsenso, 2022	Age > 46	Age > 46
Caze, 2022	Age > 50	Age > 50
Lui, 2021	Age > 50	Age > 50
Zulu, 2022	Age > 50	Age > 50
Loosen, 2022	Age > 60	Age > 60
Peghin, 2021	Age >60	Age > 60
Zuschlag, 2022	Age > 80	Age > 80
Zayet, 2021	Age > 90	Age > 90
Helmsdal, 2022	Age > 67	Age >= 65
Hossain, 2021	Age > 70	Age >= 65
Messin, 2021	Age > 71	Age >= 65
Senjam, 2021	Age > 71	Age >= 65
Silverberg, 2022	Age >=70	Age >= 65
Subramania, 2022	Age >=70	Age >= 65
Sudre, 2021	Age > 70	Age >= 65
Wu, 2022	Age >= 65	Age >= 65
Yellumahanthi, 2022	Age >=65	Age >= 65
Caze, 2022	Age 15-29	Age 15-34
Helmsdal, 2022	Age 18-34	Age 15-34
Hossain, 2021	Age 18-30	Age 15-34
Kim, 2022	Age 17-29	Age 15-34
Kim, 2022	Age 20-29	Age 15-34
Messin, 2021	Age 18-30	Age 15-34
Senjam, 2021	Age 18-25	Age 15-34

Subramania, 2022	Age 18-29	Age 15-34
Zayet, 2021	Age 18-30	Age 15-34
Peghin, 2021	Age 18-40	Age 18-49
Silverberg, 2022	Age 18-49	Age 18-49
Sudre, 2021	Age 18-49	Age 18-49
Wu, 2022	Age 18-49	Age 18-49
Yellumahanthi, 2022	Age 18-64	Age 18-64
Zuschlag, 2022	Age 20-39	Age 20-39
Zulu, 2022	Age 20-49	Age 20-49
Senjam, 2021	Age 26-45	Age 26-45
Caze, 2022	Age 30-39	Age 30-40
Faycal, 2021	Age 30-40	Age 30-40
Hossain, 2021	Age 31-40	Age 30-40
Kim, 2022	Age 30-39	Age 30-40
Kim, 2022	Age 30-39	Age 30-40
Messin, 2021	Age 31-40	Age 30-40
Subramania, 2022	Age 30-39	Age 30-40
Zayet, 2021	Age 31-40	Age 30-40
Loosen, 2022	Age 31-45	Age 31-45
Helmsdal, 2022	Age 35-49	Age 35-49
Caze, 2022	Age 40-49	Age 40-50
Hossain, 2021	Age 41-40	Age 40-50
Kim, 2022	Age 40-49	Age 40-50
Kim, 2022	Age 40-49	Age 40-50
Messin, 2021	Age 41-50	Age 40-50
Subramania, 2022	Age 40-49	Age 40-50
Zayet, 2021	Age 41-50	Age 40-50
Zuschlag, 2022	Age 40-59	Age 40-59
Peghin, 2021	Age 41-60	Age 41-60
Loosen, 2022	Age 46-60	Age 46-60
Senjam, 2021	Age 46-60	Age 46-60
Hossain, 2021	Age 51-60	Age 50-60
Kim, 2022	Age 50-59	Age 50-60
Kim, 2022	Age 50-59	Age 50-60
Messin, 2021	Age 51-60	Age 50-60
Subramania, 2022	Age 50-59	Age 50-60
Zayet, 2021	Age 51-60	Age 50-60
Helmsdal, 2022	Age 50-67	Age 50-69
Silverberg, 2022	Age 50-69	Age 50-69

Sudre, 2021	Age 50-69	Age 50-69
Wu, 2022	Age 50-64	Age 50-69
Hossain, 2021	Age 61-70	Age 60-70
Kim, 2022	Age 60-70	Age 60-70
Kim, 2022	Age 60-70	Age 60-70
Messin, 2021	Age 61-70	Age 60-70
Senjam, 2021	Age 61-70	Age 60-70
Subramania, 2022	Age 60-69	Age 60-70
Zayet, 2021	Age 61-70	Age 60-70
Zuschlag, 2022	Age 60-79	Age 60-70
Zayet, 2021	Age 71-80	Age 71-80
Zayet, 2021	Age 81-90	Age 81-90
Augustin, 2021	Rhinitis	Allergic rhinitis
Bell, 2021	Seasonal allergies	Allergic rhinitis
Caze, 2022	Respiratory allergy	Allergic rhinitis
Kim, 2022	Allergic rhinitis	Allergic rhinitis
Messin, 2021	Allergic condition	Allergic rhinitis
Förster, 2022	Anemia	Anemia
Asadi-Pooya, 2021	Comorbidities	Any Comorbidity
Augustin, 2021	Comorbidities	Any Comorbidity
Bai, 2022	Comorbidities	Any Comorbidity
Bell, 2021	Comorbidity	Any Comorbidity
Caze, 2022	Comorbidity	Any Comorbidity
de Oliveria, 2022	Comorbidities	Any Comorbidity
Garcia-Abellan, 2021	Comorbidity	Any Comorbidity
Helmsdal, 2022	Comorbidities	Any Comorbidity
Naik, 2021	Comorbidities	Any Comorbidity
Ong, 2021	Comorbidity	Any Comorbidity
Senjam, 2021	Comorbidity	Any Comorbidity
Yellumahanthi, 2022	Comorbidities	Any Comorbidity
Zulu, 2022	Comorbidity	Any Comorbidity
Sadat Larijani, 2022	Arrhythmia	Arrhythmia
Chudzik, 2022	Leg pain	Arthralgia
Chudzik, 2022	Arthritis	Arthritis
de Oliveria, 2022	Joint stiffness	Arthritis
Förster, 2022	Osteoarthritis	Arthritis
Förster, 2022	Arthritis	Arthralgia
Sadat Larijani, 2022	Rheumatism	Arthritis
Bell, 2021	Asthma	Asthma

Cervia, 2022	Asthma	Asthma
de Oliveria, 2022	Asthma	Asthma
Desgranges, 2022	Asthma	Asthma
Loosen, 2022	Asthma	Asthma
Messin, 2021	Asthma	Asthma
Sneller, 2022	Asthma	Asthma
Subramania, 2022	Asthma	Asthma
Sudre, 2021	Asthma	Asthma
Wu, 2022	Asthma	Asthma
Zayet, 2021	Asthma	Asthma
Sadat Larijani, 2022	Attention disorder	Attention disorder
Desgranges, 2022	Chronic inflammatory disease	Autoimmune condition
Garcia-Abellan, 2021	Autoimmune diseases	Autoimmune condition
Messin, 2021	Autoimmune condition	Autoimmune condition
Naik, 2021	Connective tissue disorder	Autoimmune condition
Subramania, 2022	Multiple Sclerosis	Autoimmune condition
Wu, 2022	Autoimmune diseases	Autoimmune condition
Chudzik, 2022	Back pain	Back pain
Förster, 2022	Back pain	Back pain
Subramania, 2022	Back pain	Back pain
Bai, 2022	BMI < 30	BMI < 30
Subramania, 2022	BMI <18.5	BMI <18.5
Vimercati, 2021	BMI > 25	BMI > 25
Bai, 2022	BMI >= 30	BMI >= 30
Messin, 2021	BMI > 30	BMI >= 30
Senjam, 2021	BMI >= 30	BMI >= 30
Zayet, 2021	BMI > 30	BMI >= 30
Messin, 2021	BMI < 25	BMI 18.5-25
Senjam, 2021	BMI <= 24.9	BMI 18.5-25
Subramania, 2022	BMI 18.5-25	BMI 18.5-25
Zayet, 2021	BMI 18.5-25	BMI 18.5-25
Messin, 2021	BMI 25-30	BMI 25-30
Senjam, 2021	BMI 25-29.9	BMI 25-30
Subramania, 2022	BMI 25-30	BMI 25-30
Zayet, 2021	BMI 25-30	BMI 25-30
Cervia, 2022	History of malignancy	Cancer
Desgranges, 2022	Malignancy	Cancer
Förster, 2022	Malignancy	Cancer
Garcia-Abellan, 2021	Cancer	Cancer

Kim, 2022	Malignancy	Cancer
Loosen, 2022	Cancer	Cancer
Messin, 2021	Malignancy	Cancer
Wu, 2022	Cancer	Cancer
Zayet, 2021	Malignancy	Cancer
Subramania, 2022	Celiac disease	Celiac disease
Chudzik, 2022	Chest congestion	Chest congestion
Bai, 2022	Chest pain	Chest pain
Chudzik, 2022	Chest pain	Chest pain
Messin, 2021	Chest pain	Chest pain
Sadat Larijani, 2022	Chest pain	Chest pain
Chudzik, 2022	Chills	Chills
Cervia, 2022	Lung disease	Chronic lung disease
Chan Sui Ko, 2022	Lung disease	Chronic lung disease
de Oliveria, 2022	COPD	Chronic lung disease
Förster, 2022	Lung disease	Chronic lung disease
Garcia-Abellan, 2021	Lung disease	Chronic lung disease
Hill, 2022	Lung disease	Chronic lung disease
Hossain, 2021	Lung disease	Chronic lung disease
Khalaf, 2021	Lung disease	Chronic lung disease
Kim, 2022	COPD	Chronic lung disease
Loosen, 2022	COPD	Chronic lung disease
Naik, 2021	Pulmonary fibrosis	Chronic lung disease
Sadat Larijani, 2022	Lung disease	Chronic lung disease
Spinicci, 2022	COPD	Chronic lung disease
Subramania, 2022	COPD	Chronic lung disease
Sudre, 2021	Chronic lung disease	Chronic lung disease
Wu, 2022	Lung disease	Chronic lung disease
Zayet, 2021	COPD	Chronic lung disease
Zayet, 2021	Chronic respiratory disease	Chronic respiratory disease
de Oliveria, 2022	Conjunctival hyperemia	Conjunctival hyperemia
de Oliveria, 2022	Coryza	Coryza
Augustin, 2021	Cough	Cough
Bai, 2022	Cough	Cough
Chudzik, 2022	Cough	Cough
de Oliveria, 2022	Cough	Cough
Faycal, 2021	Cough	Cough
Messin, 2021	Cough	Cough
Ong, 2021	Cough	Cough

Sadat Larijani, 2022	Dry cough	Cough
Wu, 2022	Cough	Cough
Caze, 2022	Diabetes	Diabetes
Cervia, 2022	Diabetes	Diabetes
Chan Sui Ko, 2022	Diabetes	Diabetes
de Oliveria, 2022	Diabetes	Diabetes
Desgranges, 2022	Diabetes	Diabetes
Förster, 2022	Diabetes	Diabetes
Garcia-Abellan, 2021	Diabetes	Diabetes
Hill, 2022	Diabetes	Diabetes
Hossain, 2021	Diabetes	Diabetes
Khalaf, 2021	Diabetes	Diabetes
Kim, 2022	Diabetes	Diabetes
Loosen, 2022	DM type 1	Diabetes
Messin, 2021	Diabetes	Diabetes
Naik, 2021	Diabetes	Diabetes
Ong, 2021	Diabetes	Diabetes
Peghin, 2021	Diabetes	Diabetes
Sadat Larijani, 2022	Diabetes	Diabetes
Sneller, 2022	Diabetes	Diabetes
Spinicci, 2022	Diabetes	Diabetes
Sudre, 2021	Diabetes	Diabetes
Wu, 2022	Diabetes	Diabetes
Yoo, 2022	Diabetes	Diabetes
Zayet, 2021	Diabetes	Diabetes
Augustin, 2021	Diarrhea	Diarrhea
Chudzik, 2022	Diarrhea	Diarrhea
de Oliveria, 2022	Diarrhea	Diarrhea
Faycal, 2021	Diarrhea	Diarrhea
Messin, 2021	Diarrhea	Diarrhea
Ong, 2021	Diarrhea	Diarrhea
Wu, 2022	Diarrhea	Diarrhea
Messin, 2021	Odynophagy	Difficulty swallowing
Sadat Larijani, 2022	Difficulty swallowing	Difficulty swallowing
de Oliveria, 2022	Vertigo	Dizziness
Sadat Larijani, 2022	Dizziness	Dizziness
Buonsenso, 2022	Drinks alcohol	Drinks alcohol
de Oliveria, 2022	Alcoholism	Alcoholism
Messin, 2021	Daily alcohol	Alcoholism

Messin, 2021	Occasional alcohol	Drinks alcohol
Peghin, 2021	Drinks alcohol	Drinks alcohol
Senjam, 2021	Drinks alcohol	Drinks alcohol
Bai, 2022	Dyspnea	Dyspnea
Chudzik, 2022	Dyspnea	Dyspnea
de Oliveria, 2022	Dyspnea	Dyspnea
Faycal, 2021	Dyspnea	Dyspnea
Messin, 2021	Dyspnea	Dyspnea
Ong, 2021	Dyspnea	Dyspnea
Sadat Larijani, 2022	Dyspnea	Dyspnea
Wu, 2022	Shortness of breath	Dyspnea
Sadat Larijani, 2022	Earache	Earache
Subramania, 2022	Eating disorder	Eating disorder
Subramania, 2022	Endometriosis	Endometriosis
Zayet, 2021	ENT diseases	ENT diseases
Bai, 2022	Fatigue	Fatigue
Buonsenso, 2022	Fatigue	Fatigue
Chudzik, 2022	Fatigue	Fatigue
de Oliveria, 2022	Fatigue	Fatigue
Faycal, 2021	Fatigue	Fatigue
Messin, 2021	Fatigue	Fatigue
Sadat Larijani, 2022	Fatigue	Fatigue
Wu, 2022	Fatigue	Fatigue
Augustin, 2021	Fever	Fever
Bai, 2022	Fever	Fever
Chudzik, 2022	Temperature > = 37.5	Fever
Chudzik, 2022	Temperature < 36.6 C	Temperature < 36.6 C
de Oliveria, 2022	Fever	Fever
Faycal, 2021	Fever	Fever
Messin, 2021	Fever	Fever
Ong, 2021	Fever	Fever
Sadat Larijani, 2022	Fever	Fever
Wu, 2022	Fever	Fever
Wu, 2022	Temperature > 100.4 F	Temperature > 100.4 F
Subramania, 2022	Fibromyalgia	Fibromyalgia
Bai, 2022	Former smoker	Former smoker
Messin, 2021	Former smoker	Former smoker
Peghin, 2021	Former smoker	Former smoker
Subramania, 2022	Former smoker	Former smoker

Asadi-Pooya, 2021	GI problems	GI problems
Förster, 2022	GI problems	GI problems
Subramania, 2022	IBS	GI problems
Sadat Larijani, 2022	Hair loss	Hair loss
Wu, 2022	Hair loss	Hair loss
Augustin, 2021	Headache	Headache
Bai, 2022	Headache	Headache
Chudzik, 2022	Headache	Headache
de Oliveria, 2022	Headache	Headache
Faycal, 2021	Headache	Headache
Messin, 2021	Facial headache	Facial headache
Messin, 2021	Headache	Headache
Sadat Larijani, 2022	Headache	Headache
Subramania, 2022	Migraine	Headache
Wu, 2022	Headache	Headache
Chudzik, 2022	Impaired hearing	Hearing loss
Messin, 2021	Loss of hearing	Hearing loss
Sadat Larijani, 2022	Hearing loss	Hearing loss
Subramania, 2022	Deafness	Hearing loss
Cervia, 2022	Heart disease	Heart disease
Chan Sui Ko, 2022	Heart disease	Heart disease
de Oliveria, 2022	Heart failure	Heart disease
Desgranges, 2022	Heart disease	Heart disease
Förster, 2022	Heart disease	Heart disease
Garcia-Abellan, 2021	Heart disease	Heart disease
Hill, 2022	Heart disease	Heart disease
Hossain, 2021	Heart disease	Heart disease
Khalaf, 2021	Heart disease	Heart disease
Kim, 2022	Heart disease	Heart disease
Loosen, 2022	Heart disease	Heart disease
Messin, 2021	Heart disease	Heart disease
Naik, 2021	Heart disease	Heart disease
Ong, 2021	Heart disease	Heart disease
Peghin, 2021	Heart disease	Heart disease
Sadat Larijani, 2022	Heart Failure	Heart disease
Sneller, 2022	Heart disease	Heart disease
Spinicci, 2022	Heart disease	Heart disease
Sudre, 2021	CAD	Heart disease
Wu, 2022	Heart disease	Heart disease

Zayet, 2021	Heart disease	Heart disease
Messin, 2021	Hemoptysis	Hemoptysis
Yoo, 2022	History of organ transplant	History of organ transplant
de Oliveria, 2022	HIV	HIV
Sneller, 2022	HIV	HIV
Zulu, 2022	HIV	HIV
Asadi-Pooya, 2021	ICU Admission	Hospitalized
Augustin, 2021	Hospitalized	Hospitalized
Caze, 2022	Hospitalized	Hospitalized
Cervia, 2022	Hospitalized	Hospitalized
Helmsdal, 2022	Hospitalized	Hospitalized
Knight, 2022	Hospitalized	Hospitalized
Messin, 2021	Hospitalized	Hospitalized
Naik, 2021	Admission	Hospitalized
Sneller, 2022	Hospitalized	Hospitalized
Zayet, 2021	Hospitalized	Hospitalized
Bell, 2021	Hypertension	Hypertension
Cervia, 2022	Hypertension	Hypertension
Chan Sui Ko, 2022	Hypertension	Hypertension
de Oliveria, 2022	Hypertension	Hypertension
Desgranges, 2022	Hypertension	Hypertension
Förster, 2022	Hypertension	Hypertension
Garcia-Abellan, 2021	Hypertension	Hypertension
Hill, 2022	Hypertension	Hypertension
Hossain, 2021	Hypertension	Hypertension
Khalaf, 2021	Hypertension	Hypertension
Kim, 2022	Hypertension	Hypertension
Loosen, 2022	Hypertension	Hypertension
Messin, 2021	Hypertension	Hypertension
Naik, 2021	Hypertension	Hypertension
Ong, 2021	Hypertension	Hypertension
Peghin, 2021	Hypertension	Hypertension
Sadat Larijani, 2022	Hypertension	Hypertension
Sneller, 2022	Hypertension	Hypertension
Spinicci, 2022	Hypertension	Hypertension
Vimercati, 2021	Hypertension	Hypertension
Wu, 2022	Hypertension	Hypertension
Zulu, 2022	Hypertension	Hypertension
Sadat Larijani, 2022	Hypotension	Hypotension

Naik, 2021	Hypothyroidism	Hypothyroidism
Vimercati, 2021	ILI	ILI
Cervia, 2022	Systemic immunosuppression	Immunosupression
Khalaf, 2021	Immunosuppression	Immunosupression
Sadat Larijani, 2022	Itching	Itching
Cervia, 2022	Kidney disease	Kidney disease
Chan Sui Ko, 2022	Kidney disease	Kidney disease
de Oliveria, 2022	Kidney disease	Kidney disease
Förster, 2022	Kidney disease	Kidney disease
Hill, 2022	Kidney disease	Kidney disease
Hossain, 2021	Kidney disease	Kidney disease
Khalaf, 2021	Kidney disease	Kidney disease
Kim, 2022	Kidney disease	Kidney disease
Messin, 2021	Kidney disease	Kidney disease
Naik, 2021	Kidney disease	Kidney disease
Spinicci, 2022	Kidney disease	Kidney disease
Sudre, 2021	Kidney disease	Kidney disease
Wu, 2022	Kidney disease	Kidney disease
Zayet, 2021	Kidney disease	Kidney disease
Subramania, 2022	Learning disability	Learning disability
Loosen, 2022	Lipid Metabolism disorder	Lipid Metabolism disorder
Förster, 2022	Liver disease	Liver disease
Kim, 2022	Liver disease	Liver disease
Peghin, 2021	Liver disease	Liver disease
de Oliveria, 2022	Hyporexia	Loss of appetite
Sadat Larijani, 2022	Anorexia	Loss of appetite
Augustin, 2021	Loss of smell	Loss of smell
Bai, 2022	Loss of smell	Loss of smell
Chudzik, 2022	Loss of smell	Loss of smell
de Oliveria, 2022	Loss of smell	Loss of smell
Faycal, 2021	Loss of smell	Loss of smell
Messin, 2021	Loss of smell	Loss of smell
Sadat Larijani, 2022	Loss of smell	Loss of smell
Wu, 2022	Loss of smell	Loss of smell
Augustin, 2021	Loss of taste	Loss of taste
Chudzik, 2022	Loss of taste	Loss of taste
de Oliveria, 2022	Loss of taste	Loss of taste
Faycal, 2021	Loss of taste	Loss of taste
Messin, 2021	Loss of taste	Loss of taste

Sadat Larijani, 2022	Loss of taste	Loss of taste
Chudzik, 2022	Loss of taste or smell	Loss of taste or smell
Sadat Larijani, 2022	Memory loss	Memory loss
Förster, 2022	Anxiety	Mental health problem
Loosen, 2022	Depression	Mental health problem
Sadat Larijani, 2022	Any mental health problem	Mental health problem
Sneller, 2022	Any mental health problem	Mental health problem
Subramania, 2022	Anxiety	Mental health problem
Augustin, 2021	Body aches	Myalgia
Bai, 2022	Body pain	Myalgia
de Oliveria, 2022	Myalgia	Myalgia
Faycal, 2021	Myalgia	Myalgia
Messin, 2021	Muscle or Body aches	Myalgia
Ong, 2021	Myalgia	Myalgia
Sadat Larijani, 2022	Myalgia	Myalgia
Wu, 2022	Myalgia	Myalgia
de Oliveria, 2022	Myocarditis	Myocarditis
Messin, 2021	Nasal obstruction	Nasal obstruction
Bai, 2022	Vomiting/nausea	Nausea_Vomiting
Chudzik, 2022	Vomiting	Nausea_Vomiting
de Oliveria, 2022	Vomiting	Nausea_Vomiting
Messin, 2021	Vomiting	Nausea_Vomiting
Sadat Larijani, 2022	Nausea/vomiting	Nausea_Vomiting
Wu, 2022	Vomiting	Nausea_Vomiting
Naik, 2021	Need of ventilator	Need of ventilator
Asadi-Pooya, 2021	Neurological problems	Neurologic disease
Peghin, 2021	Psychiatric disorder	Neurologic disease
Zayet, 2021	Neurologic disease	Neurologic disease
Messin, 2021	Nosebleed	Nosebleed
Caze, 2022	Obesity	Obesity
Chan Sui Ko, 2022	Obesity	Obesity
de Oliveria, 2022	Obesity	Obesity
Desgranges, 2022	Obesity	Obesity
Hill, 2022	Obesity	Obesity
Loosen, 2022	Obesity	Obesity
Peghin, 2021	Obesity	Obesity
Sneller, 2022	Obesity	Obesity
Spinicci, 2022	Obesity	Obesity
Sudre, 2021	Obesity	Obesity

Wu, 2022	Obesity	Obesity
Sadat Larijani, 2022	Paresthesia	Paresthesia
Subramania, 2022	PCOS	PCOS
Bai, 2022	Pneumonia	Pneumonia
Bai, 2022	Ethnicity: Arabic	Race: Arabic
Bai, 2022	Ethnicity: Asian	Race: Asian
Hill, 2022	Ethnicity: Asian	Race: Asian
Sneller, 2022	Race: Asian	Race: Asian
Subramania, 2022	Ethnicity: Asian	Race: Asian
Bai, 2022	Ethnicity: Black	Race: Black
Hill, 2022	Ethnicity: Black	Race: Black
Sneller, 2022	Race: Black	Race: Black
Subramania, 2022	Ethnicity: Black	Race: Black
Wu, 2022	Race: Black	Race: Black
Yoo, 2022	Race: Black	Race: Black
Bai, 2022	Ethnicity: Hispanic	Race: Hispanic
Bell, 2021	Hispanic	Race: Hispanic
Hill, 2022	Ethnicity: Hispanic	Race: Hispanic
Wu, 2022	Hispanic	Race: Hispanic
Yoo, 2022	Race: Hispanic or Latino	Race: Hispanic
Sneller, 2022	Multiple	Race: Multiple
Bell, 2021	Non-Hispanic	Race: Non-Hispanic
Bell, 2021	Other race	Race: Other
Hill, 2022	Ethnicity: Other	Race: Other
Subramania, 2022	Ethnicity: Multi-racial	Race: Other
Wu, 2022	Other	Race: Other
Yoo, 2022	Race: Other	Race: Other
Sneller, 2022	Unknown	Race: Unknown
Bai, 2022	Ethnicity: White	Race: White
Hill, 2022	Ethnicity: White	Race: White
Sneller, 2022	White	Race: White
Subramania, 2022	Ethnicity: White	Race: White
Wu, 2022	White	Race: White
Yoo, 2022	Race: White	Race: White
Sadat Larijani, 2022	Red eye	Red eye
Peghin, 2021	Renal impairment	Renal impairment
Subramania, 2022	Chronic sinusitis	Repiratory problems
Asadi-Pooya, 2021	Respiratory disease	Respiratory problems
Chudzik, 2022	Respiratory diseases	Respiratory problems

Peghin, 2021	URTI	Respiratory problems
Wu, 2022	Respiratory problems	Respiratory problems
Faycal, 2021	Rhinorrhea	Rhinorrhea
Messin, 2021	Rhinorrhea	Rhinorrhea
Ong, 2021	Rhinorrhea	Rhinorrhea
Sadat Larijani, 2022	Runny nose	Rhinorrhea
Wu, 2022	Runny/Stuffy nose	Rhinorrhea
Kim, 2022	Asymptomatic	Severity: Asymptomatic
Kim, 2022	Asymptomatic	Severity: Asymptomatic
Peghin, 2021	Asymptomatic	Severity: Asymptomatic
Senjam, 2021	Asymptomatic	Severity: Asymptomatic
Spinicci, 2022	Asymptomatic	Severity: Asymptomatic
Zuschlag, 2022	Asymptomatic	Severity: Asymptomatic
Kim, 2022	Critical	Severity: Critical
Kim, 2022	Critical	Severity: Critical
Peghin, 2021	Critical	Severity: Critical
Spinicci, 2022	Critical	Severity: Critical
Zuschlag, 2022	Critical	Severity: Critical
Zuschlag, 2022	Critical (Multi-organ failure)	Severity: Multi-organ failure
Hossain, 2021	Mild	Severity: Mild
Khalaf, 2021	Mild	Severity: Mild
Kim, 2022	Mild	Severity: Mild
Kim, 2022	Mild	Severity: Mild
Naik, 2021	Mild	Severity: Mild
Ong, 2021	Mild	Severity: Mild
Peghin, 2021	Mild	Severity: Mild
Senjam, 2021	Mild	Severity: Mild
Spinicci, 2022	Mild	Severity: Mild
Zuschlag, 2022	Mild	Severity: Mild
Hossain, 2021	Moderate	Severity: Moderate
Khalaf, 2021	Moderate	Severity: Moderate
Kim, 2022	Moderate	Severity: Moderate
Kim, 2022	Moderate	Severity: Moderate
Naik, 2021	Moderate/Severe	Severity: Moderate
Ong, 2021	Moderate	Severity: Moderate
Peghin, 2021	Moderate	Severity: Moderate
Senjam, 2021	Moderate	Severity: Moderate
Spinicci, 2022	Moderate	Severity: Moderate
Zuschlag, 2022	Moderate	Severity: Moderate

Hossain, 2021	Severe	Severity: Severe
Khalaf, 2021	Severe	Severity: Severe
Kim, 2022	Severe	Severity: Severe
Kim, 2022	Severe	Severity: Severe
Ong, 2021	Severe	Severity: Severe
Peghin, 2021	Severe	Severity: Severe
Senjam, 2021	Severe	Severity: Severe
Spinicci, 2022	Severe	Severity: Severe
Zuschlag, 2022	Severe	Severity: Severe
Asadi-Pooya, 2021	Gender (Male)	Sex (Male)
Augustin, 2021	Gender (Male)	Sex (Male)
Bai, 2022	Gender (Male)	Sex (Male)
Becker, 2021	Gender (Male)	Sex (Male)
Bell, 2021	Gender (Male)	Sex (Male)
Bliddal, 2021	Gender (Male)	Sex (Male)
Buonsenso, 2022	Gender (Male)	Sex (Male)
Caze, 2022	Gender (Male)	Sex (Male)
Cervia, 2022	Gender (Male)	Sex (Male)
Chan Sui Ko, 2022	Gender (Male)	Sex (Male)
Chudzik, 2022	Gender (Male)	Sex (Male)
de Oliveria, 2022	Gender (Male)	Sex (Male)
El Otmani, 2022	Gender (Male)	Sex (Male)
Faycal, 2021	Gender (Male)	Sex (Male)
Förster, 2022	Gender (Male)	Sex (Male)
Garcia-Abellan, 2021	Gender (Male)	Sex (Male)
Helmsdal, 2022	Gender (Male)	Sex (Male)
Hill, 2022	Gender (Male)	Sex (Male)
Hossain, 2021	Gender (Male)	Sex (Male)
Khalaf, 2021	Gender (Male)	Sex (Male)
Kim, 2022	Gender (Male)	Sex (Male)
Kim, 2022	Gender (Male)	Sex (Male)
Knight, 2022	Gender (Male)	Sex (Male)
Loosen, 2022	Gender (Male)	Sex (Male)
Lui, 2021	Gender (Male)	Sex (Male)
Messin, 2021	Gender (Male)	Sex (Male)
Naik, 2021	Gender (Male)	Sex (Male)
Nune, 2021	Gender (Male)	Sex (Male)
Ong, 2021	Gender (Male)	Sex (Male)
Peghin, 2021	Gender (Male)	Sex (Male)

Senjam, 2021	Gender (Male)	Sex (Male)
Silverberg, 2022	Gender (Male)	Sex (Male)
Sneller, 2022	Gender (Male)	Sex (Male)
Spinicci, 2022	Gender (Male)	Sex (Male)
Subramania, 2022	Gender (Male)	Sex (Male)
Sudre, 2021	Gender (Male)	Sex (Male)
Wu, 2022	Gender (Male)	Sex (Male)
Yellumahanthi, 2022	Gender (Male)	Sex (Male)
Yoo, 2022	Gender (Male)	Sex (Male)
Zayet, 2021	Gender (Male)	Sex (Male)
Zulu, 2022	Gender (Male)	Sex (Male)
Zuschlag, 2022	Gender (Male)	Sex (Male)
de Oliveria, 2022	Skin Lesion	Skin condition
Wu, 2022	Dry skin	Skin condition
Wu, 2022	Skin rash	Skin condition
Messin, 2021	Sleep apnea	Sleep disorder
Sadat Larijani, 2022	Sleep disorder	Sleep disorder
Bai, 2022	Smoker	Smoker
Bell, 2021	Smoker	Smoker
Bliddal, 2021	Smoker	Smoker
Buonsenso, 2022	Smoker	Smoker
Chan Sui Ko, 2022	Smoker	Smoker
de Oliveria, 2022	Smoking	Smoker
Desgranges, 2022	Smoker	Smoker
Förster, 2022	Smoker	Smoker
Garcia-Abellan, 2021	Smoker	Smoker
Helmsdal, 2022	Smoker	Smoker
Khalaf, 2021	Smoker	Smoker
Kim, 2022	Smoker	Smoker
Messin, 2021	Smoker	Smoker
Peghin, 2021	Smoker	Smoker
Senjam, 2021	Smoker	Smoker
Silverberg, 2022	Smoker	Smoker
Sneller, 2022	Smoker	Smoker
Subramania, 2022	Smoker	Smoker
Wu, 2022	Smoker	Smoker
Zayet, 2021	Smoker	Smoker
Messin, 2021	Sneezing	Sneezing
Wu, 2022	Sneezing	Sneezing

Augustin, 2021	Sore Throat	Sore Throat
Bai, 2022	Sore Throat	Sore Throat
de Oliveria, 2022	Sore Throat	Sore Throat
Ong, 2021	Sore throat	Sore Throat
Sadat Larijani, 2022	Dry throat	Sore throat
Wu, 2022	Sore throat	Sore Throat
Ong, 2021	Sputum	Sputum
Sadat Larijani, 2022	Sputum	Sputum
Kim, 2022	Stroke	Stroke
Subramania, 2022	Substance misuse	Substance misuse
Sadat Larijani, 2022	Sweating	Sweating
Sadat Larijani, 2022	Tachycardia	Tachycardia
Chudzik, 2022	Temperature 36.6-37.5	Temperature 36.6-37.5
Messin, 2021	Tinnitus	Tinnitus
Sadat Larijani, 2022	Weight loss	Weight loss

Table A3. Original and grouped categorical risk factors

Author, year	RiskFactorOriginal	RiskFactorNew
Asadi-Pooya, 2021	Age	Age
Augustin, 2021	Age	Age
Bai, 2022	Age	Age
Becker, 2021	Age	Age
Bell, 2021	Age	Age
Bliddal, 2021	Age	Age
Caze, 2022	Age	Age
Cervia, 2022	Age	Age
Chan Sui Ko, 2022	Age	Age
Chudzik, 2022	Age	Age
de Oliveria, 2022	Age	Age
Desgranges, 2022	Age	Age
Faycal, 2021	Age	Age
Förster, 2022	Age	Age
Garcia-Abellan, 2021	Age	Age
Helmsdal, 2022	Age	Age
Hill, 2022	Age	Age
Hossain, 2021	Age	Age
Khalaf, 2021	Age	Age
Kim, 2022	Age	Age
Knight, 2022	Age	Age
Messin, 2021	Age	Age
Naik, 2021	Age	Age
Ong, 2021	Age	Age
Sneller, 2022	Age	Age
Spinicci, 2022	Age	Age
Sudre, 2021	Age	Age
Vimercati, 2021	Age	Age
Zayet, 2021	Age	Age
Zuschlag, 2022	Age	Age
Zayet, 2021	Alanine U/L	Alanine U/L
Varghese, 2022	Albumin (g/dL)	Albumin (g/dL)
de Oliveria, 2022	Alkaline phosphatase (U/I)	Alkaline phosphatase
Varghese, 2022	Alkal. phosphatase (U/L)	Alkaline phosphatase
de Oliveria, 2022	Amylase (U/l)	Amylase (U/l)
Varghese, 2022	Antithrombin (%)	Antithrombin (%)
de Oliveria, 2022	Arterial oxygen pressure (mmHg)	Arterial oxygen pressure (mmHg)

Zayet, 2021	Aspartate U/L	Aspartate U/L
de Oliveria, 2022	AST (U/l)	AST (U/l)
Varghese, 2022	Basophile abs. (1000/uL)	Basophile abs. (1000/uL)
Varghese, 2022	Basophile rel (%)	Basophile rel (%)
de Oliveria, 2022	Direct bilirubin, mg/dl	Bilirubin direct (mg/dL)
de Oliveria, 2022	Total bilirubin, mg/dl	Bilirubin total (mg/dL)
Varghese, 2022	Bilirubin total (mg/dL)	Bilirubin total (mg/dL)
Bell, 2021	BMI	BMI
Bliddal, 2021	BMI	BMI
Caze, 2022	BMI	BMI
Cervia, 2022	BMI	BMI
Chan Sui Ko, 2022	BMI	BMI
Chudzik, 2022	BMI	BMI
Förster, 2022	BMI	BMI
Helmsdal, 2022	BMI	BMI
Messin, 2021	BMI	BMI
Sneller, 2022	BMI	BMI
Sudre, 2021	BMI	BMI
Vimercati, 2021	BMI	BMI
Varghese, 2022	Creatine kinase (U/L)	Creatine kinase (U/L)
Bai, 2022	Creatinine (mg/dL)	Creatinine
de Oliveria, 2022	Creatinine, mg/dL	Creatinine
Varghese, 2022	Creatinine, mg/dL	Creatinine
Zayet, 2021	Creatinine, umol/L	Creatinine
Bai, 2022	CRP (mg/L)	CRP
Cervia, 2022	CRP (mg/L)	CRP
Chan Sui Ko, 2022	CRP (mg/L)	CRP
de Oliveria, 2022	CRP (mg/L)	CRP
Garcia-Abellan, 2021	CRP (mg/L)	CRP
Varghese, 2022	CRP (mg/dL)	CRP
Zayet, 2021	CRP (mg/L)	CRP
Zuschlag, 2022	CRP (mg/L)	CRP
Bai, 2022	D dimer, mg/L	D-dimer
de Oliveria, 2022	D-dimer (u/ml)	D-dimer
Garcia-Abellan, 2021	D-dimer (mcg/mL)	D-dimer
Vimercati, 2021	DBP	DBP
Varghese, 2022	eGFR - CKD EPI (mL/min)	eGFR - CKD EPI (mL/min)
Varghese, 2022	Eosinoph. abs. (1000/uL)	Eosinoph. abs. (1000/uL)
Varghese, 2022	Eosiophile relative (%)	Eosiophile relative (%)

Varghese, 2022	Erythrocytes/RBC (Mio./uL)	Erythrocytes/RBC (Mio./uL)
Vimercati, 2021	Fasting glucose, mmol/L	Fasting glucose, mmol/L
Garcia-Abellan, 2021	Ferritin (ng/mL)	Ferritin
Varghese, 2022	Ferritin (ug/L)	Ferritin
Varghese, 2022	Fibrinogen (mg/dL	Fibrinogen (mg/dL
Lui, 2021	fT3, pmol/L	fT3, pmol/L
Lui, 2021	fT4, pmol/L	fT4, pmol/L
Varghese, 2022	Glucose (mg/dL)	Glucose (mg/dL)
Vimercati, 2021	HDL cholesterol, mmol/L	HDL cholesterol, mmol/L
Varghese, 2022	Hematokrit (%)	Hematocrit (%)
Bai, 2022	Hb, g/dL	Hemoglobin
de Oliveria, 2022	Hemoglobin, g/dl	Hemoglobin
Varghese, 2022	Hemoglobin, g/dl	Hemoglobin
Zayet, 2021	Hemoglobin, g/dl	Hemoglobin
Bai, 2022	GPT, UI/L	Hepatic transaminases, ALT
de Oliveria, 2022	ALT (U/l)	Hepatic transaminases, ALT
Varghese, 2022	GPT/ALT (U/L)	Hepatic transaminases, ALT
Bai, 2022	GOT, UI/L	Hepatic transaminases, AST
Varghese, 2022	GOT/AST (U/L)	Hepatic transaminases, AST
de Oliveria, 2022	GGT (U/l)	Hepatic transaminases, GGT
Varghese, 2022	Gamma-GT (U/L	Hepatic transaminases, GGT
Varghese, 2022	IgA (mg/dL)	IgA (mg/dL)
Varghese, 2022	IgG (mg/dL)	IgG (mg/dL)
Varghese, 2022	IgM (mg/dL)	IgM (mg/dL)
Cervia, 2022	IL-6, pg/mL	IL-6
Garcia-Abellan, 2021	IL-6, pg/mL	IL-6
Varghese, 2022	Interleukin-6 (pg/mL)	IL-6
Varghese, 2022	Immat. Granulocytes (1000/uL)	Immat. Granulocytes (1000/uL)
Varghese, 2022	INR (ratio)	INR (ratio)
de Oliveria, 2022	Lactate (mmol/l)	Lactate (mmol/l)
Bai, 2022	LDH, UI/L	LDH
de Oliveria, 2022	LDH (U/l)	LDH
Varghese, 2022	LDH (U/L)	LDH
Vimercati, 2021	LDL cholesterol, mmol/L	LDL cholesterol, mmol/L
Asadi-Pooya, 2021	Length of hospital stay	Length of hospital stay
Bai, 2022	Length of hospital stay	Length of hospital stay
Becker, 2021	Length of hospital stay	Length of hospital stay
de Oliveria, 2022	Length of hospital stay	Length of hospital stay
Garcia-Abellan, 2021	Length of hospital stay	Length of hospital stay

Bai, 2022	WBC x 10^3/mm^3	Leukocytes
Cervia, 2022	Leukocytes (10^9/L)	Leukocytes
de Oliveria, 2022	Leukocytes, cells/mm^3	Leukocytes
Varghese, 2022	Leukocytes, (Tsd./uL)	Leukocytes
Zayet, 2021	WBC x 10^3/mm^3	Leukocytes
Zuschlag, 2022	WBC/ mm^3	Leukocytes
de Oliveria, 2022	Lipase (U/l)	Lipase
Varghese, 2022	Lipase (U/L)	Lipase
Garcia-Abellan, 2021	Lymphocyte nadir count	Lymphocyte nadir count
Bai, 2022	Lymphocytes, cells/mm^3	Lymphocytes
Cervia, 2022	Lymphocytes (10^9/L)	Lymphocytes
de Oliveria, 2022	Lymphocytes, cells/mm^3	Lymphocytes
Lui, 2021	Lymphocytes/mm^3	Lymphocytes
Varghese, 2022	Lymphocytes abs. (1000/uL)	Lymphocytes
Varghese, 2022	Lymphocytes relative (%)	Lymphocytes relative (%)
Varghese, 2022	MCH (pg)	MCH (pg)
Varghese, 2022	MCHC (g/dL)	MCHC (g/dL)
Varghese, 2022	MCV (fL)	MCV (fL)
Varghese, 2022	Monozyten relative (%)	Monocytes
Varghese, 2022	Monocytes abs. (1000/uL)	Monocytes abs. (1000/uL)
Chan Sui Ko, 2022	Nadir lymphocyte /mm^3	Nadir lymphocyte /mm^3
Cervia, 2022	Neutrophils (10^9/L)	Neutrophil
de Oliveria, 2022	Neutrophil count, cells/mm^3	Neutrophil
Varghese, 2022	Neutrophils abs. (1000/uL)	Neutrophils abs. (1000/uL)
Varghese, 2022	Neutrophils relative (%)	Neutrophils relative (%)
Augustin, 2021	Number of symptoms	Number of symptoms
Caze, 2022	Number of symptoms	Number of symptoms
Cervia, 2022	Number of symptoms	Number of symptoms
Chudzik, 2022	Number of symptoms	Number of symptoms
Helmsdal, 2022	Number of symptoms	Number of symptoms
Sudre, 2021	Number of symptoms	Number of symptoms
Varghese, 2022	Part. Thrombopl. Time (s)	Part. Thrombopl. Time (s)
Varghese, 2022	Peudocholinesterase (U/L)	Peudocholinesterase (U/L)
Bai, 2022	PLT x 10^3/mm^3	Platelet
de Oliveria, 2022	Platlet count, cells/mm^3	Platelet
Varghese, 2022	Platelets (1000/uL)	Platelet
Varghese, 2022	Potassium (mmol/L)	Potassium (mmol/L)
Varghese, 2022	Procalcitionin (ng/mL)	Procalcitionin (ng/mL)
de Oliveria, 2022	Prothrombin time (INR)	Prothrombin time (INR)

Vimercati, 2021	SBP	SBP
Varghese, 2022	Sodium (mmol/L)	Sodium (mmol/L)
Varghese, 2022	Thrombin time (s)	Thrombin time (s)
Cervia, 2022	TNF. pg/mL	TNF. pg/mL
Vimercati, 2021	Total cholesterol, mmol/L	Total cholesterol, mmol/L
Varghese, 2022	Total Protein (g/dL)	Total Protein (g/dL)
Vimercati, 2021	Triglycerides, mmol/L	Triglycerides, mmol/L
Varghese, 2022	TSH (uU/mL)	TSH
Zayet, 2021	TSH, MIU/L	TSH
de Oliveria, 2022	Urea (mg/dL)	Urea
Varghese, 2022	Urea, mg/dl	Urea

Table A4. Assessment of study quality using QUIPS criteria adapted for this study

	Risk of Bias		
QUIPS Domains	High	Moderate	Low
Study Participation	Specialized population	Only inpatients or only outpatients	Inpatient & Outpatient
Study Attrition	Incomplete outcome ascertainment	Complete ascertainment but >10% loss to follow-up	Complete ascertainment and <10% loss to follow-up
Prognostic Factor Measurement	Unclear definition of prognostic factors or post-hoc cutoffs	Common cutoffs used, assumed prespecified	Clearly defined and prespecified cutoffs
Outcome Measurement	Outcome not defined	Unclear definition	Clear and reproducible definition
Study Confounding	No multivariate analysis performed		Multivariate analysis done
Statistical Analysis and Reporting	Selective reporting of results, no clear analytic strategy		Full reporting, analytic strategy clearly described

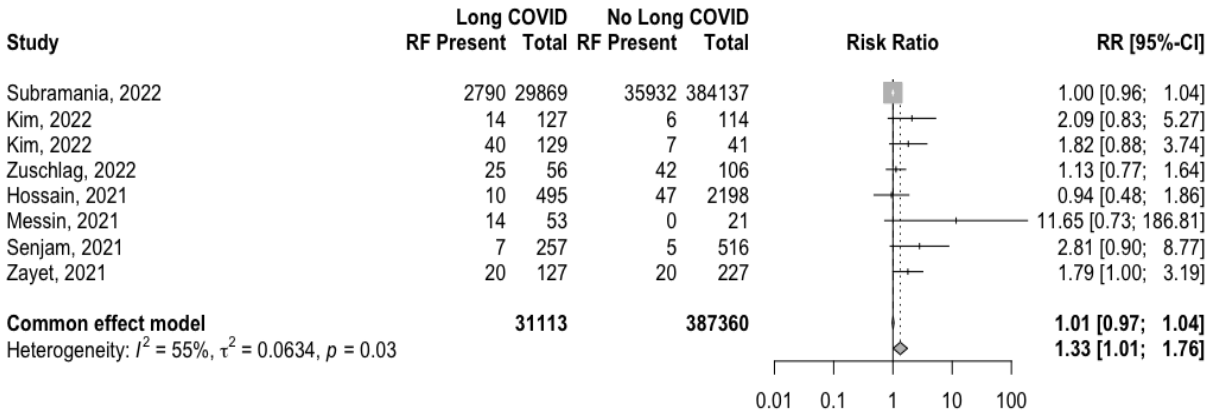

Figure A1. Forest plot for risk factor: Age 60-70 years

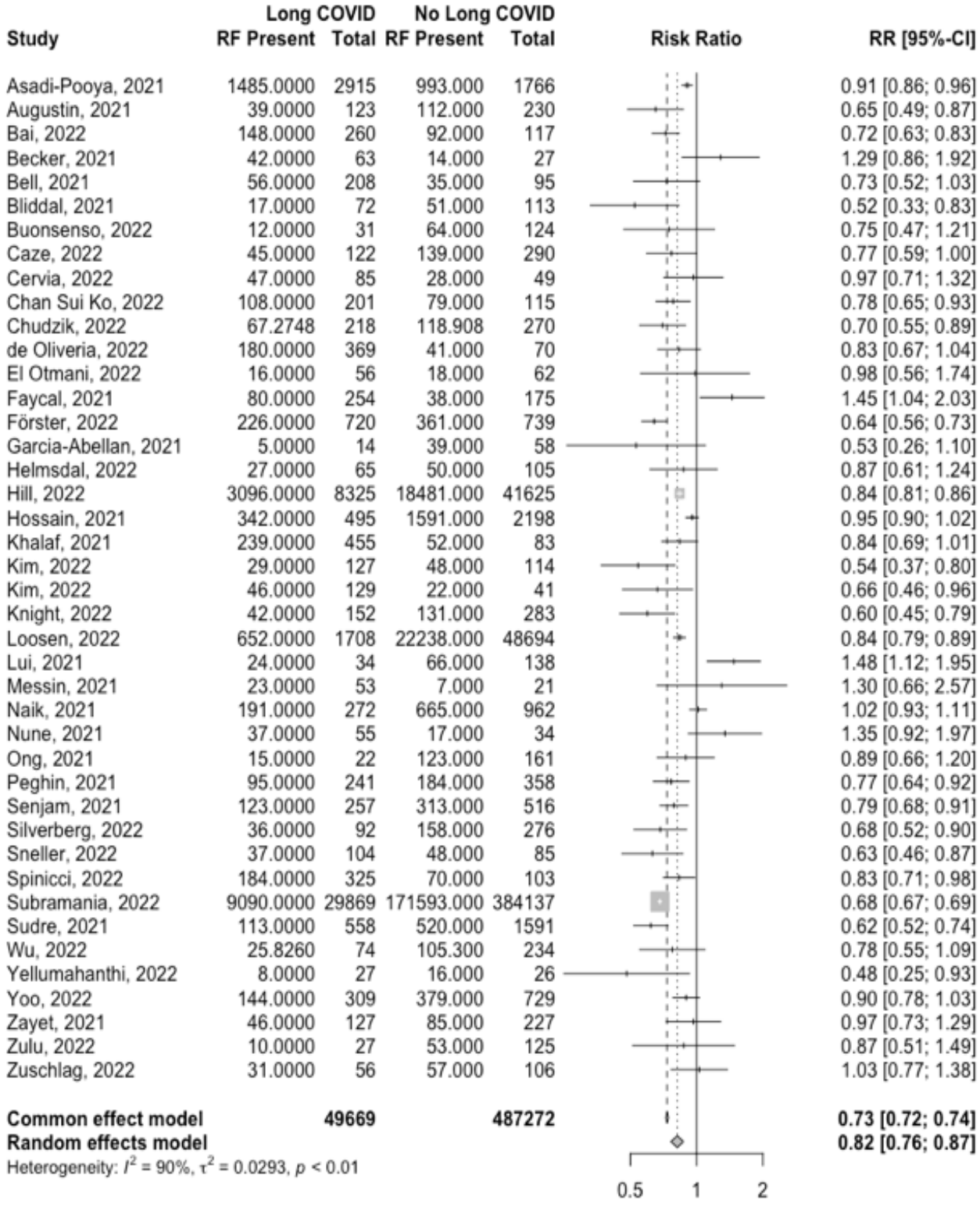

Figure A2. Forest plot for risk factor: Sex (Male)

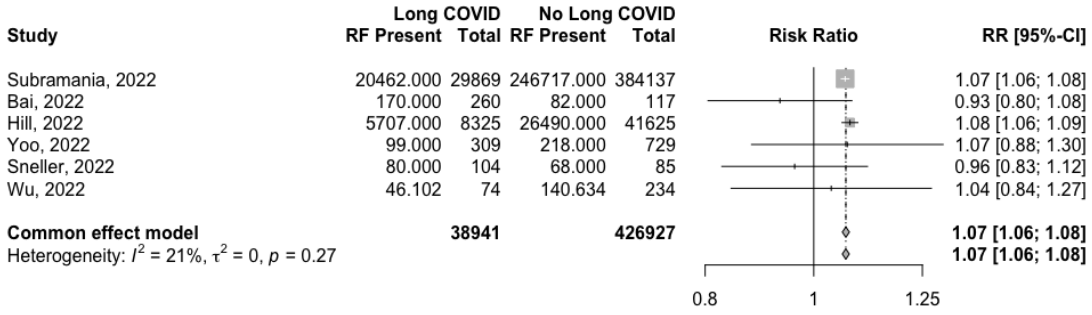

Figure A3. Forest plot for risk factor: Race-White

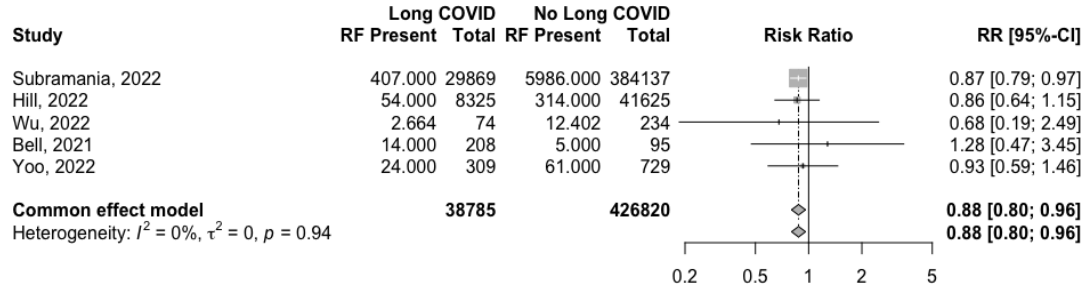

Figure A4. Forest plot for risk factor: Race-Other

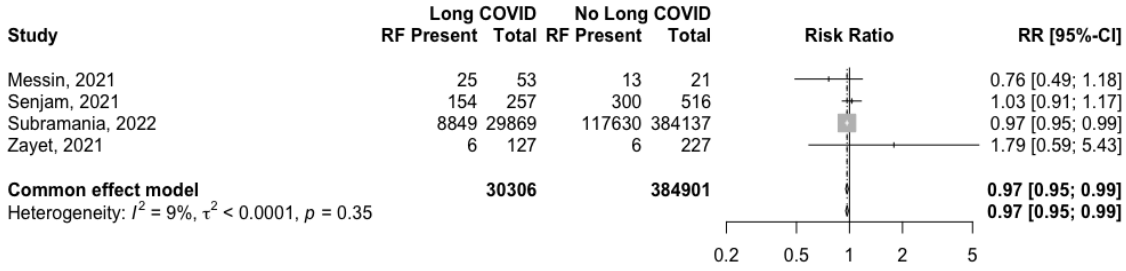

Figure A5. Forest plot for risk factor: BMI 18.5-25 kg/m^2

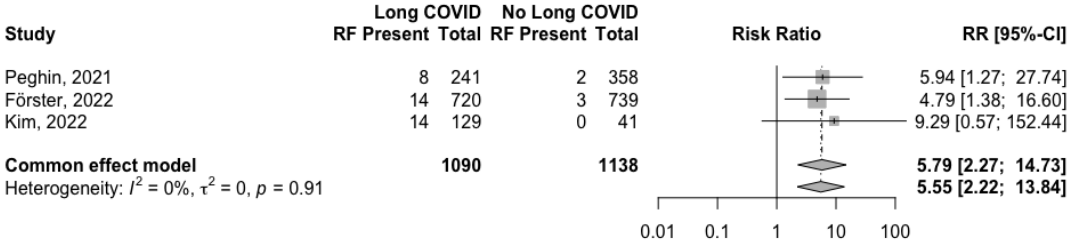

Figure A6. Forest plot for risk factor: Liver disease

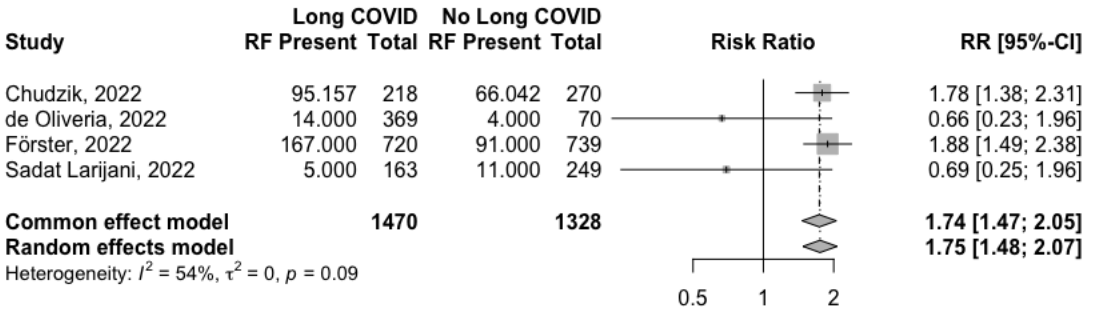

Figure A7. Forest plot for risk factor: Arthritis

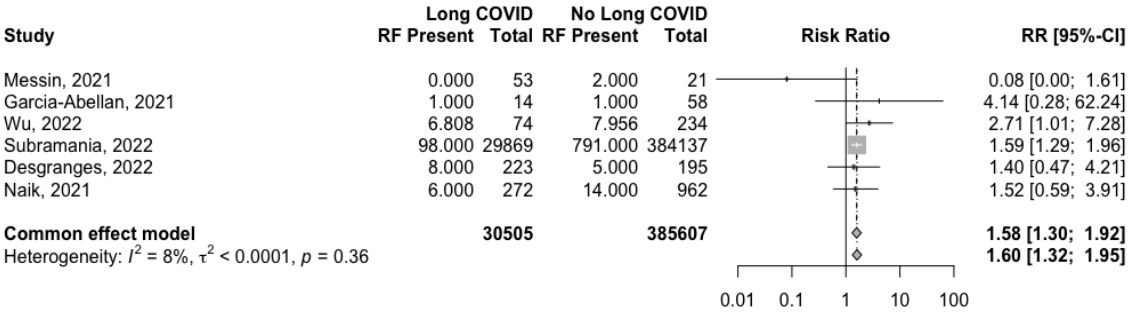

Figure A8. Forest plot for risk factor: Autoimmune disorder

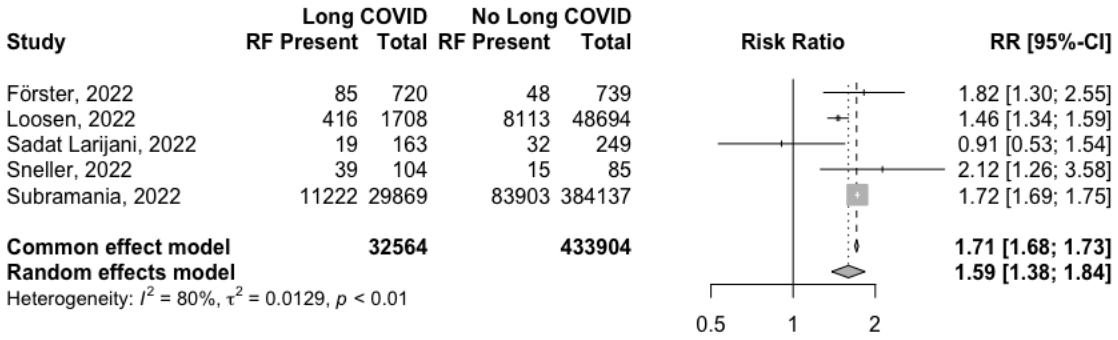

Figure A9. Forest plot for risk factor: Mental health disorder

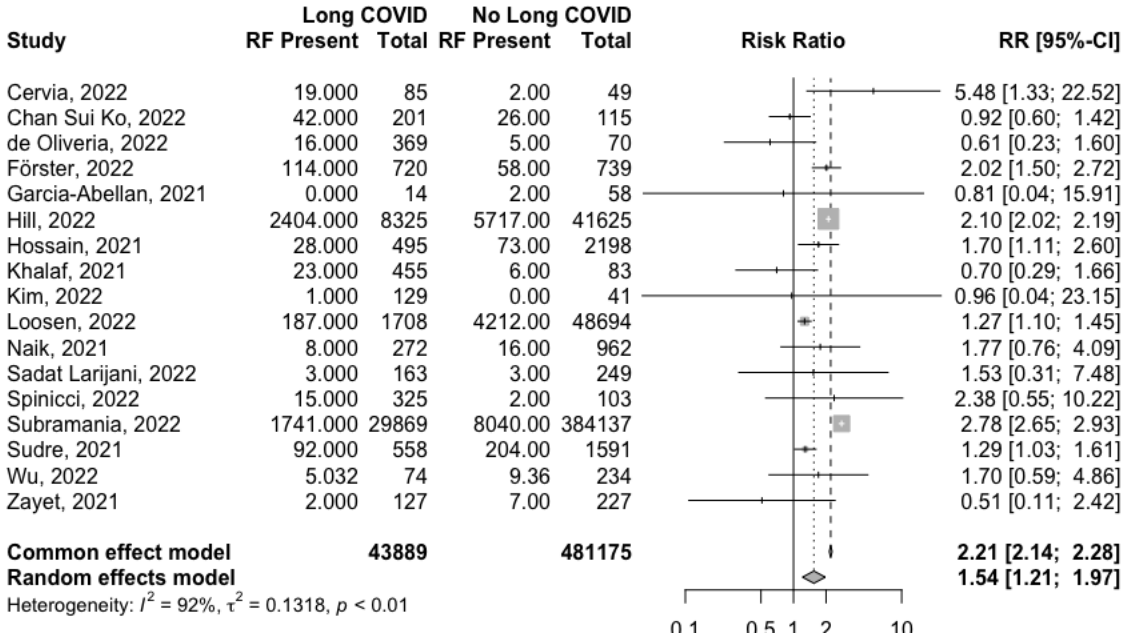

Figure A10. Forest plot for risk factor: Chronic lung disease

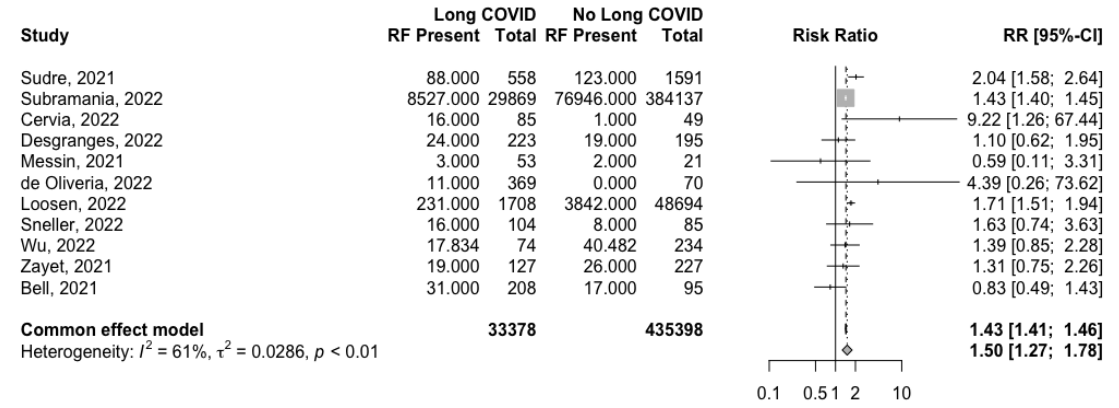

Figure A11. Forest plot for risk factor: Asthma

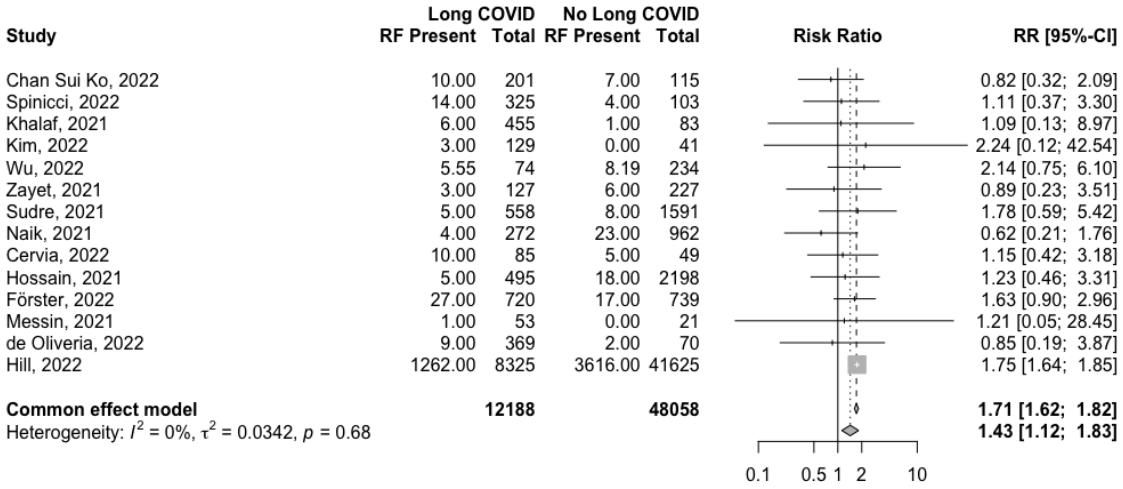

Figure A12. Forest plot for risk factor: Kidney disease

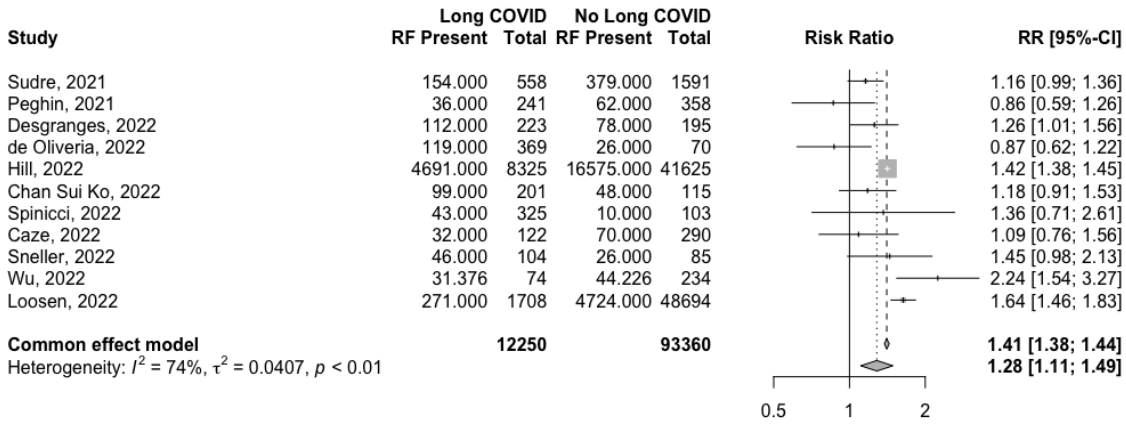

Figure A13. Forest plot for risk factor: Obesity

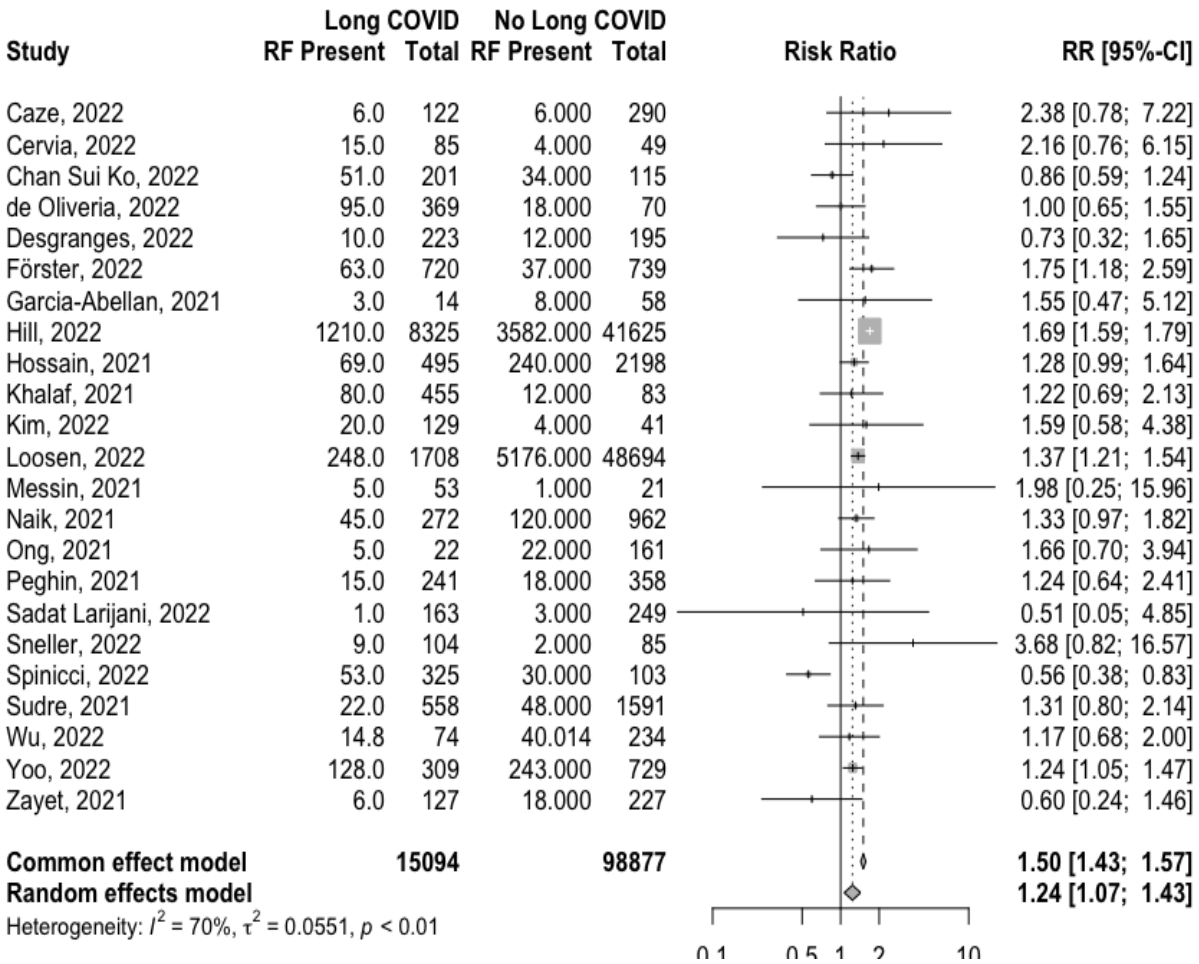

Figure A14. Forest plot for risk factor: Diabetes

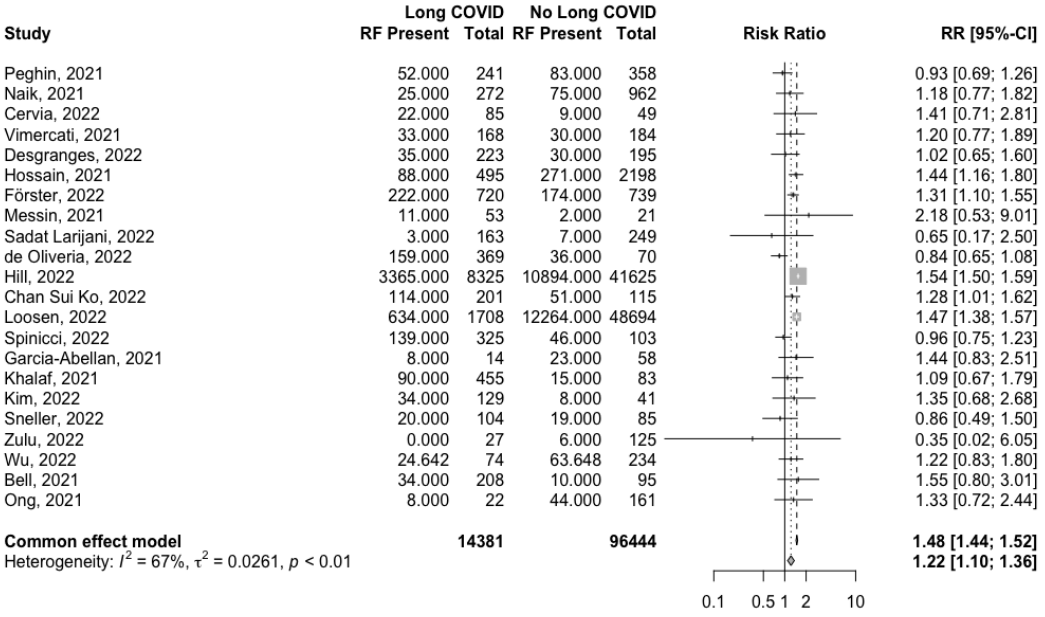

Figure A15. Forest plot for risk factor: Hypertension

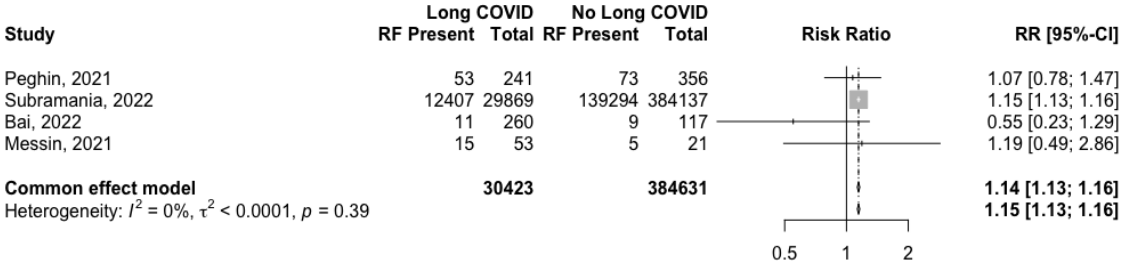

Figure A16. Forest plot for risk factor: Former smoker

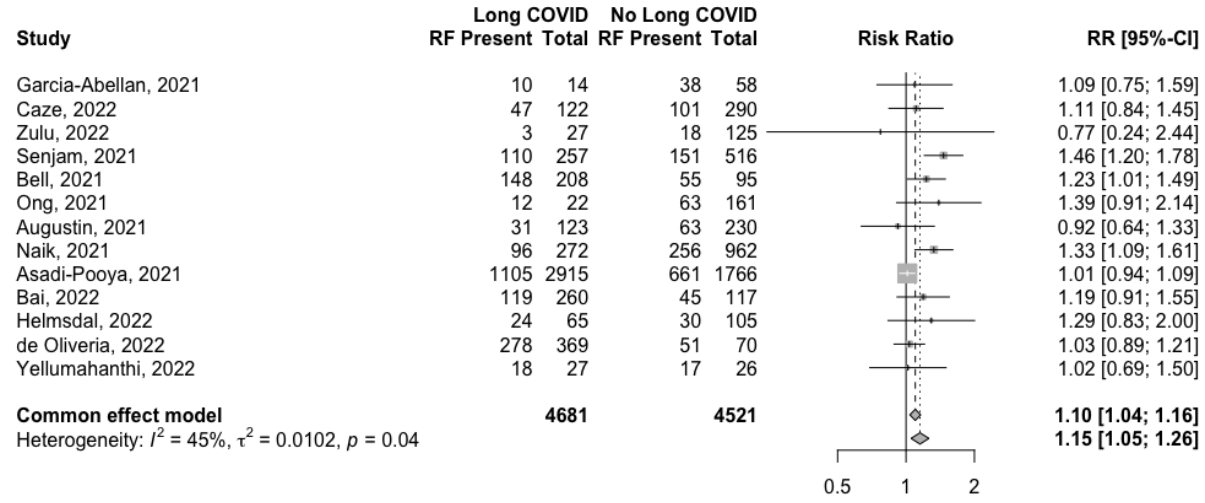

Figure A17. Forest plot for risk factor: Any comorbidity

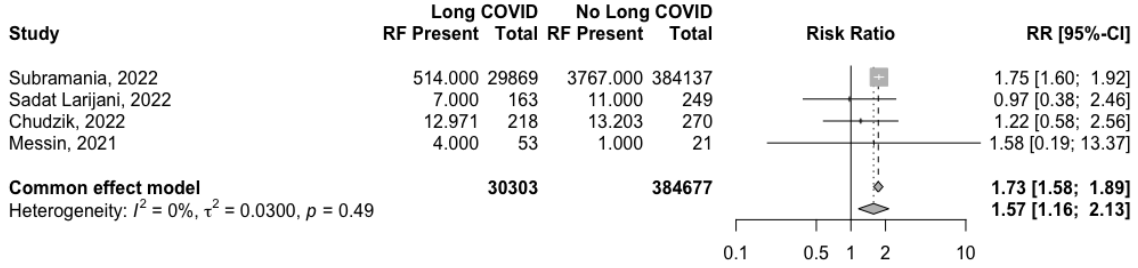

Figure A18. Forest plot for risk factor: Hearing loss

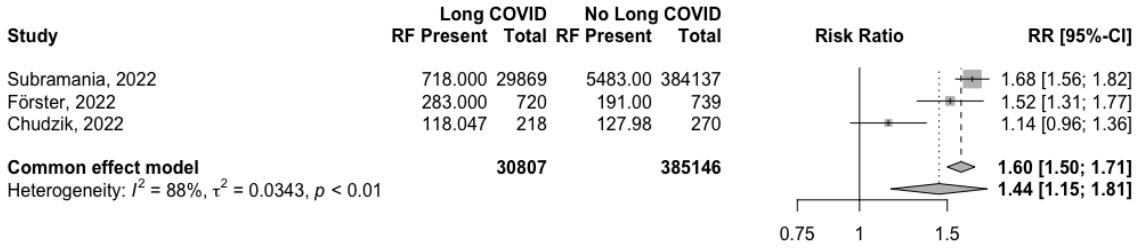

Figure A19. Forest plot for risk factor: Back pain

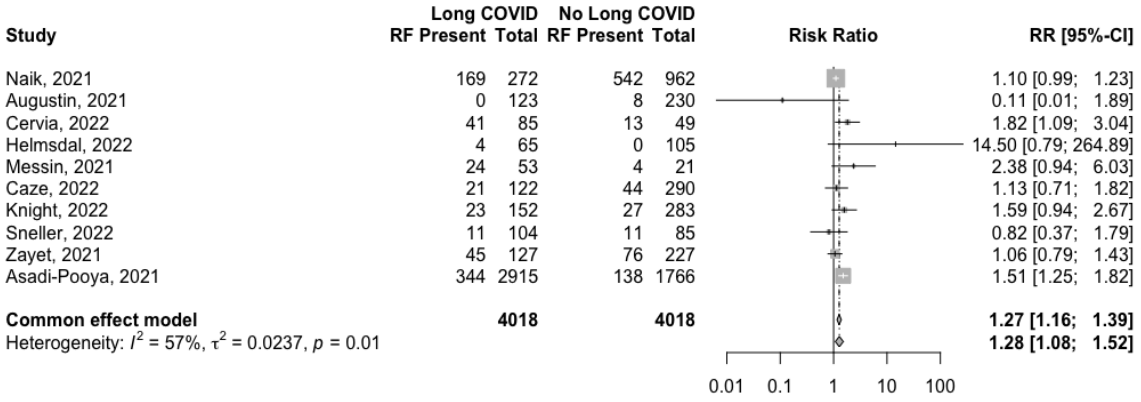

Figure A20. Forest plot for risk factor: Hospitalized

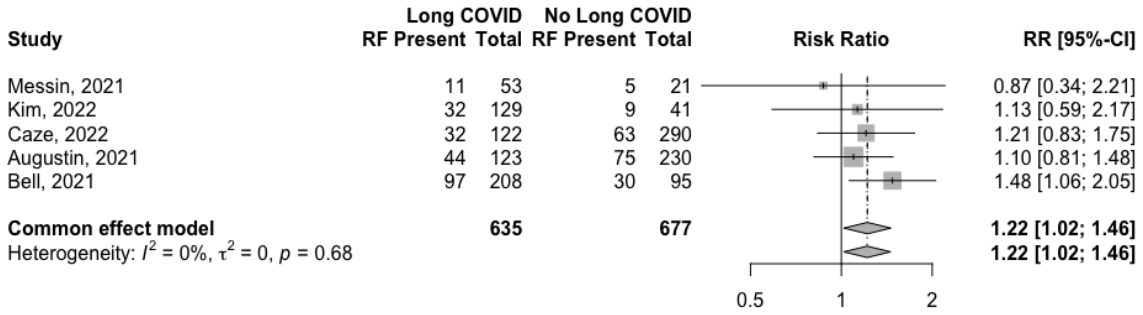

Figure A21. Forest plot for risk factor: Allergic rhinitis

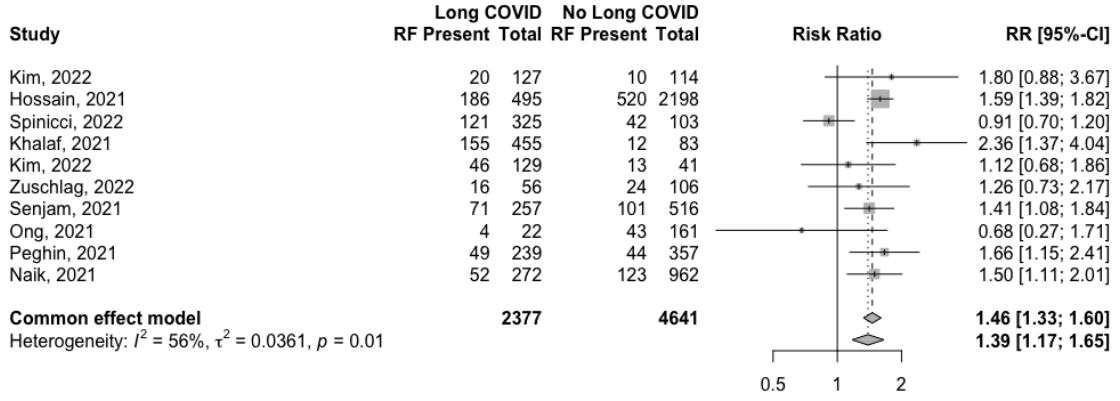

Figure A22. Forest plot for risk factor: Severity-Moderate

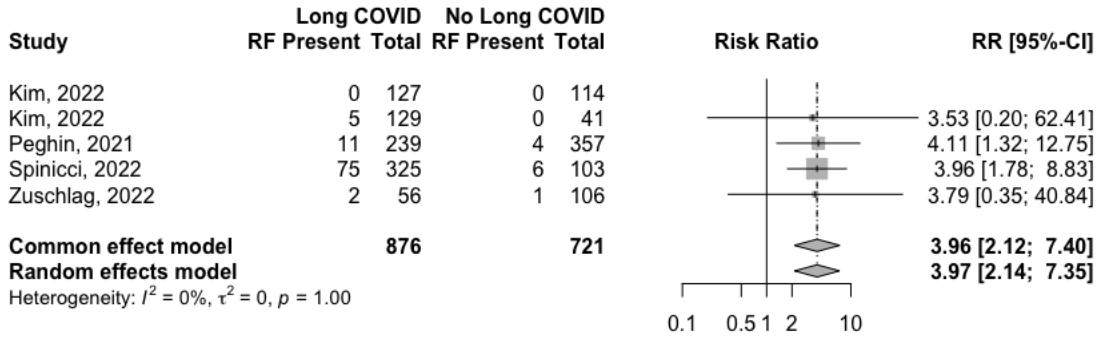

Figure A23. Forest plot for risk factor: Severity-Critical

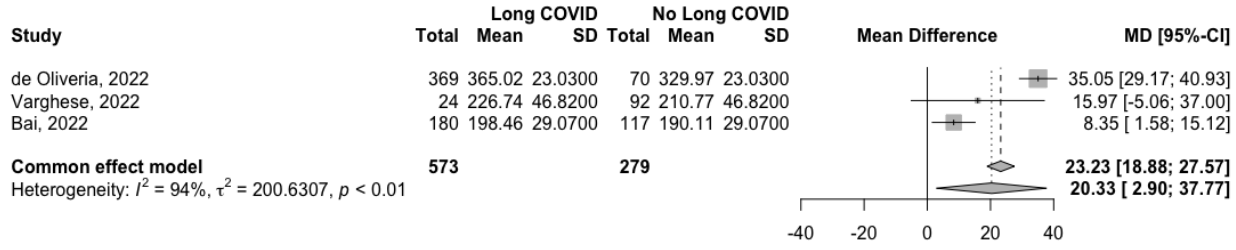

Figure A24. Forest plot for risk factor: LDH (U/L)

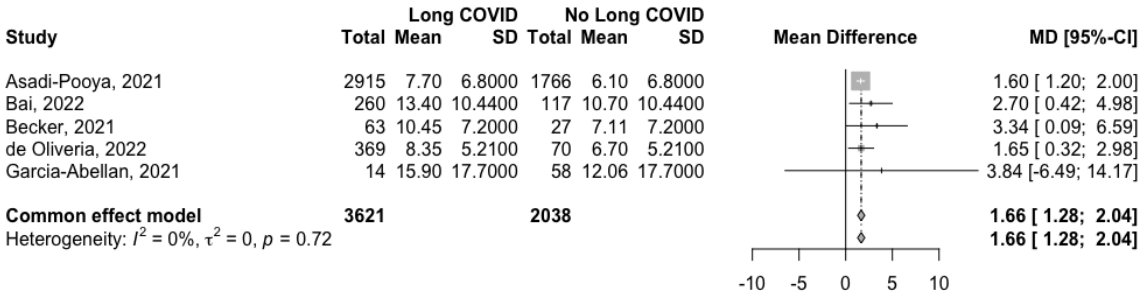

Figure A25. Forest plot for risk factor: Length of hospital stay (days)

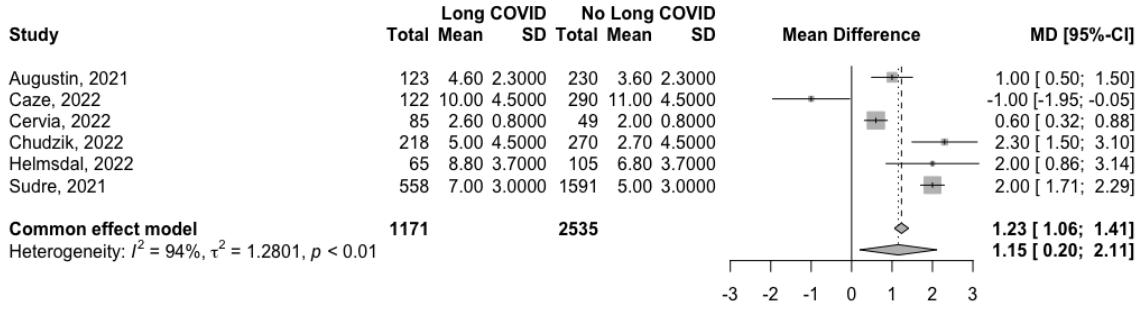

Figure A26. Forest plot for risk factor: Number of symptoms

Table A4. Categorical risk factors by study setting

Risk Factor	Studies	Patients	RR (95% CI)
Inpatient			
Sex (Male)	12	7889	0.96 (0.84-1.10)
Loss of smell	3	890	1.11 (0.85-1.45)
Fever	4	1073	1.09 (0.88-1.34)
Dyspnea	4	1073	1.05 (0.90-1.22)
Myalgia	4	1073	1.02 (0.81-1.29)
Cough	4	1073	0.95 (0.83-1.09)
Headache	3	890	0.91 (0.73-1.12)
Sore throat	3	999	0.87 (0.55-1.35)
Fatigue	3	890	0.84 (0.74-0.96)
Nausea/Vomiting	3	890	0.78 (0.45-1.35)
Hospitalized	3	5989	1.35 (0.99-1.85)
Severity: Moderate*	3	1579	1.35 (1.04-1.76)
Severity: Mild	3	1579	0.93 (0.87-0.99)
Any Comorbidity	6	6986	1.12 (1.00-1.25)
Smoker	5	1278	1.19 (0.78-1.80)
Hypertension	6	2318	1.16 (0.92-1.45)
Diabetes	6	2318	1.13 (0.88-1.44)
Autoimmune condition	3	1380	1.02 (0.18-5.80)
Chronic lung disease	4	2061	0.98 (0.67-1.43)
Heart disease	6	2318	0.88 (0.66-1.17)
Kidney disease	4	2063	0.76 (0.41-1.41)
Outpatient			
Sex (Male)	15	472261	0.78 (0.69-0.88)
Age 30-40	3	417128	1.07 (0.90-1.27)
Age 50-60	3	417128	1.07 (0.90-1.27)
Age ≥ 65*	5	419269	1.56 (1.50-1.61)
Cough	3	1270	1.04 (0.92-1.18)
Fever	3	1270	0.98 (0.87-1.09)
Headache*	4	415276	1.33 (1.10-1.59)
Loss of smell	3	1270	1.02 (0.73-1.44)
Loss of taste	3	1270	1.15 (0.68-1.93)
Any Comorbidity	4	861	1.10 (0.91-1.33)
Asthma*	6	467412	1.51 (1.18-1.94)
Cancer*	3	50954	1.53 (1.30-1.80)
Chronic lung disease*	6	469812	1.82 (1.26-2.64)
Diabetes	6	56224	1.09 (0.77-1.55)
Heart disease	6	56224	1.21 (0.86-1.68)

Hypertension*	8	54882	1.27 (1.07-1.51)
Kidney disease	4	5404	1.28 (0.76-2.16)
Obesity*	4	53397	1.35 (1.11-1.64)
Smoker*	5	415280	1.06 (1.04-1.08)
Inpatient and Outpatient			
Sex (Male)	15	56791	0.77 (0.71-0.83)
Age 15-34	6	2120	0.96 (0.47-1.95)
Age 30-40	4	1177	1.03 (0.81-1.31)
Age 40-50	4	1177	1.03 (0.81-1.31)
Age 60-70*	4	1538	1.94 (1.32-2.85)
Age ≥ 65	3	1251	1.41 (0.48-4.19)
Race: Black	4	51485	1.21 (0.67-2.19)
Race: Hispanic	3	51296	1.01 (0.79-1.31)
Race: Other	3	51296	0.87 (0.69-1.11)
Race: White*	4	51485	1.07 (1.02-1.11)
Hospitalized	5	1560	1.15 (0.92-1.43)
Any Comorbidity*	3	1355	1.30 (1.07-1.59)
Asthma	3	851	1.40 (1.00-1.95)
Cancer	4	2291	1.00 (0.97-1.04)
Chronic lung disease*	7	53191	1.59 (1.08-2.35)
Diabetes*	8	55429	1.43 (1.20-1.70)
Drinks alcohol	3	1524	0.89 (0.79-1.00)
Fatigue	3	875	1.03 (0.36-2.95)
Heart disease	9	53979	1.11 (0.71-1.73)
Hypertension*	8	53625	1.22 (1.02-1.46)
Kidney disease*	6	52779	1.74 (1.64-1.85)
Liver disease*	3	2228	5.55 (2.22-13.84)
Mental health problem	3	2060	1.54 (0.96-2.50)
Obesity*	3	51458	1.34 (1.01-1.80)
Smoker	10	4713	0.83 (0.67-1.02)
Severity: Asymptomatic	4	1780	0.17 (0.06-0.43)
Severity: Mild*	5	2318	0.95 (0.83-1.08)
Severity: Moderate	5	2318	1.53 (1.28-1.83)
Severity: Severe	5	2318	1.45 (0.85-2.46)
Severity: Critical*	3	1007	4.02 (1.40-11.54)

*indicates risk factors significantly associated with long COVID

Table A5. Continuous risk factors by setting

Risk factor	Studies	Patients	WMD (95% CI)
Inpatient			
Age	10	7628	0.83 (-1.67-3.33)
LDH*	3	852	20.33 (2.90-37.77)
Platelet	3	932	6.26 (-4.24-16.76)
Length of hospital stay*	5	5659	1.66 (1.28-2.04)
Leukocytes	4	1094	0.18 (-0.10-0.45)
Creatinine	3	932	-0.00 (-0.03- 0.03)
CRP	6	1482	0.00 (-0.02-0.02)
D-dimer	3	888	0.06 (-0.07-0.18)
Lymphocytes	4	1104	-0.02 (-0.09-0.06)
Hemoglobin	3	932	-0.29 (-0.79-0.21)
Hepatic transaminases, ALT	3	932	-0.45 (-1.64-0.73)
Outpatient			
Age	11	7932	1.91 (-0.70-4.52)
Number of symptoms*	4	3124	1.44 (0.66-2.22)
BMI*	6	3611	0.95 (0.56-1.34)
Inpatient and Outpatient			
Age*	9	53677	3.03 (0.30-5.75)
BMI	4	2230	0.39 (-0.56-1.34)

WMD=Weighted Mean Difference

*indicates risk factors significantly associated with long COVID

Table A6. Categorical risk factors by long COVID definition

Risk factor	Studies	Patients	RR (95% CI)
Length of Prolonged Symptoms = 1 month			
Sex (Male)	16	8074	0.81 (0.72-0.91)
Age ≥ 65	3	3289	1.24 (0.48-3.22)
Allergic rhinitis*	3	1068	1.24 (1.03-1.50)
Fatigue	3	1280	0.66 (0.40-1.11)
Fever	4	1633	0.52 (0.13-2.12)
Headache	4	1633	0.70 (0.32-1.55)
Loss of smell	4	1633	0.96 (0.68-1.34)
Loss of taste	4	1633	1.05 (0.66-1.69)
Cough	4	1633	0.77 (0.49-1.22)
Dyspnea	3	1280	0.76 (0.49-1.17)
Myalgia	4	1633	0.57 (0.18-1.75)
Hospitalized	5	2568	1.24 (0.99-1.56)
Any Comorbidity*	7	3666	1.19 (1.05-1.35)
Asthma	4	3025	1.95 (0.77-4.94)
Chronic lung disease*	6	4796	1.36 (1.03-1.80)
Heart disease	6	4796	0.91 (0.55-1.50)
Hypertension	8	3454	1.01 (0.86-1.19)
Kidney disease	5	4384	1.05 (0.63-1.73)
Obesity	4	3428	1.10 (0.96-1.27)
Diabetes	7	5208	1.12 (0.77-1.63)
Smoker	5	2068	0.88 (0.66-1.16)
Severity: Asymptomatic	3	1363	0.37 (0.27-0.52)
Severity: Mild	4	2597	1.02 (0.89-1.18)
Severity: Moderate*	4	4010	1.61 (1.42-82)
Severity: Severe	3	1363	1.19 (0.54-2.63)
Severity: Critical	4	2597	1.24 (0.97-1.60)
Length of Prolonged Symptoms = 3 months			
Sex (Male)	16	526168	0.83 (0.75-0.92)
Age 15-34	3	416869	1.54 (0.50-4.74)
Age ≥ 65	5	417230	1.24 (0.74-2.09)
Race: Asian	3	464333	0.90 (0.71-1.13)
Race: Black	4	464641	1.33 (0.68-2.60)
Race: Hispanic	3	50635	1.09 (0.74-1.60)
Race: Other	3	464264	0.87 (0.79-0.96)
Race: White*	4	464641	1.07 (1.06-1.08)
Back pain*	3	415953	1.44 (1.15-1.81)
Fatigue*	3	1173	1.13 (1.01-1.26)
Fever	4	1356	1.09 (0.85-1.41)
GI problems	3	420146	1.44 (0.99-2.10)
Headache*	4	415179	1.51 (1.25-1.83)
Mental health problem*	3	465867	1.62 (1.42-1.85)

Nausea/Vomiting	3	1173	0.61 (0.12-3.08)
Respiratory problems	4	6076	0.97 (0.81-1.17)
Sore throat	3	868	1.08 (0.56-2.08)
Cough	4	1356	1.06 (0.91-1.24)
Myalgia	3	868	1.14 (0.90-1.46)
Any Comorbidity	5	5464	1.09 (0.96-1.24)
Asthma*	4	465134	1.50 (1.30-1.74)
Autoimmune condition*	3	414732	1.62 (1.33-1.98)
Cancer*	4	52587	1.35 (1.01-1.80)
Chronic lung disease*	7	519356	1.79 (1.35-2.37)
Former smoker*	3	414980	1.15 (1.13-1.16)
Heart disease*	9	106550	1.57 (1.33-1.85)
Hypertension*	9	106550	1.34 (1.19-1.50)
Kidney disease*	5	54948	1.74 (1.64-1.85)
Neurologic disease	3	5634	1.05 (0.92-1.19)
Obesity*	5	101677	1.42 (1.10-1.83)
Severity: Mild	4	4010	0.84 (0.71-1.00)
Severity: Severe	4	4010	1.46 (0.91-2.34)
Smoker	8	417873	1.01 (0.80-1.26)
Diabetes*	9	106550	1.43 (1.25-1.64)
Dyspnea*	4	1356	1.35 (1.12-1.63)
Length of Prolonged Symptoms = 6 months			
Sex (Male)	5	911	0.75 (0.55-1.01)
Age 15-34	4	839	0.81 (0.54-1.23)
Age 40-50	4	839	0.83 (0.59-1.17)
Age 50-60	4	839	1.05 (0.73-1.52)
Age 60-70*	4	839	1.93 (1.29-2.88)
Cancer	4	670	1.00 (0.96-1.04)
Chronic lung disease	3	596	0.61 (0.17-2.17)
Heart disease	4	670	0.74 (0.52-1.05)
Hypertension	3	316	1.46 (0.97-2.20)
Kidney disease	3	598	1.07 (0.34-3.40)
Smoker	4	670	1.71 (0.42-7.03)
Diabetes	4	670	1.12 (0.59-2.12)

*indicates risk factors significantly associated with long COVID

Table A7. Continuous risk factors by long COVID definition

Risk factor	Studies	Patients	WMD (95% CI)
Length of Prolonged Symptoms = 1 month			
Age	13	7015	1.45 (-1.34-4.25)
BMI	6	3535	0.75 (-0.01-1.51)
Number of symptoms	4	3048	-0.71 (-0.47-1.88)
Lymphocytes	3	689	-0.13 (-0.43-0.17)
Leukocytes	4	851	0.25 (-0.02-0.53)
CRP	4	851	9.81 (-1.00-20.62)
Length of Prolonged Symptoms = 3 months			
Age*	10	60957	2.95 (1.14-4.77)
BMI	3	2117	0.84 (0.02-1.65)
Length of Prolonged Symptoms = 6 months			
Age	4	670	2.86 (-2.02-7.74)

WMD= Weighted Mean Difference
*indicates risk factors significantly associated with long COVID

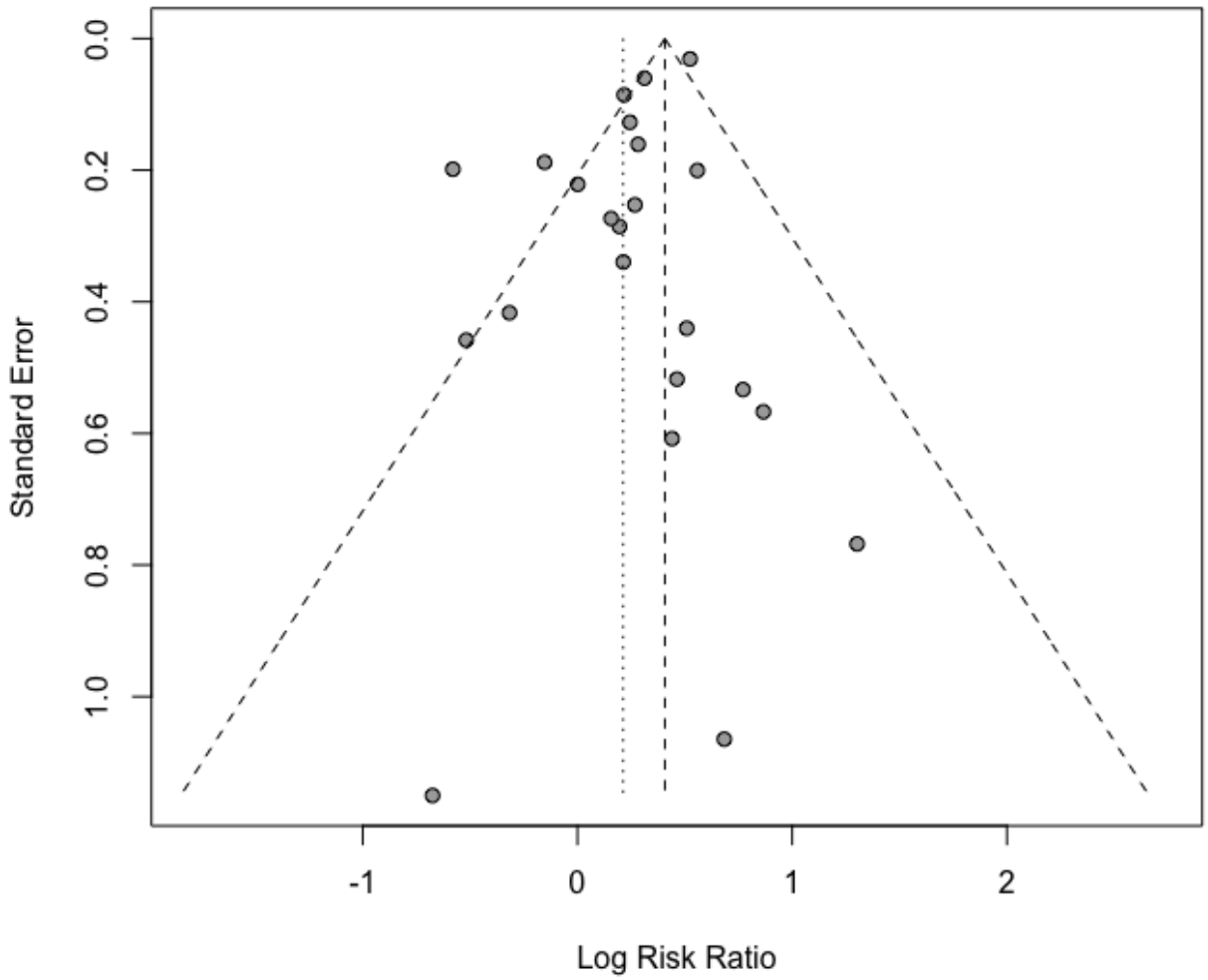

Figure A27. Funnel plot for risk factor: Diabetes

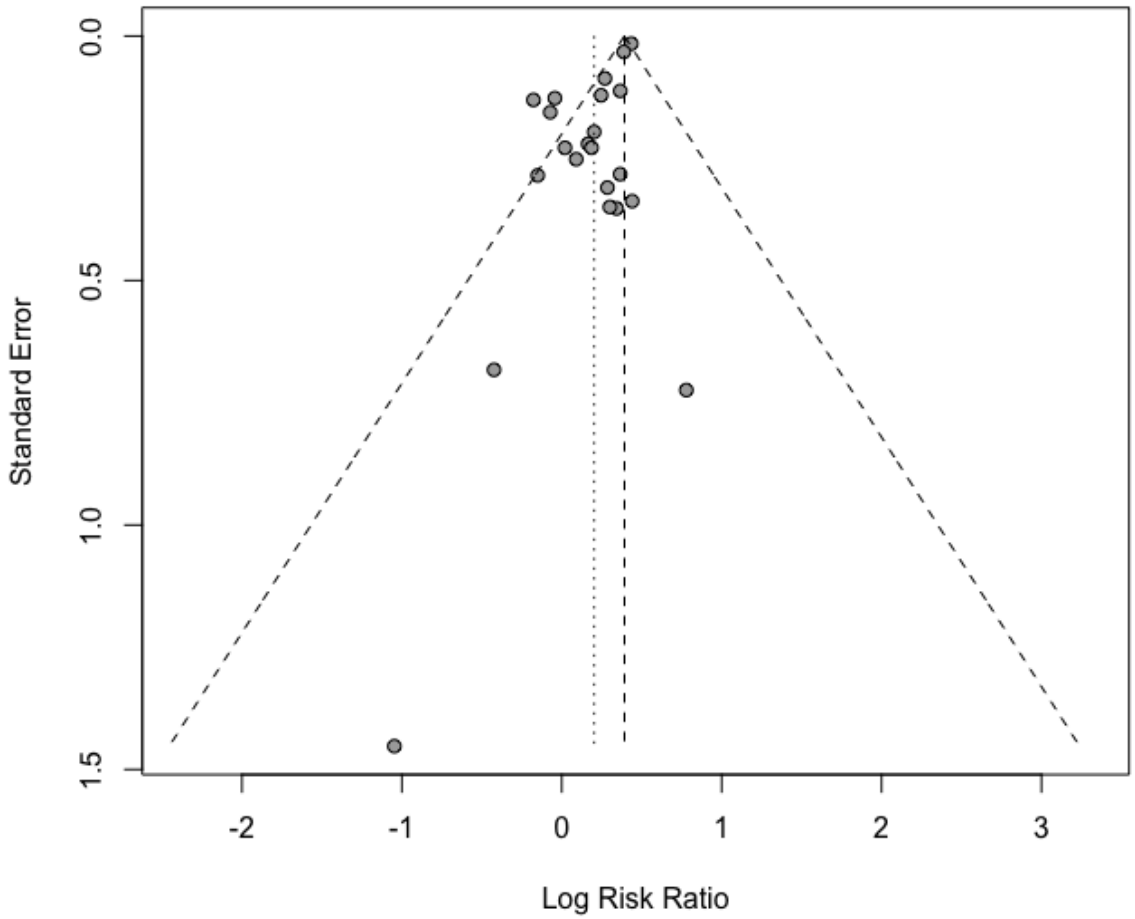

Figure A28. Funnel plot for risk factor: Hypertension

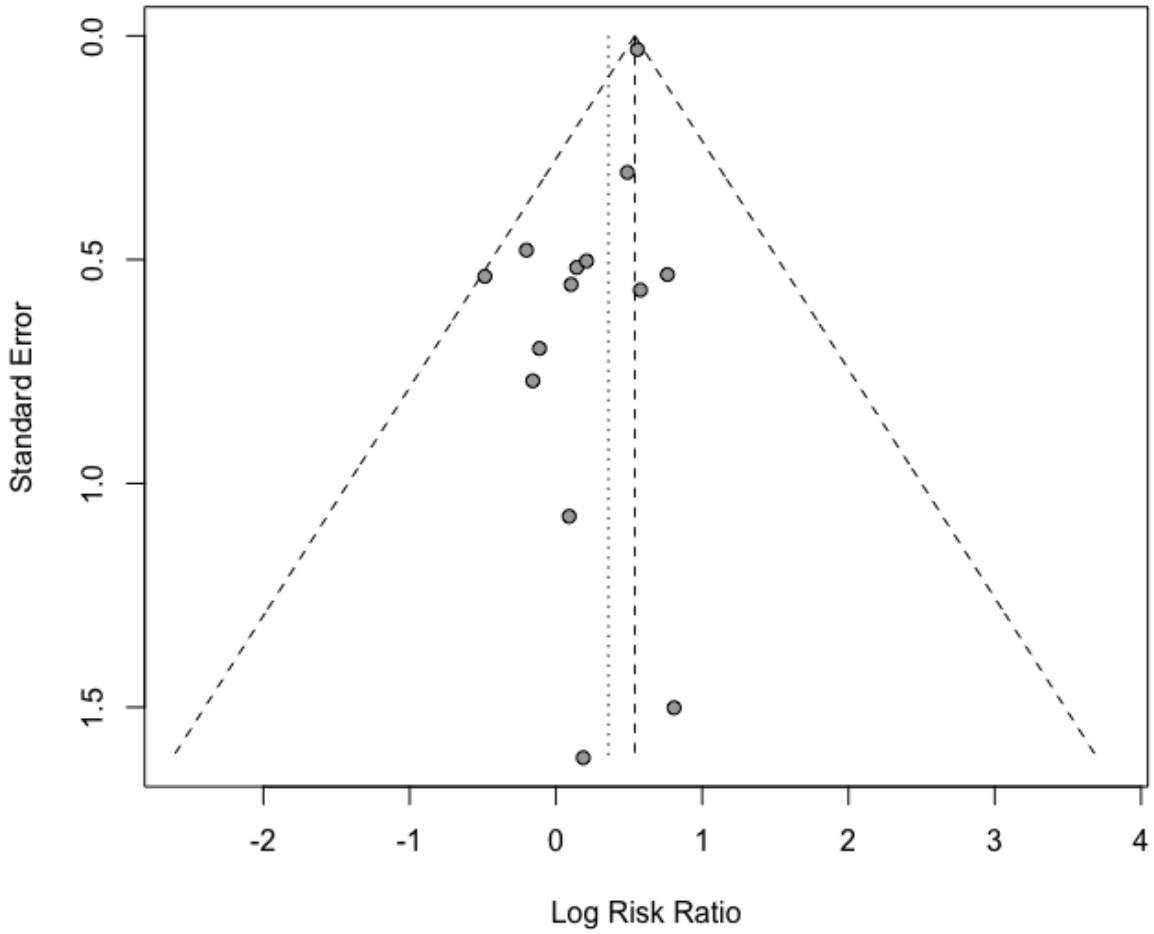

Figure A29. Funnel plot for risk factor: Kidney disease

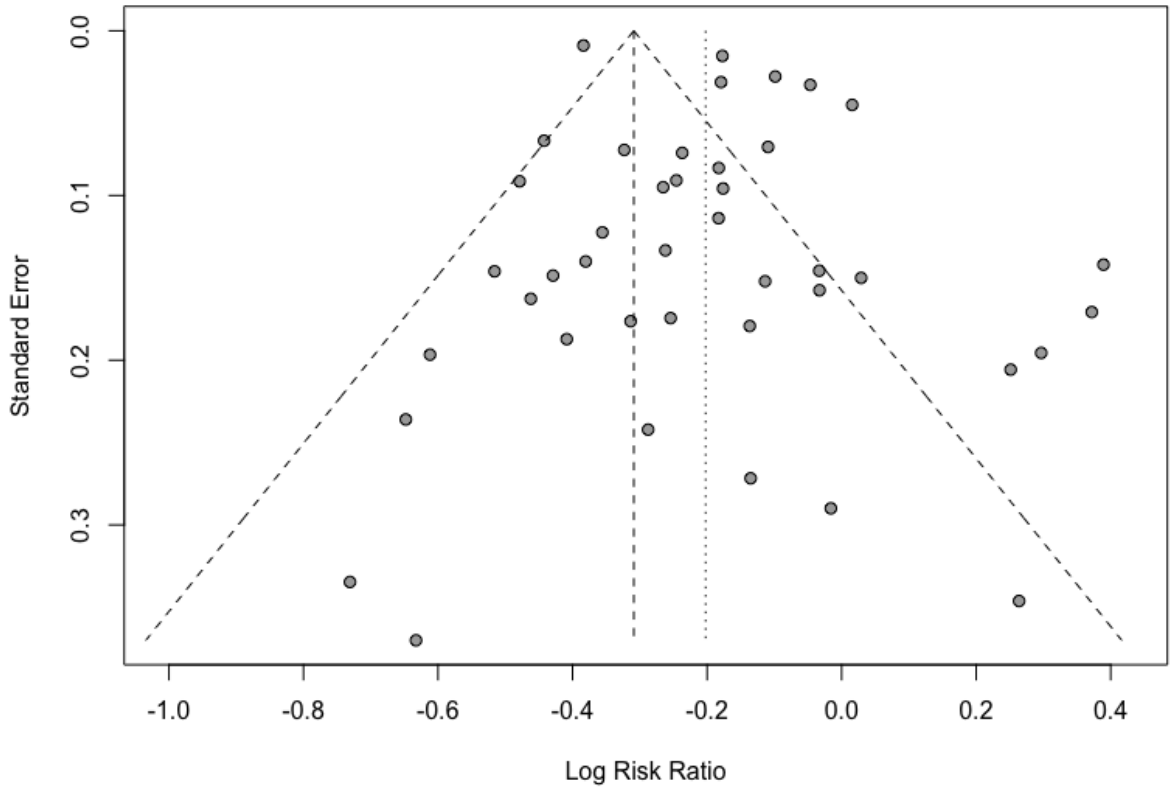

Figure A30. Funnel plot for risk factor: Sex (Male)

APPENDIX B

UNIVERSITY OF GEORGIA
CONSENT LETTER

UGA Student COVID-19 Symptom Survey

Dear Participant,

My name is Michelle Bentivegna and I am a doctoral candidate in the Department of Epidemiology at the University of Georgia under the supervision of Dr. Mark Ebell, Professor of Epidemiology. I am inviting you to take part in a research study. I am doing research on COVID-19, particularly on understanding the experience of young adults after COVID-19 diagnosis and why some people experience symptoms longer than others. Anyone who is between the ages of 18 and 30 years old and has had a positive COVID-19 test since August 1, 2021 can participate. We are interested in recruiting participants who experienced no prolonged symptoms as well as those who did.

If you agree to participate in this study, you will be asked to complete a survey that takes about 10 minutes to complete. All responses will be kept confidential and used only for the purposes of this research study. We will ask you to provide your email address to enter you in a drawing for a $100 Amazon gift card for your participation. Providing this information is voluntary, and this information along with your responses will be stored securely. We will not use your email address to contact you other than described above, and once the study is over we will delete the email address.

Participation is voluntary. You can refuse to take part or stop at any time without penalty. We do not think that any of the questions will make you uncomfortable. However, you are free to skip any questions if you do not wish to answer them.

There are no direct benefits to you in participating, but we hope that your responses will help doctors and researchers identify useful tools and treatments based on your experience and that of others completing this survey.

If you are interested in participating or have questions about this research, please feel free to contact me at Michelle.Bentivegna@uga.edu or Dr. Mark Ebell at ebell@uga.edu. If you have any complaints or questions about your rights as a research volunteer, contact the IRB at 706-542-3199 or by email at IRB@uga.edu.

By completing this survey, you are consenting to participate in this study.

Sincerely,

Michelle Bentivegna, MPH

Study Survey

Risk Factors for Long COVID-19 in Young Adults: A Case-Control Study

Thank you for considering participation in our survey. The purpose of this survey is to better understand the experience of prolonged symptoms as a result of COVID-19 infection. Participation is voluntary and you may stop or choose not to answer a question at any time throughout the survey. Please be assured that all answers will be kept strictly confidential and used only for the purpose of this research. In this survey you will be asked to answer a series of questions about your experience during and after COVID-19 diagnosis that should take about 10 minutes to complete. There are no expected risks or benefits to you for participating in this survey.

What is your gender?

- ○ Male
- ○ Female
- ○ Other
- ○ Prefer not to answer

What is your race?

- ☐ Alaska Native / American Indian / Native Hawaiian
- ☐ Asian / Pacific Islander
- ☐ Black / African-American
- ☐ White / Caucasian
- ☐ Prefer not to answer

What is your ethnicity?

- ○ Hispanic
- ○ Non-Hispanic
- ○ Prefer not to answer

What best describes your year in school as of the Fall 2022 Semester?

- ○ Year 1
- ○ Year 2
- ○ Year 3
- ○ Year 4
- ○ Year 5
- ○ Graduate/Professional
- ○ Other _____
- ○ Prefer not to answer

What was the approximate date of your positive COVID-19 test? (Please format as mm/dd/yyyy)

_____

Did you experience any of the following symptoms when you had COVID-19? (Select all that apply)

- [ ] Fever >100.4
- [ ] Chills
- [ ] Muscle or body aches
- [ ] Loss of taste or smell
- [ ] Runny nose/congestion
- [ ] Sore throat
- [ ] New or worsening cough
- [ ] Shortness of breath or difficulty breathing
- [ ] Chest pain
- [ ] Fatigue or tiredness
- [ ] Nausea or vomiting
- [ ] Headache
- [ ] Abdominal pain
- [ ] Diarrhea
- [ ] Brain fog/difficulty concentrating
- [ ] New anxiety/depression
- [ ] Other (Please specify)

_____

☐ No symptoms

☐ Prefer not to answer

Rate the severity of your symptoms when you had COVID-19. If you did not have that symptom please select no problem.

	No problem	Mild	Moderate	Severe
Fever >100.4	○	○	○	○
Chills	○	○	○	○
Muscle or body aches	○	○	○	○
Loss of taste or smell	○	○	○	○
Runny nose/congestion	○	○	○	○
Sore throat	○	○	○	○
New or worsening cough	○	○	○	○

Shortness of breath or difficulty breathing	○	○	○	○
Chest pain	○	○	○	○
Fatigue or tiredness	○	○	○	○
Nausea or vomiting	○	○	○	○
Headache	○	○	○	○
Abdominal pain	○	○	○	○
Diarrhea	○	○	○	○
Brain fog/difficulty concentrating	○	○	○	○
New anxiety/depression	○	○	○	○
Other (Please specify)	○	○	○	○
No symptoms	○	○	○	○
Prefer not to answer	○	○	○	○

Are you still experiencing any symptoms from COVID-19?

- ○ Yes
- ○ No
- ○ Prefer not to answer

If you answered yes to the previous question, rate the severity of the symptoms you are still experiencing. If no please select no problem for all symptoms.

	No problem	Mild	Moderate	Severe
Fever >100.4	○	○	○	○
Chills	○	○	○	○
Muscle or body aches	○	○	○	○
Loss of taste or smell	○	○	○	○
Runny nose/congestion	○	○	○	○
Sore throat	○	○	○	○
New or worsening cough	○	○	○	○
Shortness of breath or difficulty breathing	○	○	○	○
Chest pain	○	○	○	○
Fatigue or tiredness	○	○	○	○
Nausea or vomiting	○	○	○	○
Headache	○	○	○	○
Abdominal pain	○	○	○	○

Diarrhea	○	○	○	○
Brain fog/difficulty concentrating	○	○	○	○
New anxiety/depression	○	○	○	○
Other (Please specify)	○	○	○	○
No symptoms	○	○	○	○
Prefer not to answer	○	○	○	○

If you are not still experiencing symptoms, what is the approximate date your symptoms ended? (Please format as mm/dd/yyyy)

_____

Have you received a COVID-19 vaccine?
No, I am not vaccinated
Yes, 1 dose
Yes, 2 doses
Yes, 3 doses
Yes, 4 doses
Prefer not to answer

If yes, what was the approximate date of your first COVID-19 vaccine? (Please format as mm/dd/yyyy)

_____

What is your current height (in feet)?

_____

What is your current weight (in pounds)?

_____

Before being diagnosed with COVID-19 were you being treated for any of the following health conditions? (Select all that apply)

- [ ] Diabetes
- [ ] Cardiovascular/heart disease
- [ ] History of heart attack
- [ ] History of congestive heart failure
- [ ] High blood pressure
- [ ] High cholesterol
- [ ] History of stroke
- [ ] Autoimmune disorder (Lupus, Multiple Sclerosis, etc.)
- [ ] HIV
- [ ] Hepatitis C
- [ ] Asthma
- [ ] Chronic lung disease (COPD, emphysema, etc.)
- [ ] Cancer diagnosis
- [ ] Depression
- [ ] Pregnant
- [ ] Overweight or obese
- [ ] Anxiety or other mental health conditions

☐ Other medical condition (Specify)

_____

☐ None

☐ Prefer not to answer