Redefining and unravelling acute and long-term COVID-19 symptoms and complications: A large-scale multinational network cohort study

Document Status	
Date of final version of the study report	

List of abbreviations

Responsible parties

Principal Investigators	Junqing Xie Annika Jödicke
Data Scientists	Kim López Güell Martí Català Sabaté
Clinical Domain Expert	Daniel Prieto Alhambra

Abstract

Background and Objectives

Many people suffer from long-term symptoms after acute infection with SARS-CoV-2, usually called "long COVID", the most prevalent ones being extreme tiredness, shortness of breath, loss of smell and muscle aches[1]. Similarly, post-acute complications (PASC) including cardiovascular and thromboembolic events have also been observed following COVID-19. Despite increasing data on these public health concerns, the concept is still very broad and no proper characterisation of the condition or its prevalence and incidence has been thoroughly studied.

Our aims are:

1. To study the descriptive epidemiology of long COVID and PASC in multiple European countries and healthcare settings. We will study the prevalence and incidence of these conditions in the general population as well as in the COVID-19 infection and comparison cohorts.
2. To characterise people with long COVID or PASC, using test-negative and influenza cohorts as a benchmark. We will quantify demographic baseline characteristics, drug utilisation and treatment patterns in long COVID and PASC cohorts, COVID-19 infection cohorts and comparison cohorts.
3. To model temporal trajectories and define clinical subgroups of long COVID. We will first derive and validate relevant temporal disease trajectories for long COVID, and then identify clusters of conditions/symptoms among long COVID using machine learning.

Methods

This will be an international network study involving several data sources from different countries. The study population will consist of different COVID-19 cohorts (infection, reinfection), comparison cohorts (test negative, influenza) and outcomes (long COVID symptoms and diagnoses, PASC events, other specific medical conditions).

We will perform a large-scale characterisation of all cohorts at baseline, and study drug utilisation (before, on and after index date) and treatment patterns (after index date). We will calculate incidence and prevalence of all outcomes in the base cohorts (i.e., COVID-19 cohorts and comparison cohorts), and of all base and outcome cohorts in the general population. We will also assess the temporal

trajectories of long COVID and identify clusters of long COVID symptoms with machine learning techniques.

Study Analysis

We will use several validated analytical tools for the characterisation part and incidence and prevalence calculations. The characterisation will include population factors, comorbidities, vaccination and frequency of healthcare utilisation. The drug utilisation and treatment patterns will be studies for pre-specified drugs. Prevalence will be computed as annually and monthly period prevalence, and incidence will also be calculated monthly and annually.

Temporal disease trajectory will be described using a validated analytical tool, "Trajectories" [https://github.com/EHDEN/Trajectories] which was previously developed for OMOP mapped data[2]. The clustering will be performed with LCA or k-means methods in the patient space or projecting the data in a network structure using binary variable distances and performing community detection or using visualisation tools[3].

Amendments and Updates

NA

Milestones

	Date planned	Date actual
Protocol submission to local IRB boards	Jan 28th 2023	
Development of analysis code	Jan 30th – Mar 15th 2023	
Distributed Network analyses	Mar 15th – Apr 15th 2023	
Manuscript writing	Apr 17th -21st 2023	

Background and Rationale

Ongoing or post COVID-19 syndrome after acute infection, commonly referred to as "long COVID", is a new global health challenge. Up to Oct 2022, over 640 million individuals have been infected with SARS-COV-2 globally[4], but the number of long COVID cases remains unknown, with prevalence estimated from initial published reports varying dramatically. It was suggested that long COVID could affect around 6% of persons infected with coronavirus[5]; or up to as many as half[6]. Recently, the UK's Office for National Statistics has estimated that, on average, 1 in 5 people have symptoms beyond 5 weeks, while 1 in 10 persists over 12 weeks, accounting for 3.1% of the whole population[7].

The difference in the terminology and nomenclature is one of the major sources of heterogeneity in previous research of long COVID. Many similar terms relating to the broad concept of long COVID have been proposed and developed by patients and various clinical societies and health agencies, such as post-COVID-19 condition (WHO[8], US-CDC[9]), ongoing symptomatic COVID-19 (NICE[10]), post-COVID-19 syndrome (NICE), post-acute sequelae of SARS CoV-2 infection (US-NIH[11]), and post-intensive care syndrome (Society of Critical Care Medicine[12]). Beyond the names, lacking consensus on definitions adds another layer of complexity toward a clearer picture of long COVID. As a result, most published studies used persistent varying signs and symptoms, measured at variable time points after an acute episode, as a practical approach for phenotyping long COVID, and an array of up to 287 unique clinical conditions, in accordance with the Human Phenotype Ontology, has been put forward, covering multi-

organs and systems primarily including cough, fatigue, shortness of breath, and cognitive dysfunction[13]. Recently, a call has been reinforced for a consensus of long COVID terminology and definitions across clinicians, researchers, patients, and policymakers[14].

Long COVID is still a new disease with many unanswered questions, although acute infection is becoming much less fatal due to the availability of effective vaccines, treatments and accumulated understanding of COVID-19. It is expected that this uncertainty poses challenges for patients and clinicians, and immediate efforts and actions are needed to clarify previously few and conflicting evidence. To achieve this, assessing epidemiology, characteristics and outcomes of long COVID are crucial to establishing the scale and health burden long COVID has imposed and enabling the identification of risk factors that may predispose people more likely to develop persistent conditions. Study findings can not only have direct implications for patients and healthcare professionals on the prevention and management of long COVID, but also shed light on hypothesis generation and further research into mechanisms. On the other hand, no drugs are currently available to treat long COVID, risking clinicians and people with the condition to turn to untested and self-prescribed therapies. Data are urgently needed to decipher commonly utilised medications for patients with long COVID to minimise the harms of inappropriate drug use while maximising the positive impacts of these post-marketing drugs that could potentially ease the symptoms of long COVID.

Informed by knowledge gaps as mentioned above, this study will focus on long COVID epidemiology, clinical presentations, progression and therapeutic management, using a federated network of data partners in Europe to help us to understand long-term consequences of COVID-19.

Research Questions and Objectives

1 To understand the epidemiology of PASC and long COVID-19
   a To estimate incidence and prevalence of PASC and long COVID relevant conditions in different COVID-19 (first infection, new infection and reinfection cohorts) and other comparison cohorts (test negative and influenza cohorts).
   b To study the variations of epidemiology of long COVID and PASC across different European populations, healthcare settings, and periods/stages of the COVID-19 pandemic.
2 To characterise populations of COVID-19
   a To characterise people with acute COVID-19 infections, post-acute sequelae SARS-CoV-2 (PASC), and long COVID.
   b To study the utilisation of pre-specified drugs of interest among people with COVID-19, PASC and long COVID.
   c To study treatment patterns/trajectories of all drugs among people with COVID-19, PASC and long COVID.
3 To study temporal trajectories and define clinical subgroups of long COVID
   a To derive and validate relevant temporal disease trajectories for long COVID.
   b To identify clusters of conditions/symptoms among long COVID using machine learning.

Study design

Cohort study.

Data sources

This will be an international network study involving several data sources from different countries. We will collaborate with at least 6 other databases from 5 countries. Observation period: From 01/09/2020 (study start date) to the latest availability of participating databases.

Study population

All individuals registered in participating databases for at least 365 days at the index date, defined as below:

We will define the following "base" cohorts:

a First infection cohort: We will include persons with a first positive COVID-19 PCR or LFD test or clinical COVID-19 diagnosis during the study period. The index date will be defined as date of the first confirmed infection. No prior history of COVID-19 will be allowed.

b Reinfection cohorts: We will include persons with a positive COVID-19 PCR test or clinical COVID-19 diagnosis during the study period. The index date for the second/third/fourth or more infection episode will be defined according to the chronological order of date of the reinfection if an individual had multiple infection records. A subsequent infection will only be considered reinfection relative to the immediate past infection if the time window between them is greater than 42 days. See Figure 1 for a visual definition of the cohort.

c Test negative cohort: We will include persons with a negative COVID-19 PCR or LFD test during the same observational period as for the infection cohorts. The index date will be the first negative testing result date.

d Influenza cohort: We will include persons with a diagnosis of flu diagnosed between 01/03/2017 and 01/03/2020 (historical influenza control). The index date will be the first influenza diagnosis date.

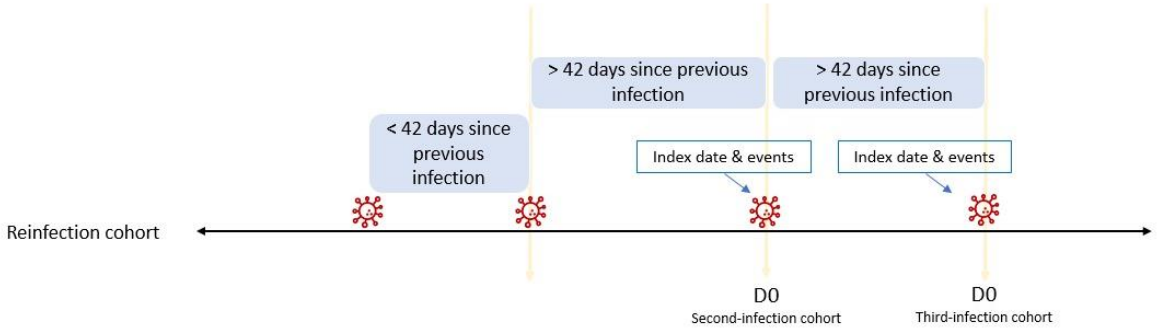

Figure 1: Visual definitions of all base cohorts, and additional details of the Reinfection cohorts.

Among the "base" cohort, we will further define the following "outcome" cohorts:

i Long COVID: We will define persons who experience any pre-specified symptoms and disease diagnoses in the period of >90 days – 365 days after the index event (the base cohort event). The list of symptoms and disease diagnoses consisting of 25 conditions based on the World Health Organisation definition. *(Technical Note: this adds up to a total of 25 "one symptom" cohorts, 1 "any symptom" cohort and 1 "Long COVID diagnosis cohort", specified in the Study Analysis section).* We will only require one occurrence of the symptom of interest. The index date will be selected based on the base cohort.

ii PASC cohort: We will select persons suffering from PASC (more information on the definition can be found in the Statistical Analysis section). This consists of 10 condition cohorts and 1 "any condition" cohort made from all of the other 10. We will only require one PASC event >90 days to 365 days after the index event. We will define the cohort index date as the index event of the base cohort.

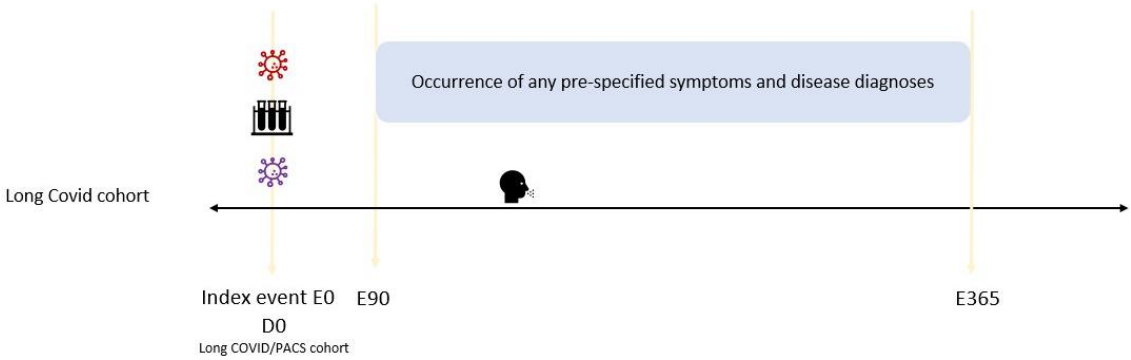

Figure 2: Visual definition of Long Covid cohort.

We will also define a specific medical conditions cohort (including dementia, cancer, type 1 diabetes, renal failure (acute kidney injury or new onset chronic kidney disease), liver injury/failure, autoimmune diseases (rheumatoid, lupus, other inflammatory arthritis, inflammatory bowel disease), MISC (for children only), myocarditis/pericarditis, dysauthonomia) for the objectives 1a and 1b.

Exclusion criteria

For all "base" cohorts, we will exclude people who have a historical influenza 3 months before the index date.

For the “base” cohort a, c and d, we will additionally exclude people who have a historical COVID-19 infection before the index date. Cohorts c and d will further be censored if they develop COVID-19 during the follow-up period.

For all “base” cohorts, we will exclude people who have less than 120 days of data availability after the index date.

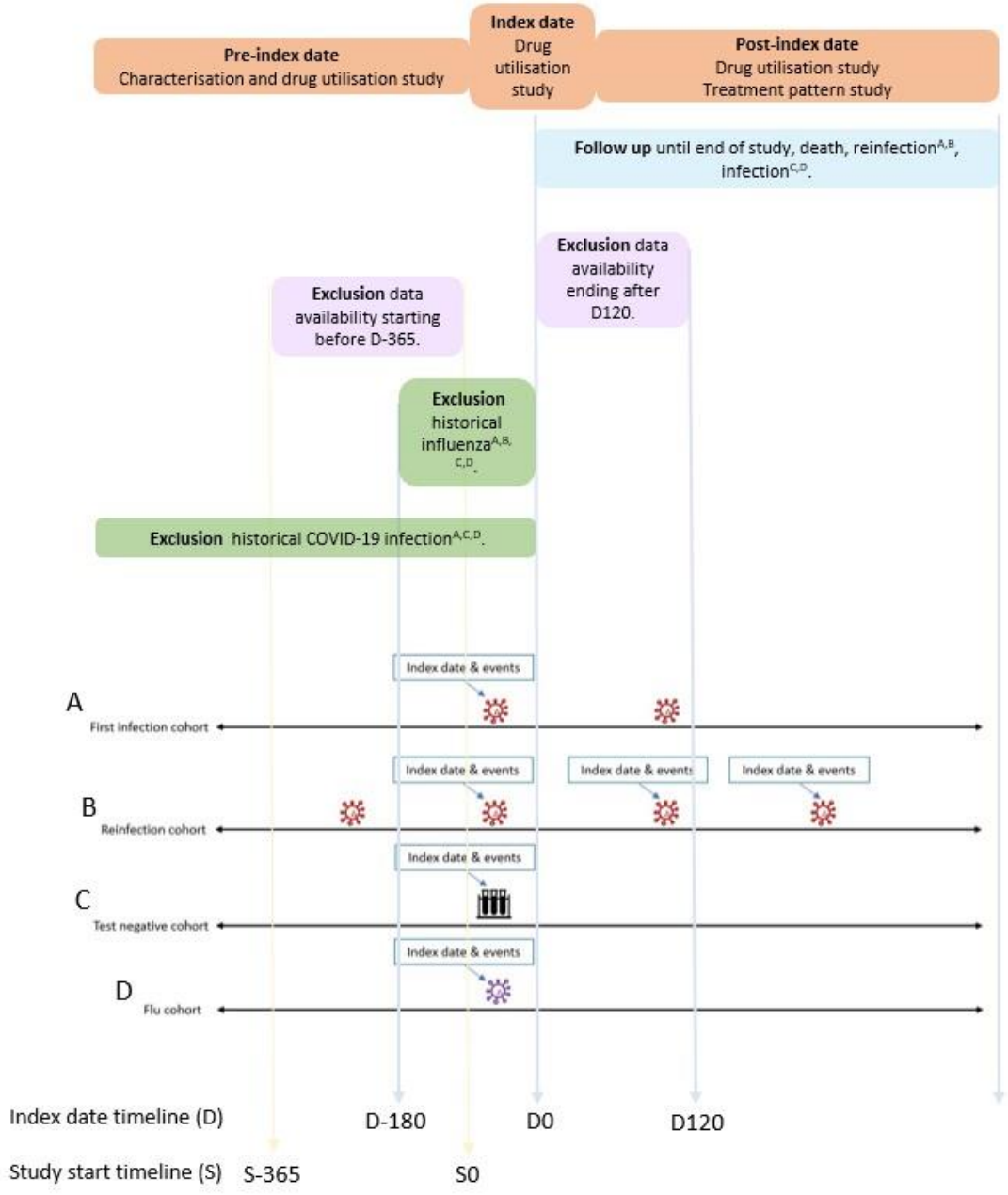

Figure 3: Timeline of the inclusion criteria of the study.

Exposure, outcome and covariates

The exposure, covariates, and outcomes will vary depending on the objectives. The follow-up for outcomes will be censored at the end of study, death, reinfection for cohorts a and b, infection for cohorts c and d.

Objective 1a:

Cohorts: Infection, reinfection, negative test, influenza.

Outcome: Incidence/Prevalence of PASC, long COVID-19, other specific medical conditions (all base cohorts)

Covariate for Stratification: Age (5 ranges for children, 18-40, 40-65, 65+ for adults), vaccination (yes/no), sex, calendar period

Objective 1b:

Cohort: Source population

Outcome: Incidence/Prevalence of Infection, reinfection, negative test, influenza, PASC, long COVID-19, other specific medical conditions

Covariate for Stratification: Age (5 ranges for children, 18-40, 40-65, 65+ for adults), vaccination (yes/no), sex, calendar period

Objective 2a:

Cohorts:

- Infection, reinfection, negative testing result, influenza
- PASC (only "any condition" cohort), long COVID-19 (only "any symptom" cohort and long COVID code cohort)
- All the "intersections" of the previous three "outcome" cohorts with the base cohorts (e.g. "any condition" PASC with previous COVID-19 infection / test negative / reinfection / influenza diagnose)

Outcome: Large-scale patients characteristics

Stratification: Age (5 ranges for children, 18-40, 40-65, 65+ for adults), vaccination (yes/no), sex, calendar period

Objective 2b:

Exposure: pre-specified drugs of interest

Cohorts: Same as Objective 1a

Outcome: Use of prespecified drugs of interest.

Stratification: Age (5 ranges for children, 18-40, 40-65, 65+ for adults), vaccination (yes/no), sex, calendar period

Objective 2c:

Exposure: pre-specified drugs of interest

Cohorts: Same as Objective 1a

Outcome: Treatment pattern of prespecified drugs of interest (COI).

Stratification: Age (5 ranges for children, 18-40, 40-65, 65+ for adults), vaccination (yes/no), sex, calendar period

Objective 3a:

Temporal disease trajectories among the long COVID patients (only “any symptom” and long COVID code cohorts). Exploration of all possible code trajectories post COVID-19 infection and onset of long COVID symptom or diagnoses.

Stratification: Age (5 ranges for children, 18-40, 40-65, 65+ for adults), vaccination (yes/no), sex, calendar period

Objective 3b:

Clustering of patients of long COVID cohorts.

Method:

(1) clustering in the patient subgroups (we will cluster long COVID symptoms and account for covariates like age and sex) using methods like Latent Class Analysis or k-means

(2) clustering algorithm in the symptom subclasses, creating a network with symptoms as nodes and weighted edges with partial correlation. Analysis using community detection algorithms for clustering and other network visualisation tools

Stratification: Age (5 ranges for children, 18-40, 40-65, 65+ for adults) and sex for method (2). Vaccination (yes/no) and calendar period for both

Data analysis

We will follow the subsequent steps:

1 Defining cohorts

COVID-19 cohorts

We will re-use code lists from previous studies done by our group.

We will define two mutually exclusive COVID-19 cohorts: first COVID-19 infections (C1), and COVID-19 re-infections (C2); and 1 tested-negative comparator cohorts first SARS-CoV-2 negative tests (C3).

All COVID-19 infections will be identified using SARS-CoV-2 antigen and RT-PCR positive tests, with the index date as the date of the test. First COVID-19 infections (C1) will be defined as new COVID-19 infections without any prior history of a COVID-19 infection. COVID-19 re-infections (C2) were defined as COVID-19 infections not identified as first infection (C1) infections without a record of a COVID-19 infection 42 days (washout period) prior to the index date.

The tested-negative comparator cohorts will be identified using SARS-CoV-2 antigen and RT-PCR negative tests, with the index date as the date of the test. Individuals included in these cohorts will be required to have a SARS-CoV-2 negative test result without a clinical COVID-19 diagnosis or positive SARS-CoV-2 test result prior to the index date and up to 120 days after the index date. First SARS-CoV-2 negative tests (C3) will be defined as SARS-CoV-2 negative tests without any prior history of a negative test.

Concept lists used to define the COVID-19 and test-negative cohorts are provided in Supplement Tables S1 and S2.

All cohorts will include individuals with at least 365 days of medical history available prior to the index date. Individuals with a clinical diagnosis or positive test result for influenza 180 days prior to or on

the index date were excluded. To ensure sufficient follow-up to develop long-COVID-related symptoms, we will only include individuals with ≥120 days of follow-up.

All cohorts will be followed until the occurrence of the event of interest, death, new COVID-19 infection or record of a COVID-19 clinical diagnosis, influenza infection (positive test result or clinical diagnosis), one year of follow-up, or end of data availability.

Long COVID cohorts

We will identify long COVID symptoms as defined by the WHO clinical case definition of "post COVID-19 syndrome" based on SNOMED codes in the respective datasets. A total of 25 symptoms will be included, namely abdominal pain, allergy, altered smell and/or taste, anxiety, blurred vision, chest pain, cognitive dysfunction, cough, depression, dizziness, dyspnoea, fatigue or malaise, gastrointestinal issues (acid reflux, constipation, or diarrhoea), headache, intermittent fever, joint pain, memory issues, menstrual problems, muscles spasms or pain, neuralgia, pins and needles sensation, sleep disorder, tachycardia, post-exertional fatigue and tinnitus and hearing problems. Code lists were developed separately for each symptom and reviewed independently by 3 clinicians [Supplement Tables S4-S28].

The WHO definition will subsequently be operationalized to identify long COVID for primary care data: Long COVID will be defined as having at least one record of any of the pre-defined symptoms between ≥90 days after the date of COVID-19 infection and 365 days. In addition, we will require the person to not have a record of the respective symptom within a washout window of 180days prior to index date.

PASC cohorts

The events to be included in the PASC definition can be found in Tables S29 - S38 (Appendix). Code lists will be reused from previous studies, e.g. Li X et al. 2021[15] and Prieto-Alhambra D et al., 2021[16].

Cohorts of post COVID outcomes

The code lists for "post COVID outcomes" can be found in Tables S39 - S58 (Appendix). Only diabetes type 1, lupus, rheumatoid arthritis, IBD, MISC and cancer will be considered for children. The cohorts for dementia and cancer will be reused from previous studies and have already been reviewed by clinical experts.

Influenza cohort

This cohort will be reused from a previous study and is being clinically reviewed. The codelist can be found with the other base cohorts, in Table S3.

Feasibility counts

The feasibility counts for all the cohorts have been performed in CPRD AURUM and the results are in the Appendix 3.

All data partners will also be asked to generate feasibility counts prior to execution of the actual study analyses, to confirm eligibility for the project.

2 Characterisation of cohorts

We will perform a large-scale characterisation of all the "base" cohorts a-d and subset of the "outcome" cohorts i-ii (specified in Objective 1a), and characterisation of intersection cohorts (a-d ∩

i-ii, each of the outcome cohorts with each of the base cohorts) using a validated analytical tool. The characterisation will include base population factors and comorbidities (we will look back to different windows depending on the event of interest, for instance any time before for diagnoses or a year for prescriptions). We will also look at other variables, if available, like:

- Vaccination
- Frequency of Healthcare utilisation: Hospitalisation, ICU, Ventilation
- Sick leave

We will study drug utilisation for all base cohorts and the intersection of the three Long COVID and PASC cohorts of interest with the base cohorts at three different periods: before the index date, at index date and after the index date. We will partition the time before and after index date in different length windows.

We will study the treatment patterns of pre-specified drugs in all base cohorts and the intersection of the three aforementioned Long COVID and PASC cohorts with the base cohorts after the index date.

For this, we will use a validated analytical tool called "TreatmentPatterns", [https://github.com/ darwin-eu / TreatmentPatterns] which allows the user to partition the time after index date in different length windows and study the relationship between the prescription of different drugs, in a naïve temporal trajectory, for all the cohorts.

The characterisations will be stratified by age groups, sex, calendar time (used as a proxy for SARS-COV-2 variant of concern) and vaccination status (defined previously to the index date).

3 Statistical analysis of Incidence and Prevalence

In this section, the term "outcome cohorts" will also include the additional specific medical conditions cohort.

We will estimate incidence and prevalence of the outcomes defined by the base and outcome cohorts among the general population. We will also estimate incidence and prevalence of the outcome cohorts among the base cohorts.

Prevalence will be calculated as annual and monthly period prevalence which summarises the total number of individuals who have the outcome of interest during a given year or month divided by the population at risk of getting exposed during that year or month. Therefore, period prevalence gives the proportion of individuals exposed at any time during a specified interval.

Annual or monthly incidence rates will be calculated as the of number of new cases of the outcome of interest per 100,000 person-years of the population at risk of getting exposed during the period for each calendar year or month. Any study participants who had the outcome already prior to the date at which they would have otherwise satisfied the criteria to enter the denominator population (as described above) will be excluded. Those study participants who enter the denominator population will then contribute time at risk up their first outcome during the study period. Or, if they do not have the outcome, they will contribute time at risk up until study end, end of observation period, death, or other censoring as specified in the Study Population section.

Prevalence will be estimated using a binomial model with 95% confidence intervals. All incidence rates will be estimated per 100,000 person-years using a Poisson or negative-binomial model with 95% confidence intervals.

The estimations will be also stratified by age, sex, calendar time and vaccination status.

5. Temporal disease trajectories

A standard method developed in 2014[17] will be used, with a validated analytical tool for OMOP data[2]. This method creates all event pairs and analyses significance of risk ratios by matching and tests directionality by a binomial test. Longer trajectories are then constructed.

We would use the tool for the long COVID cohort, defined as an infection of COVID-19 plus any symptom or the code for long COVID (I.e., the "intersection" cohort of the base cohort of infection and the outcome cohorts "any symptom" and code of long COVID).

The output of this part of the analysis would be all the significant directional disease trajectories (from single codes to single codes) of the form COVID-19 infection -> long COVID symptom/diagnose -> other. The results would include sequences of codes, network-type figures, counts and significance tests.

We would then focus on the ones with at least one COVID-19 infection and one long COVID symptom/diagnose related code in them.

The trajectory studies will be stratified by age, sex, calendar time (used as a proxy for SARS-COV-2 variant of concern) and vaccination status (defined previously to the index date).

6. Clustering of patients and symptoms

Method 1: we will construct a bipartite graph of symptoms and patients, and project it on the symptom space (with weighted edges of a distance metric between symptoms, like Phi Coefficient or Jaccard distance or Hamming distance, for a set of binary variables). We will then use a clustering algorithm such as Markov Chain Algorithm to assess clusters within the network[17], and also use other visualisation tools to study the network structure further[3].

Method 2: we will use the raw data to cluster directly in the patient space, using a soft clustering algorithm like Latent Class Analysis or a hard clustering algorithm like K-means, while also accounting for other covariates like age and sex[18].

In any case we will then study the clusters obtained and assess the profile of the patients belonging to them (characterisation at baseline). We could also look at outcomes, like frequency of Healthcare Utilisation.

The estimations will be also stratified by calendar time and vaccination status.

7. Network study and Shiny app

All the analyses will be conducted in different databases mapped to the OMOP CDM. This enables the use of standardised analytics and tools across the network since the structure of the data and the terminology system is harmonised. The OMOP CDM is developed and maintained by the Observational Health Data Sciences and Informatics (OHDSI) initiative and is described in detail on the wiki page of the CDM: https://ohdsi.github.io/CommonDataModel and in The Book of OHDSI: http://book.ohdsi.org.

This analytic code for this study will be written in R. Each data partner will execute the study code locally against their database containing patient-level data and will then return the results set which will only contain aggregated data. The results from each of the contributing data sites will then be combined in tables and figures for dissemination. All the results will be displayed and easily visualisable on a Shiny App.

8. Sensitivity analyses

We will have different variations for each definition of Long Covid (or PASC), which can be visually inspected in Figure 4. We will define index date based on the symptom or event instead of the base cohort or ask for more than one symptom or event in the window of interest. We will also restrict the window to ask for symptoms up to 120 days, instead of a year.

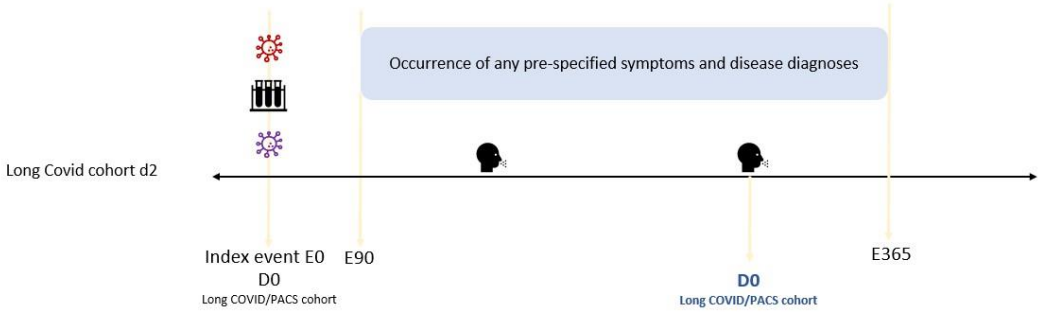

Figure 4: Alternative definition of long COVID and PASC.

Limitations of the research methods

The proposed study is observational in nature. Hence, it cannot address causality but rather describes associations. While we will comprise a list of diagnoses/symptoms characterising long COVID based on WHO Delphi consensus, some of the symptoms may not be recorded in primary care data. Therefore, we might not include them into the long COVID definition used for this study.

Protection of Human Subjects

Cell suppression will be applied as required by databases to protect people's privacy. Cell counts < 5 will be clouded.

Plans for Disseminating and Communicating Study Results

The results from this study will be disseminated through submission of research abstracts, and, if accepted, subsequent presentations (poster or oral presentation) at scientific congresses.

Additionally, we plan to prepare manuscript/s for journal publication, describing the results of the respective work packages. When reporting the data, we will not directly report events with <5 counts but 'protect' these counts with secondary suppression where necessary.

Authorship of any publications resulting from this study will be determined on the basis of the International Committee of Medical Journal Editors (ICJME) Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals.

In addition to scientific abstracts and manuscripts, we will disseminate our research findings to a lay audience and will present our findings to long COVID patient groups and to the wider public.

References

1. Long-term effects of coronavirus (long COVID). *nhs.uk* https://www.nhs.uk/conditions/coronavirus-covid-19/long-term-effects-of-coronavirus-long-covid/ (2021).
2. Künnapuu, K. *et al.* Trajectories: a framework for detecting temporal clinical event sequences from health data standardized to the Observational Medical Outcomes Partnership (OMOP) Common Data Model. *JAMIA Open* **5**, ooac021 (2022).
3. Kuan, V. *et al.* Identifying and visualising multimorbidity and comorbidity patterns in patients in the English National Health Service: a population-based study. *Lancet Digit. Health* **5**, e16–e27 (2023).
4. WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int.
5. Estimated Global Proportions of Individuals With Persistent Fatigue, Cognitive, and Respiratory Symptom Clusters Following Symptomatic COVID-19 in 2020 and 2021 | Neurology | JAMA | JAMA Network. https://jamanetwork.com/journals/jama/fullarticle/2797443.
6. Sugiyama, A. *et al.* Long COVID occurrence in COVID-19 survivors. *Sci. Rep.* **12**, 6039 (2022).
7. Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK - Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/prevalenceofongoingsymptomsfollowingcoronaviruscovid19infectionintheuk/5january2023.
8. Coronavirus disease (COVID-19): Post COVID-19 condition. https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-(covid-19)-post-covid-19-condition.
9. CDC. Post-COVID Conditions. *Centers for Disease Control and Prevention* https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html (2022).
10. Venkatesan, P. NICE guideline on long COVID. *Lancet Respir. Med.* **9**, 129 (2021).
11. RECOVER: Researching COVID to Enhance Recovery. *RECOVER: Researching COVID to Enhance Recovery* https://recovercovid.org.

12. SCCM | Post-intensive Care Syndrome. *Society of Critical Care Medicine (SCCM)* https://sccm.org/MyICUCare/THRIVE/Post-intensive-Care-Syndrome.
13. Deer, R. R. *et al.* Characterizing Long COVID: Deep Phenotype of a Complex Condition. *eBioMedicine* **74**, (2021).
14. Munblit, D. *et al.* Long COVID: aiming for a consensus. *Lancet Respir. Med.* **10**, 632–634 (2022).
15. Li, X. *et al.* Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study. *BMJ* **373**, n1435 (2021).
16. Prieto-Alhambra, D. *et al. Comparative effectiveness and safety of homologous two-dose ChAdOx1 versus heterologous vaccination with ChAdOx1 and BNT162b2: a cohort analysis*. https://www.researchsquare.com/article/rs-1074858/v1 (2021) doi:10.21203/rs.3.rs-1074858/v1.
17. Jensen, A. B. *et al.* Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nat. Commun.* **5**, 4022 (2014).
18. Miaskowski, C. *et al.* Latent Class Analysis Reveals Distinct Subgroups of Patients Based on Symptom Occurrence and Demographic and Clinical Characteristics. *J. Pain Symptom Manage.* **50**, 28–37 (2015).