Pattern of Long COVID Symptoms and Conditions: Clustering Analysis Based on Large Multinational Cohorts from an EHDEN Study-A-Thon


Abstract

We aimed to categorize and validate clusters of long COVID symptoms, as defined by WHO’s “post-COVID-19 condition”, through a broad international study. Results revealed numerous single-symptom clusters. However, when analyzing patients with multiple concurrent symptoms, we identified repeated clusters (anxiety-depression, cough-dyspnea, gastrointestinal-abdominal pain) across various databases. Still, considerable variation was noted among different databases and healthcare settings.

Introduction and Methods

Background

Long COVID, a condition with wide-ranging and heterogeneous symptoms, remains vaguely defined and difficult to diagnose. Clustering algorithms could potentially classify patients based on symptom combinations, aiding in understanding the disease’s mechanisms and requirements.

Objectives

Our goal was to identify and confirm clusters of conditions and symptoms among patients with long COVID, using the WHO’s clinical definition of “post-COVID-19 condition” in an extensive international network study.

Data

We conducted a multinational network study including databases from different healthcare settings: primary care, hospital databases and a claims-based database.

Long COVID definition

We studied those with COVID-19 (positive SARS-CoV-2 test or diagnosis) and no prior diagnosis/test within 42 days, then identified any post-acute symptoms from WHO’s list of 25. A symptom was classified as long COVID if it appeared 90-365 days post-infection and wasn’t in the 180-day pre-infection history.

Statistical analysis

We performed latent class analysis on these symptoms, considering patient’s age and sex. We characterized clusters, analyzing patient baseline characteristics, healthcare use, vaccination status, and co-morbidities. We calculated the analyses for 2-7 clusters, choosing the optimal number based on statistical and clinical relevance.

Post-hoc analyses included clustering among subgroups with at least 2 or 3 different long COVID symptoms, conducted only in databases with over 500 individuals in the respective subgroup.

We performed all the analyses in R v4.2.1. Code available in https://github.com/oxford-pharmacoepi/LongCovidStudyathon_W3.

Results

Using consistent clustering algorithms, symptom lists, and data structures across databases, we found diverse symptom combinations among long COVID patients. The prevalence of single-symptom clusters may reflect common symptoms recorded post-COVID-19, or persistent complications like olfactory deficits.

Analyzing subsets with 2+ or 3+ symptoms, clusters were more reflective of long COVID patient characteristics in specialist clinics, albeit results varied across databases and healthcare settings.

Conclusions

For more information, please email kim.lopez@spc.ox.ac.uk!