It Takes a Village: Community-Driven Phenotyping to Address a Public Health Crisis

Kristin Kostka[1,2], Evan Minty[3], Antonella Delmestri[1], Barrack Omondi[1], Martí Català[1], Elena Roel[5], Edward Burn[1,5], Daniel Prieto-Alhambra[1,4], Annika M. Jödicke[1]

[1]Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, OX3 7LD, UK

[2]The OHDSI Center at the Roux Institute, Northeastern University, Portland, ME, USA

[3]O''Brien Institute for Population Health, Faculty of Medicine, University of Calgary, Canada

[4]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands

[5]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain

Background

In Fall 2021, our study team received a grant to evaluate the impact of COVID-19 vaccination on the prevention of long COVID using UK primary care data. "Post acute COVID-19 syndrome" or "long COVID" produces persistent symptoms that continue for weeks or months following the acute COVID-19 disease. As the pandemic continues, long COVID poses a significant public health issue with potential to inflict mass disability.[(1)] Describing a newly emerging condition presents a significant challenge. Clinicians have varying opinions on symptoms associated with long COVID, creating inconsistency in defining and measuring this issue at scale.[(2–4)]

To start, we consulted with the OHDSI Community Phenotype Development & Evaluation Workgroup – a resource for investigators navigating real world data and study design providing best practices for improving the reproducibility and quality of cohort definitions.[(5)]

Together, we aimed to use OHDSI best practices for operationalising a long COVID definition and test the community methodology with UK OMOP CDM-mapped primary care data.

Methods

Following OHDSI best practice to begin from a clinical case definition[(6)], we conducted a systematic review of over 6500 original studies to summarise symptoms and diagnoses characterising long COVID based on existing literature (as of September 2021).[(7)] In October 2021, the World Health Organization (WHO) issued a Delphi consensus of the clinical case definition of post COVID-19 condition.[(8)] We opted to use the WHO definition to begin the OHDSI phenotyping process[(6)] while completing the full systematic review.

We joined the "Long COVID" subgroup meetings to solicit OHDSI collaborators input. We hosted an intensive 1-day hackathon on December 7, 2021, to take the 25 symptoms recognized by the WHO definition through the OHDSI phenotype workflow. Figure 1 is a screenshot of the inventory of Phenotype Library assets and other OHDSI network study artifacts.

Symptoms	Description	Synonyms	Status	Source	ATLAS	JSON	Concepts to consider
Abdominal pain	Abdominal pain condition record of any type; successive records with > 90 day gap are considered independent episodes	stomach pain**		LEGEND	N/A	https://github.com/OHDSI/Legend/blob/master/inst/cohorts/Abdominal%20pain.json	
Menstrual and period problems	Regularity of menstrual cycle observation record of any type; pattern of menstrual cycle condition record of any type; *see "Notes" Tab **TO BE CONFIRMED WITH AZZA**						https://athena.ohdsi.org/search-terms/terms/4265552 https://athena.ohdsi.org/search-terms/terms/4267967
Altered smell/taste	Altered smell/taste condition record of any type with a fixed cohort end date (3 days after cohort start date). Events will be combined into cohort eras if they are within 30 days of each other.	Anosmia OR Hyposmia OR Dysgeusia OR Ageusia, smell disturbance**, taste disturbance**		Phenotype Library		https://raw.githubusercontent.com/OHDSI/PhenotypeLibrary/develop/inst/cohorts/8.json	
Anxiety	The first condition record of anxiety, which is followed by another anxiety condition record or a drug used to treat anxiety			LEGEND	N/A	https://github.com/OHDSI/Legend/blob/master/inst/cohorts/Anxiety.json	
Blurred vision	Blurred vision condition record of any type	visual disturbance**				https://athena.ohdsi.org/search-terms/terms/44790749	
Chest pain	The first condition record of chest pain or angina	Angina		LEGEND	N/A	https://github.com/OHDSI/Legend/blob/master/inst/cohorts/Chest%20pain%20or%20angina.json	
Cognitive dysfunction/brain fog	[TO BE WRITTEN] **TO BE CONFIRMED WITH AZZA**	Impaired cognition				https://athena.ohdsi.org/search-terms/terms/45539345 https://athena.ohdsi.org/search-terms/terms/443432	
Cough	Cough condition record of any type with a fixed cohort end date (3 days after cohort start date). Events will be combined into cohort eras if they are within 30 days of each other.			Phenotype Library	N/A	https://github.com/OHDSI/PhenotypeLibrary/blob/develop/inst/cohorts/3.json	
Depression	The first condition record of depression, which is followed by another depression condition record, at least two drugs used to treat depression without another indication, or two psychotherapy procedures	low mood**		LEGEND	N/A	https://github.com/OHDSI/Legend/blob/master/inst/cohorts/Depression.json	
Dizziness	[TO BE WRITTEN]	Vertigo, light-headedness				https://athena.ohdsi.org/search-terms/terms/4223938 https://athena.ohdsi.org/search-terms/terms/439383	
Fatigue	**TO BE CONFIRMED WITH AZZA** Malaise or fatigue condition record of any type; permits the use of condition source concepts	Malaise, tiredness, lack of energy, lethargy, chronic fatigue syndrome (CFS)/MS*		Phenotype Library		https://github.c	https://athena.ohdsi.org/search-terms/terms/4074624
Intermittent fever	Any of the following: A Fever (38.0°C or higher) pre-coodinated condition record of any type OR a measurement of Fever (38.0°C or higher) record using a pre-coordinated term (where value is inclusive to code used) OR a measurement of Fever (38.0°C or higher) record between 38 and 42 degrees in Celsius OR a measurement of Fever (38.0°C or higher) record between 100.4 and 120 (inclusive) degrees in Fahrenheit OR an observation of Fever (38.0°C or higher) record between 38 and 42 (inclusive) degrees in Celsius OR an observation of Fever (38.0°C or higher) record between 100.4 and 120 (inclusive) degrees in Fahrenheit with continuous observation of at least 0 days prior and 0 days after event index date, and limit initial events to: all events per person.			Phenotype Library		https://raw.githubusercontent.com/OHDSI/PhenotypeLibrary/develop/inst/cohorts/6.json	
Gastrointestinal diseases (diarrohea, constipation, acid reflux)	Diarrhea condition record of any type; successive records with > 30 day gap are considered independent episodes OR Constipation condition record of any type; OR Acid reflux condition reecord of any type (Should restrict for IBD -- not sure if we should include IBS) **TO BE CONFIRMED WITH AZZA**	GERD - gastroesphageal reflux disorder		LEGEND	N/A	https://github.c	https://athena.ohdsi.org/search-terms/terms/4178369 https://athena.ohdsi.org/search-terms/terms/44783954 https://athena.ohdsi.org/search-terms/terms/46271381
Headache	Headache condition record of any type; successive records with > 30 day gap are considered independent episodes			Phenotype Library		https://raw.githubusercontent.com/OHDSI/PhenotypeLibrary/develop/inst/cohorts/7.json	
Memory issues	[TO BE WRITTEN]	Memory loss, memory impairment, impaired cognition*				https://athena.ohdsi.org/search-terms/terms/4304008 https://athena.ohdsi.org/search-terms/terms/77074	
Joint pain	Joint pain condition record of any type	Musculoskeletal pain*, Arthralgia**				https://athena.ohdsi.org/search-terms/terms/4150129	
Muscle pain/spams	[TO BE WRITTEN]	Spasticity, musculoskeletal pain*, Myalgia**				https://athena.ohdsi.org/search-terms/terms/4329728 https://athena.ohdsi.org/search-terms/terms/442752	
Neuralgia	[TO BE WRITTEN]						
New onset allergies	The first condition record of allergic condition or allergic disorder, followed by another allergic condition or allergic disorder condition record or drug used to treat allergic conditions	Allergic condition, allergic disorder, allergic process				https://athena.ohdsi.org/search-terms/terms/36683564 https://athena.ohdsi.org/search-terms/terms/43021170	
Pins and needles sensations	Pins and needles condition or observation record of any type	Paresthesia, tingling				https://athena.ohdsi.org/search-terms/terms/4266682	
Post-exertional malaise	Post-extertional malaise condition record of any type	Tired on least exertion				https://athena.ohdsi.org/search-terms/terms/40484614	

Figure 1. Inventory of OHDSI Community Assets

We concurrently divided the code list generation of individual symptoms amongst attendees (Figure 2).

Figure 2. Workflow for Phenotype Generation of Individual Symptoms

Each concept set expression was iteratively assessed through PHOEBE[9], PheValuator[10] and review of the literature.

Using CohortDiagnostics[11], we conducted cohort review on a large database of UK primary care electronic health records, Clinical Practice Research Datalink (CPRD) Aurum mapped to OMOP CDM V5.3. The study period started on 1 January 2020 and ended at the last available date (11 Mar 2021). We generated 125 symptom cohorts. To enter any cohort, persons were required to be over 18 years of age, have a qualifying COVID diagnosis or positive PCR test and at least 180 days of prior observation time. *(Note: Acute COVID entry criteria were reused from prior Oxford research by Burn et al.)* Additional inclusion criteria consisted of no history of the specific symptom prior to index (- 90 days, -180 days) and a time window of symptom persistence (+28 days, +90 days after diagnosis or test). In a subset of symptoms, we explored the use of a run-in time window (-7 days, -14 days) where symptoms may present prior to clinical confirmation of acute COVID-19. After reviewing the individual symptoms phenotypes, a composite long COVID phenotype was assembled.

Results

The initial hackathon produced 7 clinical symptom concept set expressions meeting the OHDSI best practices of “done”, 9 drafted clinical symptom concept set expressions for further review with OHDSI diagnostics and 9 clinical symptom concept set expressions to be developed. 1 WHO symptom (post-exertional malaise/fatigue) was dropped from the concept set process due to

insufficient use of concepts in primary care data. The 18 concept sets were run through PHOEBE and reviewed by clinical input. Iterative results are stored on OHDSI MS Teams.

In subsequent CohortDiagnostics[11] review, a total of 458,975 persons with COVID-19 diagnosis or a positive test met the cohort entry criteria (C124 in Figure 3). The most common persistent symptoms included shortness of breath (n=4005; C45 in Figure 3), anxiety (n=3378; C6 in Figure 3), joint pain (n=3340; C14 in Figure 3), cough (n=3275; C32 in Figure 3), and abdominal pain (n=2651; C1 in Figure 3). Cohort counts were impacted by prior history, symptom persistence, and run-in windows.

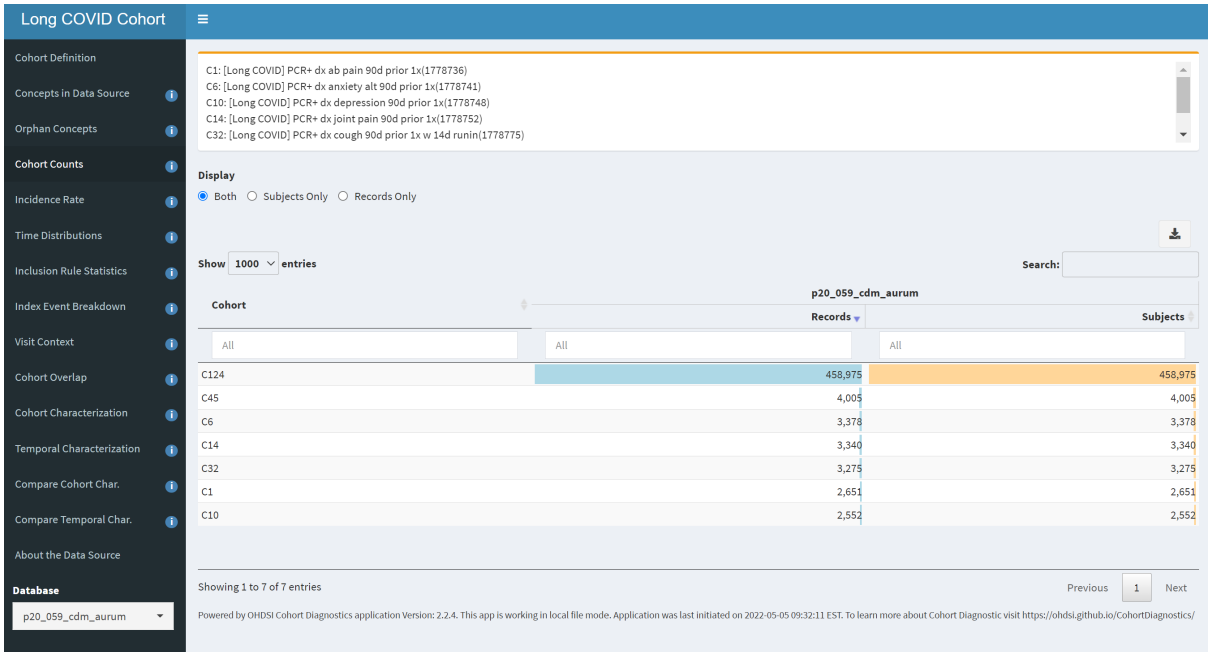

Figure 3. CohortDiagnostics Shiny Application for Study

The full public RShiny application (as shown in Figure 3) using the CohortDiagnostics framework is stored online (https://tinyurl.com/OHDSI-longCOVID). *(DISCLAIMER: Any use of this Shiny application is done in concert with ISAC approval and the involvement of the study team.)* A network study package will be made available via GitHub (https://github.com/oxford-pharmacoepi/NIHR-LongCOVID).

Conclusion

We worked to implement the OHDSI Community best practices for developing and evaluating study phenotypes for long COVID. We found gaps in the OHDSI Phenotype Library and worked to develop new definitions for these clinical symptoms.

Given strong heterogeneity in COVID-19 practice patterns in diagnosing and treating post acute COVID-19, we avoided creating a universal phenotype that could be applied to any OMOP CDM. Our work is intended for use in primary care databases. We are partnering with EHDEN Data Network to continue to validate this phenotype in other primary care databases (such as The Information System for Research in Primary Care (SIDIAP) database from Catalonia, Spain).

Our learnings provide an interesting case study in the application of OHDSI best practices at scale. The hackathon significantly strained the OHDSI Community infrastructure, including crashing PHOEBE due to the size of concept set expressions. We adopted practices to work around these system limitations and utilised other internally developed study team tools, such as an alternative

solution to PHOEBE for concept set creation (OmopConceptSetLibrary), to experiment with augmenting community best practices. Results differed across these tools suggesting a blended approach may be most effective. As we continue this effort into a full OHDSI network study, we intend to share our learnings with the community at large for the betterment of all phenotyping. It is our hope this project can provide guidance to future efforts to phenotype complex medical conditions and create resources for other public health researchers.

References

1. U.S. Government Accountability Office. Science & Tech Spotlight: Long COVID [Internet]. [cited 2022 Jun 16]. Available from: https://www.gao.gov/products/gao-22-105666
2. Greenhalgh T, Knight M, A'Court C, Buxton M, Husain L. Management of post-acute covid-19 in primary care. BMJ [Internet]. 2020 Aug 11;370:m3026. Available from: http://dx.doi.org/10.1136/bmj.m3026
3. CDC. Post-COVID Conditions [Internet]. Centers for Disease Control and Prevention. 2022 [cited 2022 May 4]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Flong-term-effects.html
4. Nalbandian A, Sehgal K, Gupta A, Madhavan MV, McGroder C, Stevens JS, et al. Post-acute COVID-19 syndrome. Nat Med [Internet]. 2021 Apr;27(4):601–15. Available from: http://dx.doi.org/10.1038/s41591-021-01283-z
5. Rao G. Phenotype Development and Evaluation Workgroup OHDSI Community Call Update [Internet]. OHDSI Community Call; 2021 Oct 5; Virtual. Available from: https://www.ohdsi.org/wp-content/uploads/2021/10/Phenotype-Development-and-Evaluation-Workgroup-Read-Only.pdf
6. Rao G. The Science of Phenotyping: Framework for Generating Content to the OHDSI Phenotype Library. Janssen R&D; 2020.
7. Annika Jödicke, Kristin Kostka, Victoria Strauss, Daniel Prieto-Alhambra. An updated systematic review to characterise Long COVID symptoms as recorded in real world data vs. self-reported data [Internet]. PROSPERO International prospective register of systematic reviews. 2021 [cited 2022 May 4]. Available from: https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=292122
8. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV, WHO Clinical Case Definition Working Group on Post-COVID-19 Condition. A clinical case definition of post-COVID-19 condition by a Delphi consensus. Lancet Infect Dis [Internet]. 2022 Apr;22(4):e102–7. Available from: http://dx.doi.org/10.1016/S1473-3099(21)00703-9
9. Ostropolets A, Hripcsak G. PHOEBE: PHenotype Observed Entity Baseline Endorsements [Internet]. OHDSI Application Library. 2020 [cited 2021 Mar 12]. Available from: https://data.ohdsi.org/PHOEBE/
10. Swerdel JN, Hripcsak G, Ryan PB. PheValuator: Development and evaluation of a phenotype algorithm evaluator. J Biomed Inform [Internet]. 2019 Sep;97:103258. Available from: http://dx.doi.org/10.1016/j.jbi.2019.103258
11. Rao G, Schuemie M, Ryan P, Weaver J, Gilbert J. CohortDiagnostics: Diagnostics for OHDSI Cohorts [Internet]. 2022 [cited 2022 May 5]. Available from: https://ohdsi.github.io/CohortDiagnostics/