Big data and long COVID

PHOSP-COVID study collaborative group

DOI:
10.1016/S2589-7500(22)00113-3

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
PHOSP-COVID study collaborative group 2022, 'Big data and long COVID', *The Lancet Digital Health*, vol. 4, no. 7, pp. e477. https://doi.org/10.1016/S2589-7500(22)00113-3

Big data and long COVID

Andrii Vodolazhskyi/shutterstock

In this issue, Emily Pfaff and colleagues show that machine learning analysis of electronic health records could be crucial in diagnosing patients with long COVID. This is the latest in a plethora of studies to use big data to determine prevalence, symptoms, or risk factors for long COVID. But is big data helping to treat patients with this heterogeneous condition, or should we be redirecting our efforts elsewhere?

Big data studies have been essential in gathering crucial information that could help reduce the prevalence of long COVID. Of note, Pfaff and colleagues report that COVID-19 vaccination after acute disease decreased the likelihood of patients being identified as having persistent symptoms after acute COVID-19, a result confirmed by a community-based cohort study in the UK. Using a deidentified electronic health record dataset from more than 90 000 patients with confirmed COVID-19 in the USA to train the algorithm, Pfaff and colleagues identified about 100 000 potential long COVID cases in the National COVID Cohort Collaborative database (representing nearly 5 million COVID-19-positive cases) as of October, 2021. This estimation of the prevalence of long COVID cases in the USA differs from other predictions and might present the most accurate to date.

Big data studies have also helped glean insight into the characteristics of patients with long COVID. Carole Sudre and colleagues analysed data from 1653 users of the COVID Symptom Study application to identify six distinct groupings of symptoms reported by patients with severe COVID. The data suggested that long COVID was about twice as common in women and that the average age of someone presenting with long COVID was about 4 years older than those without the condition. *The Lancet Respiratory Medicine* published the latest findings from the post-hospitalisation COVID-19 observational study, which aimed to recruit 10 000 people hospitalised with COVID-19 for long-term follow-up. The findings suggest that systematic inflammation and obesity are potentially treatable traits that warrant investigation.

Pfaff and colleagues reported long-awaited results from the highly anticipated National Institutes of Health RECOVER initiative; however, there has been much criticism about this initiative from patient advocacy groups and COVID-19 experts with concerns about the scope of the research goals, the clinical use of the findings, and the output speed. One of the criticisms was that too much focus was placed on observational analysis of big data, which could not lead to robust actionable clinical outcomes without additional clinical validation and instead clinical trials for therapeutics for long COVID are necessary. However, absence of consensus on terminology effects long COVID research and risks the success of trials for treatment. Implementation of the recently published classification of outcomes for long COVID is crucial.

A range of approaches will be key to progress, including hypothesis-driven prospective studies that aim to discover and test treatments. Big data analysis is an essential first step to generate reliable evidence to understand and subsequently treat long COVID. Pfaff and colleagues intend for their long COVID patient classifier algorithm to be used for clinical trial recruitment to enable the discovery of treatments.

During a pandemic, it is unsurprising that researchers' scientific goals sometimes differ from those of patients; however, there has been discomfort with the poor engagement of patients in long COVID research. In an open letter, patient advocacy groups have highlighted that meaningful involvement of patient advocates and post-viral illness experts must be integrated into the RECOVER initiative to produce clinically meaningful research. As the need to find treatments for long COVID becomes increasingly urgent in the USA, patient engagement is crucial in the development of the next generation of COVID treatments.

With such a heterogeneous range of patient characteristics, persistent symptoms, disease severities, management approaches, and SARS-CoV-2 strain infection and vaccination histories, teasing out the underlying mechanisms of disease and identifying treatment targets and strategies for management continues to be a challenge. Progress requires input from a range of health-care professionals, researchers from a variety of disciplines, and people living with long COVID. ■ *The Lancet Digital Health*

For more on the **study by Pfaff and colleagues** see **Articles** page e532

For more on the **trajectory of long COVID symptoms after COVID-19 vaccination community based cohort study** see *BMJ* 2022; **377:** e069676

For more on the **study by Sudre and colleagues** see *Sci Adv* 2021; **7:** eabd4177

For more on the **Post-hospitalisation COVID-19 study** see **Articles** *Lancet Respir Med* 2022; published online April 23. https://doi.org/10.1016/S2213-2600(22)00127-8

For more on the **criticisms of the NIH long COVID study** see https://www.statnews.com/2022/03/29/nih-long-covid-sluggish-study/

For more on **consensus for terminology and definitions in long COVID research** see **Comment** *Lancet Respir Med* 2022; published online May 4. https://doi.org/10.1016/S2213-2600(22)00135-7

For more on the **need for data in diagnosing long COVID** see https://www.statnews.com/2022/04/29/long-covid-machine-learning-n3c/

For more on the **core outcome set for post-COVID-19 condition** see **Position Paper** *Lancet Respir Med* 2022; published online June 13. https://doi.org/10.1016/S2213-2600(22)00169-2

For more on the **open letter by patient advocates** see https://docs.google.com/document/d/1OmF6PAUja_wW8skhyYOZbyDtnmyYxxqH5LUNfZQjy0o/edit#