BMJ Open

Association of Long COVID with Mental Health Disorders: Analysis of Real-world Data

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-079267
Article Type:	Original research
Date Submitted by the Author:	27-Aug-2023
Complete List of Authors:	Zhang, Yue; The Pennsylvania State University College of Medicine, Department of Public Health Sciences Chinchilli, Vernon; The Pennsylvania State University College of Medicine, Department of Public Health Sciences Ssentongo, Paddy; Penn State Health Milton S Hershey Medical Center, Ba, Djibril M.; The Pennsylvania State University College of Medicine, Department of Public Health Sciences
Keywords:	MENTAL HEALTH, Post-Acute COVID-19 Syndrome, COVID-19, EPIDEMIOLOGIC STUDIES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Association of Long COVID with Mental Health Disorders: Analysis of Real-world Data

Yue Zhang, MPH[1], Vernon M. Chinchilli, PhD[1], Paddy Ssentongo, MD, PhD[1,2*], Djibril M. Ba, PhD, MPH[1*]

*Co-Senior Authors

[1]Penn State College of Medicine, Department of Public Health Sciences, Hershey, PA

[2]Department of Medicine, Penn State Milton S. Hershey Medical Center, Hershey, PA

To whom correspondence should be addressed:

Yue Zhang, MPH

Department of Public Health Sciences

Penn State College of Medicine

90 Hope Drive, Hershey, PA 17033

E: ymz5503@psu.edu

Abstract Word Count: 299 (300 max)

Text word Count: 4039

Number of Tables: 3

Number of Figures: 2

References: 22

Running title: Long COVID and Mental Health Disorders

Conflict of Interest: None

Abstract

Objectives: Mental health disorders (MHD) rank third for US adult hospitalizations. With over half of SARS-CoV-2 survivors experiencing 'long COVID', this study aims to assess its association with increased MHD risk using extensive real-world data.

Design: A retrospective cohort study with propensity score matching was conducted. We used the international classification of 10th revision (ICD-10) codes to identify individuals with long COVID status and COVID histories. Multivariable stratified Cox proportional hazards regression analysis was conducted to determine the association of long COVID status with MHD.

Setting: Data was sourced from the TriNetX database, spanning records from 1 October 2021 to 16 April 2023.

Participants: Two distinct cohorts were established: one comprising individuals diagnosed with long COVID and another comprising individuals with no history of long COVID or COVID-19. At the start of the study, none of the participants had a recorded mental health disorder (MHD).

Primary and secondary outcome measures: The main outcome of interest was a composite diagnosis of MHD. Secondary outcomes were individual mental health conditions.

Results: The study included 43,060 control participants without long COVID and 4,306 long COVID participants, demonstrating well-balanced distribution across all covariates. After adjusting for 4 demographic factors and 10 comorbidities, long COVID was associated with MHD (aHR, 2.60; 95% CI, 2.37, 2.85). In subgroup analysis, long COVID was associated with major depression disorder (aHR, 3.36; 95% CI, 2.82, 4.00) and generalized anxiety disorder (aHR, 3.44; 95% CI, 2.99, 3.96). When individual MHD diagnosis was examined as outcomes, depression and anxiety were strongly associated with a 3-fold increased risk of MHD.

Conclusions: In this retrospective large real-world cohort study, long COVID was associated with an increased risk of incident MHD. Considering the vast number of 'long haulers', the MHD impact is significant. Enhanced MHD screening among COVID-19 survivors should be a priority.

Strengths and limitations of this study

- Large sample size of patients with 4,306 long COVID patients and 43,060 controls were included. We were able to control for confounders and comorbidities, thereby minimizing bias.
- To the best of our knowledge, it is the first large national-real-world analysis to examine the association between long COVID and MHD using a comprehensive propensity score matching approach.
- While this observational study draws from US TriNetX data, direct causality cannot be established. Nonetheless, our methodology aimed to address and exclude potential confounders and timing-related issues.
- We acknowledge that electronic health record databases can misclassify patients based on misreporting or underreporting of diagnoses codes or medications.
- Some control group members might have undiagnosed mild or asymptomatic COVID-19, as they weren't tested. This non-differential exposure misclassification could lead to underestimating the strength of the COVID-19 and MHD association.

Keywords: Mental Health Disorders; long COVID; real-world data; US

Introduction

The coronavirus disease-2019 (COVID-19) pandemic, caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), has profoundly impacted individual health and well-being globally[1]. While the effects of COVID-19 vary widely, from asymptomatic or mild disease to multi-organ failure and death, most people make a full recovery from the virus. Unfortunately, however, a substantial proportion of survivors continue to report persistent symptoms often referred to as post-acute sequelae of COVID-19 (PASC) in the literature or "long COVID," presenting a significant and ongoing public health crisis[2]. A recent meta-analysis found an increased incidence of anxiety, depression, and appetite problems among post-COVID-19 infected children, compared to those without a previous infection[3].

Mental health disorders (MHD) such as depression and anxiety disorders are the third leading common cause of hospitalization in the US[4,5]. In 2020, nearly 21 million American adults (8.4% of the adult population) suffer from a major depressive disorder[6].

Previous studies that have assessed the associations between individuals with COVID-19 and MHD were limited to special populations such as children or US veterans[7]. Studies so far have examined the impact of the pandemic on the incidence and prevalence of mental health disorders. However, the direct effect of COVID-19 among survivors is scarce and has not been delineated in diverse populations using large real-world data. Therefore, in this study, we aim to use very large real-world data to estimate the effect of long COVID using newly developed diagnosis codes on incident MHD. We hypothesize that since long COVID presents with an array of symptoms that affect daily quality of life, including sleep disorders, chronic dyspnea, and myalgia/arthralgia, it could have a negative impact on mental health outcomes.

Methods

Data source

This analysis was conducted using a TriNetX database with data extracted from October 1st, 2020 to April 16th, 2023. TriNetX serves as a federated, multi-institutional health research network that compiles de-identified data derived from Electronic Health Records (EHRs) across a broad spectrum of healthcare organizations. These include academic medical centers, specialized physician practices, and community hospitals, collectively encompassing over 250 million patients from more than 120 healthcare organizations (HCOs). This vast network enables longitudinal patient-level healthcare claims tracking, and delivers a wealth of data including, but not limited to, demographic attributes (age, race, gender, geographic location), medical histories, and prescription details. TriNetX is compliant with the Health Insurance Portability and Accountability Act (HIPAA), the US federal law which protects the privacy and security of healthcare data, and any additional data privacy regulations applicable to the contributing HCO[8].

Cohort derivation and assessment of exposure

Long COVID patients, aged greater than 18, were identified from October 1, 2021, to April 16, 2023, utilizing the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes (Supplemental Table 1). The Centers for Disease Control and Prevention (CDC) officially characterizes long COVID as a post-COVID condition manifesting in patients with a confirmed or probable history of SARS-CoV-2 infection. The corresponding ICD-10-CM code for this condition became effective on October 1, 2021[9].

We identified 9,034 participants with long COVID and 500,002 random participants without long COVID, each with at least one-month of follow-up after Oct 1st, 2021. The index dates of

long COVID group were the diagnosis dates of long COVID, and random pseudo index dates were assigned to the non-long COVID group between Oct 1st, 2021, and the maximum recorded dates for each participant, using data drawn from diagnosis records, lab test results, and medication records. None of the participants had a history of MHD prior to the index dates, and those in the non-long COVID group had no record of long COVID or COVID-19 diagnoses or testing positive for SARS-CoV-2. After all the exclusions (Figure 1), we identified a total of 4,306 eligible participants with long COVID and 186,903 eligible participants without COVID and without long COVID. The nearest-neighbor propensity score matching methods were further utilized at a 1:10 matching ratio, with the "without replacement" sampling method. This led to the final selection of 4,306 participants with long COVID and a corresponding control group of 43,060 participants for in-depth analysis. (Figure 1). This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline for cohort studies[10].

Assessment of outcomes

The primary outcome was the composite of any MHD, defined using ICD-10 codes for mental health diagnosis or substance use disorders as done in previous studies (Supplemental Table 1) that occurred after index dates during the follow-up period[11,12]. As the secondary outcomes (major depression, anxiety, other mental health conditions; Table 3), the association of long COVID with these individual MHD groups also was examined.

Assessment of potential covariates

To ensure a similar distribution of covariates at baseline and mitigate potential confounding effects, propensity score matching was conducted considering 4 demographic variables and 10 comorbidities. Demographic data on age (years), sex (male/female), race (white/ black or African

American/ unknown/ others), and US regional location (South/ West/ Midwest/ Northeast/ Unknown) were extracted directly from TriNetX patients' databases. For the identification of comorbidity covariates, we employed the Charlson Comorbidity Index, and selected the ten most prevalent comorbidities as matching variables[13]. These included myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, chronic pulmonary disease, rheumatoid disease, mild liver disease, diabetes, renal disease, and cancer (any malignancy)[14]. All these comorbidities were identified at baseline, defined as the 12 months preceding the index date, utilizing their corresponding ICD-10-CM codes (Supplemental Table 1).

Vaccination status was also taken into consideration, defined as any history of COVID-19 vaccination prior to the index date, determined by Current Procedural Terminology (CPT) codes (Supplemental Table 1)[15].

Statistical Analysis

We summarized baseline participants characteristics across long COVID and non-long COVID groups, with mean (SD) values for continuous variables, and number and percentage for categorical variables, which was presented as before matching and after 1:10 matching. Standardized mean difference (SMD) was regarded as a measure to evaluate the matching results, and an SMD greater than 0.1 is a threshold recommended for declaring imbalance[16].

We calculated the person-time of follow-up for each participant following the index date to the first occurrence of an outcome of interest (MHD), death date (if applicable), or the maximum follow-up date (the latest date recorded in diagnosis records, lab test results, and medication records), whichever took place first. Long COVID status was deemed as the primary exposure. We conducted subgroup analyses by calculating the unadjusted incidence rates and corresponding 95% CI per 1000 person-years of follow-up for the two cohorts within each

subgroup. Having confirmed no violations of the proportional hazards assumption (Supplemental Figure 1), we initially applied a stratified Cox proportional hazards regression model adjusting for age and sex to calculate the hazard ratio (HR) and 95% confidence interval (95% CI). Subsequently, we implemented another stratified Cox proportional hazards regression model using a matching ID constructed from the propensity scores with the 4 demographic factors and 10 comorbidities, which provided an adjusted HR and 95% CI. Secondary analyses were additionally performed using stratified Cox proportional hazards regression models, focusing specifically on different dimensions of MHD, including major depression, anxiety, and other mental health conditions. In addition, we assessed effect modification effects of baseline vaccination status, which was accomplished by examining the significance of the interaction term between long COVID status and vaccination within the confines of the stratified Cox proportional hazards regression model. Data were analyzed in R software version 3.6.2 (R Foundation for Statistical Computing, Vienna, Austria) and SAS Software version 9.4 (SAS Institute Inc; Cary, NC) using a two-tailed alpha level of 0.05.

Results

The present analysis incorporated 43,060 non-long COVID participants (mean [SD] age 54.43 [16.66] years; 58.3% female; 69.5% white) and 4,306 long COVID participants (mean [SD] age 54.62 [15.34] years; 58.2% female; 69.6% white). Prior to matching, imbalances in distribution were noted for patient regional location (SMD = 0.265), baseline congestive heart failure status (SMD = 0.131), baseline chronic pulmonary disease status (SMD = 0.396), and baseline diabetes status (SMD = 0.147). After the matching process, we observed balanced distributions across all covariates, with all SMDs falling below 0.1 (Table 1).

The overall cumulative incidence rate of MHD was higher among those with long COVID than those without long COVID (Figure 2). Similar results were observed when individual MHD diagnosis was examined as outcomes (Figure 2). The overall unadjusted incidence rate of MHD was higher among long COVID (251.1 per 1000 persons-years [PY])) compared to those without long COVID (99.5 per 1000-PY) (Table 2). In age and sex-adjusted stratified Cox model, long COVID was associated with higher risk of incident MHD (adjusted Hazard Ratio (aHR), 2.53; 95% CI, 2.32 - 2.75). Additional adjustment of the model with 4 demographic factors and 10 comorbidities did not dilute the association (aHR, 2.60; 95% CI, 2.37, 2.85) (Table 2).

In subgroup analyses, we observed a significant effect of long COVID on major depression (aHR, 3.36; 95% CI, 2.82, 4.00), generalized anxiety disorder (aHR, 3.44; 95% CI, 2.99, 3.96), in the full adjusted stratified Cox regression models. Conversely, while the influence of long COVID on other mental health conditions was statistically significant, the magnitude of the effect was smaller than the overall effects (aHR, 1.31; 95% CI, 1.08, 1.60) (Table 3). Baseline vaccination status did not modify the association between long COVID and MHD (P-value for interaction = 0.96).

Discussion

In the present analysis of real-world data using propensity score matching, survivors of COVID-19 with long COVID were more than twice as likely to develop MHD. The impact was largest for major depressive disorders and generalized anxiety disorders. In 2010, the global economic costs of mental disorders were estimated at US$2.5 trillion[17]. The pandemic has worsened the MHD given the millions of survivors of COVID-19 globally, the cost of MHD is likely to climb on dramatically.

The effect estimates from our study was stronger than in the study by veteran study by Xie et al that found a nearly 40% increase in depression and anxiety disorders. The weakness of the prior study is the lack of generalizability due to the demographic composition of the cohort that were mostly older white men and the limited time study time frame of nearly 30 days following acute COVID-19 illness and yet there is a temporal dynamic nature of the epidemiology of mental health outcomes in the post-acute phase of COVID-19.

The pathophysiological of MHD among long haulers is not fully understood. However, there is emerging evidence of both direct and indirect effect of the virus on the brain and phycological outcomes respectively. Starting with the direct effect, the SARS-CoV-2 virus is neurotropic, indicating a direct invasion of the nervous system causing inflammation and gliosis[18]. These changes in the neuronal vascular coupling could lead to further break down in the blood brain barrier allowing for the worsening of the inflammatory cases ace and the influx of inflammatory cells that causes injury to the neurons. Such a direct and neurotropic effect of the virus has been seen in HIV, EBV, and CMV[19]. Indirectly, the pandemic been associated with psychological distress[20] due to increased isolation, increased rates of domestic violence[21], a disruption of social

networks and unemployment[22], all risk factors of MHD. But much more research is needed to get a clear picture of how all of these variables may have contributed to MHD.

The findings of the current study have great public health and clinical implications. First, the stigma attached to MHD could hinder health care utilization and thereby worsen the prognosing of the outcomes with a shift towards suicidality and homicidally. Next, given the disproportionality in lack of access to care in the marginalized communities including people of color and the poor communities and yet these are the same communities that were mostly affected by severe COVID, the burden of untreated mental health is likely to yield negative outcomes in such communities. For the health care providers particularly primary health care providers, they should have a lower threshold to screen and treat mental health disorders in this survivor of COVID-19.

Study strengths and limitations

Strengths of our study include an analysis based on longitudinal data of a large sample of long COVID individuals using most recently approved ICD-10 code and clean controls without any diagnosis of long COVID or COVID-19 diagnoses or testing positive for SARS-CoV-2. To the best of our knowledge, it also is the first large national-real-world analysis to examine the association between long COVID and MHD using comprehensive propensity score matching approach. Nevertheless, our study has several limitations that should be addressed when interpreting the results. This is an observational study that used US TriNetX data and therefore causality cannot be inferred. We acknowledge that electronic health record databases can misclassify patients based on misreporting or underreporting of diagnoses codes or medications. Moreover, even though we excluded any patients with long COVID or COVID-19 diagnoses or testing positive for SARS-CoV-2, some individuals in the control group might still have

undetected mild or asymptomatic COVID-19 because they had not been tested. Such non-differential misclassification of the exposure may underestimate the strength of the association of COVID-19 with the onset of MHD.

Conclusions

Using a large real-world, nationwide, propensity score matched cohort, we found that long COVID was associated with an increased risk of new onset of MHD. The increase was independent of demographics, lifestyle factors, and major chronic medical conditions. These findings reinforce the importance of integrating mental health screening and services in the treatment and management of long COVID to prevent related chronic diseases, suicidal thoughts, and attempts.

Funding statement: There was no external or internal funding to support this study.

Competing interests statement: The authors have no conflicts of interest to disclose.

Ethics approval statement: The protocol of this study was reviewed and received a determination of non-human subjects' research by the Penn State Institutional Review Board. The individual informed consent requirement was waived for this secondary analysis of de-identified data

Authors' contributions: Designed research (project conception, development of overall research plan): YZ and DMB. Data extraction and study oversight: DMB. Analyzed data: YZ, PS, and DMB. Performed statistical analysis: YZ and DMB. Wrote the first draft of the manuscript: YZ, PS and DMB. Review and editing: All authors. All authors have read and approved the final manuscript.

Data sharing statement: Data is available from third-party partners "TriNetX"

Acknowledgments: None

References

1. Ssentongo P, Fronterre C, Chinchilli VM. Spatial Epidemiology of COVID-19 Pandemic: Disease Risk, Prognosis, and Complications. In: *Frontiers of COVID-19*. Springer International Publishing; 2022:241-257. doi:10.1007/978-3-031-08045-6_11

2. 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):e2128568. doi:10.1001/jamanetworkopen.2021.28568

3. Hassan NM, Salim HS, Amaran S, et al. Prevalence of mental health problems among children with long COVID: A systematic review and meta-analysis. *PLOS ONE*. 2023;18(5):e0282538. doi:10.1371/journal.pone.0282538

4. Parks, Joe, Svendsen, Dale, Singer, Patricia, Mary, Fotu. Morbidity and Mortality in People with Serious Mental Illness. *Natl Assoc State Ment Health Program Dir NASMHPD Med Dir Counc*. Published online October 2006.

5. KESSLER RC, ANGERMEYER M, ANTHONY JC, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry*. 2007;6(3):168-176.

6. Major Depression. National Institute of Mental Health (NIMH). Accessed July 15, 2023. https://www.nimh.nih.gov/health/statistics/major-depression

7. Xie Y, Xu E, Al-Aly Z. Risks of mental health outcomes in people with covid-19: cohort study. *BMJ*. 2022;376:e068993. doi:10.1136/bmj-2021-068993

8. Publication Guidelines. TriNetX. Accessed July 1, 2023. https://trinetx.com/real-world-resources/publications/trinetx-publication-guidelines/

9. Healthcare Workers. Centers for Disease Control and Prevention. Published February 11, 2020. Accessed July 2, 2023. https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/post-covid-conditions.html

10. Elm E von, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*. 2007;335(7624):806-808. doi:10.1136/bmj.39335.541782.AD

11. Busch AB, Huskamp HA, Raja P, Rose S, Mehrotra A. Disruptions in Care for Medicare Beneficiaries With Severe Mental Illness During the COVID-19 Pandemic. *JAMA Netw Open*. 2022;5(1):e2145677. doi:10.1001/jamanetworkopen.2021.45677

12. Lee SW, Yang JM, Moon SY, et al. Association between mental illness and COVID-19 susceptibility and clinical outcomes in South Korea: a nationwide cohort study. *Lancet Psychiatry*. 2020;7(12):1025-1031. doi:10.1016/S2215-0366(20)30421-1

13. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis*. 1987;40(5):373-383. doi:10.1016/0021-9681(87)90171-8

14. Quan H, Li B, Couris CM, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol*. 2011;173(6):676-682. doi:10.1093/aje/kwq433

15. Find your COVID-19 Vaccine CPT® Codes. American Medical Association. Accessed July 2, 2023. https://www.ama-assn.org/find-covid-19-vaccine-codes

16. Austin PC. Using the Standardized Difference to Compare the Prevalence of a Binary Variable Between Two Groups in Observational Research. *Commun Stat - Simul Comput*. 2009;38(6):1228-1234. doi:10.1080/03610910902859574

17. Trautmann S, Rehm J, Wittchen H. The economic costs of mental disorders. *EMBO Rep*. 2016;17(9):1245-1249. doi:10.15252/embr.201642951

18. Bauer L, Laksono BM, Vrij FMS de, Kushner SA, Harschnitz O, Riel D van. The neuroinvasiveness, neurotropism, and neurovirulence of SARS-CoV-2. *Trends Neurosci*. 2022;45(5):358-368. doi:10.1016/j.tins.2022.02.006

19. Hotta H. [Neurotropic viruses--classification, structure and characteristics]. *Nihon Rinsho Jpn J Clin Med*. 1997;55(4):777-782.

20. Brooks SK, Webster RK, Smith LE, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *The Lancet*. 2020;395(10227):912-920. doi:10.1016/S0140-6736(20)30460-8

21. Leslie E, Wilson R. Sheltering in place and domestic violence: Evidence from calls for service during COVID-19. *J Public Econ*. 2020;189:104241. doi:10.1016/j.jpubeco.2020.104241

22. Elmer T, Mepham K, Stadtfeld C. Students under lockdown: Comparisons of students' social networks and mental health before and during the COVID-19 crisis in Switzerland. *PLOS ONE*. 2020;15(7):e0236337. doi:10.1371/journal.pone.0236337

Table 1. Baseline Characteristics Before and After Matching Stratified with Long COVID Status

Variables	Before matching			After 1:10 matching		
	Non-Long COVID n=186,903	Long COVID n=4,306	SMD*	Non-Long COVID n=43,060	Long COVID n=4,306	SMD*
Age (mean (SD))	55.22 (18.39)	54.62 (15.34)	0.036	54.43 (16.66)	54.62 (15.34)	0.012
Sex (Male) (n (%))	77,742 (41.6)	1,774 (41.2)	0.008	17,938 (41.7)	1,774 (41.2)	0.009
Race (n (%))			0.084			0.013
Black or African American	24,815 (13.3)	511 (11.9)		5,002 (11.6)	511 (11.9)	
Others	7,888 (4.2)	137 (3.2)		1,347 (3.1)	137 (3.2)	
Unknown	25,128 (13.4)	662 (15.4)		6,786 (15.8)	662 (15.4)	
White	129,072 (69.1)	2,996 (69.6)		29,925 (69.5)	2,996 (69.6)	
Patient Regional Location (n (%))			0.265			0.017
Midwest	27,387 (14.7)	892 (20.7)		8,955 (20.8)	892 (20.7)	
Northeast	64,765 (34.7)	1,226 (28.5)		12,536 (29.1)	1,226 (28.5)	
South	69,226 (37.0)	1,362 (31.6)		13,409 (31.1)	1,362 (31.6)	
Unknown	2,215 (1.2)	26 (0.6)		240 (0.6)	26 (0.6)	
West	23,310 (12.5)	800 (18.6)		7,920 (18.4)	800 (18.6)	
Baseline Comorbidities (Yes)						
Myocardial Infarction (n (%))	2,045 (1.1)	102 (2.4)	0.098	837 (1.9)	102 (2.4)	0.029
Congestive Heart Failure (n (%))	5,247 (2.8)	233 (5.4)	0.131	2,033 (4.7)	233 (5.4)	0.031
Peripheral Vascular Disease (n (%))	4,850 (2.6)	139 (3.2)	0.038	1,157 (2.7)	139 (3.2)	0.032
Cerebrovascular Disease (n (%))	4,720 (2.5)	132 (3.1)	0.033	1,142 (2.7)	132 (3.1)	0.025
Chronic Pulmonary Disease (n (%))	8,965 (4.8)	727 (16.9)	0.396	7,360 (17.1)	727 (16.9)	0.006
Rheumatoid Disease (n (%))	2,337 (1.3)	110 (2.6)	0.096	902 (2.1)	110 (2.6)	0.031
Mild Liver Disease (n (%))	3,951 (2.1)	153 (3.6)	0.087	1,324 (3.1)	153 (3.6)	0.027
Diabetes (n (%))	13,064 (7.0)	483 (11.2)	0.147	4,635 (10.8)	483 (11.2)	0.014
Renal Disease (n (%))	6,538 (3.5)	235 (5.5)	0.095	2,024 (4.7)	235 (5.5)	0.034
Cancer (any malignancy) (n (%))	9,749 (5.2)	169 (3.9)	0.062	1,523 (3.5)	169 (3.9)	0.020

*SMD, standardized mean difference; An SMD greater than 0.1 is a threshold recommended for declaring imbalanced

Table 2. Incidence rates and stratified Cox proportional hazard models hazard ratio (95% CI) for the association between Long COVID and mental health disorders

Variable	Non-Long COVID	Long COVID
Person-years, y	23,586	2,773
Mental health disorders cases, n	2,347	695
Incidence rate, 95% CI per 1000 person-years*	99.5 (95.5, 103.6)	251.1 (232.3, 270.2)
Model 1	(reference)	2.53 (2.32, 2.75)
Model 2	(reference)	2.60 (2.37, 2.85)

*Unadjusted incidence rate per 1000-person-years
Model 1: stratified by age categories (18-34, 35-44, 45-54, 55-64, 65+) and sex (men/women)
Model 2: stratified by fully propensity score matching ID.

Table 3. Stratified Cox proportional hazard models hazard ratio (95% CI) for the association between Long COVID and four mental health disorders

	Non-Long COVID (n, %)	Long COVID (n, %)	HR (95% CI)*
Any Mental Health Condition	2,347 (100%)	695 (100%)	2.60 (2.37, 2.85)
Major Depression[1]	575 (24.5%)	213 (30.6%)	3.36 (2.82, 4.00)
Anxiety[2]	892 (38.0%)	341 (49.1%)	3.44 (2.99, 3.96)
Other mental health condition[3]	880 (34.5%)	141 (20.3%)	1.31 (1.08, 1.60)

1 Major depression (F32, F33)
2 Anxiety (F40 - F48)
3 Other mental health conditions (F10 - F31, F34 - F39, F49 - F99)
* stratified by full propensity score matching ID
** NA due to the small sample size of psychotic disorder cases

FIGURES:

Figure 1 Long COVID and Non-Long COVID participants inclusion from TriNetX datasets

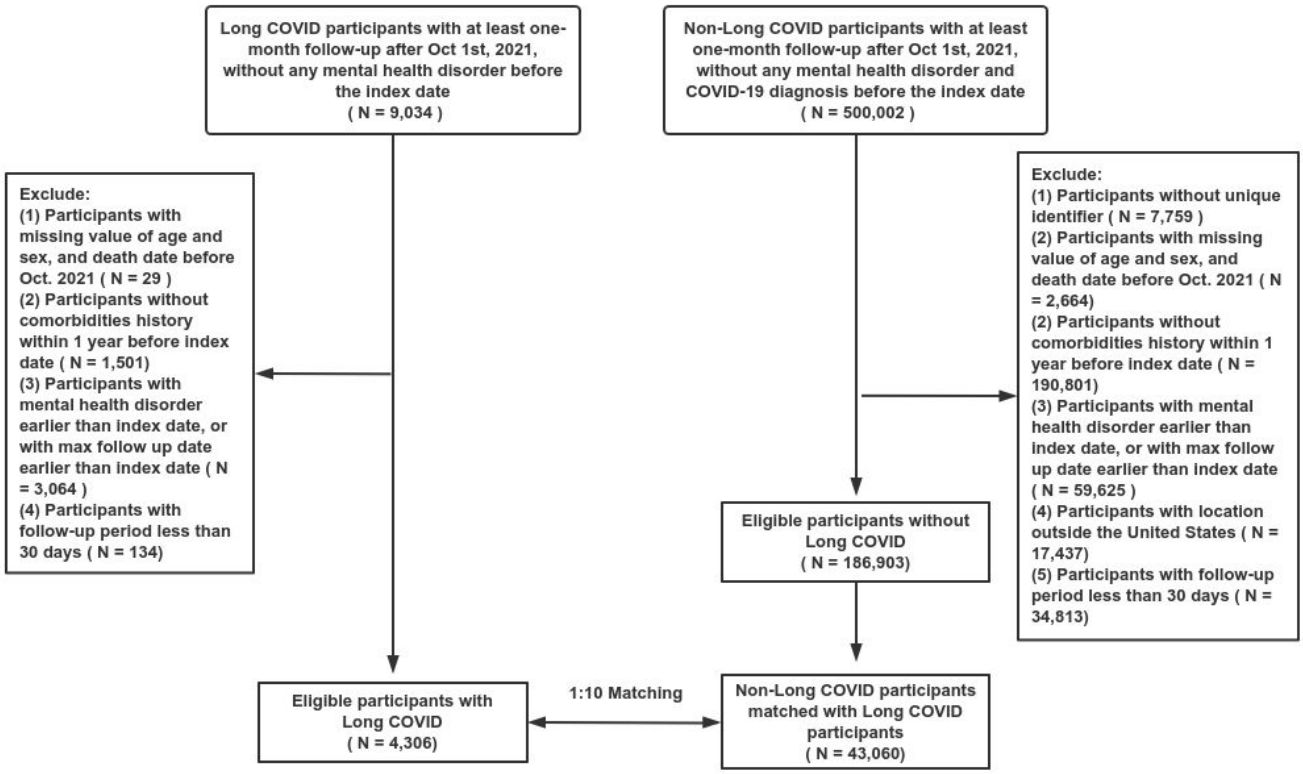
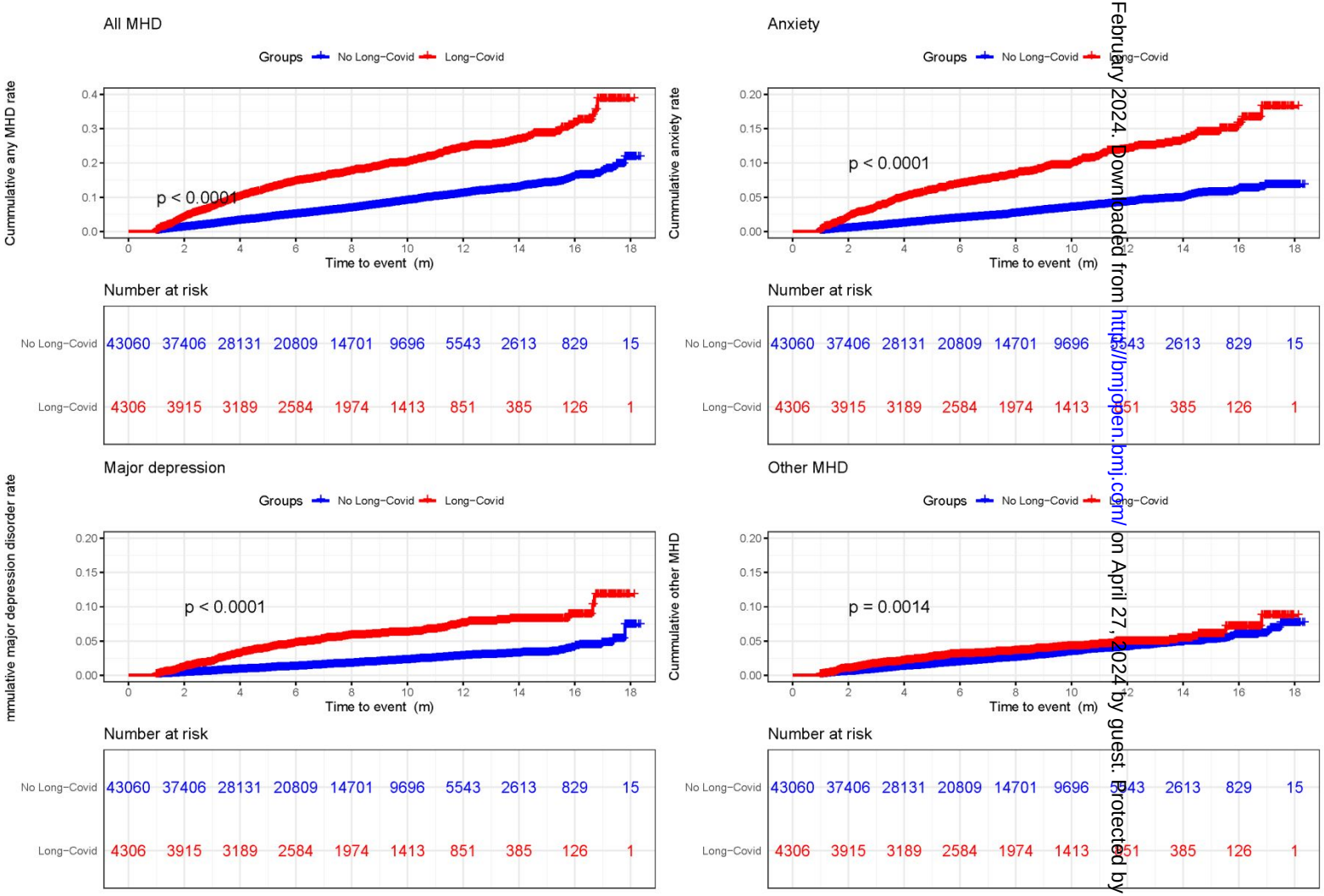


Figure 2. Cumulative Incidence Rates of Total and Categorized Major Health Disorders in Patients with Long COVID vs. Those Without Long COVID

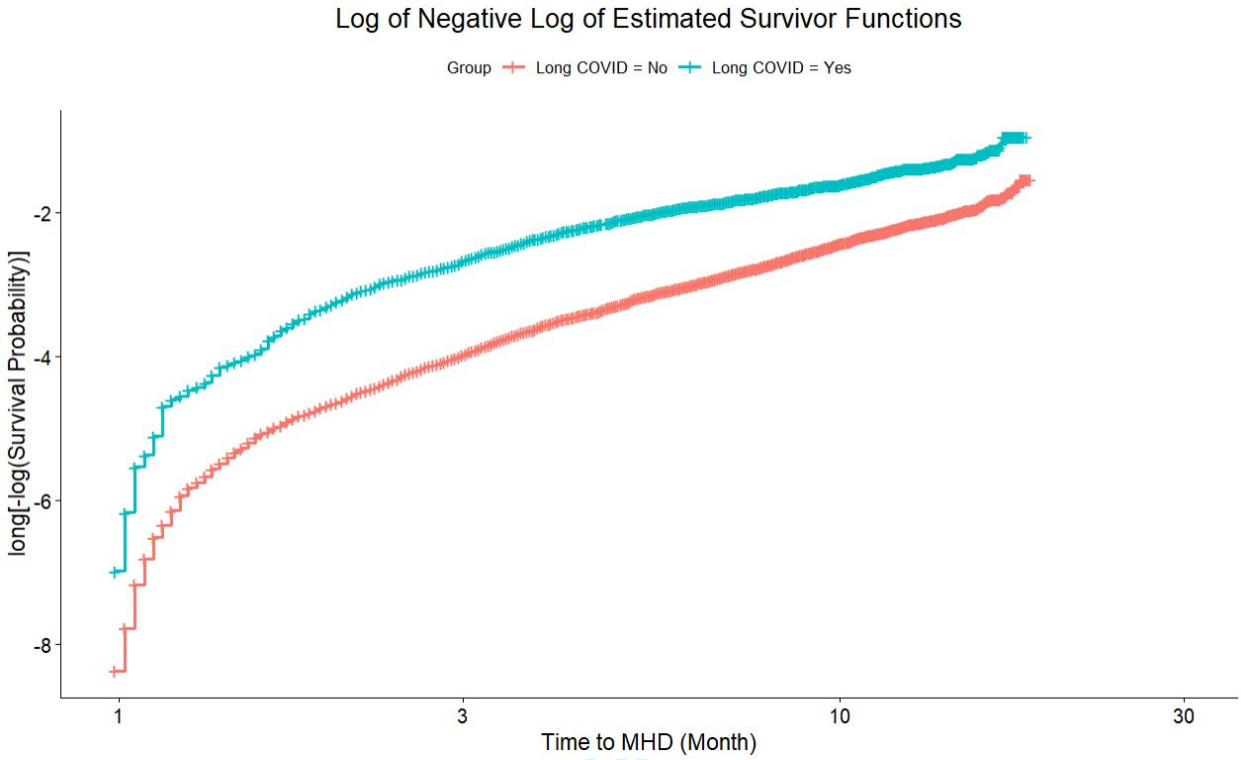


Supplementary Table 1: Diagnosis codes of the International Classification of Diseases 10th editions, Clinical Modification (ICD-10-CM), used to describe baseline/preexisting clinical medical conditions; and Current Procedural Terminology (CPT) code for COVID-19 vaccination status

Variables	ICD-10 codes
Long COVID diagnosis	U09.9
Mental health disorders	F10.x-F16.x, F17.x-F69.x, F80.x-F89.x, F90.x-F99.x
Myocardial Infarction	I21.x, I22.x, I25.x,
Congestive Heart Failure	I42.x, I43.x, I50.x,
Peripheral Vascular Disease	I70.x, I71.x, I73.x, I79.x, K55.x, Z95.x
Cerebrovascular Disease	G45.x, G46.x, I60.x-I69.x
Chronic Pulmonary Disease	I27.x, J40.x-J47.x, J60.xJ68.x, J70.x
Mild Liver Disease	B18.x, K70.x K71.x, K73.x, K74.x, K76.x, Z94.x
Diabetes	E10.x-E14.x
Renal Disease	N03.x, N05.x, N18.x, N19.x, Z49.x
Cancer (any malignancy)	C00.x-C26.x, C30.x-C34.x, C37.x-C41.x, C43.x, C45.x-58.x, C60.x-C76.x, C81.x-C85.x, C88.x, C90.x-C97.x

Variables	CPT codes
COVID-19 vaccination	"91300", "91305", "91307", "91308", "91301", "91306", "91311", "91309", "91317", "91315","91312", "91314", "91316", "91313"

Supplemental Figure 1. Log-NegLog Survival Plot for checking proportional hazard assumption



STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy
		(*e*) Describe any sensitivity analyses

Continued on next page

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
		(b) Indicate number of participants with missing data for each variable of interest
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time
		Case-control study—Report numbers in each exposure category, or summary measures of exposure
		Cross-sectional study—Report numbers of outcome events or summary measures
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		(*b*) Report category boundaries when continuous variables were categorized
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Association of Long COVID with Mental Health Disorders: A Retrospective Cohort Study Using Real-World Data from the United States

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-079267.R1
Article Type:	Original research
Date Submitted by the Author:	29-Dec-2023
Complete List of Authors:	Zhang, Yue; The Pennsylvania State University College of Medicine, Department of Public Health Sciences Chinchilli, Vernon; The Pennsylvania State University College of Medicine, Department of Public Health Sciences Ssentongo, Paddy; Penn State Health Milton S Hershey Medical Center, Ba, Djibril M.; The Pennsylvania State University College of Medicine, Department of Public Health Sciences
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Epidemiology, Infectious diseases
Keywords:	MENTAL HEALTH, Post-Acute COVID-19 Syndrome, COVID-19, EPIDEMIOLOGIC STUDIES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Association of Long COVID with Mental Health Disorders: A Retrospective Cohort Study Using Real-World Data from the United States

Yue Zhang, MPH[1], Vernon M. Chinchilli, PhD[1], Paddy Ssentongo, MD, PhD[1,2*], Djibril M. Ba, PhD, MPH[1*]

*Co-Senior Authors

[1]Penn State College of Medicine, Department of Public Health Sciences, Hershey, PA

[2]Department of Medicine, Penn State Milton S. Hershey Medical Center, Hershey, PA

To whom correspondence should be addressed:

Yue Zhang, MPH

Department of Public Health Sciences

Penn State College of Medicine

90 Hope Drive, Hershey, PA 17033

E: ymz5503@psu.edu

Abstract Word Count: 281 (300 max)

Text word Count: 2923

Number of Tables: 3

Number of Figures: 2

References: 37

Running title: Long COVID and Mental Health Disorders

Conflict of Interest: None

Abstract

Objectives: Mental health disorders (MHD) rank third for US adult hospitalizations. Given the substantial prevalence of 'Long COVID' in SARS-CoV-2 survivors, this study aims to assess its association with increased MHD risk using extensive real-world data.

Design: A retrospective cohort study with propensity score matching was conducted. We used the international classification of 10th revision (ICD-10) codes to identify individuals with Long COVID status and COVID histories. Multivariable stratified Cox proportional hazards regression analysis was conducted to determine the association of Long COVID status with MHD.

Setting: Data was sourced from the TriNetX database, spanning records from 1 October 2021 to 16 April 2023.

Participants: Two distinct cohorts were established: one comprising individuals diagnosed with Long COVID and another comprising individuals with no history of Long COVID or COVID-19. At the start of the study, none of the participants had a recorded mental health disorder (MHD).

Primary and secondary outcome measures: The main outcome of interest was a composite diagnosis of MHD. Secondary outcomes were individual mental health conditions.

Results: The study included 43,060 control participants without Long COVID and 4,306 Long COVID participants, demonstrating well-balanced distribution across all covariates. After adjusting for 4 demographic factors and 10 comorbidities, Long COVID was associated with MHD (aHR, 2.60; 95% CI, 2.37, 2.85). In subgroup analysis, Long COVID was associated with major depression disorder (aHR, 3.36; 95% CI, 2.82, 4.00) and generalized anxiety disorder (aHR, 3.44; 95% CI, 2.99, 3.96).

Conclusions: In this retrospective large real-world cohort study, Long COVID was associated with an increased risk of incident MHD. The MHD impact is significant considering the vast number of patients with Long COVID. Enhanced MHD screening among COVID-19 survivors should be a priority.

Strengths and limitations of this study

- Included a large sample size of 4,306 patients with Long COVID and 43,060 controls, with comprehensive measures to control for confounders and comorbidities.
- Utilized a comprehensive propensity score matching approach to analyze the association between Long COVID and MHD.
- Employed a retrospective cohort design using US TriNetX data, with methodologies implemented to mitigate potential confounders and address timing-related issues.
- Acknowledged potential misclassification in electronic health records due to misreporting or underreporting of MHD or Long COVID diagnosis codes.
- Considered the possibility of undiagnosed mild or asymptomatic COVID-19 in untested controls, which may underestimate the strength of the Long COVID and MHD association.

Keywords: Mental Health Disorders; Long COVID; real-world data; US

Introduction

The coronavirus disease-2019 (COVID-19) pandemic, caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), has profoundly impacted individual health and well-being globally[1]. While the effects of COVID-19 range from asymptomatic or mild disease to multi-organ failure and death, a notable proportion of the survivors with the virus experience persistent symptoms. These are commonly referred to in the literature as post-acute sequelae of COVID-19 (PASC) or "Long COVID" [2]. This condition represents a significant and ongoing public health crisis, as indicated by data suggesting that 10–30% of non-hospitalized cases and 50–70% of hospitalized cases report long-term effects[3], [4]. A recent meta-analysis found an increased incidence of anxiety, depression, and appetite problems among post-COVID-19 infected children, compared to those without a previous infection[5].

Mental health disorders (MHD) such as depression and anxiety disorders are the third leading common cause of hospitalization in the US[6], [7]. In 2020, nearly 21 million American adults (8.4% of the adult population) suffer from a major depressive disorder[8].

Previous studies that have assessed the associations between individuals with COVID-19 and MHD were limited to special populations such as children or US veterans[9]. Studies so far have examined the impact of the pandemic on the incidence and prevalence of mental health disorders. However, the direct effect of COVID-19 among survivors is scarce and has not been delineated in diverse populations using large real-world data. Therefore, in this study, we aim to use very large real-world data to estimate the effect of Long COVID using newly developed diagnosis codes on incident MHD. We hypothesize that since Long COVID presents with an array of symptoms that affect daily quality of life, including sleep disorders, chronic dyspnea, and myalgia/arthralgia, it could have a negative impact on mental health outcomes.

Methods

Data source

This analysis was conducted using a TriNetX database with data extracted from October 1st, 2020 to April 16th, 2023. TriNetX serves as a federated, multi-institutional health research network that compiles de-identified data derived from Electronic Health Records (EHRs) across a broad spectrum of healthcare organizations. These include academic medical centers, specialized physician practices, and community hospitals, collectively encompassing over 250 million patients from more than 120 healthcare organizations (HCOs). This vast network enables longitudinal patient-level healthcare claims tracking, and delivers a wealth of data including, but not limited to, demographic attributes (age, race, gender, geographic location), medical histories, and prescription details. TriNetX is compliant with the Health Insurance Portability and Accountability Act (HIPAA), the US federal law which protects the privacy and security of healthcare data, and any additional data privacy regulations applicable to the contributing HCO[10].

Cohort derivation and assessment of exposure

Patients with Long COVID, aged greater than 18, were identified from October 1, 2021, to April 16, 2023, utilizing the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes (Supplemental Table 1). The Centers for Disease Control and Prevention (CDC) officially characterizes Long COVID as a post-COVID condition manifesting in patients with a confirmed or probable history of SARS-CoV-2 infection. The corresponding ICD-10-CM code for this condition became effective on October 1, 2021[11].

We identified 9,034 participants with Long COVID and 500,002 random participants without Long COVID, each with at least one month of follow-up after Oct 1st, 2021. The index dates of Long COVID group were the diagnosis dates of Long COVID, and random pseudo index dates were assigned to the non-Long COVID group between Oct 1st, 2021, and the maximum recorded dates for each participant, using data drawn from diagnosis records, lab test results, and medication records. None of the participants had a history of MHD prior to the index dates, and those in the non-Long COVID group had no record of Long COVID or COVID-19 diagnoses or testing positive for SARS-CoV-2. After all the exclusions (Figure 1), we identified a total of 4,306 eligible participants with Long COVID and 186,903 eligible participants without COVID and without Long COVID. The nearest-neighbor propensity score matching methods were further utilized at a 1:10 matching ratio, with the "without replacement" sampling method. This led to the final selection of 4,306 participants with Long COVID and a corresponding control group of 43,060 participants for in-depth analysis. (Figure 1). This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline for cohort studies[12].

Patient and public involvement

Patients and the public were not involved in the design or planning of this secondary data analysis.

Assessment of outcomes

The primary outcome was the composite of any MHD, defined using ICD-10 codes for mental health diagnosis or substance use disorders as done in previous studies (Supplemental Table 1) that occurred after index dates during the follow-up period[13], [14]. As the secondary outcomes

(major depression, anxiety, and other mental health conditions), the association of Long COVID with these individual MHD groups also was examined.

Assessment of potential covariates

To ensure a similar distribution of covariates at baseline and mitigate potential confounding effects, propensity score matching was conducted considering 4 demographic variables and 10 comorbidities. Demographic data on age (years), sex (male/female), race (white/ black or African American/ unknown/ others), and US regional location (South/ West/ Midwest/ Northeast/ Unknown) were extracted directly from TriNetX patients' databases. For the identification of comorbidity covariates, we employed the Charlson Comorbidity Index, and selected the ten most prevalent comorbidities as matching variables[15]. These included myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, chronic pulmonary disease, rheumatoid disease, mild liver disease, diabetes, renal disease, and cancer (any malignancy)[16]. All these comorbidities were identified at baseline, defined as the 12 months preceding the index date, utilizing their corresponding ICD-10-CM codes (Supplemental Table 1). Vaccination status was also taken into consideration, defined as any history of COVID-19 vaccination prior to the index date, determined by Current Procedural Terminology (CPT) codes (Supplemental Table 1)[17].

Statistical Analysis

We summarized baseline participants characteristics across Long COVID and non-Long COVID groups, with mean (SD) values for continuous variables, and number and percentage for categorical variables, which was presented as before matching and after 1:10 matching.

Standardized mean difference (SMD) was regarded as a measure to evaluate the matching results, and an SMD greater than 0.1 is a threshold recommended for declaring imbalance[18].

We calculated the person-time of follow-up for each participant following the index date to the first occurrence of an outcome of interest (MHD), death date (if applicable), or the maximum follow-up date (the latest date recorded in diagnosis records, lab test results, and medication records), whichever took place first. Long COVID status was deemed as the primary exposure. We conducted subgroup analyses by calculating the unadjusted incidence rates and corresponding 95% CI per 1000 person-years of follow-up for the two cohorts within each subgroup. Having confirmed no violations of the proportional hazards assumption (Supplemental Figure 1), we initially applied a stratified Cox proportional hazards regression model adjusting for age and sex to calculate the hazard ratio (HR) and 95% confidence interval (95% CI). Subsequently, we implemented another stratified Cox proportional hazards regression model using a matching ID constructed from the propensity scores with the 4 demographic factors and 10 comorbidities, which provided an adjusted HR and 95% CI. Secondary analyses were additionally performed using stratified Cox proportional hazards regression models, focusing specifically on different dimensions of MHD, including major depression, anxiety, and other mental health conditions. In addition, we assessed effect modification effects of baseline vaccination status, which was accomplished by examining the significance of the interaction term between Long COVID status and vaccination within the confines of the stratified Cox proportional hazards regression model. Data were analyzed in R software version 3.6.2 (R Foundation for Statistical Computing, Vienna, Austria) and SAS Software version 9.4 (SAS Institute Inc; Cary, NC) using a two-tailed alpha level of 0.05.

Results

The present analysis incorporated 43,060 non-Long COVID participants (mean [SD] age 54.43 [16.66] years; 58.3% female; 69.5% white) and 4,306 Long COVID participants (mean [SD] age 54.62 [15.34] years; 58.2% female; 69.6% white). Prior to matching, imbalances in distribution were noted for patient regional location (SMD = 0.265), baseline congestive heart failure status (SMD = 0.131), baseline chronic pulmonary disease status (SMD = 0.396), and baseline diabetes status (SMD = 0.147). After the matching process, we observed balanced distributions across all covariates, with all SMDs falling below 0.1 (Table 1).

The overall cumulative incidence rate of MHD was higher among those with Long COVID than those without Long COVID (Figure 2). Similar results were observed when individual MHD diagnosis was examined as outcomes (Figure 2). The overall unadjusted incidence rate of MHD was higher among Long COVID (251.1 per 1000 persons-years [PY])) compared to those without Long COVID (99.5 per 1000-PY) (Table 2). In the age and sex-adjusted stratified Cox model, Long COVID was associated with higher risk of incident MHD (adjusted Hazard Ratio (aHR), 2.53; 95% CI, 2.32 - 2.75). Additional adjustment of the model with 4 demographic factors and 10 comorbidities did not dilute the association (aHR, 2.60; 95% CI, 2.37, 2.85) (Table 2).

In subgroup analyses, we observed a significant effect of Long COVID on major depression (aHR, 3.36; 95% CI, 2.82, 4.00), generalized anxiety disorder (aHR, 3.44; 95% CI, 2.99, 3.96), in the full adjusted stratified Cox regression models. Conversely, while the influence of Long COVID on other mental health conditions was statistically significant, the magnitude of the effect was smaller than the overall effects (aHR, 1.31; 95% CI, 1.08, 1.60) (Table 3). Baseline

vaccination status did not modify the association between Long COVID and MHD (p-value for interaction = 0.96).

Discussion

In the present analysis of real-world data using propensity score matching, survivors of COVID-19 with Long COVID were more than twice as likely to develop MHD. The impact was largest for major depressive disorders and generalized anxiety disorders. Our results are specifically applicable to the new incidence of MHD following a Long COVID diagnosis.

Compared with a previous retrospective cohort study conducted in the U.S., our study demonstrated stronger effect estimates. Xie et al. reported a moderately increased risk of depression (aHR, 1.39; 95% CI, 1.34, 1.43) and anxiety disorders (aHR, 1.35; 95% CI, 1.30, 1.39)[9]. The weakness of the prior study is the lack of generalizability due to the demographic composition of the cohort that were mostly older white men and the limited study time frame of nearly 30 days following acute COVID-19 illness, and yet there is a temporal dynamic nature of the epidemiology of mental health outcomes in the post-acute phase of COVID-19. Moreover, our study extends these findings by concentrating not merely on COVID-19 infection, but specifically on the incidence of mental health disorders following a diagnosis of Long COVID.

Our results are consistent with findings from studies conducted outside the United States. For instance, Murata et al. observed an increased likelihood of mood, anxiety, and psychotic disorders (odds ratio [OR], 1.39; 95% CI, 1.05, 1.85) among 662 COVID-19 patients in Japan, observed from March 2020 to July 2021[19]. Similarly, in Thailand, Phu et al. reported that individuals experiencing Long COVID symptoms were at a substantially higher risk of depression (OR, 4.00) and anxiety (OR, 6.93) in a study involving 939 hospitalized patients

between January 2021 and May 2022[20]. The variation in increased risk may be attributable to differences in participant characteristics; for example, their cohorts were hospital-based, whereas our study utilized general EHR data. Additionally, our approach involved survival analysis to account for time-to-event data and censoring, in contrast to these studies that employed logistic regression models for binary outcomes. In Europe, similar findings have been observed. A descriptive study in France reported a high incidence of cognitive impairment (61 out of 159 patients) among hospitalized COVID-19 patients, as well as a notable prevalence of depression (17 out of 94 patients) and anxiety (22 out of 94 patients) in individuals admitted to the ICU, with these observations made four months post COVID-19[21]. Additionally, a qualitative study conducted in Spain corroborates our findings. Samper-Pardo et al. noted that patients with Long COVID reported diminished self-perceived well-being attributable to persistent symptoms. These patients expressed concerns such as anguish and anxiety about the future, fear of reinfection or relapse, and apprehension regarding return to work. Notably, suicidal thoughts were also reported by several individuals in this cohort[22]. We believe that our study represents a valuable contribution to the field and possesses the potential for global generalizability.

The pathophysiological of MHD among patients with Long COVID is not fully understood. However, there is emerging evidence of both direct and indirect effects of the virus on the brain and psychological outcomes respectively. Starting with the direct effect, the SARS-CoV-2 virus is neurotropic, indicating a direct invasion of the nervous system causing inflammation and gliosis[23]. These changes in the neuronal vascular coupling could lead to further breakdown in the blood-brain barrier, worsening the inflammatory cascade and the influx of inflammatory cells that cause injury to the neurons. Such a direct and neurotropic effect of the virus has been seen in HIV, EBV, and CMV[24]. Indirectly, the pandemic has been associated with psychological

distress[25] due to increased isolation, increased rates of domestic violence[26], disruption of social networks, and unemployment[27], all risk factors of MHD. However, much more research is needed to get a clear picture of how all of these variables may have contributed to MHD.

Moreover, patients with Long COVID often experience unique psychological distress. Fatigue is the leading symptom for Long COVID[28]. The persistent fatigue experienced in Long COVID can limit an individual's ability to engage in daily activities, leading to feelings of frustration, helplessness, and even depression[29]. Additionally, cognitive impairments, commonly referred to as 'Brain Fog', are frequently observed. These impairments, which can include memory, concentration, and decision-making difficulties, may significantly affect an individual's self-esteem and their capacity to work or study, potentially leading to anxiety and depressive symptoms[30]. Sleep disturbances were also commonly observed, with a reported prevalence of 34% to 50% among patients with Long COVID[31]. Issues with sleeping can further aggravate mental health problems, leading to a cycle of insomnia and increased psychological distress[32]. Beyond physical symptoms, there is also a societal aspect to consider. Stigma and misunderstanding about Long COVID are prevalent, as evidenced by a qualitative study showing many participants encountering discrimination in healthcare settings[22]. This lack of understanding and the resultant stigma or dismissive attitudes can be particularly distressing for those suffering from Long COVID[33]. Finally, the unpredictable nature of Long COVID, including uncertainty about recovery, potential long-term impacts and possible effective treatments, can lead to significant anxiety and stress[34].

The findings of the current study have great public health and clinical implications. First, the stigma attached to MHD could hinder health care utilization and thereby worsen the prognosing of the outcomes with a shift towards suicidality and homicidally. Next, given the

disproportionality in lack of access to care in the marginalized communities including people of color and the poor communities, and yet these are the same communities that were mostly affected by severe COVID, the burden of untreated mental health is likely to yield negative outcomes in such communities. For the health care providers, particularly primary health care providers, they should have a lower threshold to screen and treat mental health disorders in this survivor of COVID-19. Furthermore, we anticipate an increase in the economic costs associated with Mental Health Disorders (MHD). In 2010, the global economic burden of mental disorders was estimated at US$2.5 trillion[35]. Considering the extensive impact of the COVID-19 pandemic and the millions of survivors worldwide, it is reasonable to project a dramatic escalation in the costs attributed to MHD.

Study strengths and limitations

Strengths of our study include an analysis based on longitudinal data of a large sample of Long COVID individuals using most recently approved ICD-10 code and clean controls without any diagnosis of Long COVID or COVID-19 diagnoses or testing positive for SARS-CoV-2. To the best of our knowledge, it also is the first large national-real-world analysis to examine the association between Long COVID and MHD using comprehensive propensity score matching approach. Nevertheless, our study has several limitations that should be addressed when interpreting the results. This is an observational study that used US TriNetX data and therefore causality cannot be inferred. We acknowledge that electronic health record databases can misclassify patients based on misreporting or underreporting of diagnoses codes or medications. Moreover, although we excluded any patients with Long COVID or COVID-19 diagnoses or testing positive for SARS-CoV-2, some individuals in the control group might still have undetected mild or asymptomatic COVID-19 because they had not been tested. Such non-

differential misclassification of the exposure may underestimate the strength of the association of COVID-19 with the onset of MHD. Lastly, we acknowledge the potential for information bias in the diagnosis of MHD for two key reasons: (1) There may be an underestimation of MHD prevalence due to limited access to healthcare services. This concern applies to both participants with Long COVID and those without, as constraints in healthcare access could lead to undiagnosed cases[36]. (2) Participants diagnosed with Long COVID are more likely to be identified with MHD. This is because they typically undergo more frequent follow-ups and screenings with healthcare professionals post-diagnosis, increasing the likelihood of MHD detection[37].

Conclusions

Using a large real-world, nationwide, propensity score matched cohort, we found that Long COVID was associated with an increased risk of new onset of MHD. The increase was independent of demographics, lifestyle factors, and major chronic medical conditions. These findings reinforce the importance of integrating mental health screening and services in the treatment and management of Long COVID to prevent related chronic diseases, suicidal thoughts, and attempts.

Funding statement: There was no external or internal funding to support this study.

Competing interests statement: The authors have no conflicts of interest to disclose.

Ethics approval statement: The protocol of this study was reviewed and received a determination of non-human subjects' research by the Penn State Institutional Review Board. The individual informed consent requirement was waived for this secondary analysis of de-identified data

Authors' contributions: Designed research (project conception, development of overall research plan): YZ and DMB. Data extraction and study oversight: DMB. Analyzed data: YZ, PS, and DMB. Performed statistical analysis: YZ and DMB. Wrote the first draft of the manuscript: YZ, PS and DMB. Review and editing: YZ, VMC, PS and DMB. All authors have read and approved the final manuscript.

Data sharing statement: Data is available from third-party partners "TriNetX"

Acknowledgments: None

References

[1] P. Ssentongo, C. Fronterre, and V. M. Chinchilli, "Spatial Epidemiology of COVID-19 Pandemic: Disease Risk, Prognosis, and Complications," in *Frontiers of COVID-19*, Springer International Publishing, 2022, pp. 241–257. doi: 10.1007/978-3-031-08045-6_11.

[2] D. Groff *et al.*, "Short-term and Long-term Rates of Postacute Sequelae of SARS-CoV-2 Infection: A Systematic Review," *JAMA Netw. Open*, vol. 4, no. 10, p. e2128568, Oct. 2021, doi: 10.1001/jamanetworkopen.2021.28568.

[3] H. E. Davis, L. McCorkell, J. M. Vogel, and E. J. Topol, "Long COVID: major findings, mechanisms and recommendations," *Nat. Rev. Microbiol.*, Jan. 2023, doi: 10.1038/s41579-022-00846-2.

[4] L. Bull-Otterson *et al.*, "Post–COVID Conditions Among Adult COVID-19 Survivors Aged 18–64 and ≥65 Years — United States, March 2020–November 2021," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 71, no. 21, pp. 713–717, May 2022, doi: 10.15585/mmwr.mm7121e1.

[5] N. M. Hassan *et al.*, "Prevalence of mental health problems among children with long COVID: A systematic review and meta-analysis," *PLOS ONE*, vol. 18, no. 5, p. e0282538, May 2023, doi: 10.1371/journal.pone.0282538.

[6] Parks, Joe, Svendsen, Dale, Singer, Patricia, and Mary, Fotu, "Morbidity and Mortality in People with Serious Mental Illness," *Natl. Assoc. State Ment. Health Program Dir. NASMHPD Med. Dir. Counc.*, Oct. 2006.

[7] R. C. KESSLER *et al.*, "Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative," *World Psychiatry*, vol. 6, no. 3, pp. 168–176, Oct. 2007.

[8] "Major Depression," National Institute of Mental Health (NIMH). Accessed: Jul. 15, 2023. [Online]. Available: https://www.nimh.nih.gov/health/statistics/major-depression

[9] Y. Xie, E. Xu, and Z. Al-Aly, "Risks of mental health outcomes in people with covid-19: cohort study," *BMJ*, vol. 376, p. e068993, Feb. 2022, doi: 10.1136/bmj-2021-068993.

[10] "Publication Guidelines," TriNetX. Accessed: Jul. 01, 2023. [Online]. Available: https://trinetx.com/real-world-resources/publications/trinetx-publication-guidelines/

[11] "Healthcare Workers," Centers for Disease Control and Prevention. Accessed: Jul. 02, 2023. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/post-covid-conditions.html

[12] E. von Elm, D. G. Altman, M. Egger, S. J. Pocock, P. C. Gøtzsche, and J. P. Vandenbroucke, "Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies," *BMJ*, vol. 335, no. 7624, pp. 806–808, Oct. 2007, doi: 10.1136/bmj.39335.541782.AD.

[13] A. B. Busch, H. A. Huskamp, P. Raja, S. Rose, and A. Mehrotra, "Disruptions in Care for Medicare Beneficiaries With Severe Mental Illness During the COVID-19 Pandemic," *JAMA Netw. Open*, vol. 5, no. 1, p. e2145677, Jan. 2022, doi: 10.1001/jamanetworkopen.2021.45677.

[14] S. W. Lee *et al.*, "Association between mental illness and COVID-19 susceptibility and clinical outcomes in South Korea: a nationwide cohort study," *Lancet Psychiatry*, vol. 7, no. 12, pp. 1025–1031, Dec. 2020, doi: 10.1016/S2215-0366(20)30421-1.

[15] M. E. Charlson, P. Pompei, K. L. Ales, and C. R. MacKenzie, "A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation," *J. Chronic Dis.*, vol. 40, no. 5, pp. 373–383, Jan. 1987, doi: 10.1016/0021-9681(87)90171-8.

[16] H. Quan *et al.*, "Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries," *Am. J. Epidemiol.*, vol. 173, no. 6, pp. 676–682, Mar. 2011, doi: 10.1093/aje/kwq433.

[17] "Find your COVID-19 Vaccine CPT® Codes," American Medical Association. Accessed: Jul. 02, 2023. [Online]. Available: https://www.ama-assn.org/find-covid-19-vaccine-codes

[18] P. C. Austin, "Using the Standardized Difference to Compare the Prevalence of a Binary Variable Between Two Groups in Observational Research," *Commun. Stat. - Simul. Comput.*, vol. 38, no. 6, pp. 1228–1234, May 2009, doi: 10.1080/03610910902859574.

[19] F. Murata, M. Maeda, C. Ishiguro, and H. Fukuda, "Acute and delayed psychiatric sequelae among patients hospitalised with COVID-19: a cohort study using LIFE study data," *Gen. Psychiatry*, vol. 35, no. 3, p. e100802, Jun. 2022, doi: 10.1136/gpsych-2022-100802.

[20] D. H. Phu *et al.*, "Prevalence and factors associated with long COVID and mental health status among recovered COVID-19 patients in southern Thailand," *PLOS ONE*, vol. 18, no. 7, p. e0289382, Jul. 2023, doi: 10.1371/journal.pone.0289382.

[21] Writing Committee for the COMEBAC Study Group *et al.*, "Four-Month Clinical Status of a Cohort of Patients After Hospitalization for COVID-19," *JAMA*, vol. 325, no. 15, pp. 1525–1534, Apr. 2021, doi: 10.1001/jama.2021.3331.

[22] M. Samper-Pardo, B. Oliván-Blázquez, R. Magallón-Botaya, F. Méndez-López, C. Bartolomé-Moreno, and S. León-Herrera, "The emotional well-being of Long COVID patients in relation to their symptoms, social support and stigmatization in social and health services: a qualitative study," *BMC Psychiatry*, vol. 23, p. 68, Jan. 2023, doi: 10.1186/s12888-022-04497-8.

[23] L. Bauer, B. M. Laksono, F. M. S. de Vrij, S. A. Kushner, O. Harschnitz, and D. van Riel, "The neuroinvasiveness, neurotropism, and neurovirulence of SARS-CoV-2," *Trends Neurosci.*, vol. 45, no. 5, pp. 358–368, May 2022, doi: 10.1016/j.tins.2022.02.006.

[24] H. Hotta, "[Neurotropic viruses--classification, structure and characteristics]," *Nihon Rinsho Jpn. J. Clin. Med.*, vol. 55, no. 4, pp. 777–782, Apr. 1997.

[25] S. K. Brooks *et al.*, "The psychological impact of quarantine and how to reduce it: rapid review of the evidence," *The Lancet*, vol. 395, no. 10227, pp. 912–920, Mar. 2020, doi: 10.1016/S0140-6736(20)30460-8.

[26] E. Leslie and R. Wilson, "Sheltering in place and domestic violence: Evidence from calls for service during COVID-19," *J. Public Econ.*, vol. 189, p. 104241, Sep. 2020, doi: 10.1016/j.jpubeco.2020.104241.

[27] T. Elmer, K. Mepham, and C. Stadtfeld, "Students under lockdown: Comparisons of students' social networks and mental health before and during the COVID-19 crisis in Switzerland," *PLOS ONE*, vol. 15, no. 7, p. e0236337, Jul. 2020, doi: 10.1371/journal.pone.0236337.

[28] S. Lopez-Leon *et al.*, "More than 50 long-term effects of COVID-19: a systematic review and meta-analysis," *Sci. Rep.*, vol. 11, no. 1, Art. no. 1, Aug. 2021, doi: 10.1038/s41598-021-95565-8.

[29] W. K. Hou, F. T. Lai, M. Ben-Ezra, and R. Goodwin, "Regularizing daily routines for mental health during and after the COVID-19 pandemic," *J. Glob. Health*, vol. 10, no. 2, p. 020315, Dec. 2020, doi: 10.7189/jogh.10.020315.

[30] A. A. Asadi-Pooya *et al.*, "Long COVID syndrome-associated brain fog," *J. Med. Virol.*, vol. 94, no. 3, pp. 979–984, 2022, doi: 10.1002/jmv.27404.
[31] C. Pena-Orbea *et al.*, "Sleep Disturbance Severity and Correlates in Post-acute Sequelae of COVID-19 (PASC)," *J. Gen. Intern. Med.*, vol. 38, no. 8, pp. 2015–2017, Jun. 2023, doi: 10.1007/s11606-023-08187-3.
[32] D. Freeman, B. Sheaves, F. Waite, A. G. Harvey, and P. J. Harrison, "Sleep disturbance and psychiatric disorders," *Lancet Psychiatry*, vol. 7, no. 7, pp. 628–637, Jul. 2020, doi: 10.1016/S2215-0366(20)30136-X.
[33] X. Zhou, C. Chen, Y. Yao, J. Xia, L. Cao, and X. Qin, "The scar that takes time to heal: A systematic review of COVID-19-related stigma targets, antecedents, and outcomes," *Front. Psychol.*, vol. 13, 2022, Accessed: Dec. 14, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.1026712
[34] D. Russell, N. J. Spence, J.-A. D. Chase, T. Schwartz, C. M. Tumminello, and E. Bouldin, "Support amid uncertainty: Long COVID illness experiences and the role of online communities," *SSM - Qual. Res. Health*, vol. 2, p. 100177, Dec. 2022, doi: 10.1016/j.ssmqr.2022.100177.
[35] S. Trautmann, J. Rehm, and H. Wittchen, "The economic costs of mental disorders," *EMBO Rep.*, vol. 17, no. 9, pp. 1245–1249, Sep. 2016, doi: 10.15252/embr.201642951.
[36] G. Pujolar, A. Oliver-Anglès, I. Vargas, and M.-L. Vázquez, "Changes in Access to Health Services during the COVID-19 Pandemic: A Scoping Review," *Int. J. Environ. Res. Public. Health*, vol. 19, no. 3, p. 1749, Feb. 2022, doi: 10.3390/ijerph19031749.
[37] CDC, "Post-COVID Conditions: Information for Healthcare Providers," Centers for Disease Control and Prevention. Accessed: Dec. 13, 2023. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/post-covid-conditions.html

Table 1. Baseline Characteristics Before and After Matching Stratified with Long COVID Status

Variables	Before matching			After 1:10 matching		
	Non-Long COVID n=186,903	Long COVID n=4,306	SMD*	Non-Long COVID n=43,060	Long COVID n=4,306	SMD*
Age (mean (SD))	55.22 (18.39)	54.62 (15.34)	0.036	54.43 (16.66)	54.62 (15.34)	0.012
Sex (Male) (n (%))	77,742 (41.6)	1,774 (41.2)	0.008	17,938 (41.7)	1,774 (41.2)	0.009
Race (n (%))			0.084			0.013
Black or African American	24,815 (13.3)	511 (11.9)		5,002 (11.6)	511 (11.9)	
Others	7,888 (4.2)	137 (3.2)		1,347 (3.1)	137 (3.2)	
Unknown	25,128 (13.4)	662 (15.4)		6,786 (15.8)	662 (15.4)	
White	129,072 (69.1)	2,996 (69.6)		29,925 (69.5)	2,996 (69.6)	
Patient Regional Location (n (%))			0.265			0.017
Midwest	27,387 (14.7)	892 (20.7)		8,955 (20.8)	892 (20.7)	
Northeast	64,765 (34.7)	1,226 (28.5)		12,536 (29.1)	1,226 (28.5)	
South	69,226 (37.0)	1,362 (31.6)		13,409 (31.1)	1,362 (31.6)	
Unknown	2,215 (1.2)	26 (0.6)		240 (0.6)	26 (0.6)	
West	23,310 (12.5)	800 (18.6)		7,920 (18.4)	800 (18.6)	
Baseline Comorbidities (Yes)						
Myocardial Infarction (n (%))	2,045 (1.1)	102 (2.4)	0.098	837 (1.9)	102 (2.4)	0.029
Congestive Heart Failure (n (%))	5,247 (2.8)	233 (5.4)	0.131	2,033 (4.7)	233 (5.4)	0.031
Peripheral Vascular Disease (n (%))	4,850 (2.6)	139 (3.2)	0.038	1,157 (2.7)	139 (3.2)	0.032
Cerebrovascular Disease (n (%))	4,720 (2.5)	132 (3.1)	0.033	1,142 (2.7)	132 (3.1)	0.025
Chronic Pulmonary Disease (n (%))	8,965 (4.8)	727 (16.9)	0.396	7,360 (17.1)	727 (16.9)	0.006
Rheumatoid Disease (n (%))	2,337 (1.3)	110 (2.6)	0.096	902 (2.1)	110 (2.6)	0.031
Mild Liver Disease (n (%))	3,951 (2.1)	153 (3.6)	0.087	1,324 (3.1)	153 (3.6)	0.027
Diabetes (n (%))	13,064 (7.0)	483 (11.2)	0.147	4,635 (10.8)	483 (11.2)	0.014
Renal Disease (n (%))	6,538 (3.5)	235 (5.5)	0.095	2,024 (4.7)	235 (5.5)	0.034
Cancer (any malignancy) (n (%))	9,749 (5.2)	169 (3.9)	0.062	1,523 (3.5)	169 (3.9)	0.020

*SMD, standardized mean difference; An SMD greater than 0.1 is a threshold recommended for declaring imbalance

Table 2. Incidence rates and stratified Cox proportional hazard models hazard ratio (95% CI) for the association between Long COVID and mental health disorders

Variable	Non-Long COVID	Long COVID
Person-years, y	23,586	2,773
Mental health disorders cases, n	2,347	695
Incidence rate, 95% CI per 1000 person-years*	99.5 (95.5, 103.6)	251.1 (232.3, 270.2)
Model 1	(reference)	2.53 (2.32, 2.75)
Model 2	(reference)	2.60 (2.37, 2.85)

*Unadjusted incidence rate per 1000-person-years
Model 1: stratified by age categories (18-34, 35-44, 45-54, 55-64, 65+) and sex (men/women)
Model 2: stratified by fully propensity score matching ID.

Table 3. Stratified Cox proportional hazard models hazard ratio (95% CI) for the association between Long COVID and four mental health disorders

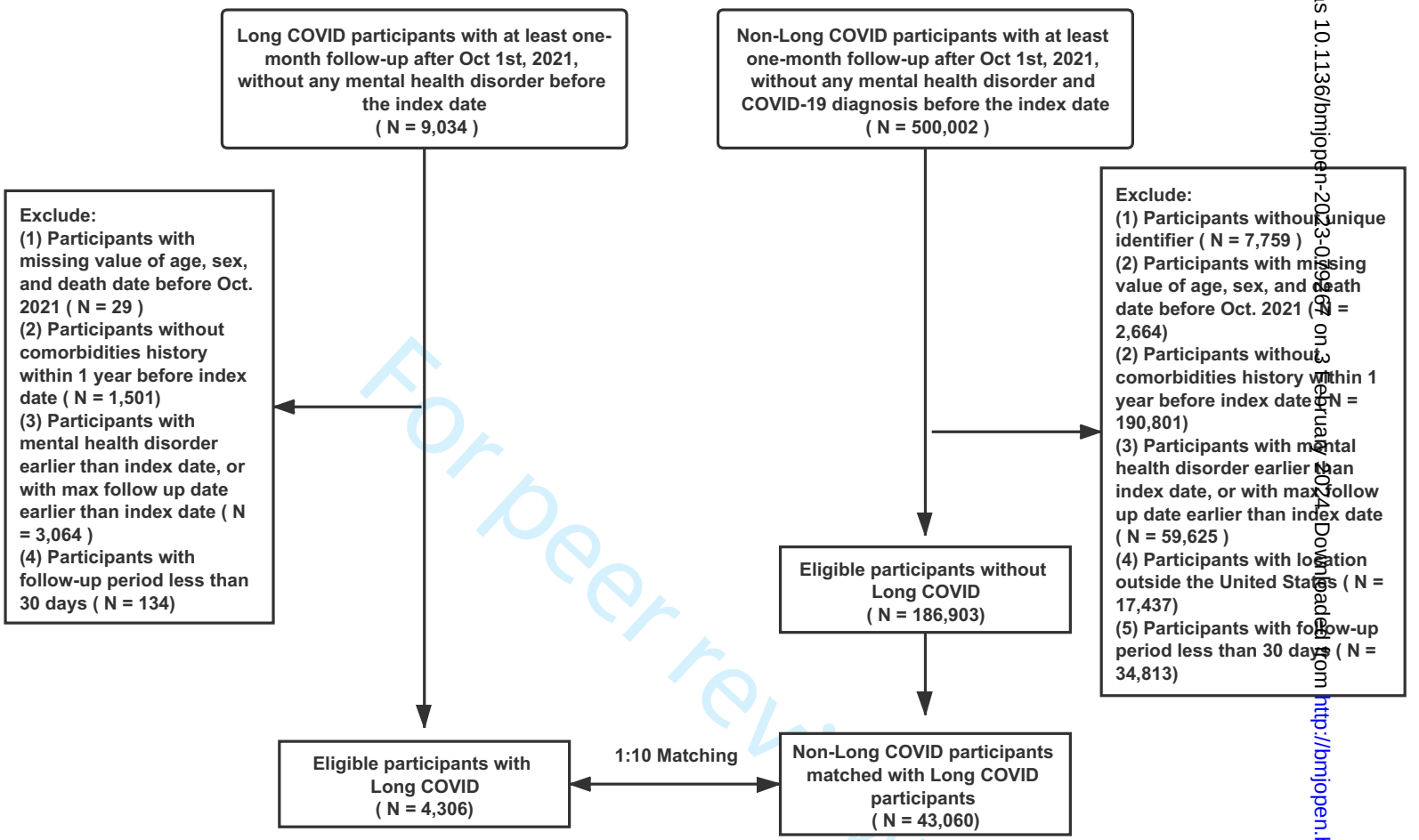
	Non-Long COVID (n, %)	Long COVID (n, %)	HR (95% CI)*
Any Mental Health Condition	2,347 (100%)	695 (100%)	2.60 (2.37, 2.85)
Major Depression[1]	575 (24.5%)	213 (30.6%)	3.36 (2.82, 4.00)
Anxiety[2]	892 (38.0%)	341 (49.1%)	3.44 (2.99, 3.96)
Other mental health condition[3]	880 (34.5%)	141 (20.3%)	1.31 (1.08, 1.60)

1 Major depression (F32, F33)
2 Anxiety (F40 - F48)
3 Other mental health conditions (F10 - F31, F34 - F39, F49 - F99)
* stratified by full propensity score matching ID
** NA due to the small sample size of psychotic disorder cases

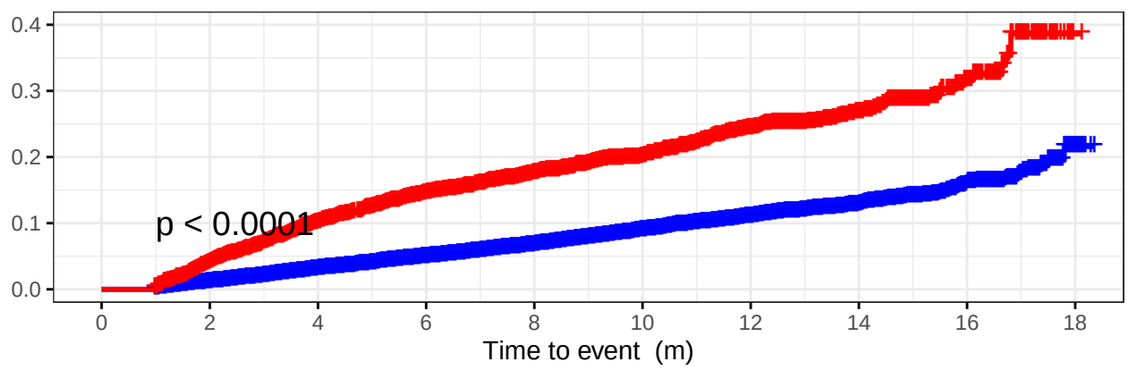
Figures:

Figure 1. Long COVID and Non-Long COVID participants inclusion from TriNetX datasets

Figure 2. Cumulative Incidence Rates of Total and Categorized Major Health Disorders in Patients with Long COVID vs. Those Without Long COVID



All MHD

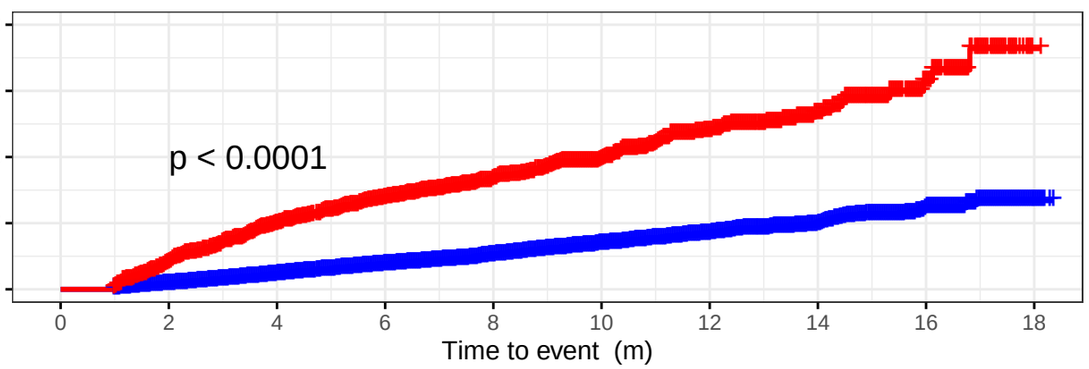


Number at risk

No Long−Covid	43060	37406	28131	20809	14701	9696	5543	2613	829	15
Long−Covid	4306	3915	3189	2584	1974	1413	851	385	126	1

Cummulative any MHD rate

Anxiety

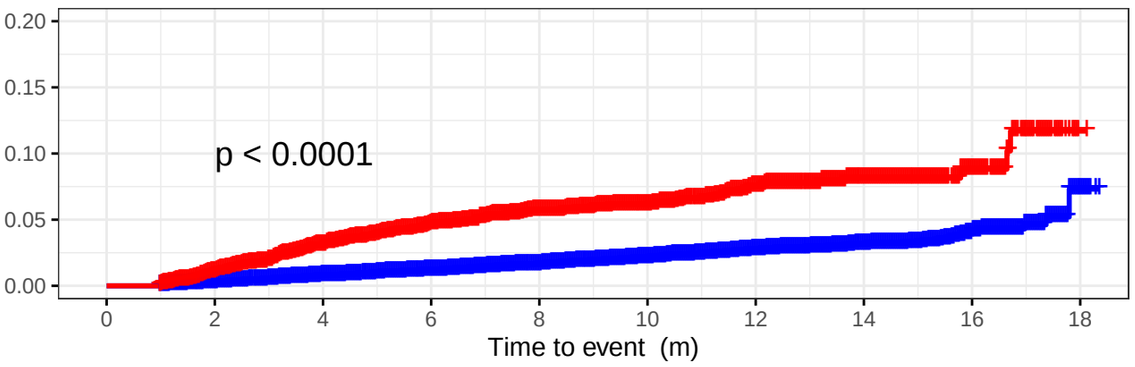


Number at risk

No Long−Covid	43060	37406	28131	20809	14701	9696	5543	2613	829	15
Long−Covid	4306	3915	3189	2584	1974	1413	851	385	126	1

Cummulative anxiety rate

Major depression

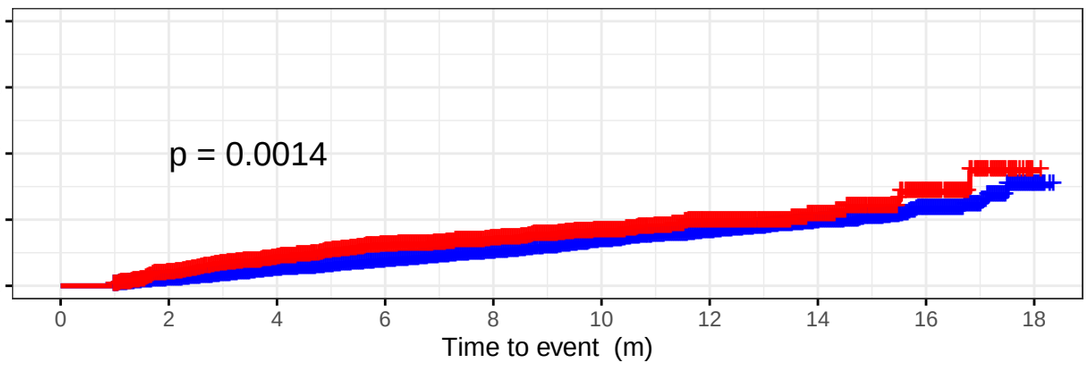


Number at risk

No Long−Covid	43060	37406	28131	20809	14701	9696	5543	2613	829	15
Long−Covid	4306	3915	3189	2584	1974	1413	851	385	126	1

Cummulative major depression disorder rate

Other MHD

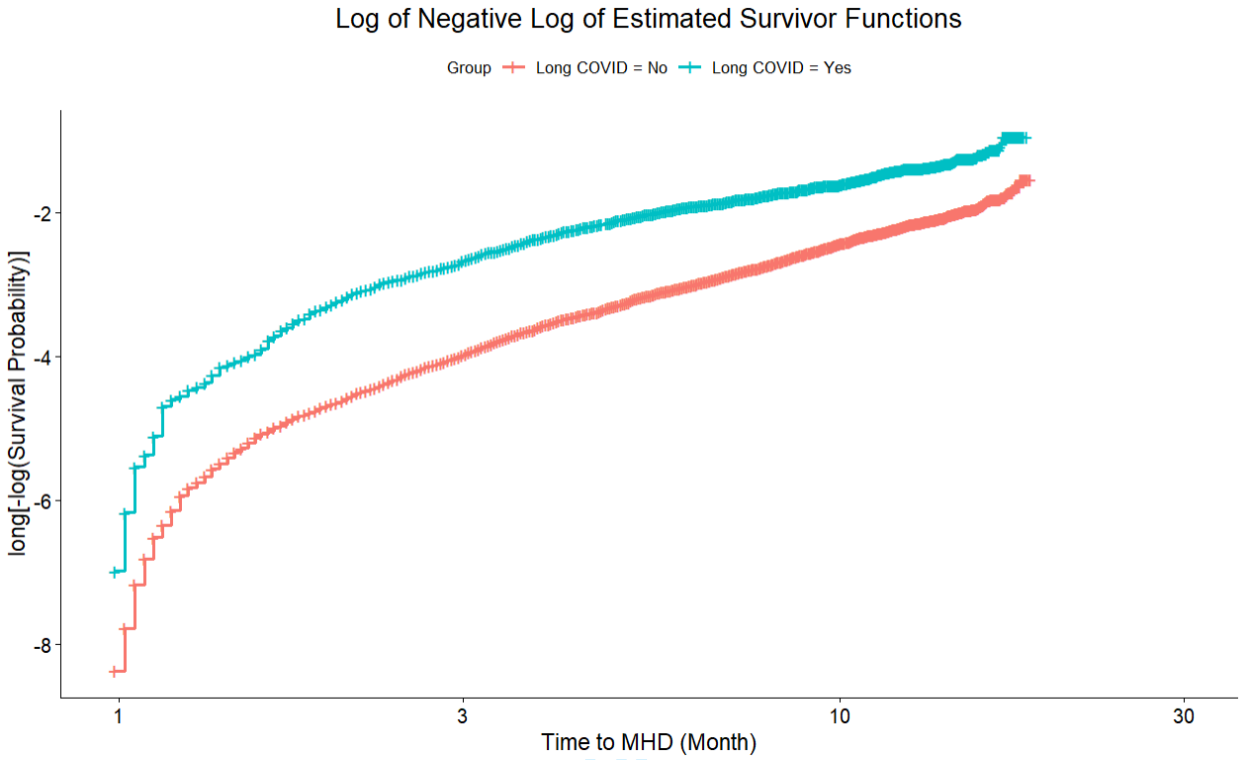


Number at risk

No Long−Covid	43060	37406	28131	20809	14701	9696	5543	2613	829	15
Long−Covid	4306	3915	3189	2584	1974	1413	851	385	126	1

Cummulative other MHD

Supplemental Figure 1. Log-NegLog Survival Plot for checking proportional hazard assumption



Supplementary Table 1: Diagnosis codes of the International Classification of Diseases 10th editions, Clinical Modification (ICD-10-CM), used to describe baseline/preexisting clinical medical conditions; and Current Procedural Terminology (CPT) code for COVID-19 vaccination status

Variables	ICD-10 codes
Long COVID diagnosis	U09.9
Mental health disorders	F10.x-F16.x, F17.x-F69.x, F80.x-F89.x, F90.x-F99.x
Myocardial Infarction	I21.x, I22.x, I25.x,
Congestive Heart Failure	I42.x, I43.x, I50.x,
Peripheral Vascular Disease	I70.x, I71.x, I73.x, I79.x, K55.x, Z95.x
Cerebrovascular Disease	G45.x, G46.x, I60.x-I69.x
Chronic Pulmonary Disease	I27.x, J40.x-J47.x, J60.xJ68.x, J70.x
Mild Liver Disease	B18.x, K70.x K71.x, K73.x, K74.x, K76.x, Z94.x
Diabetes	E10.x-E14.x
Renal Disease	N03.x, N05.x, N18.x, N19.x, Z49.x
Cancer (any malignancy)	C00.x-C26.x, C30.x-C34.x, C37.x-C41.x, C43.x, C45.x-58.x, C60.x-C76.x, C81.x-C85.x, C88.x, C90.x-C97.x

Variables	CPT codes
COVID-19 vaccination	"91300", "91305", "91307", "91308", "91301", "91306", "91311", "91309", "91317", "91315","91312", "91314", "91316", "91313"

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract Page 1 Title
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found Page 2 Abstract
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported Page 4 Introduction
Objectives	3	State specific objectives, including any prespecified hypotheses Page 4 Line 86-90
Methods		
Study design	4	Present key elements of study design early in the paper Page 5-6 Line 105 - 126
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection Page 5-6 Line 92 - 126
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up Page 5 Line 105-108
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed Page 6 Line 119 to 124 and Page 7 Line 124 to 144
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable Page 6 and 7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group Page 5
Bias	9	Describe any efforts to address potential sources of bias Page 7 and Page 8 statistics
Study size	10	Explain how the study size was arrived at Page 6 Line 112 to 126
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why Page 6 and Page 7 Assessment of outcomes and covariates
Statistical methods Page 7 and 8	12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) If applicable, explain how loss to follow-up was addressed
		(*e*) Describe any sensitivity analyses
Results		
Participants Page 9 Line 176 -182	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed

		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram Figure 1
Descriptive data Page 9 Line 176 -182 Table 1	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
		(b) Indicate number of participants with missing data for each variable of interest
		(c) Summarise follow-up time (eg, average and total amount)
Outcome data Line 183-187 Table 2	15*	Report numbers of outcome events or summary measures over time
Main results Line 187-190 Table 2	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		(*b*) Report category boundaries when continuous variables were categorized
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses Line 191 – 197 Table 3	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results Line 199 - 202	18	Summarise key results with reference to study objectives
Limitations Line 282 - 297	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation Line 203 - 262	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability Line 201 – 202; 232-233	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding Line 311	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.