Study identifies four major subtypes of long COVID

January 4 2023, by Jim Schnabel

Credit: Unsplash/CC0 Public Domain

The post-COVID syndrome known as long COVID has four major subtypes defined by different clusters of symptoms, according to a study led by researchers at Weill Cornell Medicine.

The study, published Dec. 1 in *Nature Medicine*, was the largest of its kind to examine long COVID. The researchers, who represent clinicians and informaticists, used a machine-learning algorithm to spot symptom patterns in the health records of nearly 35,000 U.S. patients who tested positive for SARS-CoV-2 infection and later developed lingering long-COVID-type symptoms.

Funded by the National Institutes of Health's Researching COVID to Enhance Recovery (RECOVER) initiative, this research is part of a one-year $9.8 million grant focusing on electronic health record cohort studies, spearheaded by principal investigator Dr. Rainu Kaushal, senior associate dean for clinical research and chair of the Department of Population Health Sciences at Weill Cornell Medicine.

"RECOVER is aiming to rapidly elucidate what is happening in long COVID," said Kaushal, co-senior author on the study. "Looking at how cases cluster can profoundly impact the prognosis and care of patients."

Of the four major patterns detected, one featured heart and kidney problems, and included a relatively high proportion of patients infected in the first few months of the pandemic in the U.S. Another pattern included respiratory problems, anxiety, sleep disorders and other symptoms including headache and chest pain; nearly two-thirds of the patients with this pattern were women.

"These results should inform ongoing research on the potential mechanisms of long COVID, and potential treatments for it," said Dr. Fei Wang, an associate professor of population health sciences, who led the study.

Viral infections sometimes leave patients with a variety of lingering, often nonspecific symptoms. For SARS-CoV-2, these post-infection syndromes are known popularly as "long COVID," and more formally as

"post-acute SARS-CoV-2 infection" (PASC). They appear to be very common; estimates of the number of Americans who have had long COVID run as high as 40% of the U.S. adult population.

"Understanding the epidemiology of long COVID allows clinicians to help patients understand their symptoms and prognoses and facilitates multispecialty treatment for patients," said Kaushal, who is also the Nanette Laitman Distinguished Professor of Population Health Sciences and physician-in-chief of population health sciences at NewYork-Presbyterian/Weill Cornell Medical Center. "Electronic health records offer a window into this condition, allowing us to better characterize long COVID symptoms, informing other types of research including foundational discoveries and clinical trials."

The health records analyzed for the study came from two large datasets amassed by the National Patient-Centered Clinical Research Network (PCORnet), which comprises eight consortiums of health care institutions from around the country. One dataset, from the INSIGHT Clinical Research Network—which Kaushal leads—included data from New York-based patients, while the other came from the OneFlorida+ network, which includes patients from Florida, Georgia and Alabama. In all, the analysis covered health records of 34,605 different patients from March 2020 to November 2021—up to but not including the first omicron wave.

Initially analyzing the New York patient dataset, the machine learning algorithm detected four major symptom patterns. The first, which accounted for about 34% of patients, was dominated by heart, kidney and circulation-related symptoms. Patients in this group, compared with those in other groups, were older on average (median age 65), more likely to be male (49%), had a relatively high rate of COVID hospitalization (61%) and had relatively more pre-existing conditions. This group also had the highest proportion (37%) of patients who were

sickened by SARS-CoV-2 during the first big U.S. wave from March to June 2020.

The second symptom pattern, comparable in frequency (33% of patients) to the first, was dominated by respiratory and sleep problems, anxiety, headache and chest pains. Patients with this pattern were mostly female (63%), with a median age of 51 years and a much lower rate (31%) of COVID hospitalization. Almost two-thirds of the patients in this group tested positive for SARS-CoV-2 during later waves, from November 2020 to November 2021. Pre-existing conditions in this cluster centered on respiratory problems such as chronic obstructive pulmonary disorder and asthma.

The other two symptom patterns were dominated, respectively, by musculoskeletal and nervous system symptoms including arthritis (23% of patients), and by a combination of digestive and respiratory symptoms (10%).

Only in the first symptom pattern was the sex ratio roughly 1-to-1; in the other three, female patients made up a significant majority (more than 60%).

"This sex difference in long-COVID risk is consistent with prior research, but so far very few studies have even tried to uncover the mechanisms underlying it," Wang said.

To validate their findings, the researchers applied their algorithm to the dataset covering patients from the three southern states, and found very similar results. The analysis also supported the overall validity of long COVID by showing that, for patients who tested negative for SARS-CoV-2, symptoms appearing in the same 30-to-180-day time interval after the test did not have such clear patterns.

The researchers currently are following up with research along several lines, including the definition of long COVID symptom patterns so that they can be identified easily from electronic health records, the identification of risk factors for different symptom patterns, and the identification of existing treatments that can be repurposed to help long COVID patients.

More information: Hao Zhang et al, Data-driven identification of post-acute SARS-CoV-2 infection subphenotypes, *Nature Medicine* (2022). DOI: 10.1038/s41591-022-02116-3

Provided by Cornell University