

## Original Research

## Characterizing Long COVID: Deep Phenotype of a Complex Condition



Rachel R Deer<sup>1,\*</sup>, Madeline A Rock<sup>1</sup>, Nicole Vasilevsky<sup>2,3</sup>, Leigh Carmody<sup>3,4</sup>, Halie Rando<sup>2,5</sup>, Alfred J Anzalone<sup>6</sup>, Marc D Basson<sup>7</sup>, Tellen D Bennett<sup>8</sup>, Timothy Bergquist<sup>9</sup>, Eilis A Boudreau<sup>10</sup>, Carolyn T Bramante<sup>11</sup>, James Brian Byrd<sup>12</sup>, Tiffany J Callahan<sup>2</sup>, Lauren E Chan<sup>3,13</sup>, Haitao Chu<sup>14</sup>, Christopher G Chute<sup>15</sup>, Ben D Coleman<sup>4,16</sup>, Hannah E Davis<sup>17</sup>, Joel Gagnier<sup>18</sup>, Casey S Greene<sup>2,5</sup>, William B Hillegass<sup>19,20</sup>, Ramakanth Kavuluru<sup>21</sup>, Wesley D Kimble<sup>22</sup>, Farrukh M Koraiшы<sup>23</sup>, Sebastian Köhler<sup>3,24</sup>, Chen Liang<sup>25</sup>, Feifan Liu<sup>26</sup>, Hongfang Liu<sup>27</sup>, Vithal Madhira<sup>28</sup>, Charisse R Madlock-Brown<sup>29</sup>, Nicolas Matentzoglou<sup>3,30,31</sup>, Diego R Mazzotti<sup>32</sup>, Julie A McMurry<sup>2,3</sup>, Douglas S McNair<sup>33</sup>, Richard A Moffitt<sup>34</sup>, Teshamae S Monteith<sup>35</sup>, Ann M Parker<sup>36</sup>, Mallory A Perry<sup>37</sup>, Emily Pfaff<sup>38</sup>, Justin T Reese<sup>3,39</sup>, Joel Saltz<sup>40,41</sup>, Robert A Schuff<sup>42</sup>, Anthony E Solomonides<sup>43,44</sup>, Julian Solway<sup>44</sup>, Heidi Spratt<sup>1</sup>, Gary S Stein<sup>45</sup>, Anupam A Sule<sup>46</sup>, Umit Topaloglu<sup>47</sup>, George D. Vavougiос<sup>48,49,50,51</sup>, Liwei Wang<sup>27</sup>, Melissa A Haendel<sup>2,3,\*</sup>, Peter N Robinson<sup>3,4,16,\*</sup>

<sup>1</sup> University of Texas Medical Branch, Galveston, TX, USA

<sup>2</sup> Center for Health AI, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

<sup>3</sup> Monarch Initiative

<sup>4</sup> The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA

<sup>5</sup> Department of Biochemistry and Molecular Genetics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

<sup>6</sup> Department of Neurological Sciences, College of Medicine, University of Nebraska Medical Center, Omaha, NE, USA

<sup>7</sup> Department of Surgery, University of North Dakota School of Medicine and Health Sciences

<sup>8</sup> Section of Informatics and Data Science, Department of Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

<sup>9</sup> Sage Bionetworks, Seattle, WA

<sup>10</sup> Department of Neurology; Department of Medical Informatics & Clinical Epidemiology, Oregon Health & Science University, Portland, OR 97239

<sup>11</sup> Departments of Internal Medicine and Pediatrics, University of Minnesota Medical School, Minneapolis, MN 55455

<sup>12</sup> Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan Medical School, Ann Arbor, MI, 48109

<sup>13</sup> College of Public Health and Human Sciences, Oregon State University, Corvallis, OR, USA

<sup>14</sup> Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN USA

<sup>15</sup> Johns Hopkins University, Schools of Medicine, Public Health, and Nursing, Baltimore, MD, USA

<sup>16</sup> Institute for Systems Genomics, University of Connecticut, Farmington, CT 06032, USA

<sup>17</sup> Patient-Led Research Collaborative

<sup>18</sup> Departments of Orthopaedic Surgery & Epidemiology, University of Michigan, Ann Arbor, MI, USA

<sup>19</sup> University of Mississippi Medical Center, University of Mississippi Medical Center, Jackson, MS, USA

<sup>20</sup> Departments of Data Science and Medicine

<sup>21</sup> Institute for Biomedical Informatics, University of Kentucky

<sup>22</sup> West Virginia Clinical and Translational Science Institute, West Virginia University, Morgantown, WV, USA

<sup>23</sup> Division of Nephrology, Department of Medicine, Stony Brook University

<sup>24</sup> Ada Health GmbH, Berlin, Germany

<sup>25</sup> Arnold School of Public Health, University of South Carolina, Columbia, SC, USA

<sup>26</sup> Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, USA

<sup>27</sup> Department of Artificial Intelligence and Informatics, Mayo Clinic, MN, USA

<sup>28</sup> Palila Software LLC, Reno, NV, USA

<sup>29</sup> Department of Diagnostic and Health Sciences, University of Tennessee Health Science Center, 920 Madison Ave. Suite 518N, Memphis TN 38613

<sup>30</sup> Semanticly Ltd

<sup>31</sup> European Bioinformatics Institute (EMBL-EBI)

<sup>32</sup> Division of Medical Informatics, Department of Internal Medicine, University of Kansas Medical Center

<sup>33</sup> Quantitative Sciences, Global Health Div., Gates Foundation, Seattle, WA 98109, USA

<sup>34</sup> Stony Brook University, Stony Brook, NY 11794, USA

<sup>35</sup> University of Miami, Miller School of Medicine, Miami, FL 33136

<sup>36</sup> Pulmonary and Critical Care Medicine, Johns Hopkins University, Schools of Medicine, Baltimore, MD, USA

<sup>37</sup> Children's Hospital of Philadelphia Research Institute, Philadelphia, PA, USA

<sup>38</sup> University of North Carolina, Chapel Hill

\* Correspondence: Dr. Rachel Deer, University of Texas Medical Branch, Division of Rehabilitation Sciences, 301 University Blvd, Galveston, TX 77555-1137, United States  
E-mail addresses: [rdeer@utmb.edu](mailto:rdeer@utmb.edu) (R.R. Deer), [melissa@tislabs.org](mailto:melissa@tislabs.org) (M.A. Haendel), [peter.robinson@jax.org](mailto:peter.robinson@jax.org) (P.N. Robinson).

<sup>39</sup> Lawrence Berkeley National Laboratory

<sup>40</sup> Stony Brook University

<sup>41</sup> Biomedical Informatics

<sup>42</sup> OCHIN, Inc Portland, OR, USA

<sup>43</sup> Outcomes Research Network, Research Institute, NorthShore University HealthSystem, Evanston, IL 60201, USA

<sup>44</sup> Institute for Translational Medicine, University of Chicago, Chicago, IL, USA

<sup>45</sup> University of Vermont Larner College of Medicine, Departments of Biochemistry and Surgery, Burlington, Vermont 05405

<sup>46</sup> St Joseph Mercy Oakland, Pontiac, MI, USA

<sup>47</sup> Wake Forest School of Medicine

<sup>48</sup> Department of Computer Science and Telecommunications, University of Thessaly, Papasiopoulou 2 – 4, P.C.

<sup>49</sup> 131 – Galaneika, Lamia, Greece

<sup>50</sup> Department of Neurology, Athens Naval Hospital 70 Deinostratou Street, P.C. 115 21 Athens, Greece

<sup>51</sup> Department of Respiratory Medicine, Faculty of Medicine, University of Thessaly, Biopolis, P.C. 41500 Larissa, Greece

## ARTICLE INFO

### Article History:

Received 10 September 2021

Revised 22 October 2021

Accepted 15 November 2021

Available online 25 November 2021

### Keywords:

COVID-19

of post-acute sequelae of SARS-CoV-2

human phenotype ontology

long COVID

phenotyping

## ABSTRACT

**Background:** Numerous publications describe the clinical manifestations of post-acute sequelae of SARS-CoV-2 (PASC or “long COVID”), but they are difficult to integrate because of heterogeneous methods and the lack of a standard for denoting the many phenotypic manifestations. Patient-led studies are of particular importance for understanding the natural history of COVID-19, but integration is hampered because they often use different terms to describe the same symptom or condition. This significant disparity in patient versus clinical characterization motivated the proposed ontological approach to specifying manifestations, which will improve capture and integration of future long COVID studies.

**Methods:** The Human Phenotype Ontology (HPO) is a widely used standard for exchange and analysis of phenotypic abnormalities in human disease but has not yet been applied to the analysis of COVID-19.

**Funding:** We identified 303 articles published before April 29, 2021, curated 59 relevant manuscripts that described clinical manifestations in 81 cohorts three weeks or more following acute COVID-19, and mapped 287 unique clinical findings to HPO terms. We present layperson synonyms and definitions that can be used to link patient self-report questionnaires to standard medical terminology. Long COVID clinical manifestations are not assessed consistently across studies, and most manifestations have been reported with a wide range of synonyms by different authors. Across at least 10 cohorts, authors reported 31 unique clinical features corresponding to HPO terms; the most commonly reported feature was Fatigue (median 45.1%) and the least commonly reported was Nausea (median 3.9%), but the reported percentages varied widely between studies.

**Interpretation:** Translating long COVID manifestations into computable HPO terms will improve analysis, data capture, and classification of long COVID patients. If researchers, clinicians, and patients share a common language, then studies can be compared/pooled more effectively. Furthermore, mapping lay terminology to HPO will help patients assist clinicians and researchers in creating phenotypic characterizations that are computationally accessible, thereby improving the stratification, diagnosis, and treatment of long COVID.

**Funding:** U24TR002306; UL1TR001439; P30AG024832; GBMF4552; R01HG010067; UL1TR002535; K23HL128909; UL1TR002389; K99GM145411.

© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## Research in context

### Evidence before this study

A majority of survivors of COVID-19 report manifestations that persist beyond the acute illness, so-called Post-Acute Sequelae of SARS-CoV-2 (PASC, or “long COVID”). Long COVID can affect even those who were initially mildly symptomatic or asymptomatic, may include a constellation of neurological, respiratory, cardiovascular, and gastrointestinal symptoms, and is debilitating in some affected individuals. Research on long COVID has been complicated due to heterogeneous study methods and lack of a standard for denoting the many phenotypic manifestations (different terms to describe the same symptom or condition).

### Added value of this study

We reviewed 303 manuscripts flagged as relevant to long COVID by CoronaCentral. From these, we identified 59 manuscripts with 81 cohorts that described 287 clinical manifestations of long COVID. Descriptions (symptoms, laboratory findings, imaging results) were mapped to Human Phenotype

Ontology (HPO) terms. We have developed layperson synonyms and definitions for each of the 287 HPO terms, which significantly improves patient and clinician accessibility.

### Implications of all the available evidence

One of the challenges in characterizing long COVID is the fact that patient-reported symptoms are often not captured by clinical evaluation or in surveys. To truly characterise long COVID and therefore stratify patients into subtypes for care decisions, it is necessary to use shared terminology. This common set of HPO definitions will promote integration of research by translating between patient and clinician descriptions of symptoms. We anticipate that this will be a critical resource for use in survey instruments and patient apps for standardizing patient-reporting in the future study of long COVID.

## 1. Introduction

Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) emerged in late 2019 as the third human coronavirus identified in the 21st century. Coronavirus disease 2019 (COVID-19) affects

diverse organ systems, including the lungs, digestive tract, kidneys, heart, and brain [1,2]. As of mid-2021, the full spectrum of the clinical consequences of COVID-19 is not completely understood. Individual symptoms and disease severity vary widely among patients during the acute infection, with some patients developing only mild symptoms or even remaining asymptomatic. In contrast, others experience acute respiratory distress syndrome (ARDS), sepsis, and other life-threatening conditions [3,4]. As more information about patient recovery has been collected, it has become clear that a wide range of outcomes can also emerge following the acute phase of the illness, with some patients experiencing residual symptoms or developing new symptoms long after the initial infection. This post-acute infection, referred to as long COVID, post-acute sequelae of COVID (PASC), or post-acute COVID-19 syndrome (PACS), represents a significant challenge for patients, physicians, and society because the causes, patient profile, and even symptom patterns remain difficult to characterise [5]. These substantial challenges in describing long COVID have led patients to self-organise and perform research to try to expedite the characterization of this disease and therefore how to best ameliorate the substantial impact that long COVID has had on their lives [6].

Long COVID, a multisystem disease, can occur following either severe, mild, or even asymptomatic SARS-CoV-2 infection [7]. There is currently no accepted definition of long COVID; however, it can be broadly defined as delayed recovery from infection with SARS-CoV-2. Long COVID can occur following cases of COVID-19 that were managed in either inpatient or outpatient settings. It is characterised by lasting effects of the infection, unexplained persistence of symptoms, or onset of new chronic diseases, for far longer than would be expected based on typical rates of viral clearance [8].

Given long COVID's recent emergence, no standard framework has yet been established for identifying and assessing associated symptoms or other clinical features. Furthermore, symptoms frequently reported by long COVID patients are not assessed consistently across studies. A systematic review available as a preprint evaluated all research on long COVID released prior to January 1, 2021 that included at least 100 patients. Based on the 15 studies that met the inclusion criteria, the authors identified 55 symptoms of long COVID [5](preprint). None of the most common symptoms were assessed by all 15 studies. The authors concluded that the symptoms of long COVID are extremely heterogeneous and that the assessment of these symptoms varies widely among studies. Another recent systematic review concluded that 73% of individuals who had acute COVID-19 experienced at least one persistent symptom [9]. However, the authors of the review concluded that the wide variation in design and quality of the studies limited the direct comparability and combinability of the data [9]. The wide range of symptoms attributed to long COVID are highlighted by an extensive patient-led survey (Patient-Led Research Collaborative). This study conducted deep longitudinal characterization of the long COVID symptoms and trajectories in suspected and confirmed COVID-19 patients who reported illness lasting more than 28 days [6]. Evaluating data from 3,762 respondents to 257 survey questions, this analysis documented 205 phenotypic features associated with long COVID. The fact that this patient-led study characterised 205 phenotypic features, while the studies cited in the aforementioned systematic review reported only 84 signs or symptoms and 19 laboratory or imaging measurements [9], suggests that the research community has not yet characterised the full spectrum of clinical manifestations of long COVID. This significant disparity in patient versus clinical characterization motivates the proposed ontological approach to specifying manifestations, which will improve capture and integration of future long COVID studies.

Deep phenotyping is the precise and comprehensive analysis of individual phenotypic abnormalities, with a focus on computational accessibility. In the field of rare disease, the Human Phenotype

Ontology (HPO) has become an international standard for deep phenotyping that enables integrated computational analysis of genotype and phenotype for diagnostics, novel disease gene discovery, and translational research [10,11]. The HPO includes a standardised vocabulary of over 16,000 terms with 37,072 synonyms that define phenotypic abnormalities associated with over 7,000 diseases. This tool enables non-exact matching of sets of phenotypic features (phenotype profile) against known diseases, other patients, and model organisms. The algorithms have been implemented for computational comparison of abnormalities and for use in genetic disease diagnostics, and they are the *de facto* standard for deep phenotyping in the field of rare disease. The HPO is used in a range of projects including the UK's National Institute for Health Research (NIHR) Rare Disease initiatives, the 100,000 Genomes project, the NIH Kidney Precision Medicine Project, and the NIH Undiagnosed Diseases Project and Network, RD-CONNECT, SOLVE-RD, and many others [10,12]. Existing publications on the clinical aspects of long COVID have not used a standard vocabulary to report phenotypic abnormalities, impeding the search, analysis, and integration of information relevant to long COVID in databases such as Medline. Ontologies such as HPO are systematic representations of knowledge that define terminology in a human-readable format and define relationships between concepts in a way that allows computational logical reasoning that supports the integration and analysis of large amounts of data [13]. The significant disparity in patient versus clinical characterization motivated the proposed ontological approach to specifying manifestations, which will improve capture and integration of future long COVID studies.

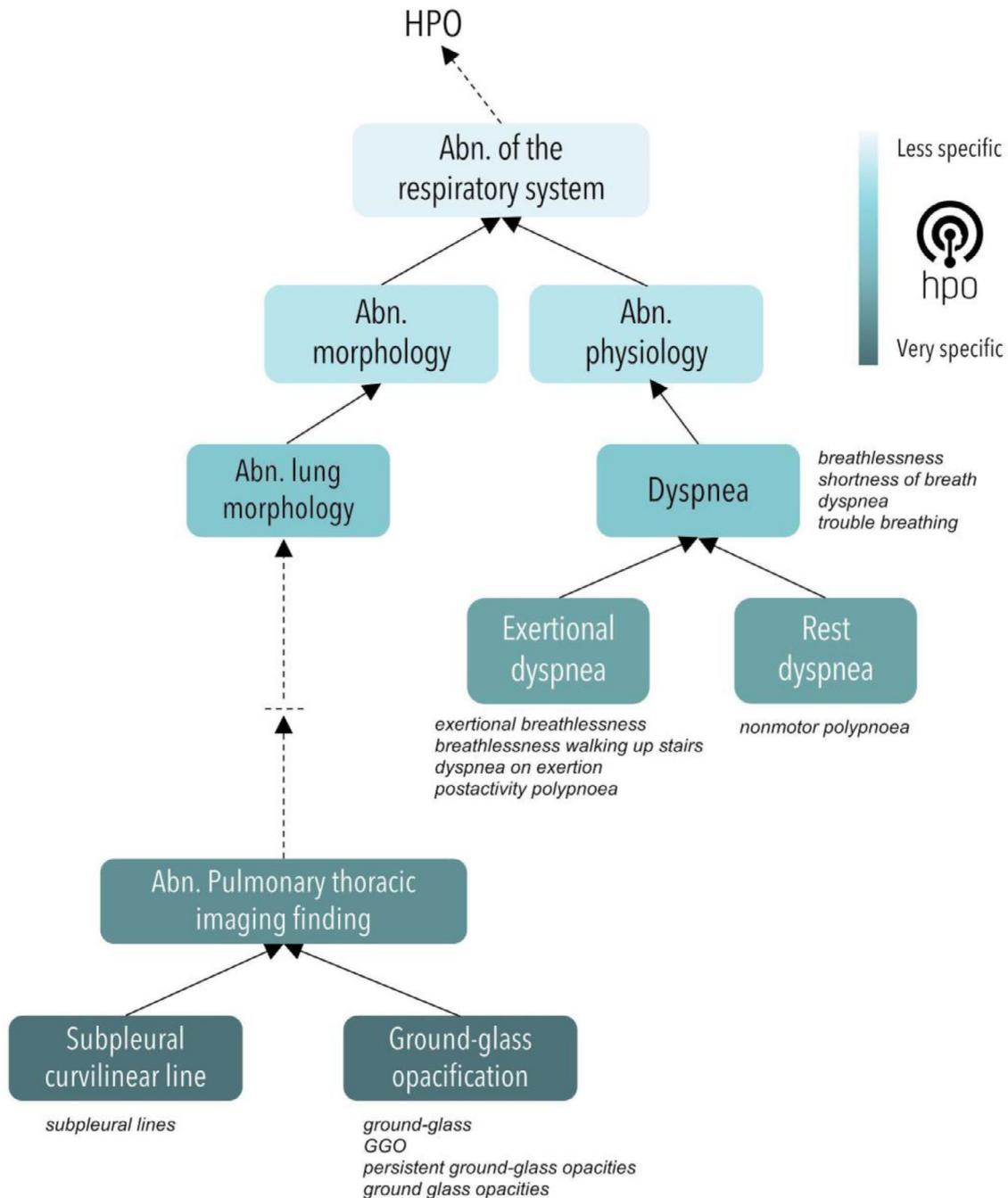
## 2. Methods

We searched for publications on long COVID using CoronaCentral, which uses machine learning to process the literature on SARS-CoV-2 [14]. We retrieved 303 articles predicted to be relevant to long COVID on April 29th, 2021. From these, 59 articles described the clinical manifestations in clinical cohorts of individuals three weeks or more following acute COVID-19. We defined three weeks or more based on the initial appearance of symptoms for outpatients or on three weeks or more after discharge for hospitalised patients. Descriptions of long COVID manifestations were mapped to Human Phenotype Ontology (<https://hpo.jax.org/app/>) terms. For this study, the HPO release 2021-06-08 was used. Four curators, one with experience in long COVID and three with extensive experience in HPO curation, manually reviewed the articles and identified HPO terms that corresponded to the description of clinical abnormalities (symptoms, signs, laboratory abnormalities, abnormal imaging findings) in the articles (Figure 1) and mapped them in a spreadsheet. Each mapping was reviewed by all four curators and discrepancies were resolved through discussion until consensus was reached. Some publications described multiple time points (e.g., early and late), or varying severities of acute illness (e.g., critical/severe, moderate, mild), which were treated as separate cohorts for the purposes of the current descriptive analysis (Supplemental Table S1). We tabulated the relationships between publications and the symptoms they reported, the mapped HPO terms, and body systems (Supplemental Table S2).

The funders had no role in study design, data collection, data analyses, interpretation, or writing of report.

## 3. Results

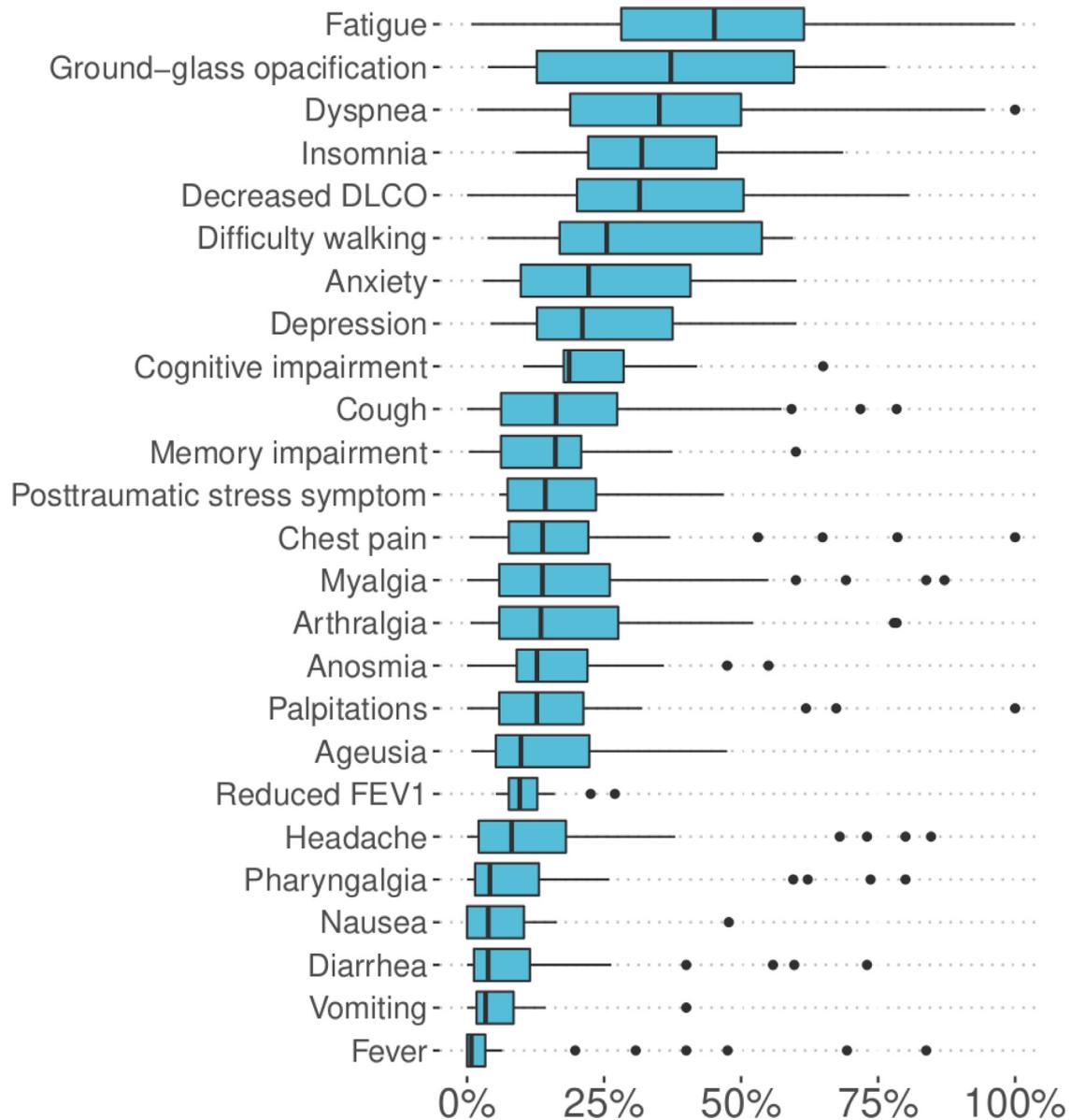
We reviewed 303 articles that were predicted to be relevant to long COVID. We excluded articles that were reviews, related only to acute-COVID timepoints, or did not provide sufficient details to extract percentages for the symptoms (i.e., only provided averages but not the number of patients affected in a cohort). Analysis of the remaining 59 articles revealed a variety of criteria were used to



**Figure 1.** The HPO is arranged in a hierarchy from general to more specific. This graph shows a representative hierarchy of a portion of the HPO 'abnormality of the respiratory system' branch. In this study, observations from 59 publications were mapped to the corresponding HPO terms (nodes). A selection of the original terminology used in the manuscripts (in italics) is shown adjacent to the HPO term to which it was mapped. A detailed list of all mapped terms is provided in Supplemental File 2.

identify and evaluate patients with long COVID. The studies included 11 cohorts of patients who had been treated in the intensive care unit (ICU) during acute COVID-19, 36 cohorts of patients who were hospitalised but not admitted to an ICU during the acute phase, 16 cohorts of patients who were not hospitalised during the acute phase, and 19 mixed cohorts. Some articles pulled data from electronic health records (EHRs), while others strictly relied on patient-reported symptoms from surveys. Studies also varied in the method of collection and instruments used. Methods of collecting data came from phone or electronic surveys, in person review, or pull from electronic medical records. For 26 cohorts, information was collected by in-person, telephone, email, or other online questionnaire. For 51 cohorts, information was collected by clinical examination, and for 5 cohorts,

information was collected by questionnaire and clinical examination (Supplemental Table S1). The time frame for data collection and follow-up also differed across studies. Some used a relatively precise window for patient assessment (e.g., 21 days after symptom onset), while others included participants at various distances from acute SARS-COV-2 infection. Some studies aimed to collect information only on patients suffering from long COVID, while others collected follow-up information on all patients that had previously had COVID-19 regardless of whether they currently or ever experienced long COVID. Studies differed in how they referred to the phenomenon studied. Some referred to it as long COVID or using a similar term such as post-acute COVID-19 syndrome, whereas others discussed the clinical course or patient recovery without mentioning long



**Figure 2.** Reported frequencies for the 25 phenotypic features identified in 12 or more cohorts. Box plots are shown for each item, displaying the minimum (1.5 times the interquartile range below the lower quartile), first quartile, median, third quartile, and maximum (1.5 times the interquartile range above the upper quartile). Outliers are shown as dots. DLCO: diffusing capacity of the lungs for carbon monoxide, FEV1: forced expiratory volume in one second; TLC: total lung capacity.

COVID specifically. Finally, studies varied widely in the terminology used to describe patient-reported symptoms.

In the 59 publications and 81 cohorts curated for this study (Supplemental Figure S1), a total of 287 phenotypic abnormalities were identified and represented as HPO terms. Of these, 132 terms were used in only one cohort (24.2%), 51 in two cohorts (9.4%), and 62 terms in at least 5 cohorts (21.6%). In most cases, multiple synonyms were mapped to the same HPO term; for instance, *Hepatic steatosis* (HP:0001397) corresponded to four descriptions used in the literature (Steatosis, Liver steatosis, Fatty infiltration of liver, Fatty liver). Among terms reported in at least 10 cohorts ( $n=31$  terms), the most commonly reported feature was *Fatigue* (median 45.1%), and the least commonly reported was *Nausea* (median 3.9%), but the reported percentages varied widely between studies (Figure 2).

Full curation details are available in Supplemental File 2. Supplemental Figures S2-S26 provide an overview of the 287 phenotypic features arranged by category, and Supplemental Note 1 provides additional commentaries. Table 1 provides details on the vast heterogeneity of symptoms, organised by the organ systems that are likely

to be involved. Few studies of long COVID to date have conducted analyses elucidating the presence or extent of organ damage. However, preliminary investigations of a number of organ systems have identified organ damage in long COVID patients. These findings are important because they highlight the possibility of asymptomatic long COVID patients sustaining organ damage due to the SARS-CoV-2 virus that does not immediately present with symptoms. Therefore, an improved understanding of organ damage as an outcome of acute COVID-19 or as a long-term sequelae of the SARS-CoV-2 virus may present new options for patients experiencing persistent symptoms or elucidate new information about how the SARS-CoV-2 virus interacts with a range of organ systems.

In our study, we curated 287 HPO terms representing clinical abnormalities observed in individuals following COVID-19. More research will be needed to determine which of the terms, and potentially which additional terms, are specifically and causally related to SARS-CoV-2 infection. For instance, 10 phenotypic abnormalities in our study have also been reported to occur in Post-Intensive Care Syndrome (PICS). While most of the manifestations were reported to

**Table 1**

Overview of abnormal phenotypic findings by category. The most commonly reported findings are shown for each category. The total number of features reported in each organ system is shown in the second column. Categories with at least 7 features are shown. The median percent column shows the median for the percentage of patients with the feature indicated in the previous column, together with the number of cohorts reporting the feature. Supplemental File S2 provides details.

Category	Total reported features	Most commonly reported feature	Median Percent (number of cohorts)
abnormalities of smell and taste	7	Anosmia (HP:0000458)	12.8% (n=44)
behavioral abnormalities	17	Anxiety (HP:0000739)	22.2% (n=24)
cardiovascular findings	16	Hypertension (HP:0000822)	20.4% (n=4)
cognitive dysfunction	8	Cognitive impairment (HP:0100543)	18.6% (n=13)
dermatological findings	10	Alopecia (HP:0001596)	18.8% (n=9)
emotion/mood abnormalities	9	Depression (HP:0000716)	21.1% (n=25)
gastrointestinal findings	9	Hepatic steatosis (HP:0001397)	26.5% (n=2)
gastrointestinal symptoms	10	Diarrhea (HP:0002014)	3.8% (n=29)
general symptoms	23	Fatigue (HP:0012378)	45.1% (n=48)
laboratory abnormalities	23	Elevated circulating D-dimer concentration (HP:0033106)	26.0% (n=6)
neuropsychiatric findings	30	Dysphagia (HP:0002015)	1.0% (n=7)
ocular abnormalities	13	Blurred vision (HP:0000622)	9.7% (n=7)
pain	9	Myalgia (HP:0003326)	13.8% (n=36)
pulmonary findings	16	Decreased DLCO (HP:0045051)	31.5% (n=18)
pulmonary imaging findings	15	Parenchymal consolidation (HP:0032177)	16.8% (n=8)
reproductive, genitourinary, endocrine, or metabolism findings	18	Fever (HP:0001945)	0.8% (n=29)
respiratory symptoms	14	Dyspnea (HP:0002094)	35.1% (n=56)
sleep impairment	7	Insomnia (HP:0100785)	31.9% (n=12)

occur at similar frequencies, Dyspnea was more commonly reported in patients following COVID-19 (Supplemental Figure S26). Additional comments are provided in Supplemental Note 2. It is conceivable that in some cases, the occurrence of these ten manifestations in COVID-19 patients is related to care in the ICU. However, all ten symptoms have also been reported in cohorts of individuals treated for COVID-19 as outpatients (Supplemental File 2).

#### 4. Discussion

The fact that some COVID-19 patients experience symptoms following recovery from acute infection is not unexpected. Other infectious diseases, including Epstein-Barr Virus, *Giardia lamblia*, *Coxiella burnetii*, *Borrelia burgdorferi* (Lyme disease) and Ross River virus are also associated with an increased risk for post-infectious sequelae. These sequelae include symptoms such as disabling fatigue, musculoskeletal pain, neurocognitive difficulties, and mood disturbance [15–17]. Chronic fatigue syndrome (CFS) is frequently preceded by a viral infection [18]. However, although these sequelae are well documented, they are still not well understood, and the molecular mechanisms underlying these post-acute presentations have yet to be elucidated.

Post-infectious sequelae have also been documented following infection by other coronaviruses. A subset of patients with severe acute respiratory syndrome (SARS), caused by the coronavirus SARS-CoV, and Middle-Eastern Respiratory Syndrome (MERS), caused by the coronavirus MERS-CoV, were observed to experience persistent or new-onset symptoms, including fatigue [19], following recovery from the acute infection [19–21]. For SARS, follow-up has been conducted up to 15 years post-infection. In addition to fatigue, studies reported effects on lung health and capacity [22–25], psychological health [19], bone health [25], and lipid metabolism [26], with the latter two attributed to treatments involving large doses of steroids [25,26]. Most of the improvement among SARS patients occurred within the first one to two years following infection [25,27,28]. Some patients continued to experience decreased quality of life for more than a decade following the acute illness [26]. Though follow-up studies in MERS patients are sparse, effects on pulmonary function were observed at one year post-infection, with patients who experienced more severe disease at greater risk for long-term effects [29].

Long-term consequences of COVID-19 comprise an unprecedented range of clinical abnormalities that we are barely beginning to understand. These symptoms appear to arise from

pathophysiologic changes that span many organ systems and tissues, potentially explained by SARS-CoV-2's interaction with the endothelium [30]. A wide range of outcomes following acute COVID-19 have emerged as more information about patient recovery has been collected and pathophysiologic mechanisms are revealed. Some patients experience residual symptoms and others develop new symptoms long after the initial infection. Given the timeline of SARS-CoV-2's emergence, studies to date have tracked patients' clinical course up to six months post-infection [31–33], but anecdotal reports are available describing patients with ongoing symptoms as long as one year post-infection [34]. Symptoms experienced after the acute illness represent a significant challenge for patients, physicians, and society as a whole. The causes, patient profile, and even symptom patterns associated with long COVID remain difficult to isolate, and the natural history of this condition remains uncharacterised. Goals of research on long COVID include understanding the natural history of the disease including the prognosis of the many individual manifestations of disease, whether there are well delineated subtypes, whether specific characteristics of the acute phase of COVID-19 predispose to long COVID, and what treatments may best accelerate recovery. Here, we have reported 287 HPO terms representing clinical anomalies reported as long COVID in persons following acute COVID-19. For some of the terms, such as those reported only once to date, further research will be required to determine if the abnormalities are specifically related to COVID-19 and their frequency. We have presented plain-language 'translations' of all terms that can be used to create patient questionnaires.

##### 4.1. Linking layperson and health-professional research

One of the challenges in characterizing long COVID is the fact that patients report symptoms that may not be captured by clinical evaluation or in standard surveys. To truly characterise long COVID and therefore stratify patients into subtypes for care decisions, it is necessary to engage patients directly in the description of their long COVID features. However, medical terminology is often perplexing to patients, making it difficult for patient-researchers to use resources like the HPO. Patients themselves are an eager and untapped source of accurate information about symptoms and phenotypes, some of which may go unnoticed by the clinician [6]. Patients and clinicians use different terms to describe the same symptoms or conditions. In many cases, the clinical term is an exact match to the layperson synonym; however, other times the layperson terminology is less precise.

HPO allows layperson synonyms to be mapped to an ontology, with more specific terms being defined as subtypes of more general terms. Mapping lay terminology to HPO for long COVID symptoms will help patients assist clinicians and researchers in creating robust computational phenotype profiles, which may improve the diagnosis and treatment of long COVID. Here, we systematically abstracted 287 long COVID manifestations including signs, symptoms, and laboratory as well as imaging abnormalities, added layperson synonyms where missing, and mapped layperson to HPO terminology. We wrote plain-language definitions for these terms to supplement the existing definitions that are aimed at healthcare professionals and researchers. [35](35). A full list is available in the supplemental material in human and computer readable form. This common set of definitions can promote integration of research by translating between patient and clinician descriptions of symptoms. We anticipate that these terms, synonyms and definitions will be a critical resource for use in survey instruments and patient apps for standardizing patient-reporting in the future study of long COVID.

In addition to using different terms, inaccurate terms have also been used to describe some symptoms of long COVID. For example, "loss of smell" is a commonly reported problem facing patients during both acute infection and long COVID. However, this term is often used to describe both true "loss of smell", anosmia, and also mistakenly used for "distorted smell," or parosmia. Parosmia is a qualitative disorder of smell that is defined by distorted olfactory perception in the presence of an existing stimulus [36]. Parosmia in COVID-19 has been reported as an unpleasant perception of odorants (troposmia) that follows anosmia and may persist for several months [37]. Post-viral parosmia has been suggested to develop in the olfactory neuron regeneration phase due to a preponderance of immature neurons during re-innervation [38]. This concept has been proposed for COVID-19 related parosmia as well due to a similar pattern of succession [39]. Parosmia was not specifically noted in the studies reviewed for this work, but has been shown to have a high prevalence in acute and post-acute COVID-19 associated parosmia [40].

The HPO layperson synonym mapping from this study will facilitate the construction of common data elements related to long COVID research. A strategic plan of the National Library of Medicine (NLM) 2017-2027 and a recent NIH strategic plan for data science identify common data elements as a means by which to improve data interoperability across studies [41]. Developing a standardised vocabulary with HPO mapped to layperson terms to characterise long COVID symptoms and findings will provide valuable guidance in building common data elements for textual and imaging annotation schema as well as creating patient-centered measurement instruments and clinical surveys.

As long COVID related data increasingly accumulates in EHRs, it is promising to generate practice-based evidence through the secondary use of EHR. However, it is challenging to conduct such research without using natural language processing (NLP), since much information is only stored in unstructured clinical narratives; additionally, because many providers are unfamiliar with the vast array of long COVID symptoms, many are not recorded in these narratives. The layperson definitions and associated synonym mapping would greatly accelerate the development and evaluation of NLP algorithms for extracting long COVID signs and symptoms from EHRs. Baseline NLP algorithms based on HPO can be implemented. Specifically, a fast trie based string matching approach can spot long COVID terms on the fly in massive clinical corpora for near real-time interactive analyses of long COVID phenomena from EHR data. A many-to-one mapping from synonyms to HPO terms identified in this effort will then facilitate more rapid long COVID analytics. Additionally, this effort would enable the rapid development of an annotation guideline for generating benchmarking data, a critical component in developing and evaluating NLP algorithms. If this annotation includes notes from several sites, even spelling mistakes (that nevertheless refer to long COVID

terms) can be spotted by first building a named entity recognition (NER) tool and then mapping mentions to long COVID HPO terms through approximate matching via neural word embeddings constructed from character-based neural language models. Another important affordance of this effort is to be able to mine social media posts for long COVID disclosures from patients and healthcare consumers to complement EHR-derived surveillance [42](preprint).

As the National COVID Cohort Collaborative (N3C) established a collaboration among multiple organizations through pandemic data sharing, i.e., Common Data Model (CDM), the long COVID concept standardization enabled by this study will play an indispensable role in achieving the semantic interoperability for the secondary use of EHR among multiple sites.

#### 4.2. Improving future research on the natural history of long COVID

All published studies analyzed in this work present their results in aggregate rather than providing row level data for each participant. This prevents most data reuse to analyze correlations between comorbidities and risk for long COVID or for specific manifestations of long COVID, and this makes it impossible to investigate potential correlations between long COVID manifestations. Therefore, future studies should present non-identifiable information about individual patients. This will allow correlations between variables. Also, studies need to use controlled vocabulary to classify patients and need to agree on a minimal set of information. Presenting (non-identifiable) data in the form of a table with one row per patient would be a great improvement over the current status. More sophisticated strategies for recording individual clinical histories are available, such as the Global Alliance for Genomics and Health phenopacket schema [43].

The majority of studies included in this analysis did not apply inclusion criteria to correspond to any specific definition of long COVID, but instead studied groups of patients who had previously undergone infection by SARS-CoV-2 with a range of manifestations in the acute phase. Long COVID can be broadly defined as delayed recovery from an episode of COVID-19 and is characterised by lasting effects of the infection, e.g., persistence of symptoms or onset of new chronic diseases, for longer than would be expected [8]. Although no firm criteria have been established to define the post-acute period or sub-categories within long COVID, several sets of guidelines have been proposed for the classification of COVID-19-related disease phenotypes, and these criteria were compared to the definitions used in the literature. For example, a recently proposed public health framework classifies SARS-CoV-2-related disease into three categories [44]. The first is acute COVID-19, or the disease most commonly associated with acute SARS-CoV-2 infection. The second category includes Multisystem Inflammatory Syndrome in Children (MIS-C) and in adults (MIS-A), a less common presentation of SARS-CoV-2 infection characterised by hyperinflammation that can appear 4-6 weeks after viral infection [45]. The third category describes late sequelae [44]. In terms of defining study cohorts, adherence with this definition would require a clinical diagnosis, rather than a SARS-CoV-2 test alone, to distinguish MIS-C/A and COVID-19. While it appears too early to propose a set of computable definitions for the various types of disease associated with SARS-CoV-2 (because we are still learning about the natural history), it would be advantageous for studies to apply a currently accepted definition of long COVID and to describe details in the methods. Studies should denote comorbidities using a standard ontology of diseases such as Mondo [46].

Additionally, in many cases it is difficult to know if a clinical abnormality was present prior to acute COVID-19 and was merely diagnosed by investigations and additional clinical tests that were performed following the diagnosis of COVID-19. For instance, Hepatic steatosis was reported to be more common in individuals following acute COVID-19 infection in a single study [47]. However, this is a

common finding in the general population and additional research will be required to characterise its precise relation to long COVID.

The studies analyzed in this work varied widely in the terminology used to describe patient-reported symptoms as well as clinical signs, laboratory abnormalities, and imaging findings. For example, the studies analyzed included a mixture of reports of ageusia [48,49], anosmia [48,49], anosmia/ageusia [50], loss of smell [51,52], loss of taste [51], loss of smell and taste [53], loss of smell or taste [54], and loss of smell and/or taste [55]. While in many cases there are parallels among studies (e.g., studies reporting anosmia and loss of smell are likely to be asking the same or similar questions of patients), the lack of a strict definition prevents straight-forward symptom matching across multiple published analyses. Different studies measure clinical manifestations in different ways. For instance, the presence of fatigue can be measured by a yes/no question in an online questionnaire or can be inferred from the results of a multidimensional study instrument such as the Short Form-36 Vitality scale [56]. In such cases, standard use of a full terminology such as HPO would be useful to create expressive and consistent meaning across studies. Future studies should make data available using either the HPO terms provided here or other terms from the full collection of over 16,000 HPO terms.

From a clinical standpoint our work demonstrates that managing long COVID patients will require a multidisciplinary effort. Given that respiratory system findings and fatigue were the most common, a pulmonologist with a pulmonary rehabilitation program will be one of the cornerstones of any long COVID management program. At the same time considering the high frequency of psychiatric and neurological symptoms, psychiatrists and neurologists with specialization in neurocognitive testing and treatment will be necessary. Furthermore given the extreme variation in symptoms and presentation, a primary care internist will have to be responsible for coordinating the care of long COVID patients with appropriate referrals when required. Taking into account the high frequency of long COVID in survivors primary care physicians will have to be made aware of the myriad presentations of long COVID with active screening for long COVID. Primary care physicians will need to refer long COVID patients to dedicated long COVID management programs.

Working toward computable long COVID phenotypes in this way will improve our ability to understand the natural history of long COVID. Such phenotypes will also allow observational analyses of factors that may reduce long COVID symptoms. The standardised phenotypic features and synonyms bundled in the HPO terms presented here are a foundation for natural language processing of EHR data, clinical decision support tools, and analytic approaches such as machine learning.

## 5. Contributors

Author contributions are as follows: MAH, PNR, RRD conceptualized and designed the study. RRD, MAR, NV, LC, HR, PNR performed the data abstraction and curation. RRD, PNR validated the data. PNR performed the data analysis. RRD, MAR, HR, PNR, MAH drafted the manuscript. RRD, MAR, NV, LC, JR, AJA, MDB, TDB, TB, EAB, CTB, JBB, TJC, LEC, HC, CGC, BDC, HED, JG, CSG, WBH, RK, WDK, FMK, SK, CL, FL, HL, VM, CRM, NM, DRM, JAM, DSM, RAM, TSM, AMP, MAP, EP, JTR, JS, RS, AES, JS, HE, GSS, AAS, UT, GDV, LW, MAH, and PNR interpreted the data and critically revised the manuscript for important intellectual content related to their area of expertise. JAM, PNR assisted with data visualization. All authors read and approved the final version to be published.

## Declaration of Competing Interest

RRD, TDB, JBB, CGC, WBH, JAM, AMP, ERP, HMR, JS, RAS, AES, JS, GS, MAH, PNR report funding from NIH. MAH and JAM are co-founders of Pryzm Health.

## Acknowledgements

This work was supported by NCATS U24 TR002306. Rachel R. Deer supported by UTMB CTSA UL1TR001439 and NIA P30AG024832; Halie M. Rando was supported by The Gordon and Betty Moore Foundation (GBMF 4552) and the National Human Genome Research Institute (R01 HG010067); Tellen D. Bennett supported by NIH UL1TR002535 03S2 and NIH UL1TR002535; James Brian Byrd supported by NIH grant K23HL128909; Christopher G. Chute supported by U24 TR002306. Julian Solway supported by UL1TR002389. Mallory A. Perry supported by K99GM145411.

## Data Sharing Statement

The data used to support the findings of this study are provided in the main text and supplemental files.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.ebiom.2021.103722](https://doi.org/10.1016/j.ebiom.2021.103722).

## REFERENCES

- [1] Puelles VG, Lütgehetmann M, Lindenmeyer MT, Sperhake JP, Wong MN, Allweiss L, et al. Multiorgan and Renal Tropism of SARS-CoV-2. *N Engl J Med* 2020;383(6):590–2 Aug 6.
- [2] Gavriatopoulou M, Korompoki E, Fotiou D, Ntanasis-Stathopoulos I, Psaltopoulou T, Kastritis E, et al. Organ-specific manifestations of COVID-19 infection. *Clin Exp Med* 2020;20(4):493–506 Nov.
- [3] Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. *Acta Biomed* 2020;91(1):157–60 Mar 19.
- [4] Gao Z, Xu Y, Sun C, Wang X, Guo Y, Qiu S, et al. A systematic review of asymptomatic infections with COVID-19. *J Microbiol Immunol Infect* 2021;54(1):12–6 Feb.
- [5] Rando HM, Bennett TD, Byrd JB, Bramante C, Callahan TJ, Chute CG, et al. Challenges in defining Long COVID: Striking differences across literature, Electronic Health Records, and patient-reported information. medRxiv [Internet] 2021 Mar 26 Available from <http://dx.doi.org/>. doi: [10.1101/2021.03.20.21253896](https://doi.org/10.1101/2021.03.20.21253896).
- [6] Davis HE, Assaf GS, McCorkell L, Wei H, Low RJ, Re'em Y, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *EClinicalMedicine* 2021;38:101019 Aug.
- [7] Greenhalgh T, Knight M, A'Court C, Buxton M, Husain L. Management of post-acute covid-19 in primary care. *BMJ* 2020;370:m3026. Aug 11.
- [8] Mahase E. Covid-19: What do we know about "long covid"? *BMJ* 2020;370:m2815. Jul 14.
- [9] Nasserie T, Hittle M, Goodman SN. Assessment of the Frequency and Variety of Persistent Symptoms Among Patients With COVID-19: A Systematic Review. *JAMA Netw Open* 2021;4(5):e2111417 May 3.
- [10] Köhler S, Gargano M, Matentzoglou N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. The Human Phenotype Ontology in 2021. *Nucleic Acids Res* 2021;49(D1):D1207–17 Jan 8.
- [11] Robinson PN, Mungall CJ, Haendel M. Capturing phenotypes for precision medicine. *Cold Spring Harb Mol Case Stud* 2015;1(1):a000372 Oct.
- [12] Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gouridine J-P, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res* 2019;47(D1):D1018–27 Jan 8.
- [13] Haendel MA, Chute CG, Robinson PN. Classification, Ontology, and Precision Medicine. *N Engl J Med* 2018;379(15):1452–62 Oct 11.
- [14] Lever J, Altman RB. Analyzing the vast coronavirus literature with CoronaCentral. *Proc Natl Acad Sci U S A* [Internet]. 2021;118(23) Jun 8 Available from <https://www.ncbi.nlm.nih.gov/pubmed/34016708>.
- [15] White PD, Thomas JM, Kangro HO, Bruce-Jones WD, Amess J, Crawford DH, et al. Predictions and associations of fatigue syndromes and mood disorders that occur after infectious mononucleosis. *Lancet* 2001;358(9297):1946–54 Dec 8.
- [16] Wensaas K-A, Langeland N, Hanevik K, Mørch K, Eide GE, Rortveit G. Irritable bowel syndrome and chronic fatigue 3 years after acute giardiasis: historic cohort study. *Gut* 2012;61(2):214–9 Feb.
- [17] Hickie I, Davenport T, Wakefield D, Vollmer-Conna U, Cameron B, Vernon SD, et al. Post-infective and chronic fatigue syndromes precipitated by viral and non-viral pathogens: prospective cohort study. *BMJ* 2006;333(7568):575. Sep 16.
- [18] Bansal AS, Bradley AS, Bishop KN, Kiani-Alikhan S, Ford B. Chronic fatigue syndrome, the immune system and viral infection. *Brain Behav Immun* 2012;26(1):24–31 Jan.
- [19] Moldofsky H, Patcai J. Chronic widespread musculoskeletal pain, fatigue, depression and disordered sleep in chronic post-SARS syndrome; a case-controlled study. *BMC Neurol* 2011;11:37. Mar 24.
- [20] Lam MH-B, Wing Y-K, Yu MW-M, Leung C-M, Ma RCW, Kong APS, et al. Mental morbidities and chronic fatigue in severe acute respiratory syndrome survivors: long-term follow-up. *Arch Intern Med* 2009;169(22):2142–7 Dec 14.

- [21] Lee SH, Shin H-S, Park HY, Kim JL, Lee JJ, Lee H, et al. Depression as a Mediator of Chronic Fatigue and Post-Traumatic Stress Symptoms in Middle East Respiratory Syndrome Survivors. *Psychiatry Investig* 2019;16(1):59–64 Jan.
- [22] Ong K-C, Ng AW-K, Lee LS-U, Kaw G, Kwek S-K, Leow MK-S, et al. 1-year pulmonary function and health status in survivors of severe acute respiratory syndrome. *Chest* 2005;128(3):1393–400 Sep.
- [23] Ngai JC, Ko FW, Ng SS, To K-W, Tong M, Hui DS. The long-term impact of severe acute respiratory syndrome on pulmonary function, exercise capacity and health status. *Respirology* 2010;15(3):543–50 Apr.
- [24] Hui DS, Ko FW, Chan DP, Fok JP, Chan MC, To KW, et al. The long-term impact of severe acute respiratory syndrome (SARS) on pulmonary function, exercise capacity, and quality of life in a cohort of survivors. *Chest* 2005;128(4):148S. Oct.
- [25] Zhang P, Li J, Liu H, Han N, Ju J, Kou Y, et al. Long-term bone and lung consequences associated with hospital-acquired severe acute respiratory syndrome: a 15-year follow-up from a prospective cohort study. *Bone Res* 2020;8:8. Feb 14.
- [26] Wu Q, Zhou L, Sun X, Yan Z, Hu C, Wu J, et al. Altered Lipid Metabolism in Recovered SARS Patients Twelve Years after Infection. *Sci Rep*. 2017;7(1):9110. Aug 22.
- [27] Yin C-H, Wang C, Wen Y, Jiang L, Lu Q, Li J-L, et al. [Prospective 2-year clinical study of patients with positive IgG-antibodies after recovering from severe acute respiratory syndrome]. *Zhongguo Wei Zhong Bing Ji Jiu Yi Xue*. 2005 Dec;17(12):740–2.
- [28] Liu Y-X, Ye Y-P, Zhang P, Chen J, Ye H, He Y-H, et al. [Changes in pulmonary function in SARS patients during the three-year convalescent period]. *Zhongguo Wei Zhong Bing Ji Jiu Yi Xue* 2007;19(9):536–8 Sep.
- [29] Il Jun K, Park WB, Kim G, Choi J-P, Rhee J-Y, Cheon S, et al. Long-term Respiratory Complication in Patients with Middle East Respiratory Syndrome: 1-year Follow-up after the 2015 Outbreak in South Korea. *Open Forum Infect Dis* 2017;4(suppl\_1):S577. Oct 4–S577.
- [30] Giordo R, Paliogiannis P, Mangoni AA, Pintus G. SARS-CoV-2 and endothelial cell interaction in COVID-19: molecular perspectives. *Vasc Biol* 2021;3(1):R15–23 Jan 12.
- [31] Sykes DL, Holdsworth L, Jawad N, Gunasekera P, Morice AH, Crooks MG. Post-COVID-19 Symptom Burden: What is Long-COVID and How Should We Manage It? *Lung* 2021;199(2):113–9 Apr.
- [32] Zhang J, Xu J, Zhou S, Wang C, Wang X, Zhang W, et al. The characteristics of 527 discharged COVID-19 patients undergoing long-term follow-up in China. *Int J Infect Dis* 2021;104:685–92 Mar.
- [33] Huang C, Huang L, Wang Y, Li X, Ren L, Gu X, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet* 2021;397(10270):220–32 Jan 16.
- [34] Kirby T. COVID-19 survivor experiencing long-term symptoms. *Lancet Respir Med* [Internet] 2021 Feb 12; Available from: <http://dx.doi.org/>. doi: 10.1016/S2213-2600(21)00092-8.
- [35] Vasilevsky NA, Foster ED, Engelstad ME, Carmody L, Might M, Chambers C, et al. Plain-language medical vocabulary for precision diagnosis. *Nat Genet* 2018;50(4):474–6 Apr.
- [36] Hawkes C. Parosmia: treatment, mechanism, and types. *BMJ* 2020;371:m4739. Dec 8.
- [37] Chary E, Carsuzaa F, Trijole J-P, Capitaine A-L, Roncato-Saberan M, Fouet K, et al. Prevalence and Recovery From Olfactory and Gustatory Dysfunctions in Covid-19 Infection: A Prospective Multicenter Study. *Am J Rhinol Allergy* 2020;34(5):686–93 Sep.
- [38] Cavazzana A, Larsson M, Münch M, Hähner A, Hummel T. Postinfectious olfactory loss: A retrospective study on 791 patients. *Laryngoscope* 2018;128(1):10–5 Jan.
- [39] Mastrangelo A, Bonato M, Cinque P. Smell and taste disorders in COVID-19: From pathogenesis to clinical features and outcomes. *Neurosci Lett* 2021;748:135694 Mar 23.
- [40] Hopkins C, Surda P, Vaira LA, Lechien JR, Safarian M, Saussez S, et al. Six month follow-up of self-reported loss of smell during the COVID-19 pandemic. *Rhinology* 2021;59(1):26–31 Feb 1.
- [41] Sheehan J, Hirschfeld S, Foster E, Ghitza U, Goetz K, Karpinski J, et al. Improving the value of clinical research through the use of Common Data Elements. *Clin Trials* 2016;13(6):671–6 Dec.
- [42] Banda JM, Singh GV S, Alser OH, Prieto-Alhambra D. Long-term patient-reported symptoms of COVID-19: an analysis of social media data [Internet]. *bioRxiv*. medRxiv 2020 Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.07.29.20164418>.
- [43] phenopacket-schema [Internet]. Github; [cited 2020 Jun 8]. Available from: <https://github.com/phenopackets/phenopacket-schema>
- [44] Datta SD, Talwar A, Lee JT. A Proposed Framework and Timeline of the Spectrum of Disease Due to SARS-CoV-2 Infection: Illness Beyond Acute Infection and Public Health Implications. *JAMA* 2020;324(22):2251–2 Dec 8.
- [45] Vogel TP, Top KA, Karatzios C, Hilmers DC, Tapia LI, Mocerri P, et al. Multisystem inflammatory syndrome in children and adults (MIS-C/A): Case definition & guidelines for data collection, analysis, and presentation of immunization safety data. *Vaccine* [Internet] 2021 Feb 24 Available from: <http://dx.doi.org/>. doi: 10.1016/j.vaccine.2021.01.054.
- [46] Shefchek KA, Harris NL, Gargano M, Matentzoglou N, Unni D, Brush M, et al. The Monarch Initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res* 2020;48(D1):D704–15 Jan 8.
- [47] Dennis A, Wamil M, Alberts J, Oben J, Cuthbertson DJ, Wootton D, et al. Multiorgan impairment in low-risk individuals with post-COVID-19 syndrome: a prospective, community-based study. *BMJ Open* 2021;11(3):e048391 Mar 30.
- [48] Garrigues E, Janvier P, Kherabi Y, Le Bot A, Hamon A, Gouze H, et al. Post-discharge persistent symptoms and health-related quality of life after hospitalization for COVID-19. *J Infect* 2020;81(6):e4–6 Dec.
- [49] Goërtz YM, Van Herck M, Delbressine JM, Vaes AW, Meys R, Machado FVC, et al. Persistent symptoms 3 months after a SARS-CoV-2 infection: the post-COVID-19 syndrome? *ERJ Open Res* [Internet] 2020;6(4) Oct Available from: <http://dx.doi.org/>. doi: 10.1183/23120541.00542-2020.
- [50] Carvalho-Schneider C, Laurent E, Lemaignen A, Beauflis E, Bourbao-Tournois C, Laribi S, et al. Follow-up of adults with noncritical COVID-19 two months after symptom onset. *Clin Microbiol Infect* 2021;27(2):258–63 Feb.
- [51] Tenforde MW, Kim SS, Lindsell CJ, Rose EB, Shapiro NI, Clark Files D, et al. Symptom Duration and Risk Factors for Delayed Return to Usual Health Among Outpatients with COVID-19 in a Multistate Health Care Systems Network – United States, March–June 2020 [Internet]. *MMWR. Morbidity and Mortality Weekly Report*. 2020;69:993–8 Available from: <http://dx.doi.org/>. doi: 10.15585/mmwr.mm6930e1.
- [52] Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and predictors of long COVID. *Nat Med* 2021;27(4):626–31 Apr.
- [53] Petersen MS, Kristiansen MF, Hanusson KD, Danielsen ME, Á Steig B, Gaini S, et al. Long COVID in the Faroe Islands - a longitudinal study among non-hospitalized patients. *Clin Infect Dis* [Internet] 2020 Nov 30; Available from: <http://dx.doi.org/>. doi: 10.1093/cid/ciaa1792.
- [54] Nehme M, Braillard O, Alcoba G, Aebischer Perone S, Courvoisier D, Chappuis F, et al. COVID-19 Symptoms: Longitudinal Evolution and Persistence in Outpatient Settings. *Ann Intern Med* [Internet] 2020 Dec 8; Available from: <http://dx.doi.org/>. doi: 10.7326/M20-5926.
- [55] Chopra V, Flanders SA, O'Malley M, Malani AN, Prescott HC. Sixty-Day Outcomes Among Patients Hospitalized With COVID-19. *Ann Intern Med* 2021;174(4):576–8 Apr.
- [56] Neuberger GB. Measures of fatigue: The fatigue questionnaire, fatigue severity scale, multidimensional assessment of fatigue scale, and short form-36 vitality (energy/fatigue) subscale of the short form health survey. *Arthritis Rheum* 2003;49(S5):S175–83 Oct 15.