Title: *EHR-based Case Identification of Pediatric Long COVID: A Report from the RECOVER EHR Cohort*

Authors:

Morgan Botdorf, PhD[1*], Kimberley Dickinson, MS[1*], Vitaly Lorman PhD[1], Hanieh Razzaghi, PhD[1], Nicole Marchesani, MS[1], Suchitra Rao, MBBS, MSCS[2], Colin Rogerson, MD[3], Miranda Higginbotham, MHA[1], Asuncion Mejias, MD, PhD, MsCS[4], Daria Salyakina, PhD[5], Deepika Thacker, MD[6], Dima Dandachi, MD, MPH[7], Dimitri A Christakis MD MPH[8], Emily Taylor, M.A[9], Hayden Schwenk, MD, MPH[10], Hiroki Morizono, PhD[11], Jonathan Cogen MD, MPH[12], Nathan M Pajor MD, MS[13], Ravi Jhaveri, MD[14], Christopher B. Forrest, MD, PhD[1], L. Charles Bailey, MD, PhD[1], *on behalf of the RECOVER Consortium*

Affiliations:

[1] Applied Clinical Research Center, Children's Hospital of Philadelphia, Philadelphia, PA

[2] Department of Pediatrics, University of Colorado School of Medicine and Children's Hospital Colorado, Denver, CO

[3] Division of Critical Care, Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN

[4] Division of Infectious Diseases, Department of Pediatrics, Nationwide Children's Hospital and The Ohio State University, Columbus, OH

[5] Center for Precision Medicine, Nicklaus Children's Hospital, Miami, FL

[6] Nemours Cardiac Center, Alfred I. duPont Hospital for Children, Wilmington, DE.

[7] Division of Infectious Diseases, Department of Medicine, University of Missouri-Columbia, Columbia, MO

[8] Center for Child Health, Behavior and Development, Seattle Children's Research Institute, Seattle, WA

[9] RECOVER Patient, Caregiver, or Community Representative New York, NY, USA

[10] Division of Pediatric Infectious Diseases, Stanford School of Medicine, Palo Alto, CA

[11] Center for Genetic Medicine Research, Children's National Hospital, Washington, DC

[12] Division of Pulmonary and Sleep Medicine, Department of Pediatrics, Seattle Children's Hospital, University of Washington, Seattle, WA

[13] Division of Pulmonary Medicine, Cincinnati Children's Hospital Medical Center and University of Cincinnati College of Medicine, Cincinnati OH

[14] Division of Infectious Diseases, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL

Authorship Statement

Authorship has been determined according to ICMJE recommendations.

Corresponding Author:

Name: Morgan Botdorf

Affiliation: Applied Clinical Research Center, Children's Hospital of Philadelphia

Email: botdorfm@chop.edu

Abbreviated Title: Identifying Children with Long COVID

Funding Source: This research was funded by the National Institutes of Health (NIH) Agreement OTA OT2HL161847-01 as part of the Researching COVID to Enhance Recovery (RECOVER) program of research.

Disclaimer: This content is solely the responsibility of the authors and does not necessarily represent the official views of the RECOVER Initiative, the NIH, or other funders.

Key Words: PEDSnet, Post-acute sequelae SARS-CoV-2 infection, Long COVID, Chronic COVID-19 Syndrome, Late sequelae of COVID-19, Long haul COVID, Long-term COVID-19, Post COVID syndrome, Post-acute COVID-19, Rule-based phenotyping, Electronic health records, Electronic phenotyping, Chart review

Abbreviations: PASC—post-acute sequelae of SARS-CoV-2 infection; COVID-19—coronavirus disease 2019; SARS-CoV-2—severe acute respiratory syndrome coronavirus 2; PCR—polymerase chain reaction; EHR—electronic health record; MIS-C—multisystem inflammatory syndrome in children; ICD-10—International Classification of Diseases, version 10.

Abstract

Objective

Long COVID, marked by persistent, recurring, or new symptoms post-COVID-19 infection, impacts children's well-being yet lacks a unified clinical definition. This study evaluates the performance of an empirically derived Long COVID case identification algorithm, or computable phenotype, with manual chart review in a pediatric sample. This approach aims to facilitate large-scale research efforts to understand this condition better.

Methods

The algorithm, composed of diagnostic codes empirically associated with Long COVID, was applied to a cohort of pediatric patients with SARS-CoV-2 infection in the RECOVER PCORnet EHR database. The algorithm classified 31,781 patients with conclusive, probable, or possible Long COVID and 307,686 patients without evidence of Long COVID. A chart review was performed on a subset of patients (n=651) to determine the overlap between the two methods. Instances of discordance were reviewed to understand the reasons for differences.

Results

The sample comprised 651 pediatric patients (339 females, M_{age} = 10.10 years) across 16 hospital systems. Results showed moderate overlap between phenotype and chart review Long COVID identification (accuracy = 0.62, PPV = 0.49, NPV = 0.75); however, there were also numerous cases of disagreement. No notable differences were found when the analyses were stratified by age at infection or era of infection. Further examination of the discordant cases revealed that the most common cause of disagreement was the clinician

reviewers' tendency to attribute Long COVID-like symptoms to prior medical conditions. The performance of the phenotype improved when prior medical conditions were considered (accuracy = 0.71, PPV = 0.65, NPV = 0.74).

Conclusions

Although there was moderate overlap between the two methods, the discrepancies between the two sources are likely attributed to the lack of consensus on a Long COVID clinical definition. It is essential to consider the strengths and limitations of each method when developing Long COVID classification algorithms.

Introduction

Long COVID, also known as post-acute sequelae of SARS-CoV-2 infection (PASC), is a significant health concern characterized by ongoing, relapsing, or new symptoms emerging four or more weeks after the acute infection phase[1]. While post-viral syndromes like chronic fatigue syndrome following mononucleosis are well-documented in children[2-3], understanding the clinical manifestations of Long COVID in pediatric patients remains incomplete. The variability of symptoms in children compared to adults complicates diagnosis and treatment[4-9]. Symptoms can range from fatigue and headache to loss of taste and smell and chest pain[4-9]. Although rare, diagnosed conditions associated with Long COVID include myocarditis, myositis, postural tachycardia syndrome (POTS), and myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS), among other conditions[10]. Despite certain symptoms and conditions clearly attributable to a SARS-CoV-2 infection, like multisystem inflammatory syndrome (MIS-C), much remains to be understood about others[11-12]. These symptoms and conditions impose a substantial burden on children and their families, leading to missed school and the need for service referrals[13-14]. This highlights the importance of improved detection and treatment strategies.

Identifying children who suffer from Long COVID in research studies is crucial to better understand this disorder and ensuring timely detection and treatments in clinical settings. However, this task is challenging due to the inconsistency and heterogeneity of associated symptoms. To address this challenge, researchers have used large observational cohort studies that use repositories of electronic health record (EHR) data to identify patients[5, 8, 9, 15, 16]. These studies have primarily relied on EHR-based diagnosis codes[15-16]. The ICD-10-CM U09.9 code, introduced in October 2021[17-18], allows clinicians to assign a Long COVID diagnosis; however, its utilization remains inconsistent and potentially biased across patients and healthcare settings [16]. Additionally, relying solely on this

code may not adequately capture all patients due to the variety of symptoms associated with Long COVID. This poses a risk of misclassification if researchers exclusively use the U09.9 code for phenotyping.

To improve identification of patients with Long COVID, computable phenotyping techniques, which involve developing a set of rules to identify patients with a disorder, have been used in Long COVID studies. Long COVID phenotypes for adult[19] and pediatric[20] patients have been developed using machine-learning approaches that leverage large numbers of clinical features. For example, in a recent pediatric study, a machine learning algorithm demonstrated high precision in classifying both general and MIS-C-specific forms of PASC, with recall rates of up to 70% [20]. Training these supervised learning models requires a labeled cohort of patients who likely have Long COVID based on healthcare utilization or Long COVID diagnosis codes. Since there is no gold-standard definition of Long COVID, it is difficult to produce an unbiased labeled training set, which limits the generalizability of the models.

In this study, we aimed to 1) identify children with Long COVID by utilizing a rules-based computable phenotype approach and 2) assess the performance of this computable phenotype for Long COVID in a subset of children. This approach involves analyzing specific diagnosis coding and symptoms that occur more frequently after a COVID-19 infection. By doing so, we can more accurately identify a larger number of children with Long COVID. In addition, we have included clinician reviews of patient charts to gain a comprehensive understanding of patients' experience with Long COVID. This combined approach represents a significant step in the automation of Long COVID clinical phenotypes using EHR data in the absence of a consensus definition.

Methods

Data Source

This retrospective cohort study is part of the NIH Researching COVID to Enhance Recovery (RECOVER) Initiative, which seeks to understand, treat, and prevent the post-acute sequelae of SARS-CoV-2 infection[21]. The RECOVER PCORnet EHR cohort includes clinical data from patients in 40 hospital systems across the United States. Data were extracted from version 6 of the pediatric RECOVER database, comprising more than 9 million children who were tested for SARS-CoV-2, diagnosed with COVID-19, or received a COVID-19 vaccine between 2019 and December 2022. Institutional Review Board (IRB) approval was obtained under Biomedical Research Alliance of New York (BRANY) protocol #21-08-508. BRANY waived the need for consent and HIPAA authorization.

Study Population

Inclusion criteria for our pediatric sample were as follows: 1) SARS-CoV-2 infection confirmed via clinical diagnosis or PCR, antigen, or qualifying serology test[22] between March 2020 and December 2022, 2) age less than 21 years at first COVID-19 infection, and 3) at least two contacts with the healthcare system (at least one being in-person or telehealth) to ensure adequate follow-up during the post-acute phase (28-179 days following infection). We defined clinically meaningful time periods surrounding the initial COVID-19 infection as shown in Figure 1. The acute phase spanned until the 27th day post-infection. The post-acute phase, which was the primary focus of our analyses, spanned from day 28 through day 179 post-infection, ensuring that symptoms were not directly related to the acute COVID-19 infection. For patients with a specific COVID-19 diagnosis or viral test, the initial infection date was the date

of diagnosis or test. For patients with diagnoses indicating “history of” or “complication of” COVID-19 or with a positive serology test, we used 28 days prior to the earliest diagnosis or test evidence of COVID-19 as a proxy for initial infection date.

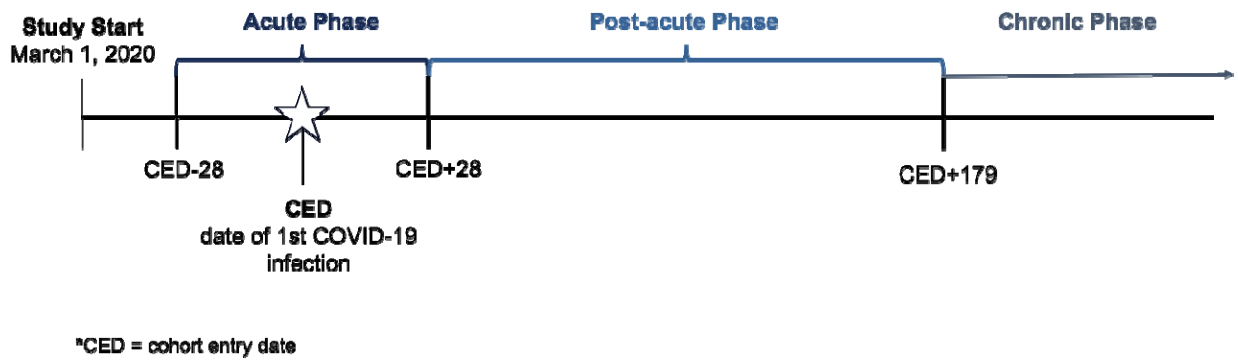

Figure 1. *Study Timeline.*

Phenotype classification

Patients were identified as having conclusive, probable, or possible Long COVID according to the algorithm described in Figure 2, which used criteria documented in the EHR in the post-acute period. The algorithm accounts for diagnoses of Long COVID (ICD-10-CM code U09.9), diagnoses of MIS-C (ICD-10-CM code M35.81), diagnoses of sequelae of specified infectious and parasitic diseases (ICD-10-CM code B94.8), and 23 diagnosis clusters identified as probable indicators of Long COVID based on our prior work[5,9] (Supplemental File 1). The diagnosis clusters were formed using a data mining approach that identified conditions more common in U09.9-diagnosed patients than in non-U09.9 diagnosed COVID-19+ patients in the post-acute period[5]. Clinicians then

reviewed the diagnosis codes to create clusters of ICD-10-CM codes. Clusters included abdominal pain, abnormal liver enzymes, acute kidney injury, acute respiratory distress syndrome, arrythmias, autonomic dysfunction, cardiovascular signs/symptoms, changes in taste/smell, chest pain, cognitive function, generalized pain, fatigue/malaise, fever, fluid/electrolyte balance, headache, heart disease, myocarditis, musculoskeletal symptoms, myositis, respiratory signs/symptoms, and thrombophlebitis/thromboembolism. Any patient with two or more diagnoses within the same cluster separated by at least 28 days during the post-acute period was labeled as having probable Long COVID, regardless of whether the patient had a specific Long COVID or MIS-C diagnosis code. Figure 2 depicts the steps applied to classify patients according to the certainty of them having Long COVID. Any patient with conclusive, probable, or possible Long COVID detected by the phenotype was labeled as “Long COVID Evidence” and all others were labelled as “No Long COVID Evidence”.

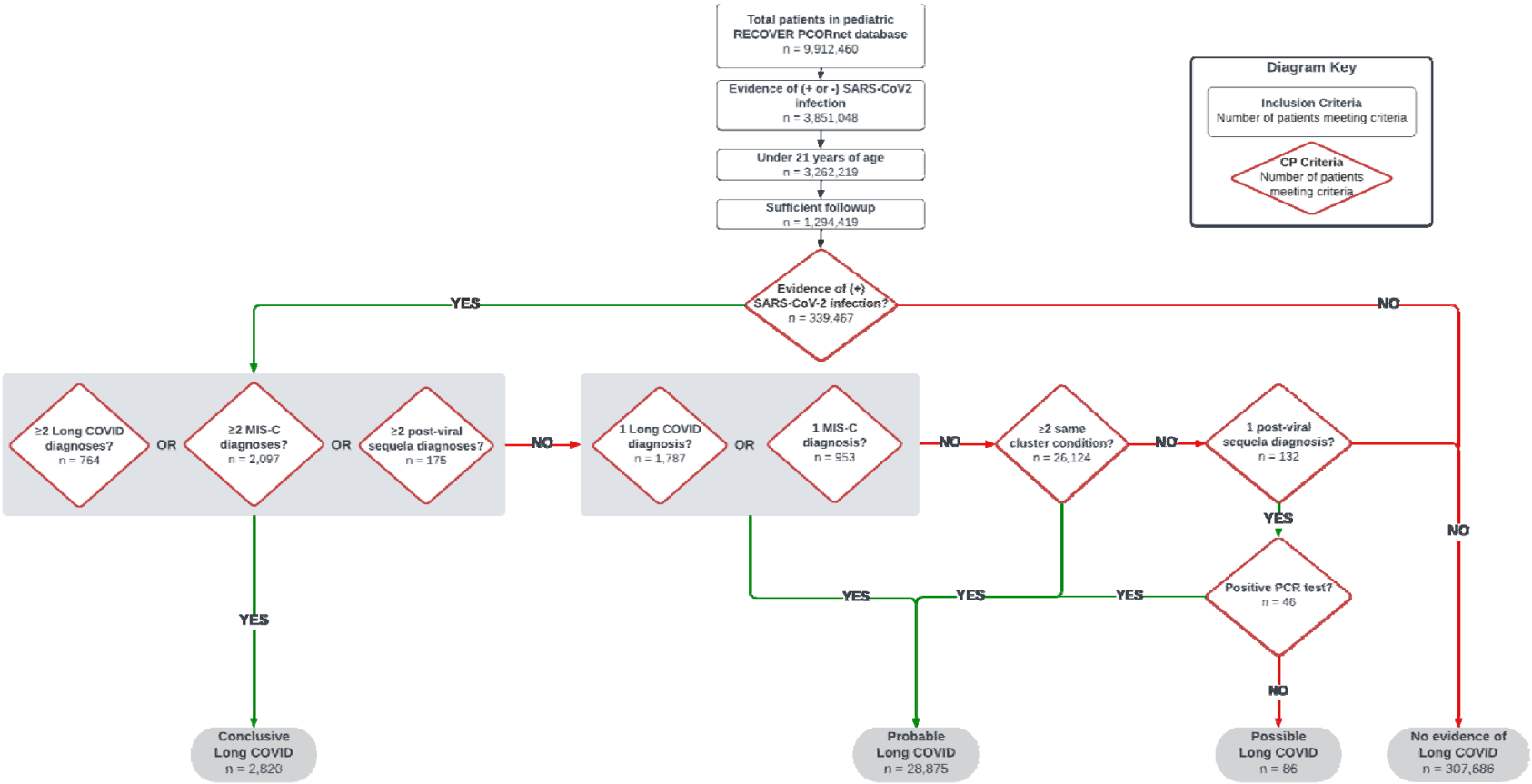

Figure 2. *Flow chart depicting attrition and algorithm used to identify patients with Long COVID.*

Chart Review Sampling

A manual chart review was performed on a subset of the study population at 16 institutions. We sampled 702 patients split between the Long COVID Evidence and No Long COVID Evidence groups to ensure there was adequate representation across sites. The sampling strategy is laid out in Figure 3. Approximately 22 Long COVID Evidence patients were randomly sampled per institution. Each Long COVID Evidence sampled patient was matched 1:1 without replacement with a No Long COVID Evidence patient using exact matching on institution, age at time of infection, calendar quarter of infection, and acute period hospitalization (yes/no). Ninety percent of the No Long COVID Evidence sample had SARS-CoV-2 infection while the remaining ten percent (35 patients) were patients with no evidence of SARS-CoV-2 but with at least two diagnoses of cluster conditions separated by 28 to 150 days. The latter group were additional patients included in the chart review to gather insight on the attribution of cluster diagnoses to conditions other than SARS-CoV-2 infection. A total of 651 children were ultimately included in analyses based on additional exclusions which will be discussed. The sampling strategy is laid out in Figure 3.

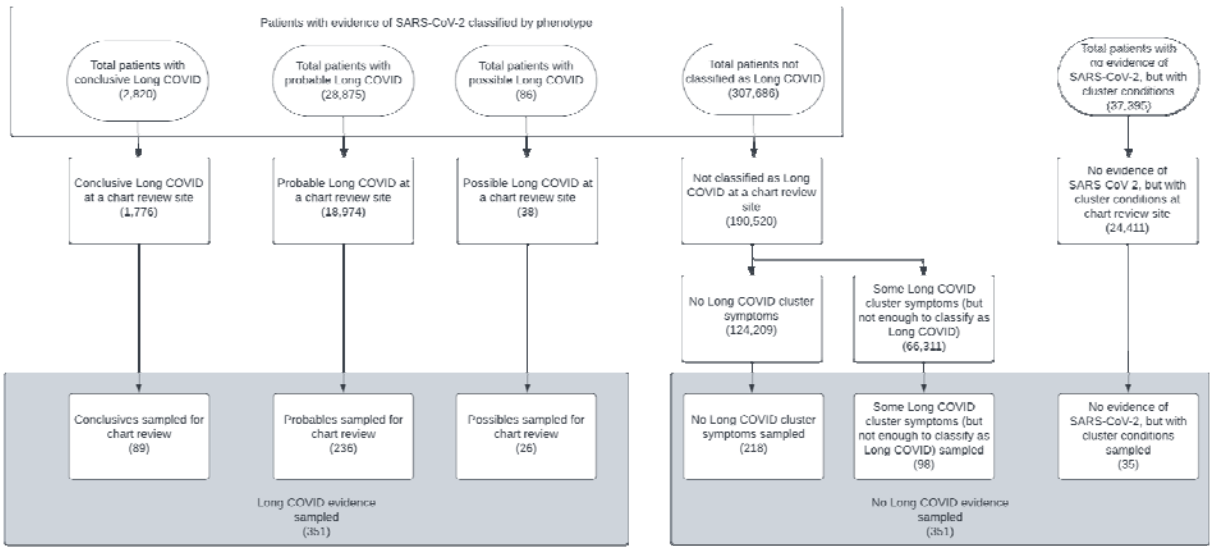

Figure 3. *Sampling strategy for the chart review cohort. The numbers reported in parentheses represent sample sizes.*

Chart Review Procedure

Clinical research teams from each participating institution conducted chart reviews using a REDCap[23] (Research Electronic Data Capture) instrument with questions including information on COVID-19 diagnoses and testing, demographics, COVID-19 prevention and treatment strategies, vaccines, functional outcomes, and conditions post COVID-19 captured in the patient's medical record. Each site had between 1 and 5 reviewers for a total of 44 reviewers across sites. Table 1 contains a summary of patient information extracted from chart review, and the full case report form is included in Supplemental File 2.

A secondary review was completed by a clinician who reviewed information extracted by the primary chart reviewer and answered questions regarding the level of confidence with which Long COVID could be assigned to the patient. The clinician was first asked if the patient met criteria for Long COVID based on the NIH definition[21] which describes Long COVID as signs, symptoms, and conditions that continue or develop after initial COVID-19 or SARS-CoV-2 infection, are present four weeks [28 days] or more after the initial phase of infection; may be multisystemic; and may present with a relapsing-remitting pattern and progression or worsening over time, with the possibility of severe and life-threatening events even months or years after infection. They were then asked if the patient met criteria for Long COVID based on the computable phenotype definition. The response to these questions (i.e., conclusive, probable, possible, no evidence) was used to assess concordance with the computable phenotype. The first

question, which analyses focused on, asked the clinician to exercise clinical judgment, while the second question was focused on assessing the validity of the structured EHR data.

For ease of assessing the performance of the computable phenotype compared with chart review, patients were collapsed into four overlapping groups: computable phenotype-positive (CP-positive), computable phenotype-negative (CP-negative), clinician review-positive (CR-positive), and clinician review-negative (CR-negative). Patients identified by the phenotype as having conclusive or probable Long COVID were placed in the CP-positive group. Conversely, patients found to have possible evidence or no evidence of Long COVID were placed in the CP-negative group. Patients with possible Long COVID were included in the CP-negative sample as the study team concluded that having only one post-viral sequelae code without a positive PCR test to confirm SARS-CoV-2 infection did not provide enough evidence to conclude that the patient's post-viral sequelae was caused by Long COVID. On the other hand, when examining the chart review, we determined that the reasons reviewers used to classify patients as possible for Long COVID were more similar to a positive than a negative Long COVID classification. Therefore, the CR-positive group consisted of patients labeled as conclusive, probable, or possible for Long COVID by the clinician reviewer. In contrast, the CR-negative group consisted of patients labeled as having no evidence of Long COVID by the clinician reviewer.

Performance Assessment

The performance of the computable phenotype was evaluated across various key metrics including sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Additionally, we examined the accuracy and F1 score of the phenotype. Accuracy assesses the proportion of CR-positive patients who were also CP-positive. The F1 score combines precision and recall providing insight into the overall effectiveness of the phenotype.

Next, we assessed whether concordance between the phenotype and clinician review classification differed by age (i.e., under vs. over 12 years old), variant period (i.e., alpha, delta, omicron), and number of symptom clusters through stratified analyses. Analyses were conducted using R version 4.1.2 (2021-11-01; 24).

We conducted an assessment to identify discrepancies between the phenotype and chart review identification of Long COVID. We reviewed cases where the phenotype identified Long COVID, but the chart review did not, and vice versa. To understand the reasons behind these discrepancies, we reviewed the chart review form for each discordant patient, along with the clinician reviewer's explanation for assigning or not assigning Long COVID. We then generated themes that accounted for the discrepancy and assigned those themes to the remaining cases. We next aimed to modify our model based the most common themes and perform a sensitivity analysis to assess whether the performance of our modified model was superior to our original model.

To describe the sample and investigate the impact of patients with complex medical histories on the performance of the phenotype, we used the Pediatric Medical Complexity Algorithm[25] (PMCA). We determined the chronic condition status of each patient by applying the more conservative version of the algorithm. This version requires that a patient have one diagnosis of a progressive or malignant condition or at least two diagnoses per body system for at least two body systems in the three years prior to the SARS-CoV-2 infection.

Results

Among patients with a positive SARS-CoV-2 infection (1,007,867), the computable phenotype detected Long COVID in 31,781 (3.2%) patients. Seven hundred and two patients were included in the chart review sample; however, sixteen charts were not completed due to

limitations in the chart reviewers' access to records and were excluded from the sample. In addition, the 35 patients with no evidence of SARS-CoV-2 infection according to the computable phenotype were not included in comparative analyses as a full chart review was not completed on them. Thus, our final sample consisted of 651 patients. Sample demographics and descriptive statistics are presented in Table 1. Sociodemographic characteristics were similar among those classified with and without Long COVID by the phenotype and by the clinician chart reviewer (Supplemental Table S1).

Table 1. Demographics of children with and without Long COVID based on the computable phenotype definition.

	CP-Detected Long COVID (N=318)	No CP-Detected Long COVID (N=333)	Overall (N=651)
Approx. CED age (years)			
Mean (SD)	10.10 (6.32)	10.10 (6.28)	10.10 (6.30)
Median [Min, Max]	11.0 [0, 21.0]	10.3 [0.1, 21.0]	10.9 [0, 21.0]
CED Age Group (years)			
<1	24 (7.5%)	25 (7.5%)	49 (7.5%)
1-4	70 (22.0%)	76 (22.8%)	146 (22.4%)
5-9	52 (16.4%)	55 (16.5%)	107 (16.4%)
10-15	98 (30.8%)	100 (30.0%)	198 (30.4%)
16-20	74 (23.3%)	77 (23.1%)	151 (23.2%)
Patient Sex			
Male	147 (46.2%)	167 (50.2%)	314 (48.2%)
Female	171 (53.8%)	166 (49.8%)	337 (51.8%)
Race			
Asian/Native Hawaiian/Pacific Islander	12 (3.8%)	14 (4.2%)	26 (4.0%)
Black	57 (17.9%)	58 (17.4%)	115 (17.7%)
White	175 (55.0%)	182 (54.7%)	357 (54.8%)
Multiracial	12 (3.8%)	9 (2.7%)	21 (3.2%)
Unknown	62 (19.5%)	70 (21.0%)	132 (20.3%)

Ethnicity			
Hispanic	78 (24.5%)	99 (29.7%)	177 (27.2%)
Non-Hispanic	214 (67.3%)	210 (63.1%)	424 (65.1%)
Unknown	26 (8.2%)	24 (7.2%)	50 (7.7%)
*Payer**			
Private	139 (43.7%)	143 (42.9%)	282 (43.3%)
Public	124 (39.0%)	134 (40.2%)	258 (39.6%)
Other/Unknown	55 (17.3%)	56 (16.8%)	111 (17.1%)

*Note. CP = computable phenotype. *=at time of COVID-19 infection.*

There was 73.33% agreement between the responses to the two questions the clinician chart reviewers answered. Table 2 presents statistics assessing the performance of the computable phenotype with chart review. The two methods had substantial but incomplete overlap. Analyses assessing whether concordance differed by selected variables showed similar results across age, era associated with infection, and number of symptom clusters (Supplemental Tables S2-4).

Table 2. Statistics comparing computable phenotype and clinician review identification of Long COVID.

	CR-Positive (N = 239)		**CR-Negative (N = 412)**			
CP-Positive (N = 318)	156		162			
CP-Negative (N = 333)	83		250			
Performance Statistics*						
	Accuracy	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**F1**
	0.624	0.653	0.607	0.491	0.751	0.560

Note. PPV = positive predictive value. NPV = negative predictive value. CP = computable phenotype. CR = chart review.
*Reported as CP relative to CR.

Computable Phenotype-Only and Clinician Review-Only Long COVID Positive Review

A review was conducted to assess the reasons for disagreement between the two methods. The initial focus was cases where the phenotype identified Long COVID, but the chart review did not (CP+/CR- cases) as there were many of these cases. Results are presented in Figure 4. In most CP+/CR- cases, the clinician reviewer agreed with the symptoms the computable phenotype identified but attributed those symptoms to another viral infection or preexisting disorder (Figure 4a). This was especially true for symptoms common to other respiratory infections and symptoms with occurrences both pre and post COVID-19 infection. In other less common cases, the reviewer did not see the diagnostic codes that the phenotype saw, or the reviewer made a conclusion based on incomplete information.

An assessment of CR+/CP- cases showed that in many cases the reviewer considered symptoms, visits, and time frames that differed from our phenotype (Figure 4b). For example, clinician reviewers considered symptoms beyond 180 days post-infection (up to 11 months in some cases) while the computable phenotype considered symptoms in the one month to 180 days following infection. In addition, reviewers considered symptoms beyond those included in the computable phenotype definition, such as mental health symptoms. For example, clinician reviewers used clinical judgment and identified anxiety as a qualifying symptom, but it is not included in the phenotype definition due to inconsistent recording in structured data.

a. CP-only Long COVID+ cases (N = 162) were grouped into the following categories:

Category	%
Reviewer agreed conditions are present but attributed them to a condition other than Long COVID	73%
Reviewer error (e.g., reviewer did not see the codes that the phenotype saw)	17%
Insufficient EHR data for the reviewer to form a conclusion	9%
Reviewer noted that the codes for Long COVID were misused by the provider	1%

Most common symptom clusters attributed to a condition other than Long COVID were:

- Respiratory symptoms
- Musculoskeletal symptoms
- Fever/chills

Symptoms were often attributed to comorbidities, including:

- Preexisting conditions (e.g., asthma)
- Co-occurring viral infections (e.g., influenza)
- Presence of complex medical conditions (e.g., chronic conditions affecting 2 body systems)

b. Chart review-only Long COVID+ cases (N = 83) were grouped into the following categories:

Category	%
Symptoms/visits identified by reviewer did not meet CP definition	36%
Reviewer considered a longer post-acute period	33%
Potential for Long COVID but inadequate evidence for phenotype	16%
Reviewer relied on unstructured EHR data to make a judgment	12%
Reviewer made an inference based on incomplete data	4%

Figure 4. *Qualitative review of a.) CP+/CR-* Long COVID *patients and b.) CR+/CP-* Long COVID *patients.*

Sensitivity Analysis

The review of CP+/CR- patients showed that comorbidities were a large factor contributing to discordance between the two methods. Given the difficulty of distinguishing symptoms due to preexisting conditions and symptoms due to Long COVID, we performed a sensitivity analysis to assess concordance with a modified model in which preexisting conditions and comorbidities were accounted for. First, we censored prior symptoms. In other words, we assessed phenotype classification when non-incident diagnoses were excluded. Patients who met criteria for the computable phenotype with only preexisting symptom clusters that persisted after their COVID-19 diagnosis were labeled as having no evidence of Long COVID (n = 32 patients). Second, given the high prevalence of symptoms reported in the setting of non-COVID-19 respiratory infections, we excluded respiratory or fever cluster diagnoses that occurred 2 weeks prior to or after a diagnosis of a non-COVID-19 respiratory infection. This led to exclusion of 31

patients who did not otherwise meet the criteria for probable long COVID. Finally, it was difficult to attribute post-acute symptoms to a COVID-19 diagnosis versus underlying medical conditions in patients with multiple chronic medical conditions. Therefore, we identified patients with complex medical histories using the PMCA[25]. We reclassified patients with a complex chronic condition as having no evidence of Long COVID (n = 67 patients) as we believed that our phenotype could not accurately identify these patients as having Long COVID.

Concordance was assessed again after incorporating these alterations into a modified model, and results showed a higher positive predictive value and specificity but lower sensitivity (Table 3). The negative predictive value remained high. Given the difficulty in adequately attributing diagnoses to Long COVID in patients with complex medical histories, we also completed a sensitivity analysis where these patients were removed from the sample, and performance was reassessed (Table 3). Results showed similar performance to the modified model overall, but a higher F1 score.

Table 3. Statistics comparing modified computable phenotype and clinician review identification of Long COVID.

	Modified	
	CR-Positive (N = 239)	**CR-Negative (N = 412)**
Modified CP-Positive (N=188)	123	65
Modified CP-Negative (N=463)	116	347
	No Medically Complex Patients	
	CR-Positive (N = 210)	**CR-Negative (N = 308)**

Modified CP-Positive (N = 188)	123	65
Modified-CP Negative (N = 330)	87	243

	Performance Statistics*					
	Accuracy	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**F1**
Modified	0.722	0.515	0.842	0.654	0.749	0.576
No Medically Complex Patients	0.707	0.586	0.789	0.654	0.736	0.618

Note. PPV = positive predictive value. NPV = negative predictive value. CP = computable phenotype. CR = chart review.
*Reported as CP relative to CR

Discussion

We conducted a study to assess the performance of a rules-based computable phenotype for identifying pediatric patients with Long COVID in a large EHR database compared to clinician chart review. Results showed moderate overlap between the two methods. Specifically, the computable phenotype was moderately sensitive in detecting patients with Long COVID and specific in detecting those without Long COVID, in comparison to chart review. However, there were several cases where the methods disagreed, with some patients being classified as having Long COVID by the phenotype but not by chart review, and vice versa. The main reason for these discrepancies was due to underlying comorbidities and subsequent respiratory infections.

Patients with comorbidities posed a challenge for the computable phenotype and the clinician reviewer. This was likely due to the lack of clinical guidelines for attribution and the difficulty in discerning exacerbation of preexisting symptoms. Clinicians were more likely to attribute post-COVID-19 symptoms to preexisting conditions when comorbidities were present, which likely resulted in the misattribution of Long COVID symptoms and may have been influenced by the provider involved in clinical care. However, the overlap between the two methods increased when the CP accounted for preexisting medical conditions by focusing on

incident diagnoses and censoring existing conditions. Nevertheless, because our phenotype was not initially designed to assess exacerbation of preexisting conditions, we caution against its use to diagnose Long COVID in patients with medical complexity. Another source of disagreement between the computable phenotype and chart review stemmed from subsequent non-COVID-19 respiratory infections, which are common in children. Although there may be an increased risk of secondary infections due to SARS-CoV-2[26-27], the symptoms are caused by a different agent. Therefore, we removed these circumstances as indicators of Long COVID in our phenotype.

An analysis of chart review-only positives (i.e., those the clinician reviewer classified as having Long COVID that the phenotype did not) showed that differences in the computable phenotype guidelines and the clinician's framework for identifying Long COVID were the main reasons for discordance. While the computable phenotype only assessed symptoms up to 180 days post SARS-CoV-2 infection, clinicians may use a longer time window in practice. This suggests an extended time frame for assessing post-acute symptoms may be necessary, but also may increase the risk of later onset of symptoms not being clearly attributable to a SARS-CoV-2 infection. In addition, clinician reviewers identified conditions beyond those included in our phenotype as providing evidence for Long COVID. For example, the computable phenotype did not consider mental health conditions due to reporting inconsistency of these conditions using diagnosis codes and the difficulty in distinguishing between biologic and social causes of mental health conditions. However, clinicians tended to include them in their framework for identifying Long COVID. Therefore, constructing computable phenotypes that incorporate subphenotypes of interest (e.g., physiological vs psychological manifestations of Long COVID) may be useful in accounting for different manifestations of Long COVID.

Many of the differences in identifying Long COVID are due to the lack of a clear and consistently used definition for Long COVID. The novelty of Long COVID in children, as well as the overlap of symptoms with other acute and chronic disorders such as headache and fatigue, contribute to these differences. Similar difficulties have been encountered when defining other post-viral syndromes. Although chart review is often viewed as the best method for detecting patients with a specific disorder, the lack of a consistent definition for Long COVID by healthcare providers poses significant challenges. Moreover, the chart review and EHR research are prone to errors such as biases, missed codes, misdiagnoses, incomplete information due to fragmented care, and a lack of availability of a unique medical code for certain Long COVID-related conditions. For example, a unique ICD code for POTS did not exist until October 2022[28], and many clinicians remain unaware of its existence, making it difficult to pick up the presence of POTS in the phenotype or chart review. Therefore, comparing our computable phenotype with chart review provided insight into the clinician's view of a patient's status, but it did not allow us to validate against a gold standard, as we cannot confidently conclude that either method accurately detects Long COVID.

Our two-pronged approach to identifying Long COVID using clinician chart review and a computable phenotype is a strength of the study as previous research that used diagnosis codes or machine learning algorithms did not incorporate a review of patient charts[19-20]. By incorporating both methods, we were able to qualitatively review cases of discordance. In addition, we focused on pediatric patients, in whom Long COVID is understudied. Research suggests that Long COVID has a lower prevalence in children; however, the current diagnostic tools may not be sensitive enough to detect all cases. Our study design is a strength as it uses data-driven symptom clusters for identifying Long COVID and specific diagnosis codes. This approach

allowed us to capture patients who may not have a clear Long COVID presentation but have Long COVID-like symptoms. Although it allowed for our phenotype to be more inclusive, it also resulted in a computable phenotype that was less inclusive than a clinician may be when subjectively assessing whether a patient has Long COVID. As more symptoms of Long COVID are identified, it may be necessary to update the symptom clusters.

Our study has limitations, but it also brings attention to some significant areas for future work to focus on. Due to the challenge of distinguishing the progression of a chronic condition from symptom exacerbation due to COVID-19 using EHR data and chart review, we were unable to evaluate the worsening of preexisting symptoms. Instead, we examined differences in concordance after excluding pre-COVID-19 symptoms. This approach provided increased certainty that symptoms were due to Long COVID but may have been too restrictive. Future research should consider cluster-specific washout windows and develop reliable methods to identify patients with Long COVID-related worsening of preexisting conditions. Additionally, our sampling strategy focused on edge cases and rare occurrences to develop and refine the phenotype. This approach was useful for identifying patients with a range of Long COVID-related symptoms and diagnoses, but limits generalizability and underestimates the performance of the phenotype. Future iterations should use random sampling to obtain a more generalizable patient sample. Finally, because our study was based on EHR data and we imposed a two-visit requirement in the post-acute period, our sample may be biased towards patients who have the means to obtain healthcare at the population level.

Conclusion

This study describes a computable phenotype approach to identify children with Long COVID in EHR data. Our study highlights the complexity of identifying and diagnosing Long

COVID due to its heterogeneity and overlap with other conditions, which leads to substantial differences observed across methods. To address this challenge, future work could include additional data sources, such as unstructured data, and further refine algorithms with clinical expertise to develop a reliable definition of Long COVID. It is also essential to develop a revised phenotype that can identify Long COVID through the worsening of pre-existing conditions. The development of a reliable CP for Long COVID in children allows for studying large data networks, which has future applications for both observational studies and clinical trials. Further research assessing the presentation of Long COVID in children and the interplay between Long COVID and comorbidities is vital to continue to understand this emerging chronic illness and evaluate interventions that can prevent or mitigate its effects.

Acknowledgements:

This study is part of the NIH Researching COVID to Enhance Recovery (RECOVER) Initiative, which seeks to understand, treat, and prevent the post-acute sequelae of SARS-CoV-2 infection (PASC). For more information on RECOVER, visit https://recovercovid.org/

We would like to thank the National Community Engagement Group (NCEG), all patient, caregiver, and community Representatives, and all the participants enrolled in the RECOVER Initiative.

Author Conflict of Interest Disclosures: Dr. Mejias reports funding from Janssen, Merck for research support, and Janssen, Merck and Sanofi-Pasteur for Advisory Board participation; Dr. Rao reports prior grant support from GSK and Biofire and is a consultant for Sequiris. Dr. Jhaveri is a consultant for AstraZeneca, Seqirus and Dynavax, and receives an editorial stipend from Elsevier. All other authors have no conflicts of interest to disclose.

Role of funder/sponsor statement: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

References

1. Davis, H. E., Assaf, G. S., McCorkell, L., Wei, H., Low, R. J., Re'em, Y., Redfield, S., Austin, J. P., & Akrami, A. (2021). Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *eClinicalMedicine*, *38*, 101019. https://doi.org/10.1016/j.eclinm.2021.101019

2. Katz, B. Z., Shiraishi, Y., Mears, C. J., Binns, H. J., & Taylor, R. (2009). Chronic Fatigue Syndrome After Infectious Mononucleosis in Adolescents. *Pediatrics*, 124(1), 189–193. https://doi.org/10.1542/peds.2008-1879

3. Sellers, S. A., Hagan, R. S., Hayden, F. G., & Fischer, W. A. (2017). The hidden burden of influenza: A review of the extra□pulmonary complications of influenza infection. *Influenza and Other Respiratory Viruses*, 11(5), 372–393. https://doi.org/10.1111/irv.12470

4. Lopez-Leon, S., Wegman-Ostrosky, T., Ayuzo Del Valle, N. C., Perelman, C., Sepulveda, R., Rebolledo, P. A., Cuapio, A., & Villapol, S. (2022). Long-COVID in children and adolescents: A systematic review and meta-analyses. *Scientific Reports*, *12*(1), 9950. https://doi.org/10.1038/s41598-022-13495-5

5. Lorman, V., Rao, S., Jhaveri, R., Case, A., Mejias, A., Pajor, N. M., Patel, P., Thacker, D., Bose-Brill, S., Block, J., Hanley, P. C., Prahalad, P., Chen, Y., Forrest, C. B., Bailey, L. C., Lee, G. M., & Razzaghi, H. (2023). Understanding pediatric long COVID using a tree-based scan statistic approach: An EHR-based cohort study from the RECOVER Program. *JAMIA Open*, *6*(1), ooad016. https://doi.org/10.1093/jamiaopen/ooad016

6. Zheng, Y.-B., Zeng, N., Yuan, K., Tian, S.-S., Yang, Y.-B., Gao, N., Chen, X., Zhang, A.-Y., Kondratiuk, A. L., Shi, P.-P., Zhang, F., Sun, J., Yue, J.-L., Lin, X., Shi, L., Lalvani, A., Shi, J., Bao, Y.-P., & Lu, L. (2023). Prevalence and risk factor for long COVID in children and adolescents: A meta-analysis and systematic review. *Journal of Infection and Public Health*, *16*(5), 660–672. https://doi.org/10.1016/j.jiph.2023.03.005

7. Pellegrino, R., Chiappini, E., Licari, A., Galli, L., & Marseglia, G. L. (2022). Prevalence and clinical presentation of long COVID in children: A systematic review. *European Journal of Pediatrics*, *181*(12), 3995–4009. https://doi.org/10.1007/s00431-022-04600-x

8. Reese J, Blau H, Bergquist T, Loomba JJ, Callahan T, Laraway B, et al. Generalizable Long COVID Subtypes: Findings from the NIH N3C and RECOVER Program. *MedRxiv Prepr Serv Health Sci* 2022:2022.05.24.22275398. https://doi.or g/10.1101/2022.05.24.22275398 PMID: 35665012

9. Rao, S., Lee, G. M., Razzaghi, H., Lorman, V., Mejias, A., Pajor, N. M., Thacker, D., Webb, R., Dickinson, K., Bailey, L. C., Jhaveri, R., Christakis, D. A., Bennett, T. D., Chen, Y., & Forrest, C. B. (2022). Clinical Features and Burden of Postacute Sequelae of SARS-CoV-2 Infection in Children and Adolescents. *JAMA Pediatrics*, *176*(10), 1000. https://doi.org/10.1001/jamapediatrics.2022.2800

10. Rao, S., Gross, R. S., Mohandas, S., Stein, C. R., Case, A., Dreyer, B., Pajor, N. M., Bunnell, H. T., Warburton, D., Berg, E., Overdevest, J. B., Gorelik, M., Milner, J., Saxena, S., Jhaveri, R., Wood, J. C., Rhee, K. E., Letts, R., Maughan, C., … Stockwell, M. S. (2024). Postacute Sequelae of SARS-CoV-2 in Children. *Pediatrics*, *153*(3), e2023062570. https://doi.org/10.1542/peds.2023-062570

11. Feldstein, L. R., Rose, E. B., Horwitz, S. M., Collins, J. P., Newhams, M. M., Son, M. B. F., Newburger, J. W., Kleinman, L. C., Heidemann, S. M., Martin, A. A., Singh, A. R., Li, S., Tarquinio, K. M., Jaggi, P., Oster, M. E., Zackai, S. P., Gillen, J., Ratner, A. J., Walsh, R. F., … Randolph, A. G. (2020). Multisystem Inflammatory Syndrome in U.S. Children and Adolescents. *New England Journal of Medicine*, *383*(4), 334–346. https://doi.org/10.1056/NEJMoa2021680

12. Wu, E. Y., & Campbell, M. J. (2021). Cardiac Manifestations of Multisystem Inflammatory Syndrome in Children (MIS-C) Following COVID-19. *Current Cardiology Reports*, *23*(11), 168. https://doi.org/10.1007/s11886-021-01602-3

13. Garai, R., Krivácsy, P., Herczeg, V., Kovács, F., Tél, B., Kelemen, J., Máthé, A., Zsáry, E., Takács, J., Veres, D. S., & Szabó, A. J. (2023). Clinical assessment of children with long COVID syndrome. *Pediatric Research*, *93*(6), 1616–1625. https://doi.org/10.1038/s41390-022-02378-0

14. Kikkenborg Berg, S., Palm, P., Nygaard, U., Bundgaard, H., Petersen, M. N. S., Rosenkilde, S., Thorsted, A. B., Ersbøll, A. K., Thygesen, L. C., Nielsen, S. D., & Vinggaard Christensen, A. (2022). Long COVID symptoms in SARS-CoV-2-positive children aged 0–14 years and matched controls in Denmark (LongCOVIDKidsDK): A

national, cross-sectional study. *The Lancet Child & Adolescent Health*, *6*(9), 614–623. https://doi.org/10.1016/S2352-4642(22)00154-7

15. Fritsche, L. G., Jin, W., Admon, A. J., & Mukherjee, B. (2023). Characterizing and Predicting Post-Acute Sequelae of SARS CoV-2 Infection (PASC) in a Large Academic Medical Center in the US. *Journal of Clinical Medicine*, *12*(4), 1328. https://doi.org/10.3390/jcm12041328

16. Pfaff, E. R., Madlock-Brown, C., Baratta, J. M., Bhatia, A., Davis, H., Girvin, A., Hill, E., Kelly, E., Kostka, K., Loomba, J., McMurry, J. A., Wong, R., Bennett, T. D., Moffitt, R., Chute, C. G., Haendel, M., The N3C Consortium, & The RECOVER Consortium. (2023). Coding long COVID: Characterizing a new disease through an ICD-10 lens. *BMC Medicine*, *21*(1), 58. https://doi.org/10.1186/s12916-023-02737-6

17. International Classification of Diseases, Tenth Revision (ICD-10), World Health Organization (WHO) 2019/2021. https://icd.who.int/browse11. Licensed under Creative Commons Attribution-NoDerivatives 3.0 IGO license (CC BY-ND 3.0 IGO).

18. Centers for Disease Control and Prevention. Post-covid conditions: Information for healthcare providers. Accessed January 5, 2024. https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/post-covid-conditions.html

19. Pfaff, E. R., Girvin, A. T., Bennett, T. D., Bhatia, A., Brooks, I. M., Deer, R. R., Dekermanjian, J. P., Jolley, S. E., Kahn, M. G., Kostka, K., McMurry, J. A., Moffitt, R., Walden, A., Chute, C. G., Haendel, M. A., Bramante, C., Dorr, D., Morris, M., Parker, A. M., … Niehaus, E. (2022). Identifying who has long COVID in the USA: A machine

learning approach using N3C data. *The Lancet Digital Health*, *4*(7), e532–e541. https://doi.org/10.1016/S2589-7500(22)00048-6

20. Lorman, V., Razzaghi, H., Song, X., Morse, K., Utidjian, L., Allen, A. J., Rao, S., Rogerson, C., Bennett, T. D., Morizono, H., Eckrich, D., Jhaveri, R., Huang, Y., Ranade, D., Pajor, N., Lee, G. M., Forrest, C. B., & Bailey, L. C. (2023). A machine learning-based phenotype for long COVID in children: An EHR-based study from the RECOVER program. *PLOS ONE*, *18*(8), e0289774. https://doi.org/10.1371/journal.pone.0289774

21. RECOVER: Researching COVID to Enhance Recovery. RECOVER: Researching COVID to Enhance Recovery. https://recovercovid.org. Accessed 4 Jan 2024.

22. Mejias A, Schuchard J, Rao S, Bennett TD, Jhaveri R, Thacker D, Bailey LC, Christakis DA, Pajor NM, Razzaghi H, Forrest CB, Lee GM. Leveraging Serologic Testing to Identify Children at Risk for Post-Acute Sequelae of SARS-CoV-2 Infection: An Electronic Health Record-Based Cohort Study from the RECOVER Program. *J Pediatr*. 2023 Jun; 257:113358. doi: 10.1016/j.jpeds.2023.02.005.

23. PA Harris, R Taylor, R Thielke, J Payne, N Gonzalez, JG. Conde, Research electronic data capture (REDCap**) –** A metadata-driven methodology and workflow process for providing translational research informatics support, *J Biomed Inform. 2009 Apr;42(2):377-81.*

24. Posit team (2023). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL http://www.posit.co/.

25. Simon, T. D., Lawrence, M., Stanford, S., Lyons, D., Woodcox, P., Hood, M., & Chen, A. Y. (2014). Pediatric Medical Complexity Algorithm: A New Method to Stratify Children by Medical Complexity. *Pediatrics*, 133(6).

26. Wang, L., Davis, P. B., Berger, N., Kaelber, D. C., Volkow, N., & Xu, R. (2023). Association of COVID-19 with respiratory syncytial virus (RSV) infections in children aged 0-5 years in the USA in 2022: a multicenter retrospective cohort study. *Family medicine and community health*, *11*(4), e002456. https://doi.org/10.1136/fmch-2023-002456

27. Xie Y, Choi T, Al-Aly Z. Long-term outcomes following hospital admission for COVID-19 versus seasonal influenza: a cohort study. *Lancet Infect Dis.* 2023 Dec 14: S1473-3099(23)00684-9. doi: 10.1016/S1473-3099(23)00684-9.

28. World Health Organization. (2019). 8D89.2 Postural orthostatic tachycardia syndrome. In *International statistical classification of diseases and related health problems* (11th ed.). https://icd.who.int/browse/2024-01/mms/en#1533647472

Supporting Information

Table S1. Demographics of children with and without Long COVID based on the computable phenotype definition.

Table S2. Statistics comparing CP and chart review identification of Long COVID presented separately for patients younger and older than 12 years of age.

Table S3. Start and end dates associated with each era of infection.

Table S4. Statistics comparing CP and chart review identification of Long COVID presented separately by era of infection.

Table S5. Statistics comparing CP and chart review identification of Long COVID presented separately by the number of clusters identified by the CP.