BMJ Open

# Prevalence and determinants of persistent symptoms after infection with SARS-CoV-2: Protocol for an observational cohort study (LongCOVID-study)



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2022-062439 |
| Article Type: | Protocol |
| Date Submitted by the Author: | 03-May-2022 |
| Complete List of Authors: | Mutubuki, Elizabeth; RIVM<br>van der Maaden, Tessa; RIVM<br>Leung, Ka Yin; RIVM<br>Wong, Albert; RIVM<br>Tulen, Anna D.; RIVM<br>de Bruijn, Siméon; RIVM<br>Haverman, Lotte; Amsterdam Universitair Medische Centra<br>Knoop, H.; Amsterdam UMC Locatie VUmc, Medical Psychology<br>Franz, Eelco; RIVM<br>van Hoek, Albert Jan; RIVM,<br>van den Wijngaard, Cees; RIVM |
| Keywords: | COVID-19, EPIDEMIOLOGY, HEALTH ECONOMICS |
| | |

# Prevalence and determinants of persistent symptoms after infection with SARS-CoV-2: Protocol for an observational cohort study (LongCOVID-study)

Elizabeth N. Mutubuki[1], Tessa van der Maaden[1], Ka Yin Leung[1], Albert Wong[1], Anna D. Tulen[1], Siméon de Bruijn[1], Lotte Haverman[2], Hans Knoop[3], Eelco Franz[1], Albert Jan van Hoek[1], Cees C. van den Wijngaard[1]

[1]National Institute for Public Health and the Environment (RIVM), Center for Infectious Disease Control, Bilthoven, the Netherlands

[2]Emma Children's Hospital, Amsterdam UMC, University of Amsterdam, Child and Adolescent Psychiatry & Psychosocial Care, Amsterdam Reproduction and Development, Amsterdam Public Health, Amsterdam, the Netherlands.

[3]Department of Medical Psychology, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, the Netherlands.

**Corresponding author:** Elizabeth N Mutubuki

email address: elizabeth.mutubuki@rivm.nl

## Abstract

**Introduction:** A substantial proportion of individuals infected with severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) report persisting symptoms weeks and months following acute infection. Estimates on prevalence vary due to differences in study designs, populations, heterogeneity of symptoms and the way symptoms are measured. Common symptoms include fatigue, cognitive impairment and dyspnea. However, knowledge regarding the nature and risk factors for developing persisting symptoms is still limited. Hence in this study we aim to determine the prevalence, severity, risk factors and impact on quality of life of persisting symptoms in the first year following acute SARS-CoV-2 infection.

**Methods and analysis:** The LongCOVID-study is both a prospective and retrospective cohort study with a one year follow up. Participants aged 5 years and above with self-reported positive or negative tests for SARS-CoV-2 will be included in the study. The primary outcome is the prevalence and severity of persistent symptoms in participants that tested positive for SARS-CoV-2 compared to controls. Symptom severity will be assessed for fatigue using the Checklist Individual Strength (CIS subscale fatigue severity), pain (Rand-36/SF-36 subscale bodily pain), dyspnea (Medical Research Council (mMRC)) and cognitive impairment using the Cognitive Failure Questionnaire (CFQ). Secondary outcomes include loss of health-related quality of life (HRQoL) and risk factors for persisting symptoms following infection with SARS-CoV-2.

**Ethics and dissemination:** The Utrecht Medical Ethics Committee (METC) declared in February 2021 that the Medical Research Involving Human Subjects Act (WMO) does not apply to this study (METC protocol number 21-124/C).

Keywords: SARS-CoV-2, post COVID-19 condition, LongCovid, prevalence, HRQoL, risk factors

**Strength and limitations of this study**

- A prospective design, allowing for detailed analysis of the prevalence and risk factors of persistent symptoms of SARS CoV-2 infection.
- Presence of control groups that allows for comparison of COVID-19 cases to controls. that have similar experiences, such as lock down measures. This is important because such factors can influence complaints
- The use of validated questionnaires with validated cut-off scores for severity is another strength of this study.
- Repeated assessment of symptoms every three months during one year of follow-up will enable assessment of the time course of symptoms, and detection of disabling symptoms at every 3 months interval.
- Furthermore, the impact of symptoms on general functioning will be assessed.
- A limitation of this study is that severity scores of only four of symptoms associated with COVID-19 will be calculated to get more insight into clinical significance.

## Introduction

During the first months of the pandemic, epidemiological research focused primarily on the spread of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) and on treatment of those with severe or fatal illness (1). The effects of SARS-CoV-2 infection vary from asymptomatic infection, through to critical and chronic disease (2). Although most individuals infected with SARS-CoV-2 fully recover, there is a growing body of evidence that suggests that a substantial number of individuals remain with long-term complications or persisting symptoms (3-5).

COVID-19 varies in clinical presentation, disease severity, recovery time as well as completeness of recovery (6). A delay in recovery whereby individuals fail to return to their normal daily routines and still report lasting effects of the infection long after the expected period of recovery has been termed ''LongCOVID'' (7), ''long-haulers" (8) and "post COVID-19 condition" (9). The term post COVID-19 condition will be used in the rest of this article. Post COVID-19 condition is reported to occur in individuals that have a history of probable or confirmed SARS-CoV-2 infection, usually 3 months from the onset of COVID-19 and symptoms with a duration of at least 2 months that cannot be explained by alternative diagnosis (9). Fatigue, shortness of breath, cognitive dysfunction are some of the common symptoms (9). Symptoms may persist from initial infection, be a new onset following initial recovery from an acute COVID-19 episode or may also fluctuate or relapse over time (9).

Over 210 million confirmed cases of COVID-19 have been reported, and of those, an estimated 10-20% are reported to experience such persisting symptoms for weeks and months following acute SARS-CoV2 infection (9). However, higher incidence rates of persisting symptoms have been reported, for example through self-surveys of patient from long COVID peer support groups (10) as well as in hospitalized patients (11). Variation in the reported incidence and prevalence rates of post COVID-19 condition can be attributed to the complexity of the syndrome, differences in population groups, heterogeneity in clinical presentation of symptoms, little knowledge regarding the natural history and clinical course (12) and in the way symptoms are measured. Common persistent symptoms are shortness of breath, fatigue, dyspnea and headaches (13, 5). Some of the initial acute symptoms such as cough, fever, and chills become less prevalent as the illness progresses, whereas cognitive dysfunction and palpitations become more prevalent later in the illness (5).

A good overview of the nature of persisting symptoms following an acute infection with SARS-CoV-2, can enable better diagnosis, management and may reduce negative consequences on HRQoL (12). Hence in this study we aim to determine the prevalence and severity of persisting symptoms in the first year of infection, in individuals infected by SARS-CoV-2 compared to individuals that were not infected. In addition risk factors for developing post COVID-19 condition and its impact on health will be analyzed.

## Methods and analysis

### Study aim and design

The LongCOVID-study is an observational cohort study consisting of prospective and retrospective data with one year of follow up. The study aims to determine the prevalence, severity, health impact and risk factors associated with persistent symptoms following a SARS-CoV-2 infection, in cases compared to population controls and test-negative controls. The study is carried out by the Dutch National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands. The Utrecht Medical Ethics Committee (METC) declared in February 2021 that the Medical Research Involving Human Subjects Act (WMO) does not apply to this study (METC protocol number 21-124/C). Patients or the public were not involved in the designing, conducting, implementing and dissemination plans of the research

### Study population

Both the prospective and retrospective cohorts include children (ages 5-17) and adults (18 years and above).

*Prospective cohort study*

Participants with a positive SARS-CoV-2 infection test result on an antigen or polymerase chain reaction (PCR) test for acute infection, are included in the study as cases, if they complete the baseline questionnaire within 7 days of testing positive. Participants that test negative to SARS-CoV-2 infection and complete their baseline questionnaire within 7 days of testing negative, are included in the study as test-negative controls. A second group of controls, population controls, consists of randomly selected participants from the Basic Registration of Persons (BRP) without a positive test for SARS-CoV-2 infection or known history of probable infections.

*Retrospective cohort study*

Participants presenting with self-reported persisting symptoms associated with SARS-CoV-2 infection with or without having had a positive test result were included in the retrospective cohort study as post COVID-19 condition cases.

**Recruitment**

Figure 1 shows the flow diagram of participant recruitment in the LongCOVID-study. Participants are recruited through the following three ways;

*Via the Community Health Service*

Individuals testing positive and negative to COVID-19 at one of the community health services (GGDs) are invited to participate in the LongCOVID-study. Registration to participate is via the LongCOVID-study website.

*Basic Registration of Persons (BRP)*

Population controls are randomly selected from the basic registration of persons and invited by letter to participate in the study.

*Self-registered participants*

Individuals interested in participating in the LongCovid-study can also self-register through the study website (longcovid.rivm.nl). Test-negative controls, cases and post COVID-19 condition cases can be included in the study this way.

*Figure 1: Recruitment of participants in the LongCOVID-study*

## Patient and public involvement

No patient involved.

## Measurements (adults)

Table 1 shows different measurement moments were data is collected in form of questionnaires. At baseline, data on demographical characteristics such as gender, education level and employment are collected. Data on comorbidities is reported at baseline and at 12 months. Information regarding testing for SARS-CoV-2, COVID-19 related complaints and vaccination data is collected at baseline and at 3, 6, 9 and 12 months.

*Health related quality of life (EQ-5D-5L and Rand-12/SF-12)*

HRQoL is assessed using the Rand-12/SF-12 and EQ-5D-5L. Additional weekly measurement using the EQ-5D-5L were carried out in individuals presenting with acute symptoms in the first 8 weeks following a positive COVID-19 test. The EQ-5D-5L questionnaire consists of five dimensions of health (mobility, self-care, usual activities, pain/discomfort, and anxiety/ depression), with five levels of response and a visual analogue scale (EQ VAS). The EQ-5D-5L scores will be converted into utility scores using the Dutch tariff (14), ranging from 0 (death) to 1 (optimal health).

The Rand-12/SF-12, a shortened version of the Rand-36/SF-36 HRQoL questionnaire consists of 12 questions from the following 8 domains; physical functioning, physical role, emotional role limitations, social functioning, physical pain, general mental health, vitality and general health perception. The 8 domains can be summarized into a physical and mental health domain (15). Health scores will be converted into utility scores using the SF-6D (Short-Form Six-Dimension). Quality adjusted life years will be calculated by multiplying the utility scores by the time a patient spends in a given health state.

*Fatigue (Checklist Individual Strength [CIS])*

Fatigue severity is assessed using the subscale fatigue severity of the Checklist Individual Strength (CIS). The CIS subscale fatigue is a 8-item fatigue questionnaire (16). Each item is scored on a 7-point Likert scale. Scores range from 8 to 56, and scores of 35 and higher indicate severe fatigue (17).

*Cognitive function (Cognitive Failure Questionnaire [CFQ])*

Cognitive function is assessed using the Cognitive Failure Questionnaire (CFQ). The CFQ ranges from 0 to a 100 with higher scores indicating more cognitive impairment (18). A score of 44 or higher indicated clinically significant complaints on cognitive function.

*Pain (bodily pain subscale of the Rand-36/SF-36 Health Status Inventory [Rand-36])*

The bodily pain subscale of the Rand-36 Health Status Inventory (Rand-36) is used to assess pain severity. The Rand-36 scores range from 0 to 100, higher scores indicate better health status.

Significant impairment due to pain is reflected by a score of 55 or lower, based on Dutch norm scores (19). The subscales physical and social functioning were also used.

*Dyspnea (Medical Research Council (dyspnea) [mMRC])*

Dyspnea is assessed using the modified Medical Research Council (dyspnea) (mMRC). The mMRC scale ranges from grade 0 to 4. Grade 0- breathless with strenuous exercise; Grade 1- short of breath when hurrying on level ground or walking up a slight hill; Grade-2 walks slower on level ground because of breathlessness or stops for a breath when walking at own pace; Grade 3- stops for breath after walking about 100 yards or after a few minutes on level ground and Grade 4- am too breathless to leave the house or I am breathless when dressing (20). A score of 1 or higher reflecting significant impairment due to dyspnea (21).

*Illness and related beliefs (The Brief Illness Perception Questionnaire [Brief IPQ])*

The Brief Illness Perception Questionnaire (Brief IPQ /IPQ-K) is a nine-item scale to assess the cognitive and emotional representations of illness including consequences, timeline, personal control, treatment control, identity, coherence, concern, emotional response and causes (22). Item scores increases, represent linear increases in the dimension measured. The Brief IPQ is reported to have good test-retest reliability (22).

*Anxiety (Hospital Anxiety and Depression Scale [HADS])*

HADS (Hospital Anxiety and Depression Scale) is a 14 item self-report questionnaire designed to measure anxious and depressive states in patients with two subscales (23). The sum score per subscale ranges from 0 to 21. Scores between 0-7 indicate no anxiety or depression, 8-10 mild cases, 11-15 moderate cases and 16 or above severe cases (Snaith 1994).

*Dyspnea (The Nijmegen Clinical Screening Instrument [NCS])*

The Nijmegen Clinical Screening Instrument (NCSI) measures health status and has the following domains, physiological functioning, symptoms, functional impairment, and quality of life as main domains (24), and 8 subdomains (25). The 8 subdomains include subjective symptoms, dyspnea emotions, fatigue, behavioral impairment, subjective impairment, general quality of life (general QoL), health related quality of life (HRQoL) and satisfaction with relations (25). Each subdomain is expressed as a single score on its own scale, with higher NCSI scores indicate more problems (24). In the study the subdomain dyspnea will be used.

*Absenteeism (iMTA Productivity Cost Questionnaire [iPCQ])*

Participants were asked to report the number of days that they had been absent from work due to illness. Absenteeism will be measured using the iPCQ.

*Unpaid Productivity losses and informal care*

Unpaid productivity losses from work, studies, voluntary work as well as informal care will be valued using the Dutch shadow price of 14,57 euros per hour (26).

**Measurements (children)**

Below are age-specific scales that were used in children (aged 5-17 years), that replaced some of the above described scales for adults (Table 1).

*Physical function (Pediatric Quality of Life Inventory [PedsQL])*

The PedsQL is a HRQoL measure consisting of 4 subscales (physical functioning, emotional functioning, social functioning and school functioning) which can be computed to two summary scores (psychosocial and physical health summary scores).Dutch norms are available which allow comparison with the general population (27). A parent proxy of the PedsQL will be used for children aged 5-7 years.

*Fatigue (Pediatric Quality of Life Inventory Fatigue Scale [PedsQL fatigue]).*

Fatigue severity in children will be assessed with the Pediatric Quality of Life Inventory Fatigue Scale (PedsQL fatigue). This 18-item PedsQL fatigue scale comprises the general fatigue scale (6 items), sleep/rest fatigue scale (6 items), and cognitive fatigue scale (6 items) and is a reliable and valid instrument to measure fatigue in children (28). Dutch norm scores are available (29). A parent proxy will be used for children aged 5-7 years.

*Illness and related beliefs (The Brief Illness Perception Questionnaire [Brief IPQ])*

The Brief Illness Perception Questionnaire (Brief IPQ /IPQ-K-parents) will be completed by a proxy (22), and by the child if they are aged 10 or older.

*Pain Visual analogue scale (VAS)*

Pain severity will be assessed using VAS (30). Scores range from 0 (no pain) to a 100 ( worst imaginable pain). A parent proxy will be used for children aged 5-7 years.

*Health related quality of life (EQ-5D-Y)*

Weekly measurement moments for up to 8 weeks in children presenting with acute symptoms follow a positive COVID-19 test, will be carried out. The EQ-5D-Y-Proxy1 will be used for children

aged 5-7 years, and the EQ-5D-Y will be used for children aged 8-17 years to measure HRQoL. The EQ-5D-Y-Proxy1 and EQ-5D-Y questionnaires consist of five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/ depression), with three levels of response and a visual analogue scale (EQ VAS)(31).

*Dyspnea (Patient-Reported Outcomes Measurement Information System [PROMIS])*

Dyspnea will be assessed in children aged 5-7 years using an adjusted Patient-Reported Outcomes Measurement Information System (PROMIS) Asthma impact Short form Proxy and in children 8-17 using an adjusted PROMIS Asthma impact Short form(32).

*Cognitive function and behaviour (PROMIS and SDQ)*

Loneliness will be assessed in children aged 5-7 years using the PROMIS short form proxy depressive symptoms and in children 8-17 using the PROMIS short form depressive symptoms, for which norm scores are available which allow comparison with the general population (32, 33) . In addition the strengths and difficulties questionnaire (SDQ) will be used as well to assess the level of depressive symptoms, with a proxy for parents in the 5-11 years of age (34).

**Measurements (acute cohort)**

Data on HRQoL and acute symptoms will be collected weekly in the first 8 weeks following infection. Data collection will stop when the symptoms stop or end at 8 weeks following infection.

## Outcome measures

*Primary outcome*

1. The first primary outcome measure is the prevalence and severity of persistent symptoms in patients that tested positive for COVID-19 infection compared to both test-negative controls and population controls. Severity of symptoms will be assessed for fatigue, pain, dyspnea, and cognitive impairment using standardized questionnaires, with population-based norm cut-off scores for clinically significant severity.

*Secondary outcomes include:*

1. Factors that predict post COVID-19 condition following an acute SARS-CoV-2 infection at different follow up moments.
2. Healthcare utilization in the first year following infection with SARS-CoV-2 in cases compared to controls (test-negative controls and population controls) will be assessed.
3. Health related quality of life in cases will be compared to that of controls (test-negative controls and population controls) in the first year following infection. Additionally a

comparison will also be made between post COVID-19 condition individuals and individuals that test positive for COVID-19 but do not develop post COVID-19 condition.

*Table 1: Measurement moments*

| | **Baseline** | *Acute symptoms (in weeks)* | | | | | | | | **3 months** | **6 months** | **9 months** | **12 months** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | | | | |
| Informed consent | X | | | | | | | | | | | | |
| Baseline characteristics | X | | | | | | | | | | | | |
| Vaccination data *(status, type, date)* | X | | | | | | | | | X | X | X | X |
| Health utilization *(contact with healthcare providers, medication use)* | X | | | | | | | | | X | X | X | X |
| Symptoms data | X | *X* | *X* | *X* | *X* | *X* | *X* | *X* | *X* | X | X | X | X |
| HRQoL (EQ-5D*/EQ-5D-Y**) | | *X* | *X* | *X* | *X* | *X* | *X* | *X* | *X* | | | | |
| Health utilization *(contact with healthcare providers, medication use)* | X | | | | | | | | | X | X | X | X |
| Vaccination data *(status, type, date)* | X | | | | | | | | | X | X | X | X |
| Co-morbidities | X | | | | | | | | | | | | X |
| **Adults** | | | | | | | | | | | | | |
| HRQoL *(Rand-12/EQ-5D)* | X | | | | | | | | | X | X | X | X |
| Pain, Physical function and Social function *(Rand-36/SF-36 subscales pain, social functioning and physical functioning)* | X | | | | | | | | | X | X | X | X |
| Cognitive function *(CFQ)* | X | | | | | | | | | X | X | X | X |
| Fatigue *(CIS)* | X | | | | | | | | | X | X | X | X |
| Illness and related beliefs *(Brief IPQ)* | X | | | | | | | | | X | X | | |
| Anxiety and depression *(HADS)* | X | | | | | | | | | X | X | | |
| Dyspnea *(mMRC)* | X | | | | | | | | | X | X | X | X |
| Dyspnea *(NCSI )* | X | | | | | | | | | X | X | X | X |
| **Children** | | | | | | | | | | | | | |
| Physical function *(PedsQL subscale physical health)* | X | | | | | | | | | X | X | X | X |
| Fatigue *(PedsQL fatigue)* | X | | | | | | | | | X | X | X | X |
| Illness and related beliefs *(Brief IPQ / brief IPQ-parents)* | X | | | | | | | | | X | X | | |
| Pain *(VAS)* | X | | | | | | | | | X | X | X | X |
| Dyspnea *(adjusted PROMIS Asthma)* | X | | | | | | | | | X | X | X | X |
| Depressive symptoms (PROMIS, SDQ) | X | | | | | | | | | X | X | | X |

## Statistical analysis

Baseline characteristics of the participants in all groups will be presented using descriptive statistics mean (standard deviation), median (range), or proportion to assess if there is a balance in the groups regarding distribution of prognostic factors such as age, gender, co-morbidity and education.

***Prospective study***

**Primary outcome analysis**

1. Prevalence and severity of persistent symptoms in COVID-19 patients

Descriptive epidemiological statistical methods will be used to analyze prevalence of persistent symptoms at 3, 6, 9 and 12 months in cases compared to both control groups (test-negative controls and population controls). Persisting symptoms are defined as symptoms in cases with a duration of at least 2 months. Such symptoms markedly elevated in cases compared to controls (test-negative controls and/or population controls) during follow up are likely to be associated with COVID-19, and cases with these symptoms are in this study defined as cases with possible post COVID-19 condition (yes/no). Severity scores of fatigue, dyspnea, cognitive functioning, and pain will be calculated. Scores of individuals with confirmed COVID-19 will be compared to those of controls, per follow-up moment (baseline, 3, 6, 9 and 12 months follow-up).

**Secondary outcome analysis**

1. *Predictors of post-COVID 19 condition*

A prediction model will be built to identify predictors of post COVID-19 condition. The outcome will be having possible post COVID-19 condition as defined above. To determine the prediction model that best suits our data, the prediction model will be constructed using super learning (35). The prediction model will be evaluated using the ROC-AUC metric (36).

2. Predictors of healthcare utilization in post COVID-19 condition

A second prediction model will be performed to identify predictors of healthcare utilization in post COVID-19 condition. Healthcare utilization is defined as contact (visit to the general practitioner, telephone call, hospitalization, emergency healthcare services, other medical health professionals/services) with a health provider regarding symptoms attributed by the patient to COVID-19 or post COVID-19 condition (yes/no). The prediction model will be performed as mentioned above.

3. Quality-adjusted life-years

HRQoL will be assessed using EQ-5D-5L and Rand-12/SF-6D. Quality-adjusted life-years (QALYs), which takes into account both the impact of length and the quality-of-life will be calculated and be compared between cases and controls.

***Retrospective study***

Descriptive epidemiological statistical methods will be used to analyze the prevalence of persistent symptoms at baseline in cases compared to both control groups (test-negative controls and population controls). Moreover, prevalence of co-morbidities will be quantified in cases and control groups. Additionally, an assessment into healthcare utilization for cases will be performed according to the aforementioned definition.

***Acute data following infection***

Descriptive epidemiological statistical methods will be used to describe the prevalence and the type of symptoms present following acute infection as well as health related quality of life.

**Missing data**

The fraction of missing questionnaires at each time points and per period during the study (e.g., per 3 months) in all patients with confirmed Covid-19 will be tabulated. Scenarios of dealing with missing data include a complete case analysis, multiple imputation, and linear interpolation combined with carry forward.

## Discussion

The LongCOVID-study aims to determine the prevalence and severity of persistent symptoms following acute SARS-CoV2 infection in cases compared to controls, as well as to investigate the risk factors of developing persistent symptoms. Previous studies have explored prevalence of long-term symptoms and risk factors in various populations, i.e., in previously hospitalized patients (37), patients with diabetes type 1 and 2 (38, 39), in home-isolated patients with milder symptoms and in the young (40).

Blomberg reported that 61% of all the patients had persisting symptoms at 6 months (40). This included patients with a mild to moderate illness following initial illness as well as young patients (16-30 years). Persisting symptoms included loss of taste and or smell, fatigue, dyspnea, impaired concentration and memory problems. In a hospitalized population (37), fatigue, muscle weakness, sleep difficulties and anxiety or depression are the most prevalent symptoms at 6 months. Due to severe illness during hospital stay and impaired pulmonary function, the hospitalized population is a target group for long-term recovery (37, 41). Our study includes both adults and children from the

age of 5 with mostly mild to moderate acute symptoms and a much smaller group of patients that were hospitalized in the acute phase of the infection. We expect a possible bias against the number of hospitalized patients due to the design of the study, which requires questionnaires to be completed no more than seven days following a positive test for COVID-19.

**Strengths and weakness**

Strengths of the current study include the prospective design, allowing for detailed analysis of the prevalence and risk factors of persistent symptoms of SARS CoV-2 infection. In addition this study is one of a few studies (42) that allows for comparison of COVID-19 cases to control groups that have similar experiences, such as lock down measures. This is important because such factors can influence complaints. The availability of the population control group in this study allows us to control for background prevalence of symptoms. In addition the use of test-negative controls allows for assessment of the impact of COVID-19 compared to other respiratory infections. The use of validated questionnaires with validated cut-off scores for severity is another strength of this study. Repeated assessment of symptoms every three months during one year of follow-up will enable assessment of the time course of symptoms, and detection of disabling symptoms at every 3 months interval. Furthermore, the impact of symptoms on general functioning will be assessed. A limitation of this study is that severity scores of only four of symptoms associated with COVID-19 will be calculated to get more insight into clinical significance. This is because only four standardized questionnaires for symptom severity were included in the study. Hence the severity of other possible symptoms related to COVID-19 will not be taken into account.

In conclusion, the LongCOVID-study is expected to provide additional insights into the prevalence and severity of persistent symptoms after SARS CoV-2 infection to the international body of literature. In the Netherlands this is the first large scale study on persisting symptoms following SARS CoV2 infection.

**Declarations**

**Ethics approval and consent to participate**: All participants gave consent to participate in the study. Written consent is obtained online prior to being able to complete the study questionnaires, which are also completed online. The Utrecht Medical Ethics Committee (METC) declared in February 2021 that the Medical Research Involving Human Subjects Act (WMO) does not apply to this study (METC protocol number 21-124/C).

**Authors' contributions**: ENM wrote the manuscript. KYL contributed to the methods. CCW, TM, AJH, ADT, KYL, AW contributed to the design of the study, LH and HK advised on the design of the

questionnaires, TM, AJH, ADT, KYL, ENM, EF, SB, CCW contributed to implementation of the data collection, all authors reviewed and contributed to drafts of the manuscript. All authors read, contributed to refinement of the study protocol and approved the manuscript.

**Funding:** The study is executed by the National Institute for Public Health by order of the Ministry of Health. This means the study is not the result of a competitive grant. The Dutch Ministry of Health, Welfare and Sport does not have a role in the design of this study, its execution, analyses and interpretation of results.

**Competing interests**: The authors have no conflict of interest to declare.

**Word Count**: 3706

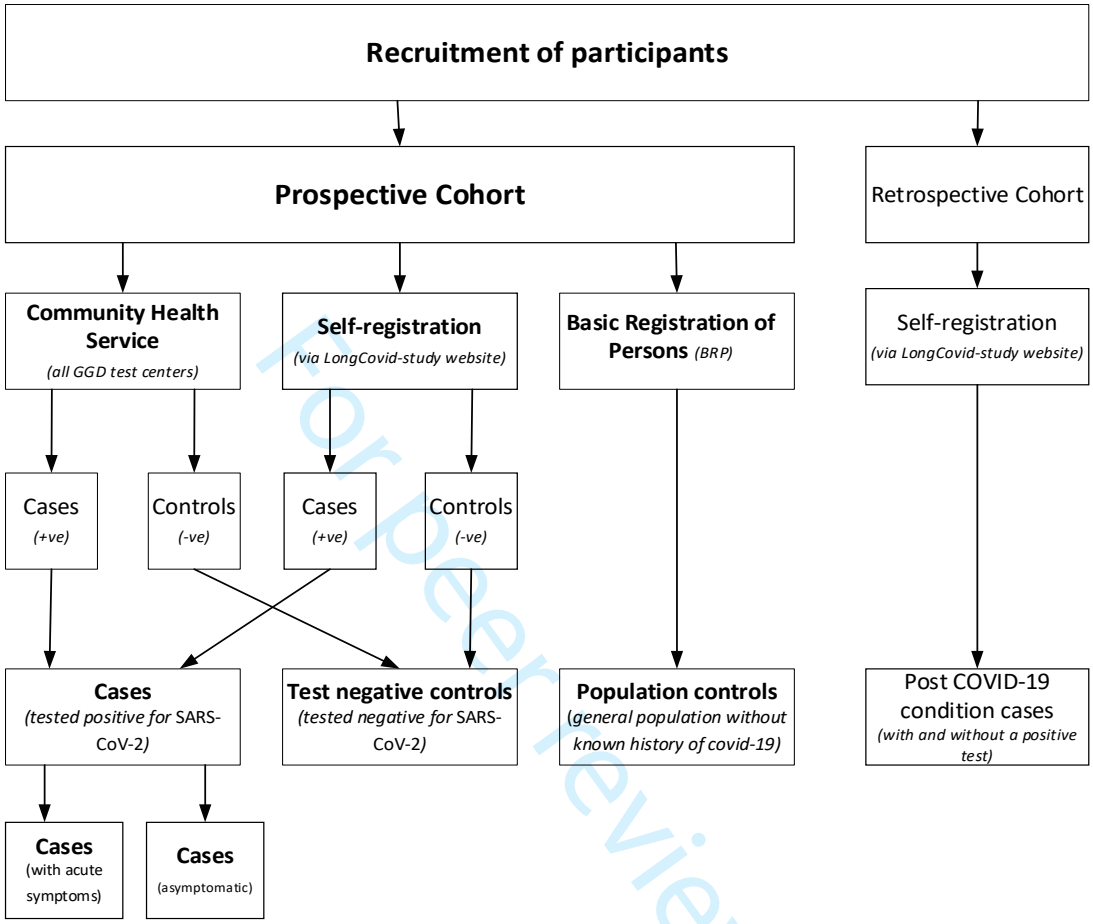
## References


1. Alwan NA. Surveillance is underestimating the burden of the COVID-19 pandemic. Lancet. 2020;396(10252):e24.
2. Ricke DO. Two Different Antibody-Dependent Enhancement (ADE) Risks for SARS-CoV-2 Antibodies. Front Immunol. 2021;12:640093.
3. Cirulli ET, Schiabor Barrett KM, Riffle S, Bolze A, Neveux I, Dabe S, et al. Long-term COVID-19 symptoms in a large unselected population. medRxiv. 2020:2020.10.07.20208702.
4. Petersen MS, Kristiansen MF, Hanusson KD, Danielsen ME, B ÁS, Gaini S, et al. Long COVID in the Faroe Islands: A Longitudinal Study Among Nonhospitalized Patients. Clin Infect Dis. 2021;73(11):e4058-e63.
5. Ziauddeen N, Gurdasani D, O'Hara ME, Hastie C, Roderick P, Yao G, et al. Characteristics of Long Covid: findings from a social media survey. medRxiv. 2021:2021.03.21.21253968.
6. Holmes E, Wist J, Masuda R, Lodge S, Nitschke P, Kimhofer T, et al. Incomplete Systemic Recovery and Metabolic Phenoreversion in Post-Acute-Phase Nonhospitalized COVID-19 Patients: Implications for Assessment of Post-Acute COVID-19 Syndrome. Journal of Proteome Research. 2021;20(6):3315-29.
7. Mahase E. Covid-19: What do we know about "long covid"? Bmj. 2020;370:m2815.
8. Baig AM. Deleterious Outcomes in Long-Hauler COVID-19: The Effects of SARS-CoV-2 on the CNS in Chronic COVID Syndrome. ACS Chem Neurosci. 2020;11(24):4017-20.
9. Organization WH. A clinical case definition of post COVID-19 condition by a Delphi consensus https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.12021 [
10. Ladds E, Rushforth A, Wieringa S, Taylor S, Rayner C, Husain L, et al. Persistent symptoms after Covid-19: qualitative study of 114 "long Covid" patients and draft quality principles for services. BMC Health Serv Res. 2020;20(1):1144.
11. Carfì A, Bernabei R, Landi F, Group ftGAC-P-ACS. Persistent Symptoms in Patients After Acute COVID-19. JAMA. 2020;324(6):603-5.
12. Rando HM, Bennett TD, Byrd JB, Bramante C, Callahan TJ, Chute CG, et al. Challenges in defining Long COVID: Striking differences across literature, Electronic Health Records, and patient-reported information. medRxiv. 2021:2021.03.20.21253896.
13. Aiyegbusi OL, Hughes SE, Turner G, Rivera SC, McMullan C, Chandan JS, et al. Symptoms, complications and management of long COVID: a review. J R Soc Med. 2021;114(9):428-42.
14. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. [Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff]. Ned Tijdschr Geneeskd. 2005;149(28):1574-8.
15. Luo X, George ML, Kakouras I, Edwards CL, Pietrobon R, Richardson W, et al. Reliability, validity, and responsiveness of the short form 12-item survey (SF-12) in patients with back pain. Spine (Phila Pa 1976). 2003;28(15):1739-45.
16. Vercoulen JH, Swanink CM, Fennis JF, Galama JM, van der Meer JW, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. J Psychosom Res. 1994;38(5):383-92.
17. Worm-Smeitink M, Gielissen M, Bloot L, van Laarhoven HWM, van Engelen BGM, van Riel P, et al. The assessment of fatigue: Psychometric qualities and norms for the Checklist individual strength. J Psychosom Res. 2017;98:40-6.
18. Ponds R VBM, Jolles J. De Cognitive Failure Questionnaire als maat voor subjectief cognitief functioneren. Tijdschrift voor Neuropsychologie. 2006(2):37-45.
19. Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. J Clin Epidemiol. 1998;51(11):1055-68.
20. Williams N. The MRC breathlessness scale. Occup Med (Lond). 2017;67(6):496-7.

21. Menges D, Ballouz T, Anagnostopoulos A, Aschmann HE, Domenghino A, Fehr JS, et al. Burden of post-COVID-19 syndrome and implications for healthcare service planning: A population-based cohort study. PLoS One. 2021;16(7):e0254523.
22. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. J Psychosom Res. 2006;60(6):631-7.
23. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983;67(6):361-70.
24. Morroy G, Peters JB, van Nieuwenhof M, Bor HH, Hautvast JL, van der Hoek W, et al. The health status of Q-fever patients after long-term follow-up. BMC Infect Dis. 2011;11:97.
25. Peters JB, Daudey L, Heijdra YF, Molema J, Dekhuijzen PN, Vercoulen JH. Development of a battery of instruments for detailed measurement of health status in patients with COPD in routine care: the Nijmegen Clinical Screening Instrument. Qual Life Res. 2009;18(7):901-12.
26. Hakkaart-van Roijen L vdLN, Bouwmans C, Kanters T, Tan S. Costing manual: Methodology of costing research and reference prices for economic evaluations in healthcare. Diemen, the Netherlands2015.
27. van Muilekom MM, Luijten MAJ, van Oers HA, Conijn T, Maurice-Stam H, van Goudoever JB, et al. Paediatric patients report lower health-related quality of life in daily clinical practice compared to new normative PedsQL(TM) data. Acta Paediatr. 2021;110(7):2267-79.
28. Varni JW, Beaujean AA, Limbers CA. Factorial invariance of pediatric patient self-reported fatigue across age and gender: a multigroup confirmatory factor analysis approach utilizing the PedsQL™ Multidimensional Fatigue Scale. Qual Life Res. 2013;22(9):2581-94.
29. Gordijn M, Cremers EM, Kaspers GJ, Gemke RJ. Fatigue in children: reliability and validity of the Dutch PedsQL™ Multidimensional Fatigue Scale. Qual Life Res. 2011;20(7):1103-8.
30. Huskisson EC. MEASUREMENT OF PAIN. The Lancet. 1974;304(7889):1127-31.
31. Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. Quality of Life Research. 2010;19(6):875-86.
32. Yeatts KB, Stucky B, Thissen D, Irwin D, Varni JW, DeWitt EM, et al. Construction of the Pediatric Asthma Impact Scale (PAIS) for the Patient-Reported Outcomes Measurement Information System (PROMIS). J Asthma. 2010;47(3):295-302.
33. Irwin DE, Stucky B, Langer MM, Thissen D, Dewitt EM, Lai J-S, et al. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2010;19(4):595-607.
34. Muris P, Meesters C, van den Berg F. The Strengths and Difficulties Questionnaire (SDQ). European Child & Adolescent Psychiatry. 2003;12(1):1-8.
35. van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol. 2007;6:Article25.
36. James G WD, Hastie T, Tibshirani R. An Introduction to Statistical Learning : with Applications in R. New York Springer; 2013.
37. Huang C, Huang L, Wang Y, Li X, Ren L, Gu X, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. Lancet. 2021;397(10270):220-32.
38. Mittal J, Ghosh A, Bhatt SP, Anoop S, Ansari IA, Misra A. High prevalence of post COVID-19 fatigue in patients with type 2 diabetes: A case-control study. Diabetes Metab Syndr. 2021;15(6):102302.
39. Khunti K, Davies MJ, Kosiborod MN, Nauck MA. Long COVID - metabolic risk factors and novel therapeutic management. Nat Rev Endocrinol. 2021;17(7):379-80.
40. Blomberg B, Mohn KG, Brokstad KA, Zhou F, Linchausen DW, Hansen BA, et al. Long COVID in a prospective cohort of home-isolated patients. Nat Med. 2021;27(9):1607-13.
41. Cortinovis M, Perico N, Remuzzi G. Long-term follow-up of recovered patients with COVID-19. Lancet. 2021;397(10270):173-5.

42. Sandmann F, Tessier E, Lacy J, Kall M, Van Leeuwen E, Charlett A, et al. Long-term health-related quality of life in non-hospitalised COVID-19 cases with confirmed SARS-CoV-2 infection in England: Longitudinal analysis and cross-sectional comparison with controls. medRxiv. 2021:2021.10.22.21264701.

*Figure 1: Recruitment of participants in the LongCOVID-study*



STROBE Statement—checklist of items that should be included in reports of observational studies

| | Item No. | Recommendation | Page No. | Relevant text from manuscript |
|---|---|---|---|---|
| **Title and abstract** | 1 | (*a*) Indicate the study's design with a commonly used term in the title or the abstract | 1 | Observational cohort |
| | | (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found | 2 | Protocol for an observation cohort study, hence details of what will be done are provided in the abstract as well as in the manuscript |
| **Introduction** | | | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported | 3-4 | |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses | 4 | |
| **Methods** | | | | |
| Study design | 4 | Present key elements of study design early in the paper | 4 | |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | 4-5 | |
| Participants | 6 | (*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up<br>*Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls<br>*Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants | 4-6 | |
| | | (*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed<br>*Case-control study*—For matched studies, give matching criteria and the number of controls per case | | |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | 5-9 | |
| Data sources/ measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | *5 -9* | |

| | | | |
|---|---|---|---|
| Bias | 9 | Describe any efforts to address potential sources of bias | The study makes use of two controls groups (negative controls and population controls) in an attempt to minimize selction bias. Questionnaires are completed every 3 months in an attempt to minimise recall bias. |
| Study size | 10 | Explain how the study size was arrived at | Our study is an ongoing prospective cohort study |

Continued on next page

| | | | | |
|---|---|---|---|---|
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | 10-11 | |
| Statistical methods | 12 | (*a*) Describe all statistical methods, including those used to control for confounding | 10-11 | |
| | | (*b*) Describe any methods used to examine subgroups and interactions | 10-11 | |
| | | (*c*) Explain how missing data were addressed | 11 | |
| | | (*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed<br>*Case-control study*—If applicable, explain how matching of cases and controls was addressed<br>*Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy | | |
| | | (*e*) Describe any sensitivity analyses | 11 | |
| **Results** | | | | |
| Participants | 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed | | Protocol paper, no results yet. |
| | | (b) Give reasons for non-participation at each stage | | Protocol paper, no results yet |
| | | (c) Consider use of a flow diagram | | |
| Descriptive data | 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders | | Protocol paper, no results yet |
| | | (b) Indicate number of participants with missing data for each variable of interest | | Protocol paper, no results yet |
| | | (c) *Cohort study*—Summarise follow-up time (eg, average and total amount) | | Protocol paper, no results yet |
| Outcome data | 15* | *Cohort study*—Report numbers of outcome events or summary measures over time | | Protocol paper, no results yet |
| | | *Case-control study*—Report numbers in each exposure category, or summary measures of exposure | | |
| | | *Cross-sectional study*—Report numbers of outcome events or summary measures | | |
| Main results | 16 | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included | | Protocol paper, no results yet |
| | | (*b*) Report category boundaries when continuous variables were categorized | | Protocol paper, no results yet |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | | Protocol paper, no results yet |

Continued on next page

| | | | |
|---|---|---|---|
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses | Protocol paper, no results yet |
| **Discussion** | | | |
| Key results | 18 | Summarise key results with reference to study objectives | Protocol paper, no results yet |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias | 12 |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence | Protocol paper, no results yet |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results | Protocol paper, no results yet |
| **Other information** | | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | 13 |

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

# Prevalence and determinants of persistent symptoms after infection with SARS-CoV-2: protocol for an observational cohort study (LongCOVID-study)



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2022-062439.R1 |
| Article Type: | Protocol |
| Date Submitted by the Author: | 02-Jun-2022 |
| Complete List of Authors: | Mutubuki, Elizabeth; RIVM<br>van der Maaden, Tessa; RIVM<br>Leung, Ka Yin; RIVM<br>Wong, Albert; RIVM<br>Tulen, Anna D.; RIVM<br>de Bruijn, Siméon; RIVM<br>Haverman, Lotte; Amsterdam Universitair Medische Centra<br>Knoop, H.; Amsterdam UMC Locatie VUmc, Medical Psychology<br>Franz, Eelco; RIVM<br>van Hoek, Albert Jan; RIVM,<br>van den Wijngaard, Cees; RIVM |
| Primary Subject Heading: | Infectious diseases |
| Secondary Subject Heading: | Epidemiology, Public health |
| Keywords: | COVID-19, EPIDEMIOLOGY, HEALTH ECONOMICS |
| | |

# Prevalence and determinants of persistent symptoms after infection with SARS-CoV-2: protocol for an observational cohort study (LongCOVID-study)

Elizabeth N. Mutubuki[1], Tessa van der Maaden[1], Ka Yin Leung[1], Albert Wong[1], Anna D. Tulen[1], Siméon de Bruijn[1], Lotte Haverman[2], Hans Knoop[3], Eelco Franz[1], Albert Jan van Hoek[1], Cees C. van den Wijngaard[1]

[1]National Institute for Public Health and the Environment (RIVM), Center for Infectious Disease Control, Bilthoven, the Netherlands

[2]Emma Children's Hospital, Amsterdam UMC, University of Amsterdam, Child and Adolescent Psychiatry & Psychosocial Care, Amsterdam Reproduction and Development, Amsterdam Public Health, Amsterdam, the Netherlands.

[3]Department of Medical Psychology, Amsterdam University Medical Center, University of Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, the Netherlands.

**Correspondence to:** Elizabeth N Mutubuki

Email address: elizabeth.mutubuki@rivm.nl



## Abstract

**Introduction:** A substantial proportion of individuals infected with severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), report persisting symptoms weeks and months following acute infection. Estimates on prevalence vary due to differences in study designs, populations, heterogeneity of symptoms and the way symptoms are measured. Common symptoms include fatigue, cognitive impairment and dyspnea. However, knowledge regarding the nature and risk factors for developing persisting symptoms is still limited. Hence in this study we aim to determine the prevalence, severity, risk factors and impact on quality of life of persisting symptoms in the first year following acute SARS-CoV-2 infection.

**Methods and analysis:** The LongCOVID-study is both a prospective and retrospective cohort study being conducted in the Netherlands**,** with a one year follow up. Participants aged 5 years and above, with self-reported positive or negative tests for SARS-CoV-2 will be included in the study. The primary outcome is the prevalence and severity of persistent symptoms in participants that tested positive for SARS-CoV-2 compared to controls. Symptom severity will be assessed for fatigue (Checklist Individual Strength (CIS subscale fatigue severity)), pain (Rand-36/SF-36 subscale bodily

pain), dyspnea (Medical Research Council (mMRC)) and cognitive impairment (Cognitive Failure Questionnaire (CFQ)). Secondary outcomes include effect of vaccination prior to infection on persistent symptoms, loss of health-related quality of life (HRQoL) and risk factors for persisting symptoms following infection with SARS-CoV-2.

**Ethics and dissemination:** The Utrecht Medical Ethics Committee (METC) declared in February 2021 that the Medical Research Involving Human Subjects Act (WMO) does not apply to this study (METC protocol number 21-124/C). Informed consent is required prior to participation in the study. Results of this study will be submitted for publication in a peer-reviewed journal.

**Keywords:** SARS-CoV-2, post COVID-19 condition, LongCovid, prevalence, HRQoL, risk factors.

**Strengths and limitations of this study**

- The prospective design allows for tracking of progression of symptoms, and hence identification of persisting symptoms.
- Having control groups enables identification of symptoms in COVID-19 patients, with prevalence higher than the background prevalence, and prevalence among individuals that likely have another respiratory infection.
- Recruitment of participants from community health testing improves representation of the general population.
- Like many other studies, a limitation of this study is the inability to determine for individuals whether self-reported symptoms are not a result of other illnesses (i.e. background prevalence).
- No serological data is available for cases and controls in order to investigate infections that may go unnoticed.

## Introduction

During the first months of the pandemic, epidemiological research focused primarily on the spread of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) and on treatment of those with severe or fatal illness (1). The effects of SARS-CoV-2 infection vary from asymptomatic infection, through to critical and chronic disease (2). Although most individuals infected with SARS-CoV-2 fully

recover, there is a growing body of evidence suggesting a substantial number of individuals remain with long-term complications or persisting symptoms (3-5).

COVID-19 varies in clinical presentation, disease severity and recovery time (6). A delay in recovery whereby individuals fail to return to their normal daily routines, and still report lasting effects of the infection long after the expected period of recovery, has been termed; ''LongCovid'' (7), ''long-haulers" (8) and "post COVID-19 condition" (PCC) (9). The term PCC will be used in the remainder of this article. PCC is reported to occur in individuals that have a history of probable or confirmed SARS-CoV-2 infection, usually 3 months from the onset of COVID-19, with symptoms lasting at least 2 months that cannot be explained by an alternative diagnosis (9). Symptoms may persist from initial infection, be of new onset following initial recovery from an acute COVID-19 episode or may also fluctuate or relapse over time (10)

In December 2021, more than 263 million confirmed COVID-19 cases had been reported worldwide, and of those, an estimated 10-20% are reported to experience persisting symptoms for weeks or months following acute SARS-CoV2 infection (10). However, higher incidence rates of persisting symptoms have been reported, for example through self-surveys of patient from LongCovid peer support groups (11) as well as in hospitalized patients (12). Variations in the reported incidence and prevalence rates of post COVID-19 condition can be attributed to; the complexity of the syndrome, differences in population groups, heterogeneity in clinical presentation of symptoms, little knowledge regarding the natural history (13) and in the way symptoms are measured.

A good overview of the nature of persisting symptoms following an acute infection with SARS-CoV-2, can enable better diagnosis, management and may reduce negative consequences on health-related quality of life (HRQoL) (13). Hence in this study we aim to determine the prevalence and severity of persisting symptoms in the first year of infection, in individuals infected by SARS-CoV-2 compared to individuals that were not infected. In addition, risk factors for developing post COVID-19 condition and its impact on HRQoL will be analyzed.

## Methods and analysis

### Study aim and design

The LongCOVID-study is an observational cohort study consisting of a prospective, and a retrospective cohort with both data collected in the phase of acute illness and during one year of follow up. The study aims to determine the prevalence, severity, health impact and risk factors associated with persistent symptoms following a SARS-CoV-2 infection, in cases compared to population controls and test-negative controls. The study is carried out by the Dutch National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands. Recruitment of

participants in the study started in May 2021 and is ongoing. Currently there is no set timeline for the completion of recruitment. Participants will be followed up at 3, 6, 9 and 12 months. Follow-up questionnaires can be completed within six weeks from the invitation sent at 3, 6, 9 and 12 months. Analyses will be performed separately for the prospective and retrospective cohorts.

**Study population**

Both the prospective and retrospective cohorts include children (ages 5-17) and adults (18 years and above).

**Inclusion and exclusion criteria**

*Prospective cohort study*

Participants with a positive SARS-CoV-2 infection test result on an antigen or polymerase chain reaction (PCR) test for acute infection, are included in the study as cases, if they complete the baseline questionnaire within 7 days of testing positive regardless of whether or not they had symptoms related to SARS-CoV-2 infection. Participants that test negative to SARS-CoV-2 infection and complete their baseline questionnaire within 7 days of testing negative, are included in the study as test-negative controls. A second group of controls, population controls, consists of randomly selected participants from the Basic Registration of Persons (BRP) without a positive test for SARS-CoV-2 infection or known history of probable infections.

*Retrospective cohort study*

Participants presenting with self-reported persisting symptoms associated with SARS-CoV-2 infection with or without having had a positive test result are included in the retrospective cohort study as self-reported post COVID-19 condition cases.

**Recruitment of participants**

Figure 1 shows the flow diagram of participant recruitment in the LongCOVID-study.

**FIGURE TITLE**

**Figure 1: Recruitment of participants in the LongCOVID-study (14)**

Participants are recruited through the following three ways:

*Via Community Health Testing Services*

Individuals testing positive and negative to COVID-19 at one of the community health testing services (GGDs) in the Netherlands, are invited to participate in the LongCOVID-study. Registration to participate is via the LongCOVID-study website.

*Basic Registration of Persons (BRP)*

Population controls including pediatric controls are *frequency matched to the distribution of age and sex of the cases* randomly selected from the BRP in the Netherlands and invited by letter to participate in the study.

*Self-registered participants*

Individuals interested in participating in the LongCovid-study can also self-register through the study website (longcovid.rivm.nl). Test-negative controls, cases and post COVID-19 condition cases can be included in the study this way.

**Patient and public involvement**

Questionnaires will be tested on a lay public and adjusted according to the feedback given. There will be no further patient or public involvement.

**Measurements in adults**

Table 1 shows different measurement moments when data are collected using questionnaires. At baseline, data on demographical characteristics such as gender, education level and employment are collected. Data on comorbidities is reported at baseline and at 12 months. Information regarding testing for SARS-CoV-2, COVID-19 related complaints and vaccination data is collected at baseline and at 3, 6, 9 and 12 months.

*Health related quality of life (EuroQoL five dimensional instrument(EQ-5D-5L) and Rand-12/SF-12)*

HRQoL regarding long-term symptoms is assessed using the Rand-12/SF-12 in cases and controls. For HRQoL regarding the acute phase of disease, additional weekly measurements using the EQ-5D-5L are carried out in individuals presenting with acute symptoms in the first 8 weeks following a positive COVID-19 test. The EQ-5D-5L questionnaire consists of five dimensions of health (mobility, self-care, usual activities, pain/discomfort, and anxiety/ depression), with five levels of response and a visual analogue scale (EQ VAS). The EQ-5D-5L scores will be converted into utility scores using the Dutch tariff (15), ranging from 0 (death) to 1 (optimal health).

The Rand-12/SF-12, a shortened version of the Rand-36/SF-36 HRQoL questionnaire consists of 12 questions from the following 8 domains: physical functioning, physical role limitations, emotional role limitations, social functioning, physical pain, general mental health, vitality and general health perception. The 8 domains can be summarized into a physical and mental health domain (16). Health scores will be converted into utility scores using the SF-6D (Short-Form Six-Dimension). Quality adjusted life years (QALYs) will be calculated by multiplying the utility scores by the time a patient spends in a given health state.

*Fatigue (Checklist Individual Strength [CIS])*

Fatigue severity is assessed using the subscale fatigue severity of the CIS. The CIS subscale fatigue is a 8-item fatigue questionnaire (17). Each item is scored on a 7-point Likert scale. Scores range from 8 to 56, and scores of 35 and higher indicate severe fatigue (18).

*Cognitive function (Cognitive Failure Questionnaire [CFQ])*

Cognitive function is assessed using the CFQ. The CFQ consists of 25 items that are scored on a 5-point scale ranging from very often to never. Total scores range from 0 to a 100, with higher scores indicating more cognitive impairment (19). A score of 44 or higher indicated clinically significant complaints on cognitive function.

*Pain (bodily pain subscale of the Rand-36/SF-36 Health Status Inventory [Rand-36])*

The bodily pain subscale of the Rand-36 Health Status Inventory (Rand-36) is used to assess pain severity. The Rand-36 scores range from 0 to 100, higher scores indicate better health status. Significant impairment due to pain is reflected by a score of 55 or lower, based on Dutch norm scores (20).

*Dyspnea (Medical Research Council (dyspnea) [mMRC])*

Dyspnea is assessed using the modified mMRC (dyspnea). The mMRC scale ranges from grade 0 to 4. Grade 0- breathless with strenuous exercise; Grade 1- short of breath when hurrying on level ground or walking up a slight hill; Grade-2 walks slower on level ground because of breathlessness or stops for a breath when walking at own pace; Grade 3- stops for breath after walking about 100 yards or after a few minutes on level ground and Grade 4- am too breathless to leave the house or I am breathless when dressing (21). A score of 1 or higher reflects significant impairment due to dyspnea (22).

*Illness and related beliefs (The Brief Illness Perception Questionnaire [Brief IPQ])*

The Brief Illness Perception Questionnaire (Brief IPQ /IPQ-K) is an eight-item scale to assess the cognitive and emotional representations of illness including consequences, timeline, personal control, treatment control, identity, coherence, concern, emotional response and causes (23). Item scores increases, represent linear increases in the dimension measured. The Brief IPQ is reported to have good test-retest reliability (23).

*Anxiety (Hospital Anxiety and Depression Scale [HADS])*

HADS is a 14 item self-report questionnaire designed to measure anxious and depressive states in patients with two subscales (24). The sum score per subscale ranges from 0 to 21. Scores between 0-7 indicate no anxiety or depression, 8-10 mild cases, 11-15 moderate cases and 16 or above severe cases (Snaith 1994).

*Dyspnea (The Nijmegen Clinical Screening Instrument [NCSI])*

The NCSI has four main domains (25), and 8 subdomains (26). Each subdomain is expressed as a single score on its own scale, with higher NCSI scores indicate more problems (25). In the study the subdomain dyspnea will be used.

*Absenteeism (iMTA Productivity Cost Questionnaire [iPCQ])*

Participants will be asked to report the number of days that they have been absent from work due to illness. Absenteeism will be measured using the iPCQ.

*Unpaid Productivity losses and informal care*

Unpaid productivity losses from work, studies, voluntary work as well as informal care will be valued using the Dutch shadow price relevant for that year.

*Acute phase*

Data on HRQoL and acute symptoms will be collected weekly in the first 8 weeks following infection in the prospective cohort. Data collection will stop when the symptoms stop or end at 8 weeks following baseline measurements.

**Measurements in children**

Below are age-specific scales that will be used in children (aged 5-17 years). These differ from some of the above described scales for adults (Table 1).

*Physical function (Pediatric Quality of Life Inventory [PedsQL])*

The PedsQL is a HRQoL measure consisting of 4 subscales (physical functioning, emotional functioning, social functioning and school functioning) which can be computed **into** two summary scores (psychosocial and physical health summary scores). Dutch norms are available which allow comparison with the general population (27). A parent proxy of the PedsQL will be used for children aged 5-7 years.

*Fatigue (Pediatric Quality of Life Inventory Fatigue Scale [PedsQL fatigue]).*

Fatigue severity in children will be assessed with the PedsQL fatigue. This 18-item PedsQL fatigue scale comprises the general fatigue scale (6 items), sleep/rest fatigue scale (6 items), and cognitive fatigue scale (6 items) and is a reliable and valid instrument to measure fatigue in children (28). Dutch norm scores are available (29). A parent proxy will be used for children aged 5-7 years.

*Illness and related beliefs (The Brief Illness Perception Questionnaire [Brief IPQ])*

The Brief IPQ /IPQ-K-parents will be completed by a parent (22), and by the child if they are aged 10 or older.

*Pain Visual analogue scale (VAS)*

Pain severity will be assessed using VAS (30). Scores range from 0 (no pain) to a 100 ( worst imaginable pain). A parent proxy will be used for children aged 5-7 years.

*HRQoL (EQ-5D-Y)*

Weekly measurement moments for up to 8 weeks in children presenting with acute symptoms follow a positive COVID-19 test, will be carried out. The EQ-5D-Y-Proxy1 will be used for children aged 5-7 years, and the EQ-5D-Y will be used for children aged 8-17 years to measure HRQoL. The EQ-5D-Y-Proxy1 and EQ-5D-Y questionnaires consist of five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/ depression), with three levels of response and a visual analogue scale (EQ VAS)(31).

*Dyspnea (Patient-Reported Outcomes Measurement Information System [PROMIS])*

Dyspnea will be assessed in children aged 5-7 years using an adjusted PROMIS Asthma impact Short form Proxy and in children 8-17 using an adjusted PROMIS Asthma impact Short form(32).

*Cognitive function and behaviour (PROMIS and SDQ)*

Loneliness will be assessed in children aged 5-7 years using the PROMIS short form proxy depressive symptoms and in children 8-17 using the PROMIS short form depressive symptoms, for which norm scores are available which allow comparison with the general population (32, 33). In addition the strengths and difficulties questionnaire (SDQ) will be used as well to assess the level of depressive symptoms, with a proxy for parents in the 5-11 years of age (34).

**Outcome measures**

*Primary outcome*

1. The first primary outcome measure is the prevalence and severity of persistent symptoms in patients that test positive for COVID-19 infection compared to both, test-negative and

population controls. Severity of symptoms will be assessed for fatigue, pain, dyspnea, and cognitive impairment using standardized questionnaires, with population-based norm cut-off scores for clinically significant severity.

*Secondary outcomes include:*

1. Effect of vaccination to SARS-CoV-2 at baseline (i.e. before infection) on the prevalence and severity of persistent symptoms after SARS-CoV-2 infection.
2. Factors that predict post COVID-19 condition following an acute SARS-CoV-2 infection at different follow up moments.
3. Healthcare utilization in the first year following infection with SARS-CoV-2 in cases compared to controls (test-negative controls and population controls) will be assessed.
4. HRQoL in cases will be compared to that of controls (test-negative controls and population controls) in the first year following infection. Additionally a comparison will also be made between post COVID-19 condition individuals and individuals that test positive for COVID-19 but do not develop post COVID-19 condition.

**Table 1: Measurement timetable**

| | **Baseline** | *Acute symptoms (in weeks)* | | | | | | | | **3 months** | **6 months** | **9 months** | **12 months** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | | | | |
| Informed consent | X | | | | | | | | | | | | |
| Baseline characteristics | X | | | | | | | | | | | | |
| Vaccination data *(status, type, date)* | X | | | | | | | | | X | X | X | X |
| Health utilization *(contact with healthcare providers, medication use)* | X | | | | | | | | | X | X | X | X |
| Symptom data | X | *X* | *X* | *X* | *X* | *X* | *X* | *X* | *X* | X | X | X | X |
| HRQoL (EQ-5D*/EQ-5D-Y**) | X* | *X* | *X* | *X* | *X* | *X* | *X* | *X* | *X* | X* | X* | X* | X* |
| Health utilization *(contact with healthcare providers, medication use)* | X | | | | | | | | | X | X | X | X |
| Co-morbidities | X | | | | | | | | | | | | X |
| **Adults** | | | | | | | | | | | | | |
| HRQoL *(Rand-12)* | X | | | | | | | | | X | X | X | X |
| Pain, Physical function and Social function *(Rand-36/SF-36 subscales pain, social functioning and physical functioning)* | X | | | | | | | | | X | X | X | X |
| Cognitive function *(CFQ)* | X | | | | | | | | | X | X | X | X |
| Fatigue *(CIS)* | X | | | | | | | | | X | X | X | X |
| Illness and related beliefs *(Brief IPQ)* | X | | | | | | | | | X | X | | |
| Anxiety and depression *(HADS)* | X | | | | | | | | | X | X | | |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dyspnea *(mMRC)* | X | | | | | | | | | X | X | X | X |
| Dyspnea *(NCSI )* | X | | | | | | | | | X | X | X | X |
| **Children** | | | | | | | | | | | | | |
| Physical function *(PedsQL subscale physical health)* | X | | | | | | | | | X | X | X | X |
| Fatigue *(PedsQL fatigue)* | X | | | | | | | | | X | X | X | X |
| Illness and related beliefs *(Brief IPQ / brief IPQ-parents)* | X | | | | | | | | | X | X | | |
| Pain *(VAS)* | X | | | | | | | | | X | X | X | X |
| Dyspnea *(adjusted PROMIS Asthma)* | X | | | | | | | | | X | X | X | X |
| Depressive symptoms (PROMIS, SDQ) | X | | | | | | | | | X | X | **X** | X |

*Measured in adults

**Measured in kids

**Sample size**

The study should have sufficient power to determine whether and which long-term symptoms are more common in COVID-19 patients than in controls. Experience from similar studies shows that around 25% of the population experiences long-term symptoms to some extent (reporting a score indicating fatigue/pain/concentration problems for at least 3 months)(35). With 2000 cases and 1000 test-negative controls, a difference of 5% or more between the prevalence of 25% in controls compared to 30% in COVID cases can be detected with a power slightly above 80% (power 82%; alpha: 0.05). However, recruitment will continue even after the participant counts mentioned above are reached.

**Statistical analysis**

Baseline characteristics of the participants in all groups will be presented using descriptive statistics mean (standard deviation), median (range), or proportion to assess if there is a balance in the groups regarding distribution of prognostic factors such as age, gender, co-morbidity and education. The analyses of the children will initially be conducted separately from those of the adults.

***Prospective study***

**Primary outcome analysis**

1. Prevalence and severity of persistent symptoms in COVID-19 patients

Descriptive epidemiological statistical methods will be used to analyze prevalence of persistent symptoms at 3, 6, 9 and 12 months in cases compared to both control groups (test-negative controls and population controls). Persisting symptoms are defined as symptoms in cases with a duration of

at least 2 months. Such symptoms significantly elevated in cases compared to controls (test-negative controls and/or population controls) during follow up are likely to be associated with COVID-19, and cases with these symptoms are in this study defined as cases with possible PCC condition (yes/no). Severity scores of fatigue, dyspnea, cognitive functioning, and pain will be calculated. Scores of individuals with confirmed COVID-19 will be compared to those of controls, per follow-up moment (baseline, 3, 6, 9 and 12 months follow-up). Analyses will be controlled for age, gender, number of comorbidities and level of education. In a later stage symptom prevalence and severity in post-COVID-19 condition may be compared between different age groups including children vs adults.

**Secondary outcome analysis**

1. Effect of vaccination to SARS-CoV-2 at baseline on the prevalence and severity of persistent symptoms after SARS-CoV-2 infection

To assess the effect of vaccination for SARS-CoV-2 at baseline, prevalence of COVID-related symptoms, will be compared between fully vaccinated cases and cases that were partially vaccinated or unvaccinated at the time of their positive SARS-CoV-2 test.

2. *Predictors of post-COVID 19 condition*

A prediction model will be built to identify predictors of possible post COVID-19 condition at each follow up moment or period separately. The outcome will be having possible post COVID-19 condition as defined above. To determine the prediction model that best suits our data, the prediction model will be constructed using super learning (36). The prediction model will be evaluated using the ROC-AUC metric (37) and analyzed using explainable artificial intelligence (AI), in particular partial dependence plots and variable importance (38). For potential predictors to be included in the model see table S1.

3. Predictors of healthcare utilization in post COVID-19 condition

A second prediction model will be performed to identify predictors of healthcare utilization in post COVID-19 condition. Healthcare utilization is defined as self -reported contact (visit to the general practitioner, telephone call, hospitalization, emergency healthcare services, other medical health professionals/services) with a health provider regarding symptoms attributed by the patient to COVID-19 or post COVID-19 condition (yes/no). The prediction model will be performed as described above and with similar predictors except for contact with the GP.

4. Quality-adjusted life-years

HRQoL will be assessed using EQ-5D-5L and Rand-12/SF-6D. QALYs, which take into account both the impact of length and the quality-of-life will be calculated and be compared between cases and controls.

***Retrospective cohort***

Descriptive epidemiological statistical methods will be used to analyze the prevalence of persistent symptoms at baseline in cases in the retrospective cohort compared to both control groups (test-negative controls and population controls). Moreover, prevalence of co-morbidities will be quantified in cases and control groups. Additionally, an assessment into healthcare utilization for cases will be performed according to the aforementioned definition.

***Acute data following* SARS CoV-2 *infection***

Descriptive epidemiological statistical methods will be used to describe the prevalence and the type of symptoms present following acute infection as well as HRQoL.

**Missing data**

The fraction of missing questionnaires at each time point and per period during the study (e.g. per 3 months) in all patients with confirmed COVID-19 will be tabulated. Scenarios of dealing with missing data include a complete case analysis, multiple imputation, and linear interpolation combined with carry forward.

## Ethics and dissemination

All participants are required to consent to participate in the study. Informed consent will be obtained online prior to completing the study questionnaires. The Utrecht Medical Ethics Committee (METC) declared in February 2021 that the Medical Research Involving Human Subjects Act (WMO) does not apply to this study (METC protocol number 21-124/C). Results of this study will be submitted for publication in a peer-reviewed journal.

## Discussion

The LongCOVID-study aims to determine the prevalence and severity of persistent symptoms following acute SARS-CoV2 infection in cases compared to controls, as well as to investigate the risk factors for developing persistent symptoms. Previous studies have explored prevalence of long-term symptoms and risk factors in various populations, i.e. in previously hospitalized patients (39), patients with diabetes type 1 and 2 (40), in home-isolated patients with milder symptoms and in the young (41).

Blomberg et al. reported that 61% of all the patients had persisting symptoms at 6 months (41). This included patients with a mild to moderate illness following infection as well as young patients (16-30

years). Persisting symptoms included loss of taste and or smell, fatigue, dyspnea, impaired concentration and memory problems. In a hospitalized population (39), fatigue, muscle weakness, sleep difficulties and anxiety or depression were the most prevalent symptoms at 6 months. Due to severe illness during hospital stay and impaired pulmonary function, the hospitalized population is a target group for long-term recovery (39)). We expect that our study will include participants with mostly mild to moderate acute symptoms and fewer patients that are hospitalized. This is due to the design of the study, which enables recruitment from community health testing services, where people go when they do not have severe disease. Therefore, our study is complementary to studies with a focus on hospitalized patients, and more reflective of the impact of long-term symptoms in patients with an initially relatively mild COVID-19.

Strengths of the current study include the prospective design, allowing for detailed analysis of the prevalence and risk factors of persistent symptoms of SARS CoV-2 infection. In addition this study is one of a few studies (42) that allows for comparison between COVID-19 cases and control groups that have similar experiences, such as lock down measures. This is important because such factors can influence complaints. The availability of the population control group in this study allows us to control for background prevalence of symptoms. Although a negative COVID-19 test does not confirm infection by another respiratory pathogen, the use of test-negative controls gives us the opportunity to assess to what extent the long-term symptoms after testing positive for COVID-19 are more prevalent or severe than in a control group with acute symptoms that tests negative for COVID-19. Another strength of this study is the recruitment of participants from the nationwide community health testing centers, which enable a better representation of the general population. The use of validated questionnaires with validated cut-off scores for severity is another strength of this study. Repeated assessment of symptoms every three months during one year of follow-up will enable assessment of the time course of symptoms, and detection of disabling symptoms at every 3 months interval. Furthermore, the impact of symptoms on general functioning will be assessed. A limitation of this study is that severity scores of only four of symptoms associated with COVID-19 will be calculated to get more insight into clinical significance. This is because only four standardized questionnaires for symptom severity were included in the study. Hence the severity of other possible symptoms related to COVID-19 will not be considered. Another limitation of this study is the risk of lost to follow-up. Hence, we will perform several alternative substitution methods for missing data to check the robustness of our results. The inability to determine for individual participants whether self-reported symptoms are not as a result of other illnesses, is also a limitation in this study. In addition, no serological data is available in this study to investigate infections that may go unnoticed.

In conclusion, the LongCOVID-study is expected to provide additional insights into the prevalence and severity of persistent symptoms after SARS CoV-2 infection to the international body of literature. In the Netherlands this is the first large scale study on persisting symptoms following SARS CoV2 infection.

**Contributors:** ENM wrote the manuscript. KYL contributed to the methods. CCW, TM, AJH, ADT, KYL, ENM, AW contributed to the design of the study, LH and HK advised on the design of the questionnaires, TM, AJH, ADT, KYL, EF, SB, CCW contributed to implementation of the data collection, all authors reviewed and contributed to drafts of the manuscript. All authors read, contributed to refinement of the study protocol and approved the manuscript.

**Funding:** The study is executed by the National Institute for Public Health by order of the Ministry of Health. This means the study is not the result of a competitive grant. The Dutch Ministry of Health, Welfare and Sport does not have a role in the design of this study, its execution, analyses and interpretation of results.

**Competing interests:** The authors have no conflict of interest to declare.

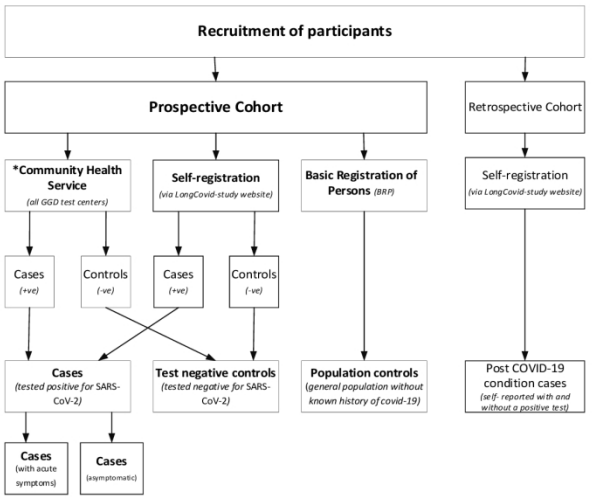
## References


1. Alwan NA. Surveillance is underestimating the burden of the COVID-19 pandemic. Lancet. 2020;396(10252):e24.
2. Ricke DO. Two Different Antibody-Dependent Enhancement (ADE) Risks for SARS-CoV-2 Antibodies. Front Immunol. 2021;12:640093.
3. Cirulli ET, Schiabor Barrett KM, Riffle S, Bolze A, Neveux I, Dabe S, et al. Long-term COVID-19 symptoms in a large unselected population. medRxiv. 2020:2020.10.07.20208702.
4. Petersen MS, Kristiansen MF, Hanusson KD, Danielsen ME, B ÁS, Gaini S, et al. Long COVID in the Faroe Islands: A Longitudinal Study Among Nonhospitalized Patients. Clin Infect Dis. 2021;73(11):e4058-e63.
5. Ziauddeen N, Gurdasani D, O'Hara ME, Hastie C, Roderick P, Yao G, et al. Characteristics of Long Covid: findings from a social media survey. medRxiv. 2021:2021.03.21.21253968.
6. Holmes E, Wist J, Masuda R, Lodge S, Nitschke P, Kimhofer T, et al. Incomplete Systemic Recovery and Metabolic Phenoreversion in Post-Acute-Phase Nonhospitalized COVID-19 Patients: Implications for Assessment of Post-Acute COVID-19 Syndrome. Journal of Proteome Research. 2021;20(6):3315-29.
7. Mahase E. Covid-19: What do we know about "long covid"? Bmj. 2020;370:m2815.
8. Baig AM. Deleterious Outcomes in Long-Hauler COVID-19: The Effects of SARS-CoV-2 on the CNS in Chronic COVID Syndrome. ACS Chem Neurosci. 2020;11(24):4017-20.
9. Organization WH. A clinical case definition of post COVID-19 condition by a Delphi consensus https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.12021 [
10. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV. A clinical case definition of post-COVID-19 condition by a Delphi consensus. Lancet Infect Dis. 2022;22(4):e102-e7.
11. Ladds E, Rushforth A, Wieringa S, Taylor S, Rayner C, Husain L, et al. Persistent symptoms after Covid-19: qualitative study of 114 "long Covid" patients and draft quality principles for services. BMC Health Serv Res. 2020;20(1):1144.
12. Carfì A, Bernabei R, Landi F, Group ftGAC-P-ACS. Persistent Symptoms in Patients After Acute COVID-19. JAMA. 2020;324(6):603-5.
13. Rando HM, Bennett TD, Byrd JB, Bramante C, Callahan TJ, Chute CG, et al. Challenges in defining Long COVID: Striking differences across literature, Electronic Health Records, and patient-reported information. medRxiv. 2021:2021.03.20.21253896.
14. (RIVM) RvVeM. Risicofactoren voor COVID-19 (CONTEST) 2021 [Available from: https://www.rivm.nl/coronavirus-covid-19/onderzoek/risicofactoren-contest-onderzoek.
15. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. [Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff]. Ned Tijdschr Geneeskd. 2005;149(28):1574-8.
16. Luo X, George ML, Kakouras I, Edwards CL, Pietrobon R, Richardson W, et al. Reliability, validity, and responsiveness of the short form 12-item survey (SF-12) in patients with back pain. Spine (Phila Pa 1976). 2003;28(15):1739-45.
17. Vercoulen JH, Swanink CM, Fennis JF, Galama JM, van der Meer JW, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. J Psychosom Res. 1994;38(5):383-92.
18. Worm-Smeitink M, Gielissen M, Bloot L, van Laarhoven HWM, van Engelen BGM, van Riel P, et al. The assessment of fatigue: Psychometric qualities and norms for the Checklist individual strength. J Psychosom Res. 2017;98:40-6.
19. Ponds R VBM, Jolles J. De Cognitive Failure Questionnaire als maat voor subjectief cognitief functioneren. Tijdschrift voor Neuropsychologie. 2006(2):37-45.
20. Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. J Clin Epidemiol. 1998;51(11):1055-68.
21. Williams N. The MRC breathlessness scale. Occup Med (Lond). 2017;67(6):496-7.

22. Menges D, Ballouz T, Anagnostopoulos A, Aschmann HE, Domenghino A, Fehr JS, et al. Burden of post-COVID-19 syndrome and implications for healthcare service planning: A population-based cohort study. PLoS One. 2021;16(7):e0254523.
23. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. J Psychosom Res. 2006;60(6):631-7.
24. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983;67(6):361-70.
25. Morroy G, Peters JB, van Nieuwenhof M, Bor HH, Hautvast JL, van der Hoek W, et al. The health status of Q-fever patients after long-term follow-up. BMC Infect Dis. 2011;11:97.
26. Peters JB, Daudey L, Heijdra YF, Molema J, Dekhuijzen PN, Vercoulen JH. Development of a battery of instruments for detailed measurement of health status in patients with COPD in routine care: the Nijmegen Clinical Screening Instrument. Qual Life Res. 2009;18(7):901-12.
27. van Muilekom MM, Luijten MAJ, van Oers HA, Conijn T, Maurice-Stam H, van Goudoever JB, et al. Paediatric patients report lower health-related quality of life in daily clinical practice compared to new normative PedsQL(TM) data. Acta Paediatr. 2021;110(7):2267-79.
28. Varni JW, Beaujean AA, Limbers CA. Factorial invariance of pediatric patient self-reported fatigue across age and gender: a multigroup confirmatory factor analysis approach utilizing the PedsQL™ Multidimensional Fatigue Scale. Qual Life Res. 2013;22(9):2581-94.
29. Gordijn M, Cremers EM, Kaspers GJ, Gemke RJ. Fatigue in children: reliability and validity of the Dutch PedsQL™ Multidimensional Fatigue Scale. Qual Life Res. 2011;20(7):1103-8.
30. Huskisson EC. MEASUREMENT OF PAIN. The Lancet. 1974;304(7889):1127-31.
31. Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. Quality of Life Research. 2010;19(6):875-86.
32. Yeatts KB, Stucky B, Thissen D, Irwin D, Varni JW, DeWitt EM, et al. Construction of the Pediatric Asthma Impact Scale (PAIS) for the Patient-Reported Outcomes Measurement Information System (PROMIS). J Asthma. 2010;47(3):295-302.
33. Irwin DE, Stucky B, Langer MM, Thissen D, Dewitt EM, Lai J-S, et al. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2010;19(4):595-607.
34. Muris P, Meesters C, van den Berg F. The Strengths and Difficulties Questionnaire (SDQ). European Child & Adolescent Psychiatry. 2003;12(1):1-8.
35. Vrijmoeth HD, Ursinus J, Harms MG, Zomer TP, Gauw SA, Tulen AD, et al. Prevalence and determinants of persistent symptoms after treatment for Lyme borreliosis: study protocol for an observational, prospective cohort study (LymeProspect). BMC Infect Dis. 2019;19(1):324.
36. van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol. 2007;6:Article25.
37. James G WD, Hastie T, Tibshirani R. An Introduction to Statistical Learning : with Applications in R. New York Springer; 2013.
38. Molnar C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. 2nd ed. ed2022.
39. Huang C, Huang L, Wang Y, Li X, Ren L, Gu X, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. Lancet. 2021;397(10270):220-32.
40. Mittal J, Ghosh A, Bhatt SP, Anoop S, Ansari IA, Misra A. High prevalence of post COVID-19 fatigue in patients with type 2 diabetes: A case-control study. Diabetes Metab Syndr. 2021;15(6):102302.
41. Blomberg B, Mohn KG, Brokstad KA, Zhou F, Linchausen DW, Hansen BA, et al. Long COVID in a prospective cohort of home-isolated patients. Nat Med. 2021;27(9):1607-13.
42. Sandmann F, Tessier E, Lacy J, Kall M, Van Leeuwen E, Charlett A, et al. Long-term health-related quality of life in non-hospitalised COVID-19 cases with confirmed SARS-CoV-2 infection in

England: Longitudinal analysis and cross-sectional comparison with controls. medRxiv. 2021:2021.10.22.21264701.

*Figure 1: Recruitment of participants in the LongCOVID-study*




*Cases and test negative controls were recruited via the community health services as part of the CONTEST study. Cases were also recruited via the community health services as part of Test and Trace.

Figure 1: Recruitment of participants in the LongCOVID-study

531x752mm (79 x 79 DPI)

**Supplementary files :** S1 potential predictors

| *Potential predictor* | *Measuring instrument/responses* |
|---|---|
| Health related quality of life (HRQoL) | EQ-5D-5L |
| Pain | SF-36 subscale bodily pain |
| Education level | low/moderate/high |
| Body mass index (BMI) | Age/weight |
| Employment | yes/no |
| Recurrent infection with coronavirus | yes/no/unknown |
| Age | years |
| Gender | male/female |
| Nationality | Dutch/non-Dutch |
| Smoking | never smoked/former smoker/current smoker |
| Region of residence | North/South/West/East |
| Number of household members | live alone/2/3-4/more than 4 |
| Self-reported COVID-19 vaccination status | fully vaccinated/not vaccinated/partly vaccinated |
| Type of vaccination received | Moderna/AstraZeneca/Pfizer/Jansen/ combination |
| Received a flu vaccination in the autumn of 2020 | yes/no/unknown |
| Received pneumococcal vaccination | yes/no/unknown |
| Received pneumococcal vaccination | yes/no/unknown |
| Hospitalized due to COVID-19 infection | yes/no |
| Use of medication | yes/no |
| Depressive symptoms | yes/no |
| Anxiety symptoms | yes/no |
| Pregnancy | yes/no/not applicable |
| Number of comorbidities | none/one/more than one |
| Respiratory comorbidities | yes/no |
| Cardiovascular comorbidities | yes/no |
| Hypertension | yes/no |
| Physical activity | IPAQ number of sitting hours |
| Test location | hospital/self-test/ GGD/other |
| *Contact with the general practitioner (GP) | yes/no |
| Illness perception | brief IPQ |

*Contact with GP is excluded as a potential predictor in the prediction model for healthcare utilization in post COVID-19 condition

STROBE Statement—checklist of items that should be included in reports of observational studies

| | Item No. | Recommendation | Page No. | Relevant text from manuscript |
|---|---|---|---|---|
| **Title and abstract** | 1 | (*a*) Indicate the study's design with a commonly used term in the title or the abstract | 1 | Observational cohort |
| | | (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found | 2 | Protocol for an observation cohort study, hence details of what will be done are provided in the abstract as well as in the manuscript |
| **Introduction** | | | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported | 3-4 | |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses | 4 | |
| **Methods** | | | | |
| Study design | 4 | Present key elements of study design early in the paper | 4 | |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | 4-5 | |
| Participants | 6 | (*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up<br>*Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls<br>*Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants | 4-6 | |
| | | (*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed<br>*Case-control study*—For matched studies, give matching criteria and the number of controls per case | | |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | 5-9 | |
| Data sources/ measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | *5 -9* | |

| | | | |
|---|---|---|---|
| Bias | 9 | Describe any efforts to address potential sources of bias | The study makes use of two controls groups (negative controls and population controls) in an attempt to minimize selction bias. Questionnaires are completed every 3 months in an attempt to minimise recall bias. |
| Study size | 10 | Explain how the study size was arrived at | Our study is an ongoing prospective cohort study |

Continued on next page

| | | | | |
|---|---|---|---|---|
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | 10-11 | |
| Statistical methods | 12 | (*a*) Describe all statistical methods, including those used to control for confounding | 10-11 | |
| | | (*b*) Describe any methods used to examine subgroups and interactions | 10-11 | |
| | | (*c*) Explain how missing data were addressed | 11 | |
| | | (*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed<br>*Case-control study*—If applicable, explain how matching of cases and controls was addressed<br>*Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy | | |
| | | (*e*) Describe any sensitivity analyses | 11 | |
| **Results** | | | | |
| Participants | 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed | | Protocol paper, no results yet. |
| | | (b) Give reasons for non-participation at each stage | | Protocol paper, no results yet |
| | | (c) Consider use of a flow diagram | | |
| Descriptive data | 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders | | Protocol paper, no results yet |
| | | (b) Indicate number of participants with missing data for each variable of interest | | Protocol paper, no results yet |
| | | (c) *Cohort study*—Summarise follow-up time (eg, average and total amount) | | Protocol paper, no results yet |
| Outcome data | 15* | *Cohort study*—Report numbers of outcome events or summary measures over time | | Protocol paper, no results yet |
| | | *Case-control study*—Report numbers in each exposure category, or summary measures of exposure | | |
| | | *Cross-sectional study*—Report numbers of outcome events or summary measures | | |
| Main results | 16 | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included | | Protocol paper, no results yet |
| | | (*b*) Report category boundaries when continuous variables were categorized | | Protocol paper, no results yet |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | | Protocol paper, no results yet |

Continued on next page

| | | | |
|---|---|---|---|
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses | Protocol paper, no results yet |
| **Discussion** | | | |
| Key results | 18 | Summarise key results with reference to study objectives | Protocol paper, no results yet |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias | 12 |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence | Protocol paper, no results yet |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results | Protocol paper, no results yet |
| **Other information** | | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | 13 |

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.