


BMJ Open 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 Wijngaard¹

To cite: Mutubuki EN, van der Maaden T, Leung KY, *et al*. Prevalence and determinants of persistent symptoms after infection with SARS-CoV-2: protocol for an observational cohort study (LongCOVID-study). *BMJ Open* 2022;**12**:e062439. doi:10.1136/bmjopen-2022-062439

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-062439>).

Received 03 May 2022

Accepted 13 June 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Elizabeth N Mutubuki;
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 dyspnoea. 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 with controls. Symptom severity will be assessed for fatigue (Checklist Individual Strength (CIS subscale fatigue severity)), pain (Rand-36/SF-36 subscale bodily pain), dyspnoea (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.

INTRODUCTION

During the first months of the pandemic, epidemiological research focused primarily

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 patients with COVID-19, 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 (ie, background prevalence).
- ⇒ No serological data are available for cases and controls in order to investigate infections that may go unnoticed.

on the spread of 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 that 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 'long Covid',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-CoV-2 infection.¹⁰ However, higher incidence rates of persisting symptoms have been reported, for example through self-surveys of patient from long COVID peer support groups¹¹ as well as in hospitalised 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 history¹³ 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 with individuals that were not infected. In addition, risk factors for developing post-COVID-19 condition and its impact on HRQoL will be analysed.

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 with 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 6 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).

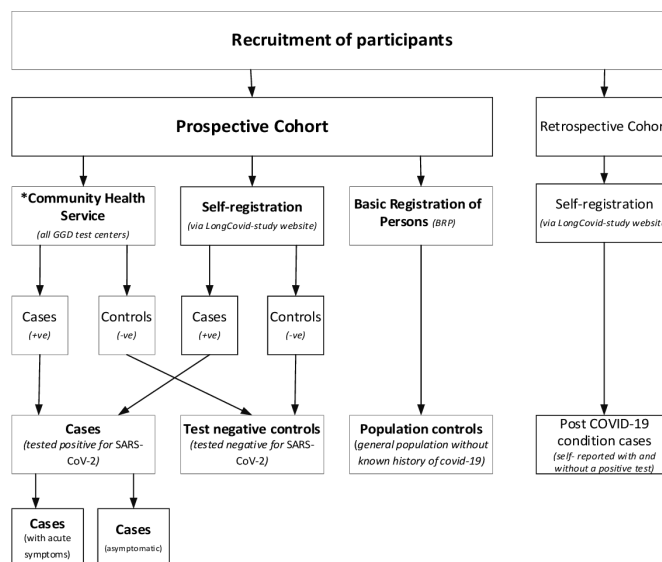


Figure 1 Recruitment of participants in the LongCOVID-study.⁴²

Inclusion and exclusion criteria

Prospective cohort study

Participants with a positive SARS-CoV-2 infection test result on an antigen or 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. 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 paediatric 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 are 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,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 eight domains: physical functioning, physical role limitations, emotional role limitations, social functioning, physical pain, general mental health, vitality and general health perception. The eight domains can be summarised 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 (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 an 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 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.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

Dyspnoea (Medical Research Council (dyspnoea) (mMRC))

Dyspnoea is assessed using the modified mMRC (dyspnoea). 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; grade 4—I am too breathless to leave the house or I am breathless when dressing.²⁰ A score of 1 or higher reflects significant impairment due to dyspnoea.²¹

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.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 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 and 7 indicate no anxiety or depression, 8–10 mild cases, 11–15 moderate cases and 16 or above severe cases.²³

Dyspnoea (The Nijmegen Clinical Screening Instrument (NCSI))

The NCSI has four main domains²⁴ and eight subdomains.²⁵ Each subdomain is expressed as a single score on its own scale, with higher NCSI scores indicating more problems.²⁴ In this study, the subdomain dyspnoea 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.

Table 1 Measurement timetable

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

Continued

Table 1 Continued

	Baseline	<i>Acute symptoms (in weeks)</i>								3 months	6 months	9 months	12 months
		1	2	3	4	5	6	7	8				
Depressive symptoms (PROMIS, SDQ)	X									X	X	X	X
*Measured in adults only. †Measured in kids only.													

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 previously described scales for adults ([table 1](#)).

Physical function (Pediatric Quality of Life Inventory (PedsQL))

The PedsQL is a HRQoL measure consisting of four 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.²⁶ 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 (six items), sleep/rest fatigue scale (six items) and cognitive fatigue scale (six items), and is a reliable and valid instrument to measure fatigue in children.²⁷ Dutch norm scores are available.²⁸ 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,²¹ and by the child if they are aged 10 or older.

Pain visual analogue scale (VAS)

Pain severity will be assessed using VAS.²⁹ 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, following 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).³⁰

Dyspnoea (Patient-Reported Outcomes Measurement Information System (PROMIS))

Dyspnoea will be assessed in children aged 5–7 years using an adjusted PROMIS Asthma impact Short form Proxy and in children aged 8–17 years using an adjusted PROMIS Asthma impact Short form.³¹

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 aged 8–17 years using the PROMIS short form depressive symptoms, for which norm scores are available which allow comparison with the general population.^{31 32} 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.³³

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

Secondary outcomes

1. Effect of vaccination to SARS-CoV-2 at baseline (ie, 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 utilisation in the first year following infection with SARS-CoV-2 in cases compared with controls (test-negative controls and population controls) will be assessed.
4. HRQoL in cases will be compared with that of controls (test-negative controls and population controls) in the first year following infection. In addition, 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.

Sample size

The study should have sufficient power to determine whether and which long-term symptoms are more common in patients with COVID-19 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).³⁴ With 2000 cases and 1000 test-negative controls, a difference of 5% or more between the prevalence of 25% in controls compared with 30% in COVID-19 cases can be detected with a power slightly above 80% (power 82%; alpha: 0.05). However, recruitment will continue even after the participant counts previously mentioned are reached.

Statistical analysis

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

Prospective study

Primary outcome analysis

Prevalence and severity of persistent symptoms in patients with COVID-19

Descriptive epidemiological statistical methods will be used to analyse prevalence of persistent symptoms at 3, 6, 9 and 12 months in cases compared with 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 with 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, dyspnoea, cognitive functioning and pain will be calculated. Scores of individuals with confirmed COVID-19 will be compared with those of controls, per follow-up moment (baseline, 3, 6, 9 and 12 months of 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

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.

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 previously. 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 metric³⁶ and analysed using explainable artificial intelligence (AI), in particular partial dependence plots and variable importance.³⁷ For potential predictors to be included in the model see online supplemental table S1.

Predictors of healthcare utilisation in post-COVID-19 condition

A second prediction model will be performed to identify predictors of healthcare utilisation in post-COVID-19 condition. Healthcare utilisation is defined as self-reported contact (visit to the general practitioner, telephone call, hospitalisation, 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 previously and with similar predictors except for contact with the GP.

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 analyse 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 comorbidities will be quantified in cases and control groups. Additionally, an assessment into healthcare utilisation 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 (eg, 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-CoV-2 infection in cases compared with controls, as well as to investigate the risk factors for developing persistent symptoms. Previous studies have explored the prevalence of long-term symptoms and risk factors in various populations, that is, in previously hospitalised patients,³⁸ patients with diabetes type 1 and 2,³⁹ in home-isolated patients with milder symptoms and in the young.⁴⁰

Blomberg *et al* reported that 61% of all the patients had persisting symptoms at 6 months.⁴⁰ 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, dyspnoea, impaired concentration and memory problems. In a hospitalised population,³⁸ 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 hospitalised population is a target group for long-term recovery.³⁸ We expect that our study will include participants with mostly mild to moderate acute symptoms and fewer patients that are hospitalised. 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 hospitalised 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 studies⁴¹ that allows for comparison between COVID-19 cases and control groups that have similar experiences, such as lockdown 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 centres,

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 3 months during 1 year of follow-up will enable assessment of the time course of symptoms, and detection of disabling symptoms at every 3-month 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 standardised 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 are 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-CoV-2 infection.

Author affiliations

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

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

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

Contributors ENM wrote the manuscript. KYL contributed to the methods. CCW, TM, AJH, ADT, KYL, ENM and AW contributed to the design of the study. LH and HK advised on the design of the questionnaires. TM, AJH, ADT, KYL, EF, SB and CCW contributed to the 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.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants, but The Utrecht Medical Ethics Committee (METC) (METC protocol number 21-124/C) exempted this study. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability

of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Elizabeth N Mutubuki <http://orcid.org/0000-0001-6471-1432>

REFERENCES

- Alwan NA. Surveillance is underestimating the burden of the COVID-19 pandemic. *Lancet* 2020;396:e24.
- Ricke DO. Two different antibody-dependent enhancement (ADE) risks for SARS-CoV-2 antibodies. *Front Immunol* 2021;12:640093.
- Cirulli ET, Schiabor Barrett KM, Riffle S. Long-Term COVID-19 symptoms in a large unselected population. *medRxiv* 2020:2020.10.07.20208702.
- Petersen MS, Kristiansen MF, Hanusson KD, et al. Long COVID in the Faroe Islands: a longitudinal study among nonhospitalized patients. *Clin Infect Dis* 2021;73:e4058–63.
- Ziauddeen N, Gurdasani D, O'Hara ME. Characteristics of long Covid: findings from a social media survey. *medRxiv* 2021:2021.03.21.21253968.
- Holmes E, Wist J, Masuda R, 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. *J Proteome Res* 2021;20:3315–29.
- Mahase E. Covid-19: What do we know about "long covid"? *BMJ* 2020;370:m2815.
- 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;11:4017–20.
- Organization WH. A clinical case definition of post COVID-19 condition by a Delphi consensus. Available: https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.12021
- Soriano JB, Murthy S, Marshall JC, et al. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis* 2022;22:e102–7.
- Ladds E, Rushforth A, Wieringa S, 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;20:1144.
- Carfi A, Bernabei R, Landi F, et al. Persistent symptoms in patients after acute COVID-19. *JAMA* 2020;324:603–5.
- Rando HM, Bennett TD, Byrd JB, et al. Challenges in defining long COVID: striking differences across literature, electronic health records, and patient-reported information. *medRxiv* 2021:2021.03.20.21253896.
- Lamers LM, Stalmeier PFM, McDonnell J. [Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff]. *Ned Tijdschr Geneesk* 2005;149:1574–8.
- Luo X, George ML, Kakouras I, et al. Reliability, validity, and responsiveness of the short form 12-item survey (SF-12) in patients with back pain. *Spine* 2003;28:1739–45.
- Vercoulen JH, Swanink CM, Fennis JF, et al. Dimensional assessment of chronic fatigue syndrome. *J Psychosom Res* 1994;38:383–92.
- Worm-Smeitink M, Gielissen M, Bloot L, et al. The assessment of fatigue: psychometric qualities and norms for the checklist individual strength. *J Psychosom Res* 2017;98:40–6.
- Ponds R, van Boxtel M, Jolles J. De cognitive failure questionnaire ALS maat voor subjectief cognitief functioneren. *Tijdschrift voor Neuropsychologie* 2006;2:37–45.
- Aaronson NK, Muller M, Cohen PD, 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;51:1055–68.
- Williams N. The MRC breathlessness scale. *Occup Med* 2017;67:496–7.
- Menges D, Ballouz T, Anagnostopoulos A, et al. Burden of post-COVID-19 syndrome and implications for healthcare service planning: a population-based cohort study. *PLoS One* 2021;16:e0254523.
- Broadbent E, Petrie KJ, Main J, et al. The brief illness perception questionnaire. *J Psychosom Res* 2006;60:631–7.
- Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361–70.
- Morroy G, Peters JB, van Nieuwenhof M, et al. The health status of Q-fever patients after long-term follow-up. *BMC Infect Dis* 2011;11:97.
- Peters JB, Daudey L, Heijdra YF, et al. 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;18:901–12.
- van Muilekom MM, Luijten MAJ, van Oers HA, et al. Paediatric patients report lower health-related quality of life in daily clinical practice compared to new normative PedsQL™ data. *Acta Paediatr* 2021;110:2267–79.
- 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;22:2581–94.
- Gordijn MS, Suzanne Gordijn M, Cremers EMP, et al. Fatigue in children: reliability and validity of the Dutch PedsQL™ multidimensional fatigue scale. *Qual Life Res* 2011;20:1103–8.
- Huskisson EC. Measurement of pain. *Lancet* 1974;2:1127–31.
- Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res* 2010;19:875–86.
- Yeatts KB, Stucky B, Thissen D, et al. Construction of the Pediatric Asthma Impact Scale (PAIS) for the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Asthma* 2010;47:295–302.
- Irwin DE, Stucky B, Langer MM, et al. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. *Qual Life Res* 2010;19:595–607.
- Muris P, Meesters C, van den Berg F. The strengths and difficulties questionnaire (SDQ). *Eur Child Adolesc Psychiatry* 2003;12:1–8.
- Vrijmoeth HD, Ursinus J, Harms MG, 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;19:324.
- van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol* 2007;6:Article25.
- James GWD, Hastie T, Tibshirani R. *An introduction to statistical learning: with applications in R*. New York: Springer, 2013.
- Molnar C. *Interpretable machine learning: a guide for making black box models explainable*. 2nd edn, 2022.
- Huang C, Huang L, Wang Y, et al. 6-Month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet* 2021;397:220–32.
- Mittal J, Ghosh A, Bhatt SP, et al. High prevalence of post COVID-19 fatigue in patients with type 2 diabetes: a case-control study. *Diabetes Metab Syndr* 2021;15:102302.
- Blomberg B, Mohn KG-I, Brokstad KA, et al. Long COVID in a prospective cohort of home-isolated patients. *Nat Med* 2021;27:1607–13.
- Sandmann F, Tessier E, Lacy J. 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.
- Rijksinstituut voor Volksgezondheid en Milieu (RIVM). Risicofactoren voor COVID-19 (contest), 2021. Available: <https://www.rivm.nl/coronavirus-covid-19/onderzoek/risicofactoren-contest-onderzoek> [Accessed 30-5-2022].