Evolving Phenotypes of non-hospitalized Patients that Indicate Long Covid

Hossein Estiri, PhD[1,2], Zachary H Strasser, MD[1,2,3], Gabriel A Brat, MD[3], Yevgeniy R Semenov, MD[4], Chirag J Patel, PhD[3], Shawn N Murphy, MD, PhD[1,5,6]

[1]Laboratory of Computer Science, Massachusetts General Hospital, Boston, MA, 02144, USA
[2]Department of Medicine, Massachusetts General Hospital, Boston, MA, 02144, USA
[3]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
[4]Department of Dermatology, Massachusetts General Hospital, Boston, MA, 02114, USA
[5]Department of Neurology, Massachusetts General Hospital, Boston, MA, 02114, USA
[6]Research Information Science and Computing, Mass General Brigham, Boston MA, USA

Abstract

Many of the symptoms characterized as the post-acute sequelae of SARS-CoV-2 infection (PASC) could have multiple causes or similarly seen in non-COVID patients. An accurate identification of phenotypes will be important to guide future research and the healthcare system to focus its efforts and resources on adequately controlled age- and gender-specific sequelae of COVID-19 infection. This retrospective electronic health records (EHR) cohort study, we applied a computational framework for knowledge discovery from clinical data, MLHO, to identify phenotypes that positively associate with a past positive PCR test for COVID-19. We evaluated the post-test phenotypes in two temporal windows at 3-6 and 6-9 months after the test and by age and gender. We utilized longitudinal diagnosis records stored in EHRs from Mass General Brigham (MGB) 57 thousand patients who tested positive or negative for COVID-19 and were not hospitalized. Statistical analyses were performed on data from March 2020 to March 2021. PCR test results and subsequent diagnosis records that were recorded for the first time two months or later after the PCR test. We identified 28 phenotypes among different age/gender cohorts or time windows that positively associated with a past SARS-CoV-2 infection. All identified phenotypes were newly recorded in patients' medical records two months or longer after a COVID-19 PCR test in non-hospitalized patients regardless of the test result. Among these phenotypes, a new diagnosis record for anosmia and dysgeusia (OR 2.17, 95% CI [1.42 - 3.25]), alopecia (OR 3.54, 95% CI [2.92 - 4.3]), chest pain (OR 1.35, 95% CI [1.16 - 1.56]), or chronic fatigue syndrome (OR 1.81-2.28, 95% CI [1.38 - 3.68]) are the most significant indicators of a past COVID-19 infection, especially among women younger than 65. Among men, edema (OR 1.83, 95% CI [1.23 - 2.66]) and disease of nail (OR 3.54, 95% CI [1.63 - 7.29]) in patients 65 and older or proteinuria (OR 2.66, 95% CI [1.61 - 4.34]) in patients under 65 are associated with a positive COVID-19 PCR test in the past few months. Our approach avoids a flood of false positive discoveries, while offering a more probabilistic flexible criterion than the standard linear phenome-wide association study (PheWAS). These findings suggest that some of the previously identified post sequelae of COVID-19 may not be accurate and that most of the PASC are observed in patients under 65 years of age.

Introduction

The onslaught of the COVID-19 pandemic in the United States has been relentless. For hundreds of thousands (if not millions), recovery from the acute phase of the SARS-CoV-2 infection, the coronavirus that causes COVID-19, will be grueling with a debilitating second act. A collection of persistent physical (e.g., fatigue, dyspnea, chest pain, cough), psychological (e.g., anxiety, depression, post-traumatic stress disorder), and neurocognitive symptoms (e.g., impaired memory and concentration) can appear and last for weeks or months in patients after acute COVID-19.[1–8] Many of the symptoms characterized as the post-acute sequelae of COVID-19 (PASC) could have multiple causes.

So far, a number of studies have been published on PASC,[1–7,9,10] but most have small samples, case-series, or rely on self-reports. Carfi et al assessed 179 hospitalized COVID patients in Italy at an average of 60 days after the onset of symptoms using a standard questionnaire.[11] Only 12.6% were completely free of all COVID-19 symptoms and 55% had 3 or more symptoms. The most common symptoms were fatigue, dyspnea, joint pain, and chest pain. Chopra et al performed an observational study of 488 patients who were hospitalized 60 days after their discharge with a phone survey.[12] The most common persistent symptoms were cough, dyspnea, persistent loss of taste or smell, and worsening difficulty completing activities of daily living. Huang et al performed one of the larger cohort studies where they analyzed 1,733 COVID patients discharged from a hospital in China with a questionnaire at 6 months.[13] They identified fatigue, muscle weakness, sleep difficulties, anxiety, and depression as the most common symptoms 6 months after the initial diagnoses. These studies are all case-series, focusing only on patients with COVID-19. Additionally, prior PASC studies often focus on patients with severe COVID-19 symptoms after hospitalization. It is unclear whether the identified persistent symptoms hold true among Covid patients not hospitalized. Further, many of the published studies are based on small cohorts (several hundred COVID-19 patients were analyzed) and relied on self-reported outcomes which can embody potential biases due to, for example, exaggeration of symptoms.[14]

There have also been a number of less commonly reported symptoms including ocular inflammation[15], cardiac involvement [16,17], autonomic instability[18], recurrent pseudomonas infections [19], persistent mucous secretion [20], micro-structural changes to the brain [21] and Guillain-Barre syndrome.[22]. A large cohort analyzing the ICD-10 diagnoses in the electronic health record between patients with and without a history of COVID could help clarify the actual association with the disease.

We present results from a retrospective cohort study of over 57,000 patients with a PCR test for COVID-19 in a Mass General Brigham (MGB) facility. We detected de novo phenotypes that appeared for the first time in EHRs at two temporal windows of 3-6 and 6-9 months after a COVID-19 test for both COVID-positive and -negative patients. We determined which phenotypes were positively associated with a recent/past SARS-CoV-2 infection. Leveraging MLHO, a computational framework developed for knowledge discovery from electronic health records (EHRs)[23–25] with a validated utility for studying and modeling post-COVID outcomes,[26,27] we identified 28 phenotypes in different age/gender groups or time windows positively associated with a past COVID-19 infection. All identified phenotypes were newly recorded in patients' medical records two months or longer after a COVID-19 PCR test in non-hospitalized patients regardless of the test result.

Methods

We utilized longitudinal EHR diagnosis records from all patients who tested for SARS-CoV-2 infection -- polymerase chain reaction (PCR) -- between March 2020 and March 2021 in a Mass General Brigham (MGB) facility. We limited the patient cohort to those who were alive and not hospitalized. To increase the confidence that a patient in our cohort would likely seek care within

MGB in the post-COVID era, we further narrowed the study population to patients who had two diagnosis records, 6 months apart, in our electronic data repositories up to three years before the COVID-19 test. We also excluded patients who had an ICD-10 code referring to past COVID-19 but having a negative PCR test in the MGB records due to our inability to approximate the infection date. Use of clinical data in this study was approved by the MGB Institutional Review Board (IRB) with a waiver of informed consent.

The EHR diagnosis records contained ICD-10 codes (the 10th revision of the International Statistical Classification of Diseases and Related Health Problems). To represent phenotypes for the analyses, we mapped the ICD-10 diagnosis codes for each medical condition to a unique phenotype code (PheCode) from the Phenome-wide association studies (PheWAS)[28,29] groups of phenotypes. We assigned a temporal buffer of two months after the PCR test as a proxy for the acute phase in COVID-19 patients and used the first observation of phenotypes that were recorded for the first time after the acute phase (Figure 1). Using this temporal segmentation, we further limited the data, but only using the first observation of the records (to minimize the problem list repetitions) and only considered the diagnosis records that for the first time appeared in a patient's medical records two months or longer after the PCR test -- see eMethods for more details.

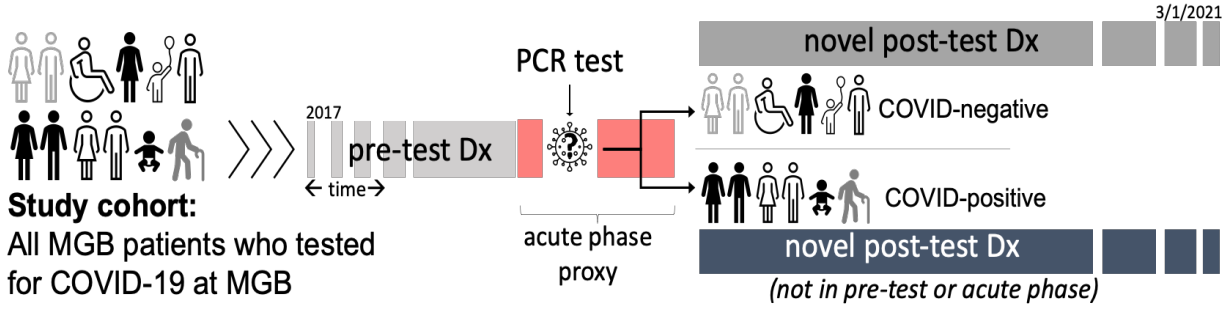


Figure 1. Study cohort, temporal segmentation, and diagnosis recording.

Computational methods

To robustly identify the phenotypes that are positively associated with a recent positive test for COVID-19, we applied a multivariate temporal approach to classify past PCR test results from the post-test clinical data. We leveraged the MLHO framework,[26] which includes a suite of computational algorithms[23,26] that were specifically designed for modeling and phenotyping clinical data and to develop multivariate regression models predicting a past positive/negative COVID-19 PCR test. We followed the same analytic process used by Estiri et al. (2021)[30] that was used to identify risk factors for COVID-19 mortality from EHR data. From the MLHO framework, the computational process to conduct multivariate PheWAS involved applying the Minimize Sparsity, Maximize Relevance (MSMR)[23,31,32] algorithm, clinical expertise, and multinomial generalized linear modeling (GLM) with component-wise functional gradient boosting, and a composite confidence score to identify the phenotypes that are positively associated with a past PCR test (see eMethods).

We also performed linear univariate analyses for the cohort and each of the sub-cohorts as a benchmark, in which we counted all new phenotypes entered into the patient's chart at least two months after the PCR test. We computed the incidence rates (Covid-positive), Odds Ratio (OR), and P-values between those with and without the novel phenotype and computed 1-sided Fisher's exact tests.

The approach applied in this study can be conceptualized as a multivariate PheWAS,[28,29] in which we expand the univariate p-value-dependant criteria for identifying phenome-wide associations

to a more comprehensive and multivariate entropy-based process. Our approach avoids a flood of false positive discoveries, while offering a more flexible probabilistic criterion than the standard PheWAS. The iterative process in MLHO provides means to an interpretable probabilistic confidence score for each phenotype associated with a past positive COVID-19 PCR test. Analyses were conducted in R statistical language.

Cohort Stratification

To increase specificity, we stratified the analyses by age and gender in a nested structure. This resulted in the following strata: 1) all patients, 2) 65 and older, 3) under 65, 4) 65 and older female, 5) 65 and older male, 6) under 65 female, and 7) under 65 male. In addition to stratifying the cohort, we controlled for age and gender (in gender-agnostic models) of the patient.

Results

From approximately 340,000 patients who tested for COVID-19 in an MGB facility with a nasal swab, 138,257 met our inclusion/exclusion criteria. After applying the approach for keeping records, 57,622 patients remained in our final study cohort, 11,491 (19.94%) of whom were positive for the SARS-CoV-2 virus (Table 1S). After the sparsity screening (i.e., removing low prevalence [<0.22%] phenotypes from sub-cohorts), on average 441 phenotypes were evaluated in each PheWAS.

Overall, we identified 28 phenotypes in different age/gender groups or time windows that positively associated with a past positive COVID-19 test, with confidence score higher than 80 (Figures 2 and 3). All identified phenotypes were newly recorded in patients' medical records two months or longer after a COVID-19 PCR test in non-hospitalized patients regardless of the test result. We present the phenotypes by the confidence scores. Univariate Odds Ratios (ORs), 95 percent confidence intervals, and Confidence Scores (CSs) are provided.

Results demonstrated extremely high confidence (>97%) in seven phenotypes, which in the overall cohort and/or one or more sub-cohorts indicate a positive past COVID-19 infection. Alopecia was identified in all iterations of the of MLHO between months three and six, in the overall cohort (OR: 3.54, 95% CI [2.91 - 4.29], CS: 100) and both younger and older than 65, specifically in women both under 65 (OR: 3.49, 95% CI [2.75 - 4.42], CS: 100) and 65 and older (OR: 4.17, 95% CI [2.7 - 6.4], CS: 100). The phenotype was not identified in the 6-9 month temporal window. Similarly, a new diagnosis record of nonspecific chest pain was indicative of a past COVID-19 infection in the 3-6 month temporal window (OR: 1.35, 95% CI [1.16 - 1.56], CS: 100), specifically among women younger than 65 (OR: 1.54, 95% CI [1.25 - 1.89], CS: 100).

Anosmia and dysgeusia was identified in more than 97 percent of the MLHO iterations, both within the 3-6 (OR: 2.17, 95% CI [1.51 - 3.07], CS: 100) and 6-9 (OR: 2.17, 95% CI [1.42 - 3.24], CS:97) months temporal windows. The phenotype was specifically indicative of past positive COVID-19 in women who were under 65 (3-6 months OR: 2.55, 95% CI [1.57 - 4.1], CS: 99).

Among other identified phenotypes with 97 and higher confidence score, was edema in men 65 and older (OR: 1.83, 95% CI [1.23 - 2.66], CS: 98) and proteinuria among patients under 65 (OR: 2.66, 95% CI [1.61 - 4.34], CS: 98), specifically in men, both phenotypes in the 3-6 months. In the 6-9 months temporal window, disease of nail, NOS was also identified by 100% of the multivariate PheWAS iterations, specifically among men 65 and older (OR: 3.54, 95% CI [1.63 - 7.29], CS: 97).

10 additional phenotypes were identified by the MLHO as an indication of a past COVID-19 infection with confidence scores between 90 and 96. Chronic fatigue syndrome was identified in both temporal windows among patients younger than 65 (3-6 months OR: 1.81, 95% CI [1.26 - 2.58], CS: 92, and 6-9 months OR: 2.28, 95% CI [1.38 - 3.68], CS: 95), particularly in women.

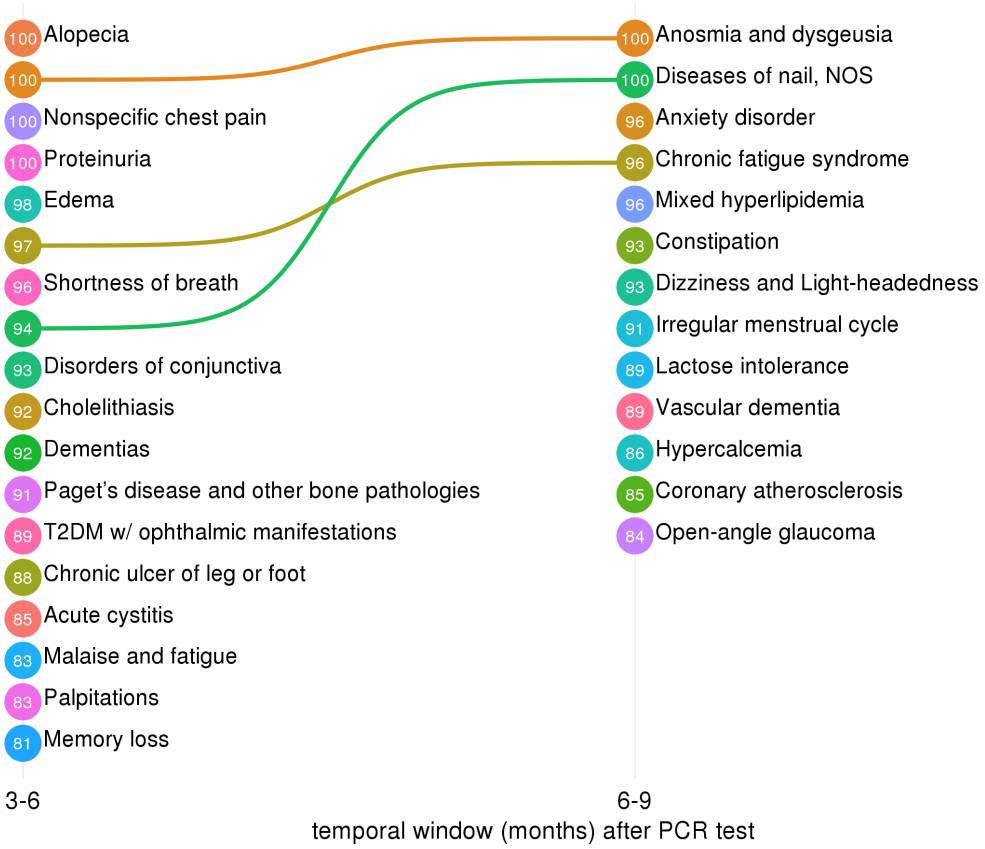
Among other COVID-19 related phenotypes identified as indicators of a past COVID-19 infection with a 90-to-96 confidence score were: within the 3-6 months post-test window, among 65 and

older, cholelithiasis in women (OR: 2.96, 95% CI [1.30 - 6.35], CS: 92) and dementias (OR: 2.58, 95% CI [1.38 - 4.65] , CS: 92), and Paget's disease and other bone pathologies (OR: 1.99, 95% CI [1.26 - 3.07], CS: 91), specially in men. In patients under 65, disorders of conjunctiva (OR: 2.34, 95% CI [1.41 - 3.83], CS: 93) was identified.

* MLHO confidence scores are plotted in white font. 100% means identified in all 100 MLHO iterations. Phenotypes included have been associated with a positive past COVID-19 test with a confidence score higher than 80 percent in at least a sub-cohort.

Figure 2. Phenotypes that positively associate with a past COVID-19 positive PCR test.

In the 6-9 month temporal window and among women 65 and older, anxiety disorder (OR: 2.12, 95% CI [1.28 - 3.36], CS: 96), constipation (OR: 1.34, 95% CI [0.66 - 2.49]), dizziness and light-headedness (OR: 1.97, 95% CI [1.23 - 3.03], CS: 93) were indications of a past positive COVID-19 PCR test. In the under-65 patients, irregular menstrual cycle in women (OR: 1.97, 95% CI [1.29 - 2.95], CS: 91), and mixed hyperlipidemia (OR: 2.14, 95% CI [1.26 - 3.53], CS: 96) in men were associated with a prior COVID-19 infection.



* phenotypes are ranked by MLHO confidence score.

Figure 3. Presentation of the PASC phenotypes across temporal windows.

11 more novel phenotypes had a 81-89% confidence score, indicating a positive association with a past SARS-CoV-2 infection (Figure 2). Notable indicators from this group of phenotypes were: a) among women under 65, lactose intolerance 6-9 months post infection (OR: 1.76, 95% CI [1.19 - 2.54], CS: 89), and malaise and fatigue in the 3-6 months post-test window (OR: 1.32, 95% CI [1.08 - 1.60], CS: 80), b) among men 65 and older, memory loss (OR: 2.12, 95% CI [1.12 - 3.79], CS: 81) and type 2 diabetes with ophthalmic manifestations (OR: 3.57, 95% CI [1.377 - 8.74], CS: 89) in the 3-6 month window, c) vascular dementia among women 65 and older in the 6-9 months post-COVID window (OR: 3.88, 95% CI [1.51 - 9.33], CS: 86), d) open-angle glaucoma among men under 65, 6-9 month after COVID test (OR: 4.05, 95% CI [1.36 - 12.09], CS: 84), and e) palpitations in the overall under 65 population (OR: 1.44, 95% CI [1.15 - 1.79], CS: 83).

Discussion

The COVID-19 pandemic in the United States has raged nearly uncontrolled since 2020. While the exact number of people afflicted by the post-acute sequelae of SARS-CoV-2 infection is

unknown, it represents a significant public health burden because of the large magnitude of the COVID-19 spread globally.

Phenotypes such as alopecia, anosmia and dysgeusia, and chest pain, have been documented as common signs and symptoms of PASC.[7,34,35] This study also shows that they are some of the earliest associations with the syndrome. Interestingly, Paget's disease was identified in several of the cohort groups by the analysis. The univariate analysis also noted an association with COVID despite it not being significant after FDR adjustment. Data exists that suggests Paget's disease is not entirely genetic, but may be associated with viral infections.[36,37] Previous studies have found clusters of exposure to unvaccinated dogs, cats, or birds as having an association with a heightened risk of developing Paget's disease.[38] While the authors are unaware of any studies showing Paget's disease as associated with COVID-19, these results suggest there could be relationship.

Another surprising finding was the strong association with disease of the nail. The disease of the nail phenotype includes a variety of diagnoses including leukonychia, onycholysis, mee's lines, muehrcke lines, and Beau's lines all of which are markers of overall well-being and have been associated with infections, renal, or hepatic dysfunction previously. Beau's lines have specifically been associated with COVID-19 infections.[39] Our results suggest this association is widespread and likely a result of the systemic infection including renal injury.

Proteinuria was also identified as having an association with COVID-19 among patients less than 65. COVID-19 has previously been associated with acute kidney injury. [40] And proteinuria is a known surrogate for kidney disease.[41] The identification of proteinuria as an association with COVID-19 in the young patient cohort suggests the insult of COVID-19 to the kidneys persists months after the infection has resolved.

Primary open angle glaucoma was identified as an indicator of past COVID-19 infection among men under 65, 6-9 months post COVID-19 PCR test. There is limited previous data of this relationship in the literature.[42] Yan et al. published a case report of a patient with an acute glaucoma attack two months after a patient's COVID-19 infection. The nucleocapsid of coronavirus was found intracellularly in the ocular tissue at the time of surgery. Our study suggests that acute glaucoma may be an underrecognized, but important, sequela of COVID in the early period of PASC.

There are a number of phenotypes that have a relatively high statistical significance (a p-value between 0.01 and 0.001), but would have been dropped in univariate PheWAS after p-value correction. The MLHO framework helps identify which of these phenotypes are most likely valid associations. Two examples are such phenotypes are palpitations and malaise and fatigue, both of which have previously been described as common symptoms of PASC.[7,34,35]

Phenotypes identified as indications of a past COVID-19 infection with a confidence score between 80 and 96 were more sporadic in identification distribution among the sub-cohorts. In fact, other than a single exception (chronic fatigue syndrome), all these phenotypes belong to sub-cohort specific groups.

MLHO's implementation in this study is similar to the standard univariate PheWAS as both offer computational solutions for high throughput association mining from clinical data. However, a challenge in standard PheWAS is to find a sensible balance between adequately applying correction to P-values in order to reduce false discovery due to multiple testing and minimizing false negatives.[33] In comparison to the deterministic P-value-based metrics in standard/linear PheWAS, our approach iteratively applies joint mutual information, performs sparsity screening, and uses gradient boosting to characterize the post-acute sequelae of COVID-19. MLHO's computational algorithms avoid a flood of false positive discoveries, while offering a more probabilistic flexible criterion than the standard PheWAS. As a result, and along with inclusion of

COVID-negative patients, this study rules out some of the phenotype associations, which were previously identified through poorly controlled observational data, such as cutaneous eruptions outside of nail changes and alopecia.

We acknowledge that this study's findings may present limitations due to the use of only diagnosis codes, which can result in missing signs and symptoms that are in clinical notes and laboratory results. In addition, given the intensity of the pandemic and spread of misinformation, EHR data may represent confirmatory bias between providers and patients. Finally, we have excluded hospitalized COVID-19 patients. On the one hand, it would be difficult to match hospitalized Coronavirus patients during the COVID era with non-COVID hospitalized patients. On the other hand, the post-COVID syndrome can still be observed in patients who were never hospitalized.[12,43–47] Regardless, future PASC studies should include hospitalized patients.

Our understanding of COVID-19 and its chronic sequelae is evolving, and new risks are unknown. We do not know who might develop post-COVID syndrome, how long symptoms last, and whether COVID-19 prompts the presentation of chronic diseases. There is a unique opportunity today to understand the post-acute effects that can follow SARS-CoV-2 infection. The ever-increasing adoption and magnitude of clinical data stored in EHR repositories over the past decade provide exceptional opportunities for instrumenting healthcare systems to study evolving pandemic byproducts. Our approach avoids a flood of false positive discoveries, while offering a more probabilistic flexible criterion than the standard phenome-wide association study (PheWAS).

Conflict of Interest Disclosures: None.

Funding/Support: This work was supported by National Human Genome Research Institute grant 3U01HG008685-05S2 and the National Library of Medicine grant T15LM007092. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH nor Massachusetts General Hospital.

Additional Contributions: We thank many colleagues in the Mass General Brigham Research Information Science & Computing team, for curating MGB COVID-mart and providing information science and computing support.

References

1. Carfì A, Bernabei R, Landi F, Others. Gemelli against COVID-19 post-acute care study group. Persistent symptoms in patients after acute COVID-19. *JAMA*. 2020;324(6):603-605.
2. Xiong Q, Xu M, Li J, et al. Clinical sequelae of COVID-19 survivors in Wuhan, China: a single-centre longitudinal study. *Clin Microbiol Infect*. 2021;27(1):89-95.
3. Goërtz YMJ, Van Herck M, Delbressine JM, et al. Persistent symptoms 3 months after a SARS-CoV-2 infection: the post-COVID-19 syndrome? *ERJ Open Res*. 2020;6(4). doi:10.1183/23120541.00542-2020
4. Meini S, Suardi LR, Busoni M, Roberts AT, Fortini A. Olfactory and gustatory dysfunctions in 100 patients hospitalized for COVID-19: sex differences and recovery time in real-life. *Eur Arch Otorhinolaryngol*. 2020;277(12):3519-3523.
5. Halpin SJ, McIvor C, Whyatt G, et al. Postdischarge symptoms and rehabilitation needs in survivors of COVID-19 infection: A cross-sectional evaluation. *J Med Virol*. 2021;93(2):1013-1022.
6. Bowles KH, McDonald M, Barrón Y, Kennedy E, O'Connor M, Mikkelsen M. Surviving

COVID-19 After Hospital Discharge: Symptom, Functional, and Adverse Outcomes of Home Health Recipients. *Ann Intern Med*. Published online November 24, 2020. doi:10.7326/M20-5206
7. Nehme M, Braillard O, Alcoba G, et al. COVID-19 Symptoms: Longitudinal Evolution and Persistence in Outpatient Settings. *Ann Intern Med*. Published online December 8, 2020. doi:10.7326/M20-5926
8. CDC. Long-Term Effects of COVID-19. Published December 9, 2020. Accessed March 11, 2021. https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects.html
9. Hopkins C, Surda P, Whitehead E, Nirmal Kumar B. Early recovery following new onset anosmia during the COVID-19 pandemic – an observational cohort study. *Journal of Otolaryngology - Head & Neck Surgery*. 2020;49(1). doi:10.1186/s40463-020-00423-8
10. Wong AW, Shah AS, Johnston JC, Carlsten C, Ryerson CJ. Patient-reported outcome measures after COVID-19: a prospective cohort study. *Eur Respir J*. 2020;56(5). doi:10.1183/13993003.03276-2020
11. Carfì A, Bernabei R, Landi F, Gemelli Against COVID-19 Post-Acute Care Study Group. Persistent Symptoms in Patients After Acute COVID-19. *JAMA*. 2020;324(6):603-605.
12. Chopra V, Flanders SA, O'Malley M, Malani AN, Prescott HC. Sixty-Day Outcomes Among Patients Hospitalized With COVID-19. *Ann Intern Med*. Published online November 11, 2020. doi:10.7326/M20-5661
13. Huang C, Huang L, Wang Y, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet*. 2021;397(10270):220-232.
14. Garcia J, Gustavson AR. The Science of Self-Report. *APS Obs*. 1997;10(1). Accessed March 22, 2021. https://www.psychologicalscience.org/observer/the-science-of-self-report
15. Bakhoum MF, Ritter M, Garg AK, Chan AX, Bakhoum CY, Smith DM. Subclinical ocular inflammation in persons recovered from ambulatory COVID-19. *medRxiv*. Published online September 23, 2020:2020.09.22.20128140.
16. Puntmann VO, Carerj ML, Wieters I, et al. Outcomes of Cardiovascular Magnetic Resonance Imaging in Patients Recently Recovered From Coronavirus Disease 2019 (COVID-19). *JAMA Cardiol*. 2020;5(11):1265-1273.
17. Brito D, Meester S, Yanamala N, et al. High Prevalence of Pericardial Involvement in College Student Athletes Recovering From COVID-19. *JACC Cardiovasc Imaging*. 2021;14(3):541-555.
18. Dani M, Dirksen A, Taraborrelli P, et al. Autonomic dysfunction in "long COVID": rationale, physiology and management strategies. *Clin Med* . 2021;21(1):e63-e67.
19. Gregorova M, Morse D, Brignoli T, et al. Post-acute COVID-19 associated with evidence of bystander T-cell activation and a recurring antibiotic-resistant bacterial pneumonia. *Elife*. 2020;9. doi:10.7554/eLife.63430
20. Manckoundia P, Franon E. Is Persistent Thick Copious Mucus a Long-Term Symptom of COVID-19? *Eur J Case Rep Intern Med*. 2020;7(12):002145.
21. Lu Y, Li X, Geng D, et al. Cerebral Micro-Structural Changes in COVID-19 Patients - An MRI-based 3-month Follow-up Study. *EClinicalMedicine*. 2020;25:100484.
22. Raahimi MM, Kane A, Moore CE, Alareed AW. Late onset of Guillain-Barré syndrome following SARS-CoV-2 infection: part of "long COVID-19 syndrome"? *BMJ Case Rep*. 2021;14(1). doi:10.1136/bcr-2020-240178
23. Estiri H, Strasser ZH, Murphy SN. High-throughput phenotyping with temporal sequences. *J Am Med Inform Assoc*. Published online December 14, 2020. doi:10.1093/jamia/ocaa288
24. Estiri H, Vasey S, Murphy SN. Transitive sequential pattern mining for discrete clinical data. *International Conference on Artificial*. Published online 2020. https://link.springer.com/chapter/10.1007/978-3-030-59137-3_37
25. Estiri H, Strasser ZH, Klann JG, McCoy TH Jr., Wagholikar KB, Vasey S, Castro VM, Murphy ME, Murphy SN. Transitive Sequencing Medical Records for Mining Predictive and

Interpretable Temporal Representations. *Patterns*. Published online June 18, 2020.
26. Estiri H, Strasser ZH, Murphy SN. Individualized prediction of COVID-19 adverse outcomes with MLHO. *Sci Rep*. 2021;11(1):5322.
27. Estiri H, Strasser ZH, Klann JG, Naseri P, Wagholikar KB, Murphy SN. Predicting COVID-19 mortality with electronic medical records. *NPJ Digit Med*. 2021;4(1):15.
28. Denny JC, Ritchie MD, Basford MA, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*. 2010;26(9):1205-1210. doi:10.1093/bioinformatics/btq126
29. Denny JC, Bastarache L, Roden DM. Phenome-Wide Association Studies as a Tool to Advance Precision Medicine. *Annu Rev Genomics Hum Genet*. 2016;17:353-373.
30. Estiri H, Strasser ZH, Klann JG, Naseri P, Wagholikar KB, Murphy SN. Predicting COVID-19 Mortality with Electronic Medical Records. *npj Digital Medicine*. doi:10.1038/s41746-021-00383-x
31. Estiri H, Vasey S, Murphy SN. Transitive Sequential Pattern Mining for Discrete Clinical Data. In: *Artificial Intelligence in Medicine*. Springer International Publishing; 2020:414-424.
32. Estiri H, Vasey S, Murphy SN. Generative transfer learning for measuring plausibility of EHR diagnosis records. *J Am Med Inform Assoc*. Published online October 12, 2020. doi:10.1093/jamia/ocaa215
33. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*. 2003;100(16):9440-9445.
34. Nalbandian A, Sehgal K, Gupta A, et al. Post-acute COVID-19 syndrome. *Nat Med*. Published online March 22, 2021. doi:10.1038/s41591-021-01283-z
35. Petersen MS, Kristiansen MF, Hanusson KD, et al. Long COVID in the Faroe Islands - a longitudinal study among non-hospitalized patients. *Clin Infect Dis*. Published online November 30, 2020. doi:10.1093/cid/ciaa1792
36. Reddy SV, Kurihara N, Menaa C, et al. Osteoclasts Formed by Measles Virus-Infected Osteoclast Precursors from hCD46 Transgenic Mice Express Characteristics of Pagetic Osteoclasts. *Endocrinology*. 2001;142(7):2898-2905.
37. Kurihara N, Reddy SV, Menaa C, Anderson D, Roodman GD. Osteoclasts expressing the measles virus nucleocapsid gene display a pagetic phenotype. *J Clin Invest*. 2000;105(5):607-614.
38. Khan SA, Brennan P, Newman J, Gray RE, McCloskey EV, Kanis JA. Paget's disease of bone and unvaccinated dogs. *Bone*. 1996;19(1):47-50.
39. Alobaida S, Lam JM. Beau lines associated with COVID-19. *CMAJ*. 2020;192(36):E1040.
40. Rudnick MR, Hilburg R. Acute Kidney Injury in COVID-19: Another Challenge for Nephrology. *Am J Nephrol*. 2020;51(10):761-763.
41. Cravedi P, Remuzzi G. Pathophysiology of proteinuria and its value as an outcome measure in chronic kidney disease. *Br J Clin Pharmacol*. 2013;76(4):516-523.
42. Yan Y, Diao B, Liu Y, Zhang W, Wang G, Chen X. Severe Acute Respiratory Syndrome Coronavirus 2 Nucleocapsid Protein in the Ocular Tissues of a Patient Previously Infected With Coronavirus Disease 2019. *JAMA Ophthalmol*. Published online October 8, 2020. doi:10.1001/jamaophthalmol.2020.3962
43. Wijeratne T, Crewther S. Post-COVID 19 Neurological Syndrome (PCNS); a novel syndrome with challenges for the global neurology community. *J Neurol Sci*. 2020;419:117179.
44. Garg P, Arora U, Kumar A, Wig N. The "post-COVID" syndrome: How deep is the damage? *J Med Virol*. Published online 2020. https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.26465
45. Marshall M. The lasting misery of coronavirus long-haulers. *Nature*. 2020;585(7825):339-341.
46. Rubin R. As Their Numbers Grow, COVID-19 "Long Haulers" Stump Experts. *JAMA*.

2020;324(14):1381-1383.
47. Sudre CH, Murray B, Varsavsky T, et al. Attributes and predictors of Long-COVID: analysis of COVID cases and their symptoms collected by the Covid Symptoms Study App. *bioRxiv*. Published online October 21, 2020. doi:10.1101/2020.10.19.20214494