PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Prevalence and determinants of persistent symptoms after infection with SARS-CoV-2: protocol for an observational cohort study (LongCOVID-study)
AUTHORS	Mutubuki, Elizabeth; van der Maaden, Tessa; Leung, Ka Yin; Wong, Albert; Tulen, Anna D.; de Bruijn, Siméon; Haverman, Lotte; Knoop, H.; Franz, Eelco; van Hoek, Albert Jan; van den Wijngaard, Cees

VERSION 1 – REVIEW

REVIEWER	Luis, Bruno Instituto Mexicano del Seguro Social
REVIEW RETURNED	19-Mar-2022

GENERAL COMMENTS	The authors suggest a protocol for studying post-COVID-19 conditions; nowadays, it is an important topic. However, to improve the scientific content of the study and provide more meaningful information beyond the already reported, various details should be improved. Methods section What is the reason to include patients that test negative? It is likely had symptoms, and the sensitivity of antigen test or even PCR test is changing throughout the days of this disease. The only way that patients who tested negative will be included as controls is with a negative antibody test, which will be difficult in vaccinated people with preexisting plasma antibodies. What is the reason to include patients with a past infection in a retrospective evaluation?. Is it needed? I suggest that to improve the study's reliability and homogeneity, only keep the prospective evaluation. How will be dealt the data of two QOL evaluations?. In the case of EQ-5D-5L, anxiety and pain levels are evaluated in other questionaries such as Rand-36 and HADS. How will these overlapping symptoms evaluated will be handled?. I suggest decreasing the number of standardized evaluations to avoid overlapping. What variables will be included in your prediction model in post-COVID-19 conditions and healthcare utilization?. These variables should be clarified before the study is performed. If the patients were hospitalized after seven days, how will the follow-up and analysis of the pátients be performed?. Would it be an exclusion criterion? Since multiple predictive factors will be evaluated in the statistical analysis, these should be adjusted. A multivariable analysis should be performed, and in the case of a cohort, an analysis upon a time such as COX regression for survival would be added. Discussion section

	Line 23-24. A negative COVID-19 test does not confirm another etiologic agent.

REVIEWER	Mondello, Stefania University of Messina, Biomedical, Dental, Morphological and Functional Imaging Sciences
REVIEW RETURNED	25-Mar-2022

GENERAL COMMENTS	Investigators aim at addressing a timely and clinically relevant problem, namely the characterization of persistent symptoms after infection with SARS-CoV-2 and the identification of factors associated with their occurrence. The protocol is well written, and for the most part, arguments are presented in a clear manner. However, there are several important facets that should be further clarified. Abstract 1. Please, clarify the infrastructure/s and center/s involved in the study and include details about the dissemination plan. Strength and limitations of this study 2. This section should contain up to five short bullet points. 3. The authors may want to mention limitations related to the risk of lost to follow-up and possible bias against the number of hospitalized patients due to the study design. Introduction 4. Please, carefully re-read and improve the flow of ideas (e.g., remove redundancies [i.e., description of common symptoms]). 5. Define abbreviations and acronyms on their first occurrence in the manuscript (i.e., “HRQoL”) Methods and analysis 6. How many subjects are expected to be involved? Was any sample size calculation performed? 7. The dates of the study should be included in the manuscript. Also, please provide participant inclusion and exclusion criteria. 8. Please, a more detailed description of the clinical setting and centers involved in patient enrollment (e.g., numbers, location) should be provided. 9. Given the need for registration on the study website, the inclusion of subjects, especially pediatric controls, could be difficult. The investigators may want to comment on that. 10. How do the investigators plan to proceed in case of unbalance in the groups regarding the distribution of prognostic factors such as age, gender, co-morbidity, and education? 11. The statistical approaches for group comparison and identifying predictors of post-COVID condition are only briefly described. In addition, it is unclear whether the longitudinal profile of the post-COVID 19 conditions will be taken into account. 12. The authors may also want to analyze potential age differences (adults vs. children) of the post-COVID 19 condition. Discussion 13. Among the study limitations, the authors may want to mention the risk of lost to follow-up. 14. Ethical and safety considerations and any dissemination plan (publications, data deposition, and curation) should be covered here. Additional comments 15. The manuscript requires revision toward improving English grammar and syntax and correction of misspellings.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: Dr. Bruno Luis, Instituto Mexicano del Seguro Social
The authors suggest a protocol for studying post-COVID-19 conditions; nowadays, it is an important topic. However, to improve the scientific content of the study and provide more meaningful information beyond the already reported, various details should be improved.

1. Methods section
   What is the reason to include patients that test negative? It is likely had symptoms, and the sensitivity of antigen test or even PCR test is changing throughout the days of this disease. The only way that patients who tested negative will be included as controls is with a negative antibody test, which will be difficult in vaccinated people with preexisting plasma antibodies.

Participants that test negative will be included in the study as negative controls under the assumption that they are not infected with SARS-CoV-2, but possibly with other respiratory pathogens. The purpose is to determine to what extent long-term symptoms after COVID-19 occur, compared to other respiratory infections. The reviewer is right in saying that, to some extent these controls may be infected with SARS-CoV-2 after all, since the PCR test is not 100% sensitive. This is why we also included the population controls, that reflect the population background prevalence of long-term symptoms. Unfortunately no serological data is available for cases and controls in our study. We have added this limitation to the limitations section in the discussion on page 14, lines 24-27. It now reads as follows:
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.

2. What is the reason to include patients with a past infection in a retrospective evaluation? Is it needed? I suggest that to improve the study's reliability and homogeneity, only keep the prospective evaluation.

The LongCOVID study consists of two separate cohorts, the prospective and the retrospective cohort. The data from the different cohorts will be analyzed separately. The reason for having a retrospective cohort was to be able to gain insight into symptom patterns and predictors among participants that reported having post COVID-19 condition already at baseline. This includes patients that were infected at the start of the epidemic. We now more explicitly state that the analyses of the different cohorts will be performed separately on page 4, line 23:
Analyses will be performed separately for the prospective and retrospective cohorts.

3. How will be dealt the data of two QOL evaluations?. In the case of EQ-5D-5L, anxiety and pain levels are evaluated in other questionnaires such as Rand-36 and HADS. How will these overlapping symptoms evaluated will be handled?. I suggest decreasing the number of standardized evaluations to avoid overlapping.

We do not intent to calculate a QALY loss using different questionnaires in the same cohort. The EQ-5D-5L is included specifically to assess HRQoL for the acute phase of COVID-19 in the prospective cohort, whereas the RAND-12 is used during the later follow-up moments to measure the impact of long-term symptoms on HRQoL in both the prospective and the retrospective cohort. Rand-36 subscale pain and HADS will be used to assess the pain and anxiety respectively, in a more descriptive manner. We now more explicitly mention that in the methods section on page 6, lines 4-8 as follows:

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.

4. What variables will be included in your prediction model in post-COVID-19 conditions and healthcare utilization?. These variables should be clarified before the study is performed.

Potential predictors to be added to the prediction model for possible post-COVID-19 condition and health utilization have been added to the manuscript. These can be found in table S1, under supplementary files.

5. If the patients were hospitalized after seven days, how will the follow-up and analysis of the patients be performed?. Would it be an exclusion criterion?

If participants are included in the study prior to hospitalization, then participants will not be excluded because they have been hospitalized. Patients that are hospitalized will continue to receive questionnaires at the stipulated moments, 3, 6, 9 and 12 months, regardless of whether they complete them or not. Missing data will be dealt with as indicated in the manuscript on page 13, lines 4-8. See below:

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.

6. Since multiple predictive factors will be evaluated in the statistical analysis, these should be adjusted. A multivariable analysis should be performed, and in the case of a cohort, an analysis upon a time such as COX regression for survival would be added.

A multivariate analysis will be performed for our study. The COX regression is not added in our study because we predict the occurrence of long-term symptoms yes/no in the cohort of patients at a given point or period in time. We have adjusted the text in the manuscript to make it more clear. The adjustments we made can be seen on page 12, lines 7-14 and are as follows:

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. Discussion section

   Line 23-24. A negative COVID-19 test does not confirm another etiologic agent.

The reviewer is right that a negative COVID-19 test does not confirm infection by another respiratory pathogen. Although it is likely that a substantial part of the test-negative controls will have a symptomatic infection by another respiratory pathogen, not all of test-negative controls will have it. Based on the reviewers comment, a more cautiously phrased text has been added on page 14, lines 7-11 as follows:

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.

Reviewer 2: Dr. Stefania Mondello, University of Messina

Investigators aim at addressing a timely and clinically relevant problem, namely the characterization of persistent symptoms after infection with SARS-CoV-2 and the identification of factors associated with their occurrence. The protocol is well written, and for the most part, arguments are presented in a clear manner. However, there are several important facets that should be further clarified.

1. Please, clarify the infrastructure/s and center/s involved in the study and include details about the dissemination plan.

Participants testing positive or negative to COVID-19 are recruited from community health testing services. Population controls are selected randomly from the population register. We have made it more clear in the manuscript on page 5, lines 12-19 and on the flow chart of participants (figure 1) as follows:

Via Community Health Testing Services

Individuals testing positive and negative to COVID-19 at 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.

Information regarding dissemination plans has been updated, see the answer to point 3 of the editor, above.

2. Strengths and limitations of this study
   This section should contain up to five short bullet points.

We have revised the strengths and limitations of the study accordingly. See the answer to point 4 of the editor above.

3. The authors may want to mention limitations related to the risk of lost to follow-up and possible bias against the number of hospitalized patients due to the study design.

The reviewer is right that we have to deal with lost to follow up and missing data. We now mention this more explicitly in the revised manuscript, page 14, lines 20-22, as follows:

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.

We also more explicitly mention that our study is more likely to recruit participants with mostly mild to moderate acute symptoms and fewer patients that are hospitalized, on page 13, lines 28-33, as follows:

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.

Introduction

Please, carefully re-read and improve the flow of ideas (e.g., remove redundancies [i.e., description of common symptoms]).

This has been done and adjusted accordingly.

4. Define abbreviations and acronyms on their first occurrence in the manuscript (i.e., “HRQoL”)

This has been changed throughout the manuscript.

5. Methods and analysis
   How many subjects are expected to be involved? Was any sample size calculation performed?

We have added a sample size calculation. See the answer to point 2 of the editor, above.

6. The dates of the study should be included in the manuscript. Also, please provide participant inclusion and exclusion criteria.

The dates of the study have been added in the manuscript on page 4, lines 19-21 as follows:

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.

Inclusion and exclusion criteria are now more explicitly mentioned in the manuscript on page 4, lines 27-32 & on Page 5, lines 1-8 as follows:

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.

7. Please, a more detailed description of the clinical setting and centers involved in patient enrollment (e.g., numbers, location) should be provided.

Participants in this study are recruited from community health testing services. As the study is an ongoing recruitment no stipulated number of participants to be recruited has been assigned to the study and/or recruitment centers. This detail has been added to the manuscript on page 5, lines 12-15 as follows:
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.

8. Given the need for registration on the study website, the inclusion of subjects, especially pediatric controls, could be difficult. The investigators may want to comment on that.

The reviewer is right that inclusion of pediatric controls can be challenging, but population controls, including pediatric, will be invited randomly from the basic registration of persons in the Netherlands and can be oversampled if needed. This has been updated in the manuscript on page 5, line 16-19 as follows:
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.

9. How do the investigators plan to proceed in case of unbalance in the groups regarding the distribution of prognostic factors such as age, gender, co-morbidity, and education?

In the analyses we will control for possible confounding factors that are unbalanced in the groups, i.e. age, gender, comorbidities, and the level of education. The text has been updated in the manuscript on page 11, lines 26-27 as follows:
Analyses will be controlled for age, gender, number of comorbidities and level of education.

10. The statistical approaches for group comparison and identifying predictors of post-COVID condition are only briefly described. In addition, it is unclear whether the longitudinal profile of the post-COVID 19 conditions will be taken into account.

The prediction analyses will both take a cross-sectional approach, hence we will look at given follow-up moments separately, but we will also take into account long-lasting episodes of persistence of symptoms. More information regarding potential predictors has been added to the manuscript, see the answer to point 4 of reviewer 1 above.

11. The authors may also want to analyze potential age differences (adults vs. children) of the post-COVID 19 condition.

The analyses of the children will initially be conducted separately from those of the adults. Symptom prevalence and severity in possible post-COVID-19 condition can be compared between different age groups as well including children vs adults. This has been made more explicit in the manuscript on page 11, lines as follows:
The analyses of the children will initially be conducted separately from those of the adults.

And on page 11, lines 26-28.
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.

12. Discussion
    Among the study limitations, the authors may want to mention the risk of lost to follow-up.

The risk of lost to follow up, has been added to the manuscript, see the answer to point 3, you raised above.

13. Ethical and safety considerations and any dissemination plan (publications, data deposition, and curation) should be covered here.

Ethical and safety considerations and dissemination plans have been updated, see the answer to answer to point 3 of the editor.

14. Additional comments
    The manuscript requires revision toward improving English grammar and syntax and correction of misspellings.

A proofread of the article has been done. Spellings and grammar errors have been corrected accordingly, see the answer to answer to point 7 of the editor. We thank the reviewer for this observation.

Other changes:
Finally, in the current version of the manuscript we now specifically mention the effect of vaccination on long-term symptoms as a secondary outcome, see page 9, lines 22-23
Effect of vaccination to SARS-CoV-2 at baseline on the prevalence and severity of persistent symptoms after SARS-CoV-2 infection.
Also see page 12 lines 1-5:
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.

VERSION 2 – REVIEW

REVIEWER	Luis, Bruno Instituto Mexicano del Seguro Social
REVIEW RETURNED	11-Jun-2022

GENERAL COMMENTS	The authors have answered the questions appropriately and performed the suggested changes. The protocol is ready to publish as well as the study to start. Finally, I would recommend the authors give priority to the prospective evaluation of the cohort; it will improve the study's conclusions.

REVIEWER	Mondello, Stefania University of Messina, Biomedical, Dental, Morphological and Functional Imaging Sciences
REVIEW RETURNED	09-Jun-2022

GENERAL COMMENTS	the authors have been responsive to the critiques and revised the manuscript accordingly. I have no other concerns.