

Characterization of Long COVID Definitions and Clinical Coding Practices

Monika Maripuri MBBS, MPH¹, Andrew T Dey MPH¹, Jacqueline P Honerlaw RN, MPH¹, Chuan Hong PhD², Yuk-Lam Ho MPH¹, Vidisha Tanukonda, MD¹, Alicia W Chen, MS¹, Vidul Ayakulangara Panickan MS², Xuan Wang PhD², Harrison G Zhang BA², Doris Yang⁵, Malarkodi Jebathilagam Samayamuthu MD³, Michele Morris BA³, Shyam Visweswaran MD, PhD³, Brendin R Beaulieu-Jones MD, MBA², Rachel B Ramoni D.M.D, Sc.D⁷, Sumitra Muralidhar Ph.D.⁷, Michael J. Gaziano MD, MPH¹, Katherine P Liao MD, MPH⁶, Zongqi Xia MD, PhD⁸, Gabriel A Brat MD, MPH^{2,4}, Tianxi Cai ScD², Kelly Cho PhD, MPH^{1,9}

1. VA Boston Healthcare System, Boston, MA, United States.
2. Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States.
3. Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, United States.
4. Department of Surgery, Beth Israel Deaconess Medical Center, Boston, MA, United States.
5. Harvard University, Boston, MA, United States
6. Division of Rheumatology, Inflammation, and Immunity, Brigham and Women's Hospital, Boston, MA, United States.
7. Office of Research and Development, U.S. Department of Veterans Affairs, Washington, D.C., United States.
8. Department of Neurology, University of Pittsburgh, Pittsburgh, PA, United States
9. Division of Aging, Mass General Brigham, Harvard Medical School, Boston, MA, United States

Abstract

Background: Long COVID characterized as post-acute sequelae of SARS-CoV-2 (PASC) has no universal clinical case definition. Recent efforts have focused on understanding long COVID symptoms and electronic health records (EHR) data provides a unique resource for understanding this condition. The introduction of the International Classification of Diseases (ICD)-10 code U09.9 for “Post COVID-19 condition, unspecified” to identify patients with long COVID has provided a method of evaluating this condition in EHRs, however, the accuracy of this code is unclear.

Objective: Our study aimed to characterize the utility and accuracy of the U09.9 code across three healthcare systems - The Veterans Health Administration (VHA), Beth Israel Deaconess Medical Center (BIDMC) and The University of Pittsburgh Medical Center (UPMC) against patients identified with long COVID via a chart review by operationalizing the World Health Organization (WHO) and Centers for Disease Control (CDC) definitions.

Methods: COVID positive patients with either a U07.1 ICD code or positive polymerase chain reaction (PCR) test within these healthcare systems were identified for chart review. Among this cohort we sampled patients based on two approaches i) with a U09.9 code and ii) without a U09.9 code but with a new onset PASC related ICD code, which allows us to assess the sensitivity of the U09.9 code. To operationalize the long COVID definition based on health agency guidelines, we grouped symptoms into a “core” cluster of 11 commonly reported symptoms among long COVID patients and an extended cluster, that captured all other symptoms by disease domain. Patients having at least 2 symptoms persisting for ≥ 60 days that were new onset after their COVID infection, with at least one symptom in the core cluster, were labeled as having long COVID per chart review. We compared the performance of the code across three health systems and across different time periods of the pandemic.

Results: A total of 900 patient charts were reviewed across 3 healthcare systems. The prevalence of long COVID among the cohort with the U09.9 ICD code, based on the operationalized WHO definition was between 23.2%-62.4% across these healthcare systems. We also evaluated a less

stringent version of the WHO definition and the Centers for Disease Control (CDC) definition and observed an increase in the prevalence of long COVID at all three healthcare systems.

Conclusions: This is one of the first studies to evaluate the U09.9 code against a clinical case definition for long COVID, as well as the first to apply this definition to EHR data using a chart review approach on a nationwide cohort across multiple healthcare systems. This chart review approach can be implemented at other EHR systems to further evaluate the utility and performance of the U09.9 code.

Keywords: Veterans; Long COVID; Post-Acute Sequelae of SARS-CoV-2 (PASC); U09.9 ICD-10 Code; Algorithm Validation; Chart Review; Electronic Health Records

Introduction

Characterizing the public health burden of post-acute sequelae of SARS-CoV-2 (PASC), also known as long COVID, has been difficult given that multiple clinical case definitions have been proposed by various international health agencies. [1-3] While the exact components of these definitions vary, they share some common underlying features such as the development of symptoms that are new onset after a COVID-19 infection and persistence of new onset symptoms for a duration of time post-acute infection period. Electronic health records (EHR) provide a uniquely rich resource for studying this condition at scale and there have been multiple efforts to describe long-COVID symptoms and estimate prevalence in various EHR systems. [4-9]

Specifically, the introduction of the International Classification of Diseases (ICD)-10 code U09.9 for “Post COVID-19 condition, unspecified” has provided an alternative method of evaluating this condition in EHRs, and its use has been described in various healthcare systems. [10-13] However, the accuracy of the U09.9 code in identifying long COVID has not yet been evaluated against any existing clinical case definitions in a multicenter setting. Clinical coding of long COVID has the potential for misclassification given the heterogeneity and ambiguity around the definition of Long COVID. [14]

Our study aims were to i) Characterize the use of ICD-10 code U09.9 across three healthcare systems and ii) evaluate the accuracy of the U09.9 code against patients identified with long COVID via chart review.

Methods

Data Sources and Study Cohort

The Consortium for Clinical Characterization of COVID-19 by EHR (4CE) is an international consortium for data driven studies on the COVID-19 pandemic. [15] Three health systems from the 4CE Consortium contributed chart review results for the current study, namely the national Veterans Health Administration (VHA), Beth Israel Deaconess Medical Center (BIDMC) and University of Pittsburgh Medical Center (UPMC). Over 15 million patients are collectively provided care across all three health systems. [16-18] The VHA is the largest integrated healthcare system in the United States, with 171 medical centers throughout the country. [16] BIDMC is an academic medical center that is a part of Beth Israel Lahey healthcare system located in Boston and UPMC is a Pittsburgh based healthcare system with 40 hospitals across Pennsylvania. [17-18] All activities were approved by the Institutional Review Boards (IRB) at each of the participating healthcare systems.

We used EHR data from the three health systems to identify COVID positive patients, define patient characteristics and obtain clinical notes for chart review. The sampling strategy for our

chart review is described in Figure 1. Patients who had their first incidence of COVID-19 diagnosis reported within the participating health systems EHR with either a U07.1 ICD code for “COVID-19”, or a positive polymerase chain reaction (PCR) test performed between March 1st, 2020, to December 31st, 2021, were identified for chart review. From this COVID positive cohort, we then sampled patients based on two approaches: i) the presence of the U09.9 ICD code, which was first introduced in the United States in October 2021 or ii) the presence of at least one new onset PASC related ICD code if the patient did not have a U09.9 code. These PASC related ICD codes were selected to enrich the chart review sample for patients who may potentially have long COVID. At the VHA we further sampled patients from two time periods, those who were COVID positive before September 1, 2021 (pre-U09.9 period) and those who were COVID positive after this date (post-U09.9 period).

The presence of PASC related feature ICD codes were identified via a data-driven process using EHR data from ten healthcare systems at 4CE. [19] Initial steps consisted of extracting longitudinal codified features such as ICD codes and mapping these codified features to Phecodes for new onset of conditions after COVID-19 infection. [20] New onset conditions were defined as those which were not present before initial COVID-19 infection. The conditions were selected such that patients with COVID-19 are associated with higher risk of a new onset of the condition after adjusting for baseline confounders such as age, sex, self-reported race, and healthcare utilization. (Supplement 1)

Chart Review Approach

The primary aim of our chart review was to operationalize the clinical case definition for Long COVID by the World Health Organization (WHO), with secondary aims to compare against a less stringent WHO definition and the Centers for Disease Control (CDC) definition.[3,21] The chart review protocol (Supplement 2) was developed at the VHA with guidance from 4CE Consortium subject matter experts to operationalize the WHO and CDC clinical case definitions. Long COVID symptoms were identified from the WHO definition as well as through a literature review, and 11 commonly occurring symptoms among long COVID patients were classified into a “core” symptom cluster. [22-25] All other symptoms were classified into an “extended” symptom cluster based on their disease domain, which included cardiovascular, neurological, dermatological, musculoskeletal, digestive, and respiratory domains. For patients to be labeled as having long COVID per the WHO definition during the chart review (reported here as “WHO-2”), at least 2 new onset symptoms after their COVID infection were required (Figure 2). These could be either i) two “core” symptoms or ii) one “core” and one “extended” cluster symptom, each of which must have persisted for 60 days or longer. All sampled patient charts had at least 6 months of clinical notes for review after the incident COVID-19 infection to allow appropriate assessment of symptoms. During chart review, all symptoms were collected based on their onset and duration of persistence for either 30 or 60 days (Supplement 2) to allow evaluation against multiple long COVID definitions. The less stringent WHO definition (reported here as “WHO-1”) was defined as a patient having just one core symptom persisting for at least 60 days or longer and the CDC definition was defined as a patient having just one core symptom persisting for at least 30 days.

Reviewers had access to all clinical notes 1 year prior to the incident COVID-19 infection to determine baseline symptoms and conditions. Any symptoms present at the time of the incident COVID-19 infection or exacerbations of existing conditions were not considered new onset and thus not captured in the review. However, symptoms that waxed and waned over time were captured. At the VHA a total of 500 patient charts were reviewed and 200 patient charts were reviewed at each of the other two sites - BIDMC and UPMC.

Characterizing U09.9 ICD Code

To characterize the use of the U09.9 ICD code in clinical practice, we investigated the following three metrics i) the frequency of the U09.9 code used over time from October 2021 to September 2022; ii) the frequency of the U09.9 code use across Veterans Integrated Service Networks (VISNs, which are regional systems of care at the VHA); and iii) the time elapsed between COVID diagnosis and U09.9 code assignment.

Results

Characteristics of Study Cohort

Demographics across the health systems varied notably, at the VHA patients who had a COVID-19 diagnosis between March 1, 2020, and December 31, 2021, were generally White and male veterans. Among those who were assigned a U09.9 code the demographics were generally similar with a few notable exceptions (Table 1). At BIDMC and UPMC however the demographics were different with a higher proportion of females who were assigned the U09.9 code. We also observed across all three health systems that a higher proportion of those assigned a U09.9 code had received at least one dose of a COVID vaccine.

We observed a substantial variation in the use of U09.9 code to diagnose long COVID over time and region. Figures 3 and 4 show the results of our characterizations of the U09.9 code use 12 months following its introduction in the United States, on October 1, 2021. Figure 3 shows the frequency of the U09.9 code diagnosis per 10,000 new COVID cases that occurred in the last 12 months from when they received the code. Frequency of the U09.9 code used to diagnose long COVID was highest from January to March 2022 at health system 1, February to March 2022 at health system 2 and December 2021 to January 2022 at health system 3.

There were also large regional differences in the use of U09.9 code across VHA Healthcare system VISNs (Figure 4). VISN 17 assigned the U09.9 code to 28.4% of all patients who received this code at the VHA, while VISN1 assigned the U09.9 code to just 2.4% of all patients who received the code at the VHA.

Chart Review Results

Chart review at the VHA was conducted by 2 clinical reviewers (M.M and J.H.) with a 1% overlap and an inter-rater reliability of 80%. At BIDMC and UPMC chart review was conducted by 1 clinical reviewer (B.B and M.S) respectively. The most common symptoms identified during chart review among patients with long COVID as per the WHO-2 definition were shortness of breath, fatigue, cough, and loss of smell or taste from the core symptom cluster. Among the extended symptom clusters, we most commonly saw symptoms across the cardiovascular, gastrointestinal, neurological, and respiratory disease domains.

Chart review was performed on a total of 900 patients infected with COVID-19 across three healthcare systems. We anonymized these three institutions to provide an unbiased interpretation of the results. The positive predictive value (PPV) of long COVID from chart review, among the sample of patients with the U09.9 ICD code per the WHO-2 definition was 29.8% at health system 1 and 62.4%, 23.2% at health system 2 and 3 respectively. (Figure 5) Among the sample of patients with a new onset PASC features the PPV of long COVID was 7% at health system 1 and 6.74%, 3% at health system 2 and 3 respectively. However, when we consider the WHO-1 and the CDC definitions, the PPV of long COVID was higher at all three health systems but at health system 2 the PPV was slightly higher for the WHO-1 definition but remained the same for the CDC definition. (Figure 5) The overall performance of the U09.9 code based on the WHO-2

definition resulted in a weighted sensitivity of 15% at health system 1 and 4.9%, 19.1% at health system 2,3 respectively.

Additionally, at the VHA we looked at the prevalence of long COVID among patients with the U09.9 ICD code across the different time periods based on their first COVID infection date. From the chart reviewed cohort 44.64% had long COVID from the pre-U09.9 period and 22.75% from the post-U09.9 period at the VHA. The positive predictive value (PPV) of patients with long COVID in the pre-U09.9 period is higher as many of these patients were back coded.

Through our chart review we observed that patients were given the U09.9 code over a wide range of time from less than 29 days to over 365 days following their initial COVID-19 infection. Most patients did not have persisting symptoms post-acute infection and waxing - waning of symptoms was frequently observed.

Discussion

Principal Results

This study provides a comprehensive, multicenter evaluation of the U09.9 code against proposed clinical case definitions for long COVID. It is also the one of the first studies to apply a clinical case definition to EHR data using a chart review approach. The use of EHR data allowed evaluation of the U09.9 code across multiple healthcare systems nationwide. The availability of ample clinical notes enabled reviewers to ascertain whether observed symptoms post COVID infection were truly new onset and to evaluate the duration of new symptoms, which are critical components of case definitions for long COVID. Another strength of our study was that we evaluated both the WHO and CDC clinical case definitions for long COVID since one universal definition is not currently available. Our symptom collection approach (Supplement 2) captured discrete symptoms by duration of 30 or 60 days, which allowed for multiple case definitions to be applied.

There were large variations in the accuracy of use of the U09.9 code for long COVID. We observed that one center had a much higher predictive value for patients with long COVID among the U09.9 cohort across all three definitions than the other two healthcare systems. This health system also had the highest average number of new onset symptoms among patients seen in long COVID clinics. The U09.9 code assignment at this health system could have been more accurate due to a higher proportion of patients being seen at long COVID clinics.

While looking at the capture of PASC symptoms using ICD codes and Natural Language Processing (NLP) data, the performance of NLP is significantly better for all the commonly occurring long COVID symptoms (Figure 5). [14] The accuracy of using either a diagnosis code or NLP had the best results with 97% accuracy for loss of smell or taste and 87% accuracy for chest pain and cough.

Limitations

There were several limitations with our study. The cohort at the VHA had a higher proportion of male patients who were generally older, and predominantly white. Incident COVID-19 infection was required for inclusion in the chart review, and it is possible that patients had an infection outside of the health systems which was not recorded in the EHR. Patients may have also had symptoms that were not reported at healthcare system visits. The regional variation in the number of long COVID clinics may have led to differential capture of symptoms for patients seen at long COVID clinics versus those seen by other care providers. We observed that symptoms among these patients were well documented as most long COVID clinics have a specific template for

evaluating and capturing of COVID-19 symptoms. [26] In some instances, it was difficult to assess whether a symptom was truly new onset due to COVID-19 infection or a result of underlying health conditions noted at baseline. While the WHO definition has been in use since 2021, long COVID is still an evolving disease, and the case definition may change over time as the condition is further characterized. We also faced some challenges in optimizing a heterogeneous and sparse data capture within the EHR systems.

Comparison with Prior Work

The use of the U09.9 code has been described in several cohorts. The NIH's National COVID Cohort Collaborative (N3C) reported on the growing use of U09.9 from October 2021 through January 2022 in a nationwide cohort of 21,072 patients with the code [10]. However, N3C did not require patients in the cohort to have a positive COVID test to evaluate the use of the U09.9 code and 32.3% of patients did not have a COVID index date. McGrath and colleagues also reported increasing use of the U09.9 code in the months following its release in the nationwide Health Verity cohort of 56,143 patients with a U09.9 code which included children under 18 [12]. Similar to N3C, this cohort did not require a COVID positive test for evaluation of the U09.9 code and only 70.4% had a documented COVID infection. Of the COVID positive patients with U09.9, the median time from infection to U09.9 diagnosis was 56 days. A study in Sweden by Bygdell et al reported 10,196 patients with the U09.9 code, [11] They also found that 2.0% of the COVID positive population in the two largest regions of Sweden had U09.9 at least 28 days post infection.

Conclusions

Our findings suggest that the U09.9 code should be used judiciously in EHR-based studies of long COVID. Given the low PPV of the U09.9 code, its usage as a proxy for long COVID is not recommended. However, the sensitivity of the code makes it useful for identifying patients who may have long COVID and thus require further clinical evaluation.

This was one of the initial efforts towards validating long COVID against a clinical case definition and the U09.9 code through a chart review on a nationwide cohort. The chart review approach developed at the VA can be implemented at other EHR systems to further evaluate the utility and performance of the U09.9 code. Further efforts to develop a more refined and reproducible phenotyping algorithm for long COVID is underway utilizing the chart review labels from our study for algorithm training and development.

Table 1 - Patient Demographics. VHA – Veterans Health Administration, BIDMC - Beth Israel Deaconess Medical Center, UPMC - University of Pittsburgh Medical Center

	All COVID positive* patients from March 1st, 2020, to December 31st, 2021			All COVID positive* patients with a U09.9 ICD code			All COVID positive* patients with a new onset PASC feature		
Health Systems	VHA	BIDMC	UPMC	VHA	BIDMC	UPMC	VHA	BIDMC	UPMC
Total	307,909	30,294	147,653	22,196	164	6,057	294,302	7,245	92,120
Age at incident COVID Diagnosis - Mean (SD)	59.2 (16.1)	47.5 (20.5)	45.7 (23.8)	61.7 (15.1)	54.7 (13.9)	55.1 (17.2)	61.2 (15.7)	54.6 (18.4)	47.5 (24.1)
Sex									
Male	272,957 (88.7%)	16,579 (54.7%)	63,421 (43%)	19,275 (86.8%)	58 (35.4%)	2219 (36.6%)	260,055 (88.4%)	2,947 (40.67%)	38,348 (41.6%)
Female	34,898 (11.3%)	13,715 (45.3%)	84,232 (57%)	2,921 (13.2%)	106 (64.6%)	3838 (63.4%)	34,214 (11.6%)	4,298 (59.3%)	53,772 (58.4%)
Race									
White	208,457 (67.7%)	8,983 (29.7%)	125,503 (85%)	16,102 (72.5%)	105 (64%)	4,965 (82%)	197,237 (67.0%)	3,096 (42.7%)	77,471 (84.1%)
Black or African American	69,067 (22.4%)	5,074 (16.7%)	14,966 (10.1%)	3,621 (16.3%)	29 (17.7%)	880 (14.5%)	69,037 (26.5%)	1,730 (23.9%)	10,564 (11.5%)
Asian	3,326 (1.1%)	1,200 (4%)	1,303 (0.9%)	249 (1.1%)	4 (2.4%)	70 (1.2%)	3,360 (1.1%)	270 (3.7%)	986 (1.1%)
American Indian or Alaska Native	3,521 (1.1%)	31 (0.1%)	704 (0.5%)	273 (1.2%)	1 (0.6%)	0 (0%)	3,282 (1.1%)	8 (0.1%)	493 (0.5%)
Native Hawaiian or Pacific Islander	3,189 (1%)	21 (0.07%)	35 (0.02%)	239 (1.1%)	0 (0%)	0 (0%)	3,018 (1.0%)	9 (0.1%)	0 (0%)
Not Reported	0 (0%)	14,985 (49.5%)	5,141 (3.5%)	0 (0%)	25 (15.2%)	142 (2.3%)	0 (0%)	2,132 (29.4%)	2,606 (2.8%)
Vaccination Status									
Received at least one dose of COVID vaccine	199,235 (64.7%)	9,649 (31.9%)	58,244 (39.4%)	15,295 (68.9%)	112 (68.3%)	3,521 (58.1%)	203,569 (69.2%)	4,470 (61.7%)	41,271 (44.8%)
Did not receive at least one dose of COVID vaccine	104,907 (34.1%)	20,645 (68.1%)	89,409 (60.6%)	6,684 (30.9%)	52 (31.7%)	2,536 (41.9%)	87,569 (29.8%)	2,775 (38.3%)	50,849 (55.2%)

*COVID positive is defined as a patient having either a U07.1 ICD-10 Code or a documented Positive PCR Test

Figure 1- Patient sampling strategy for chart review

EHR-Electronic Health Records, ICD-International Classification of Diseases, PCR - Polymerase Chain Reaction, PASC – Post Acute Sequelae SARS-CoV-2, VHA – Veterans Health Administration, BIDMC – Beth Israel Deaconess Medical Center, UPMC – University of Pittsburgh Medical Center

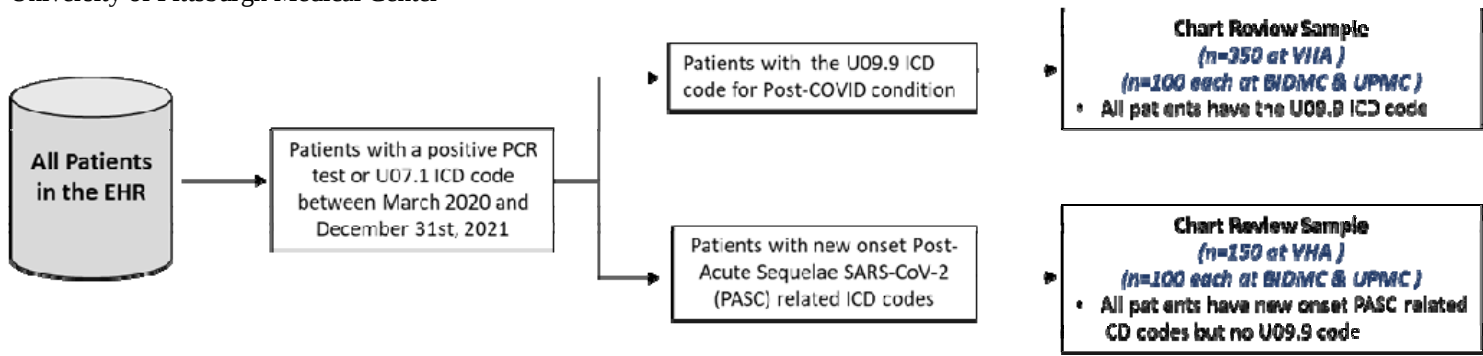


Figure 2 – Chart Review Approach.

WHO-World Health Organization, CDC-Centers for Disease Control

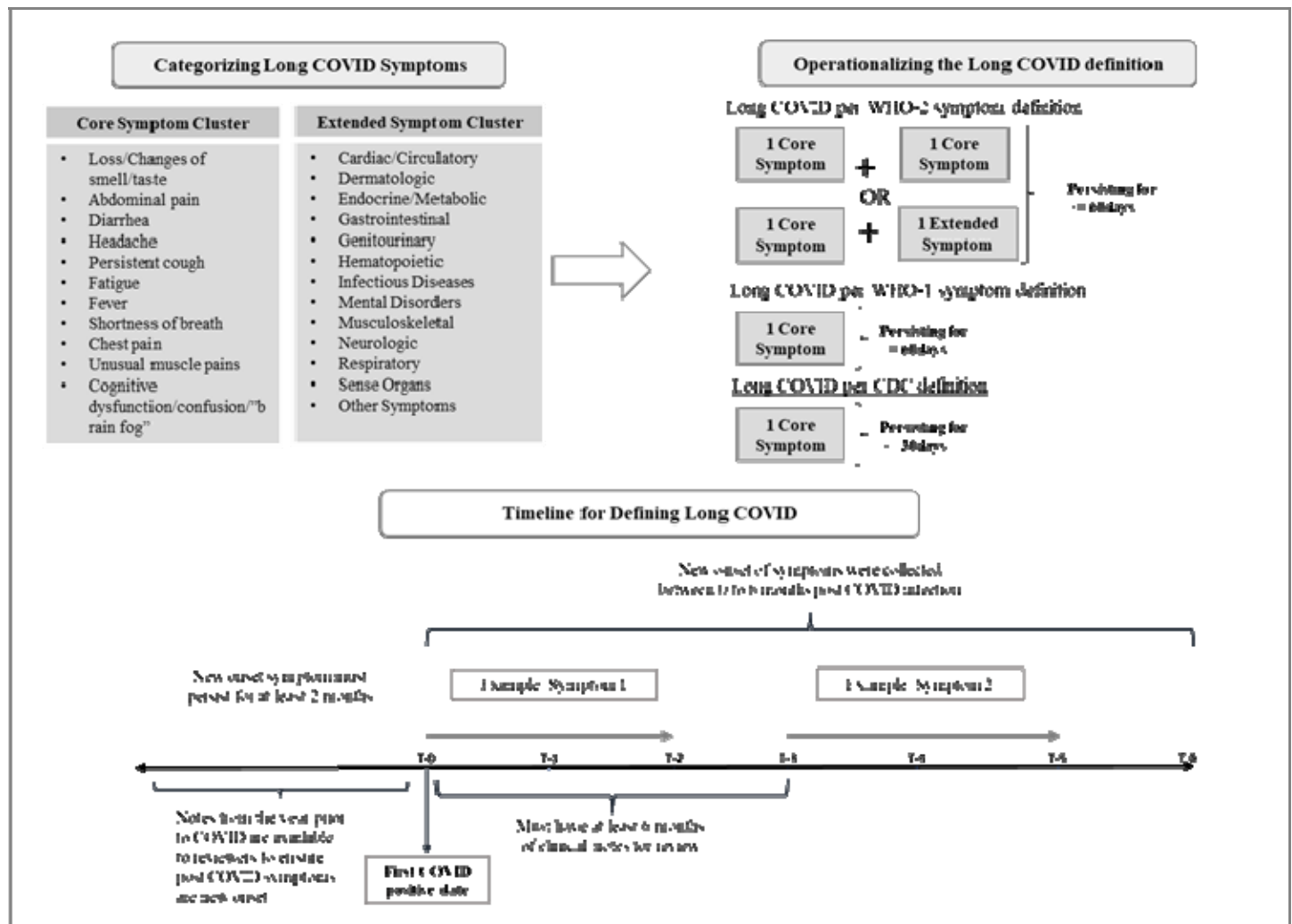


Figure 3 - Frequency of a new U09.9 ICD code assignment per 10,000 new COVID cases that occurred within the previous 12 months.

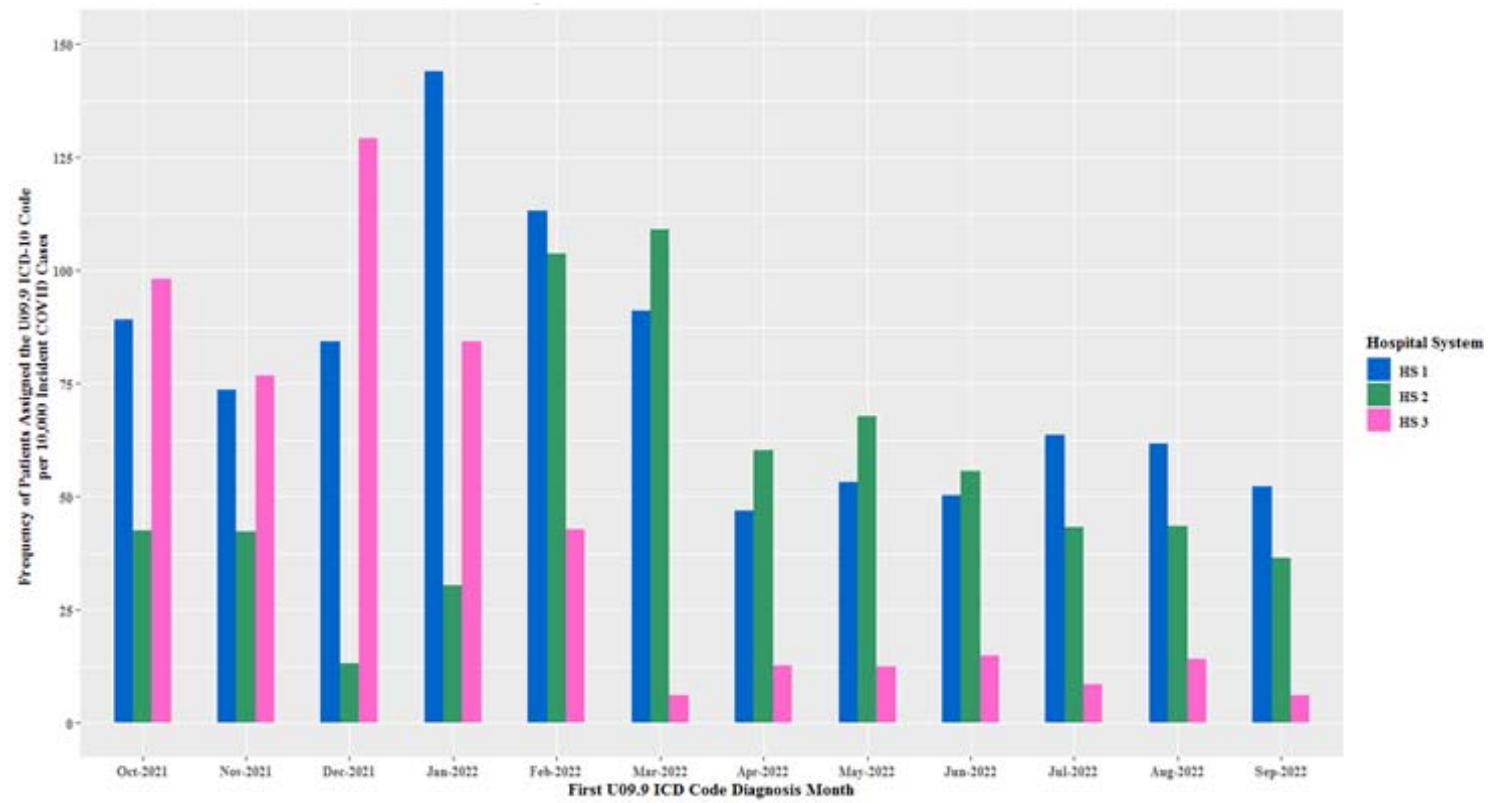


Figure 4 – Prevalence of the U09.9 ICD code by region - VA Veterans Integrated Service Networks (VISN)

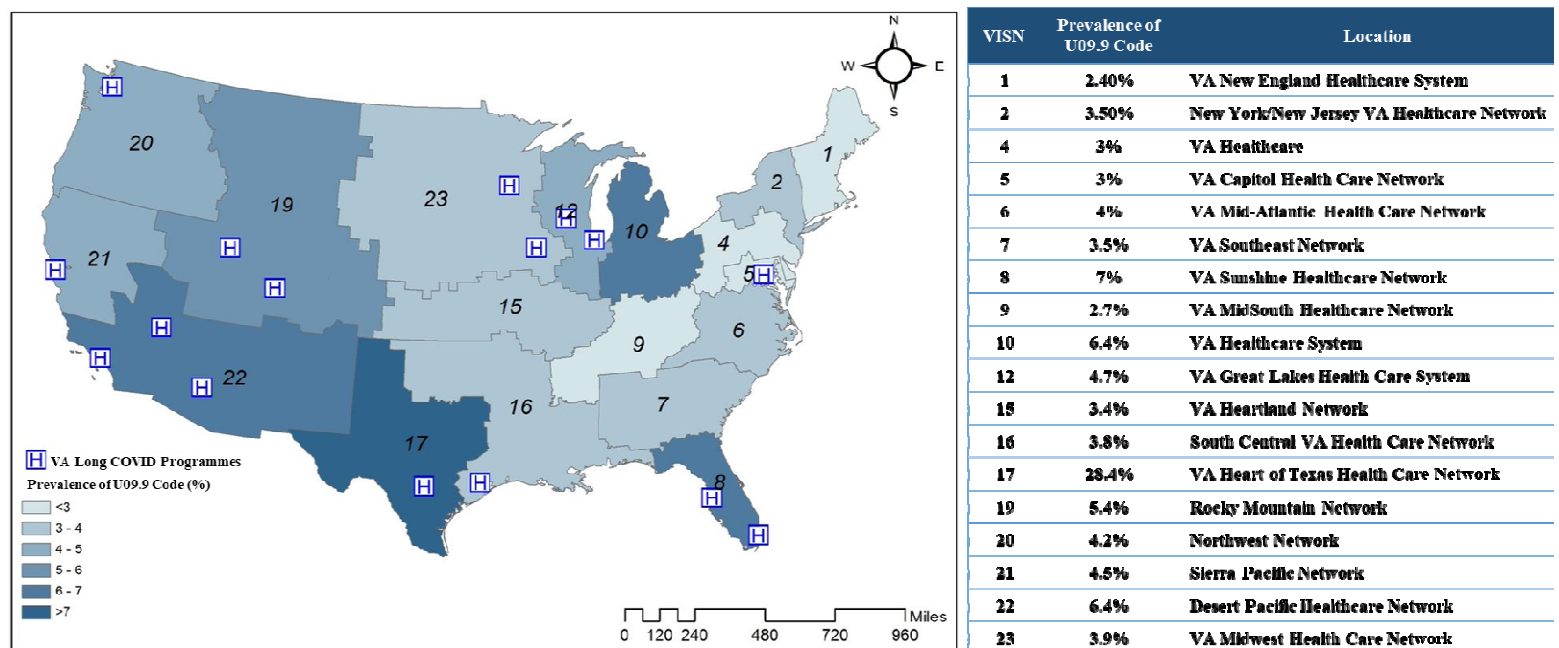
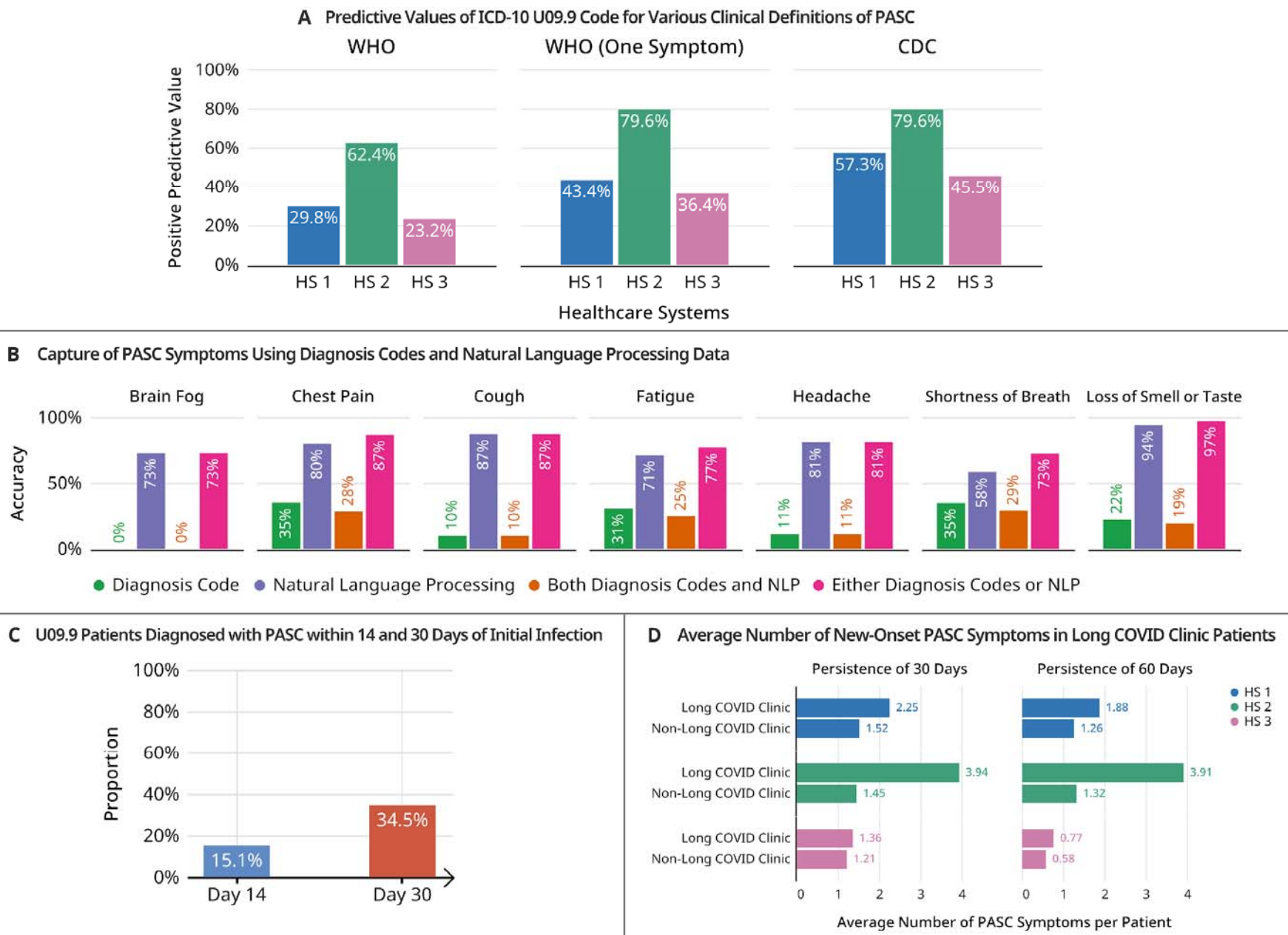


Figure 5 – Comparison of results among all sites.

ICD – International Classification of Diseases, WHO- World Health Organization, CDC- Centers for Disease Control, PASC- Post Acute Sequelae of SARS-CoV-2



References

1. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV. WHO Clinical Case Definition Working Group on Post-COVID-19 Condition. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis.* 2022 Apr;22(4):e102-e107 [FREE Full text] [doi: 10.1016/S1473-3099(21)00703-9.] [Medline: 34951953]
2. Overview: Covid-19 rapid guideline: Managing the long-term effects of covid-19: Guidance (no date) NICE. Available at: <https://www.nice.org.uk/guidance/NG188> [accessed 2022-05-11].
3. Centers for Disease Control and Prevention: Long COVID or Post-COVID Condition. 2022. URL:<https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html> [accessed 2022-05-11]
4. Fritsche LG, Jin W, Admon AJ, Mukherjee B. Characterizing and Predicting Post-Acute Sequelae of SARS CoV-2 Infection (PASC) in a Large Academic Medical Center in the US. *J Clin Med.* 2023 Feb 7;12(4):1328. [FREE Full text] [doi: 10.3390/jcm12041328.] [Medline: 36835863]
5. Pfaff ER, Girvin AT, Bennett TD, Bhatia A, Brooks IM, Deer RR, Dekermanjian JP, Jolley SE, Kahn MG, Kostka K, McMurry JA, Moffitt R, Walden A, Chute CG, Haendel MA; N3C Consortium. Identifying who has long COVID in the USA: a machine learning approach using N3C data. *Lancet Digit Health.* 2022 Jul;4(7):e532-e541. [FREE Full text] [doi: 10.1016/S2589-7500(22)00048-6.] [Medline: 35589549]
6. Thompson EJ, Williams DM, Walker AJ, Mitchell RE, Niedzwiedz CL, Yang TC, Huggins CF, Kwong ASF, Silverwood RJ, Di Gessa G, Bowyer RCE, Northstone K, Hou B, Green MJ, Dodgeon B, Doores KJ, Duncan EL, Williams FMK; OpenSAFELY Collaborative; Steptoe A, Porteous DJ, McEachan RRC, Tomlinson L, Goldacre B, Patalay P, Ploubidis GB, Katikireddi SV, Tilling K, Rentsch CT, Timpson NJ, Chaturvedi N, Steves CJ. Long COVID burden and risk factors in 10 UK longitudinal studies and electronic health records. *Nat Commun.* 2022 Jun 28;13(1):3528. [FREE Full text] [doi: 10.1038/s41467-022-30836-0.] [Medline: 35764621]
7. Wulf Hanson S, Abbafati C, Aerts JG, Al-Aly Z, Ashbaugh C, Ballouz T, Blyuss O, Bobkova P, Bonsel G, Borzakova S, Buonsenso D, Butnaru D, Carter A, Chu H, De Rose C, Diab MM, Ekbohm E, El Tantawi M, Fomin V, Frithiof R, Gamirova A, Glybochko PV, Haagsma JA, Javanmard SH, Hamilton EB, Harris G, Heijenbroek-Kal MH, Helbok R, Hellemons ME, Hillus D, Huijts SM, Hultström M, Jassat W, Kurth F, Larsson IM, Lipcsey M, Liu C, Loflin CD, Malinovschi A, Mao W, Mazankova L, McCulloch D, Menges D, Mohammadifard N, Munblit D, Nekliudov NA, Ogbuonji O, Osmanov IM, Peñalvo JL, Petersen MS, Puhon MA, Rahman M, Rass V, Reinig N, Ribbers GM, Ricchiuto A, Rubertsson S, Samitova E, Sarrafzadegan N, Shikhaleva A, Simpson KE, Sinatti D, Soriano JB, Spiridonova E, Steinbeis F, Svistunov AA, Valentini P, van de Water BJ, van den Berg-Emons R, Wallin E, Witzernath M, Wu Y, Xu H, Zoller T, Adolph C, Albright J, Amlag JO, Aravkin AY, Bang-Jensen BL, Bisignano C, Castellano R, Castro E, Chakrabarti S, Collins JK, Dai X, Daoud F, Dapper C, Deen A, Duncan BB, Erickson M, Ewald SB, Ferrari AJ, Flaxman AD, Fullman N, Gamkrelidze A, Giles JR, Guo G, Hay SI, He J, Helak M, Hulland EN, Kereselidze M, Krohn KJ, Lazzar-Atwood A, Lindstrom A, Lozano R, Magistro B, Malta DC, Månsson J, Mantilla Herrera AM, Mokdad AH, Monasta L, Nomura S, Pasovic M, Pigott DM, Reiner RC, Reinke G, Ribeiro ALP, Santomauro DF, Sholokhov A, Spurlock EE, Walcott R, Walker A, Wiysonge CS, Zheng P, Bettger JP, Murray CJ, Vos T. A global systematic analysis of the occurrence, severity, and recovery pattern of long COVID in 2020 and 2021. *medRxiv*

- [Preprint]. 2022 May 27:2022.05.26.22275532. [FREE Full text] [doi: 10.1101/2022.05.26.22275532.] [Medline: 35664995]
8. Xie Y, Bowe B, Al-Aly Z. Burdens of post-acute sequelae of COVID-19 by severity of acute infection, demographics and health status. *Nat Commun.* 2021 Nov 12;12(1):6571. [FREE Full text] [doi: 10.1038/s41467-021-26513-3.] [Medline: 34772922]
 9. Al-Aly Z, Xie Y, Bowe B. High-dimensional characterization of post-acute sequelae of COVID-19. *Nature.* 2021 Jun;594(7862):259-264. [FREE Full text] [doi: 10.1038/s41586-021-03553-9.] [Medline: 33887749]
 10. Pfaff ER, Madlock-Brown C, Baratta JM, Bhatia A, Davis H, Girvin A, Hill E, Kelly E, Kostka K, Loomba J, McMurry JA, Wong R, Bennett TD, Moffitt R, Chute CG, Haendel M; N3C Consortium; RECOVER Consortium. Coding long COVID: characterizing a new disease through an ICD-10 lens. *BMC Med.* 2023 Feb 16;21(1):58. [FREE Full text] [doi: 10.1186/s12916-023-02737-6.] [Medline: 36793086]
 11. Bygdell M, Leach S, Lundberg L, Gyll D, Martikainen J, Santosa A, Li H, Gisslén M, Nyberg F. A comprehensive characterization of patients diagnosed with post-COVID-19 condition in Sweden 16 months after the introduction of the International Classification of Diseases Tenth Revision diagnosis code (U09.9): a population-based cohort study. *Int J Infect Dis.* 2023 Jan;126:104-113. [FREE Full text] [doi: 10.1016/j.ijid.2022.11.021.] [Medline: 36410693]
 12. McGrath LJ, Scott AM, Surinach A, Chambers R, Benigno M, Malhotra D. Use of the Postacute Sequelae of COVID-19 Diagnosis Code in Routine Clinical Practice in the US. *JAMA Netw Open.* 2022 Oct 3;5(10):e2235089. [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.35089.] [Medline: 36201207]
 13. Ioannou GN, Baraff A, Fox A, Shahoumian T, Hickok A, O'Hare AM, Bohnert ASB, Boyko EJ, Maciejewski ML, Bowling CB, Viglianti E, Iwashyna TJ, Hynes DM. Rates and Factors Associated With Documentation of Diagnostic Codes for Long COVID in the National Veterans Affairs Health Care System. *JAMA Netw Open.* 2022 Jul 1;5(7):e2224359. [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.24359.] [Medline: 35904783]
 14. Zhang HG, Honerlaw JP, Maripuri M, Samayamuthu MJ, Beaulieu-Jones BR, Baig HS, L'Yi S, Ho YL, Morris M, Panickan VA, Wang X, Weber GM, Liao KP, Visweswaran S, Tan BWQ, Yuan W, Gehlenborg N, Muralidhar S, Ramoni RB; Consortium for Clinical Characterization of COVID-19 by EHR (4CE); Kohane IS, Xia Z, Cho K, Cai T, Brat GA. Potential pitfalls in the use of real-world data for studying long COVID. *Nat Med.* 2023 Apr 13. [FREE Full text] [doi: 10.1038/s41591-023-02274-y.] [Medline: 37055567]
 15. Consortium for Clinical Characterization of COVID-19 by EHR 2022. URL: <https://covidclinical.net/> [accessed 2022-05-11]
 16. Administration, V.H. (2008) Veterans Affairs. Available at: <https://www.va.gov/health/> [accessed 2023-03-22]
 17. By the numbers: UPMC Facts and Figures. (2023). Available at: <https://www.upmc.com/about/facts/numbers> [accessed 2023-03-22]
 18. About BIDMC|BIDMC of Boston (n.d). Available at:<https://www.bidmc.org/about-bidmc> [accessed 2023-03-22]
 19. Liu, M., Katsevich, E., Janson, L., & Ramdas, A. (2021, June 4). Fast and Powerful Conditional Randomization Testing via Distillation. *ArXiv.org*. <https://doi.org/10.48550/arXiv.2006.03980>
 20. Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300. <http://www.jstor.org/stable/2346101>
 21. World Health Organization. “Post COVID-19 Condition (Long COVID).” *Www.who.int*, 7 Dec. 2022, www.who.int/europe/news-room/fact-sheets/item/post-covid-19-condition.

22. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, Pujol JC, Klaser K, Antonelli M, Canas LS, Molteni E, Modat M, Jorge Cardoso M, May A, Ganesh S, Davies R, Nguyen LH, Drew DA, Astley CM, Joshi AD, Merino J, Tsereteli N, Fall T, Gomez MF, Duncan EL, Menni C, Williams FMK, Franks PW, Chan AT, Wolf J, Ourselin S, Spector T, Steves CJ. Attributes and predictors of long COVID. *Nat Med.* 2021 Apr;27(4):626-631. [FREE Full text] [doi: 10.1038/s41591-021-01292-y.] [Medline:33692530]
23. Cervia C, Zurbuchen Y, Taeschler P, Ballouz T, Menges D, Hasler S, Adamo S, Raeber ME, Bächli E, Rudiger A, Stüssi-Helbling M, Huber LC, Nilsson J, Held U, Puhon MA, Boyman O. Immunoglobulin signature predicts risk of post-acute COVID-19 syndrome. *Nat Commun.* 2022 Jan 25;13(1):446. [FREE Full text] [doi: 10.1038/s41467-021-27797-1.] [Medline:35078982]
24. Lambert N; Survivor Corps; El-Azab SA, Ramrakhiani NS, Barisano A, Yu L, Taylor K, Esperança Á, Mendiola C, Downs CA, Abraham HL, Hughes T, Rahmani AM, Borelli JL, Chakraborty R, Pinto MD. The other COVID-19 survivors: Timing, duration, and health impact of post-acute sequelae of SARS-CoV-2 infection. *J Clin Nurs.* 2022 Sep 30;10.1111/jocn.16541. [FREE Full text] [doi: 10.1111/jocn.16541.] [Medline:36181315]
25. Su Y, Yuan D, Chen DG, Ng RH, Wang K, Choi J, Li S, Hong S, Zhang R, Xie J, Kornilov SA, Scherler K, Pavlovitch-Bedzyk AJ, Dong S, Lausted C, Lee I, Fallen S, Dai CL, Baloni P, Smith B, Duvvuri VR, Anderson KG, Li J, Yang F, Duncombe CJ, McCulloch DJ, Rostomily C, Troisch P, Zhou J, Mackay S, DeGottardi Q, May DH, Taniguchi R, Gittelman RM, Klinger M, Snyder TM, Roper R, Wojciechowska G, Murray K, Edmark R, Evans S, Jones L, Zhou Y, Rowen L, Liu R, Chour W, Algren HA, Berrington WR, Wallick JA, Cochran RA, Micikas ME; ISB-Swedish COVID-19 Biobanking Unit; Wrin T, Petropoulos CJ, Cole HR, Fischer TD, Wei W, Hoon DSB, Price ND, Subramanian N, Hill JA, Hadlock J, Magis AT, Ribas A, Lanier LL, Boyd SD, Bluestone JA, Chu H, Hood L, Gottardo R, Greenberg PD, Davis MM, Goldman JD, Heath JR. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell.* 2022 Mar 3;185(5):881-895.e20. [FREE Full text] [doi: 10.1016/j.cell.2022.01.014.] [Medline:35216672]
26. Gustavson AM, Eaton TL, Schapira RM, Iwashyna TJ, Adly M, Purnell A. Approaches to long COVID care: the Veterans Health Administration experience in 2021. *BMJ Mil Health.* 2022 Jul 1:e002185. [FREE Full text] [doi: 10.1136/military-2022-002185.] [Medline:35777859]