

Implementing the OHDSI Community Approach to Phenotype a Complex Medical Condition in European Primary Care Data

Authors: Kristin Kostka^{1,2}, Evan Minty³, Antonella Delmestri¹, Barrack Omondi¹, Martí Català¹, Edward Burn^{1,5}, Daniel Prieto-Alhambra^{1,4}, Annika M. Jödicke¹

Affiliations: ¹Pharmaco- and Device Epidemiology, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, OX3 7LD, UK; ²The OHDSI Center at the Roux Institute, Northeastern University, Portland, ME, US; ³O'Brien Centre for Population Health, Faculty of Medicine, University of Calgary, CA; ⁴Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, NL; ⁵Fundació Institut Universitari per a la recerca a l'Atenció Primària, Barcelona, ES

Introduction

Background: “Post-acute COVID-19 syndrome” or “long COVID” are persistent symptoms that continue for weeks or months following the acute COVID-19 disease. As the COVID-19 pandemic continues, long COVID poses a significant public health issue with potential to inflict mass disability [1]. Clinicians have varying familiarity in the characteristic symptoms associated with long COVID, creating challenges in defining and measuring this issue at scale.



Figure 1. A sampling of media coverage of long COVID from Spring 2022

Objective: To follow OHDSI best practices for developing a long COVID phenotype and apply them to UK OMOP CDM-mapped primary care data.

Methods

Engaging the Community: We partnered with the OHDSI Phenotype Development & Evaluation Workgroup to run a Long COVID phenotyping hackathon on December 7, 2021. In the hackathon, we used the World Health Organization (WHO) Delphi consensus of the clinical case definition of post COVID-19 condition [2]. We assembled concept sets for the 25 individual symptoms using a consistent process (Figure 2). Each concept set expression was inspected through use of PHOEBE⁴, PheValuator⁵, and available literature.

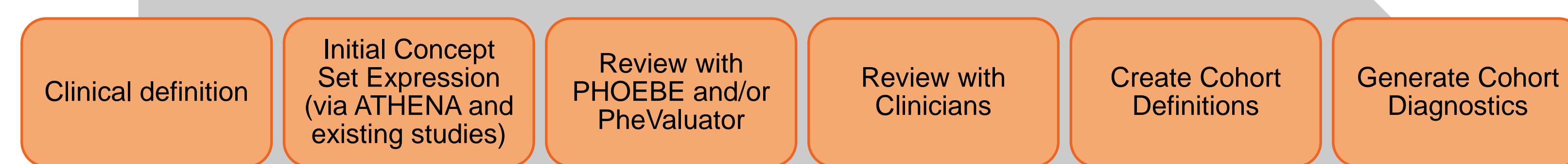


Figure 2. Workflow for Phenotype Generation of Individual Symptoms

Tapping into Cohort Diagnostics: We ran an initial CohortDiagnostics[3] package 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 then used the symptom code lists to iteratively constructed cohort definition parameters to generate 125 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 initial clinical review of the individual phenotypes, a composite long COVID phenotype was assembled.

Results

- The 1-day community hackathon produced:
 - 7 final clinical symptom** concept set expressions meeting the OHDSI best practices
 - 9 drafted clinical symptom** concept set expression for further review with OHDSI diagnostics
 - 9 clinical symptom** concept set expressions to be developed.(... and Gowtham broke PHOEBE. 😊)
- 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 later run through PHOEBE and reviewed by clinical input. Iterative results are stored in the OHDSI Phenotype WG Long COVID channel on OHDSI MS Teams.
- In the CohortDiagnostics review, a total of **458,975 persons** with COVID-19 diagnosis or a positive test met the cohort entry criteria (C124).
- The most common persistent symptoms included **shortness of breath (n=4005; C45)**, **anxiety (n=3378; C6)**, **joint pain (n=3340; C14)**, **cough (n=3275; C32)**, **abdominal pain (n=2651; C1)** and **depression (n=2552; C10)**.
- Cohort counts were impacted by prior history, symptom persistence, and run-in windows.

Cohort	Records	Subjects
C124	458,975	458,975
C45	4,005	4,005
C6	3,378	3,378
C14	3,340	3,340
C32	3,275	3,275
C1	2,651	2,651
C10	2,552	2,552

Figure 3. CohortDiagnostics Shiny Application for Study



Let's collaborate!
Visit our study repo

Conclusion

The OHDSI community approach to phenotyping **provides a robust framework to evaluating a complex medical condition**, such as long COVID. We observed differences in cohorts based on logic changes in prior follow-up time, time for symptom persistence, gender, and age. Our findings can help researchers **understand the impact of fluctuating clinical logic** on describing and measuring long COVID at scale.

Reference: [1] U.S. Government Accountability Office. Science & Tech Spotlight: Long COVID. [cited 16 Jun 2022]. Available: <https://www.gao.gov/products/gao-22-105666>. [2] Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV, WHO Clinical Case Definition Working Group on Post-COVID-19 Condition. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis.* 2022;22:e102–e107. doi:10.1016/S1473-3099(21)00703-9. [3] Rao G, Schuemie M, Ryan P, Weaver J, Gilbert J. CohortDiagnostics: Diagnostics for OHDSI Cohorts. 2022 [cited 5 May 2022]. Available: <https://ohdsi.github.io/CohortDiagnostics/>