

Using a Computer Program to Identify Who Has Long COVID

Research summary published
May 2023

This research at a glance

Main Questions

What patterns are more common in people who have Long COVID?

Could the computer program use these patterns to correctly identify people with Long COVID?

Answers

The computer program found that people who have Long COVID have patterns in their health care visits, age groups, symptoms and health conditions, and the medicines they take.

Yes, the computer program could correctly identify more than 8 out of 10 people with Long COVID.

The researchers concluded their computer program could be used to identify people with possible Long COVID. This could help connect people with Long COVID to health care or invite them to join research studies.



An Initiative Funded by the National Institutes of Health

Contents

Why was this research needed? _____	Page 2
What happened during this research? _____	Page 3
What were the main results? _____	Page 4
How has this research helped? _____	Page 5
Where can I learn more about this research? _____	Page 5
About RECOVER _____	Page 5



Why was this research needed?

Researchers are looking for better ways to identify people with Long COVID. Long COVID is a condition in which a person is sick many months after getting COVID. It can cause a range of new, repeating, or ongoing symptoms, such as trouble breathing or cough, and health problems, such as with the heart. Because Long COVID is a newer health condition with a range of symptoms and health problems, it can be hard to identify people with Long COVID. This means people with Long COVID may not get the health care they need.

In this study, researchers wanted to learn if they could create and use a computer program to identify people with Long COVID based on their health records.

The goal of this research was to answer these questions:

- What patterns are more common in people who have Long COVID?
- Could the computer program use these patterns to correctly identify people with Long COVID?



What happened during this research?

Researchers used data in electronic health records (EHRs) from 65 hospitals that were a part of the National COVID Cohort Collaborative (N3C). EHRs are digital versions of patients' medical records.

- 1 The researchers created a computer program to compare 2 groups of people:
 - **People with Long COVID:** People who visited a Long COVID clinic
 - **People who didn't have Long COVID:** People who did not visit a Long COVID clinic

The researchers used the computer program to compare the 2 groups for differences in patterns, such as patterns in their:

- Symptoms
- Health conditions
- Medicines
- Doctor visits and hospital stays
- Other data, such as their age

- 2 The researchers used the computer program to look at EHRs from a larger group of people to identify people who may have Long COVID based on people who match the patterns.

Whose data was included in this research?

Researchers used data from EHRs from in the United States. Researchers started with data from about 1,800,000 adults who:



About
1,800,000
Adults



Were ages 18 years or older



Had COVID more than 3 months ago, based on a positive COVID test result or being diagnosed while in the hospital



What were the main results?

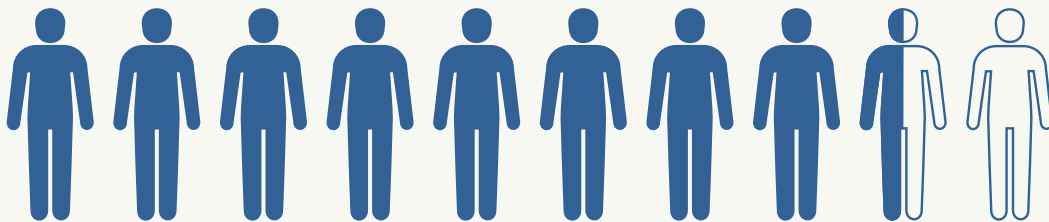
What patterns are more common in people who have Long COVID?

The computer program found these patterns are more common in people who have Long COVID, compared to people who didn't have Long COVID:

- Having more doctor visits and hospital stays after having COVID
- Being in certain age groups
- Having shortness of breath (trouble breathing) after having COVID
- Having other health conditions, such as diabetes or kidney disease
- Using certain medicines after having COVID, such as medicines that help with breathing

Could the computer program use these patterns to correctly identify people with Long COVID?

Yes, the computer program could correctly identify **more than 8 out of every 10 people with Long Covid.**



The researchers learned this by looking at the number of people who:

- **Had** Long COVID that the computer program **correctly identified** as having Long COVID
- **Didn't have** Long COVID that the computer **incorrectly identified** as having Long COVID



How has this research helped?

The computer program found patterns in the EHRs of people who have Long COVID, including:

- Age
- Prescribed medicines after having COVID
- Shortness of breath after having COVID
- How often they needed health care after having COVID

The computer program was able to correctly identify more than 8 out of every 10 people with Long COVID using EHRs. The computer program could be used to identify people with possible Long COVID and connect them to health care or invite them to join research studies.



Where can I learn more about this research?

Read the full paper published in *The Lancet Digit Health* on May 16, 2022 at [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(22\)00048-6/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(22)00048-6/fulltext)

Full paper title: Identifying who has long COVID in the USA: a machine learning approach using N3C data

Authors: Emily R. Pfaff, Andrew T. Girvin, Tellen D. Bennett, Abhishek Bhatia, Ian M. Brooks, Rachel R. Deer, Jonathan P. Dekermanjian, Sarah Elizabeth Jolley, Michael G. Kahn, Kristin Kostka, Julie A. McMurry, Richard Moffitt, Anita Walden, Christopher G. Chute, Melissa A. Haendel, The N3C Consortium

About RECOVER

RECOVER is a research project that aims to learn about the long-term health effects of COVID. We're hopeful that this project will help us better prevent and treat Long COVID in the future. RECOVER is paid for by the National Institutes of Health (NIH).

Learn more at: RECOVERcovid.org