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

Kristin Kostka¹,², Evan Minty³, Antonella Delmestri¹, Barrack Omondi¹, Martí Català¹, Elena Roel⁵, Edward Burn¹,⁵, Daniel Prieto-Alhambra¹,⁴, Annika M. Jödicke¹

¹Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, OX3 7LD, UK
²The OHDSI Center at the Roux Institute, Northeastern University, Portland, ME, USA
³O’Brien Institute for Population Health, Faculty of Medicine, University of Calgary, Canada
⁴Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands
⁵Fundació 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.
We concurrently divided the code list generation of individual symptoms amongst attendees (Figure 2).

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.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shortness of breath</td>
<td>4005</td>
</tr>
<tr>
<td>Anxiety</td>
<td>3378</td>
</tr>
<tr>
<td>Joint pain</td>
<td>3340</td>
</tr>
<tr>
<td>Cough</td>
<td>3275</td>
</tr>
<tr>
<td>Abdominal pain</td>
<td>2651</td>
</tr>
</tbody>
</table>

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](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](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


