# TITLE:

A multicentre international prospective study of the validity and reliability of a COVID-19 specific health-related quality of life questionnaire

## Authors:

Cecilie Delphin Amdal 1, 2, Ragnhild Sørum Falk 1, Susanne Singer 3, Madeline Pe 4, Claire Piccinin 4, Andrew Bottomley 4, Lambert Tetteh Appiah 5, Juan Ignacio Arraras 6, Oliver Bayer 3, Eirik Alnes Buanes 7,8, Anne Sophie Darlington 9, Gracia Dekanic Arbanas 10, Kristin Hofsø 11, 12, Bernard Holzner 13, Pernilla Sahlstrand-Johnson 14, Dagmara Kuliś 4,, Ghansyam Parmar 15, Niveen ME Abu Rmeileh 16, Melanie Schranz 3, Samantha Sodergren 9, Kristin Bjordal 1, 17.

## Affiliations:

1. Research support services, Oslo University Hospital, Oslo, Norway
2. Department of Oncology, Oslo University Hospital, Oslo, Norway
3. Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Centre of Johannes Gutenberg University Mainz, Mainz, Germany
4. Quality of Life Department, European Organisation for Research and Treatment of Cancer, Brussels, Belgium
5. Komfo Anokye Teaching Hospital, Kumasi, Ghana
6. Servicio de Navarro de Salud; Pamplona, Spain
7. Department of Anaesthesia and Intensive Care, Haukeland University Hospital, Bergen, Norway
8. Norwegian Intensive Care and Pandemic Registry, Haukeland University Hospital, Bergen, Norway
9. School of Health Sciences, University of Southampton, Southampton, United Kingdom
10. Clinical Hospital Centre Rijeka, Rijeka, Croatia
11. Department of Research and Development, Division of Emergencies and Critical Care, Oslo University Hospital, Oslo, Norway
12. Lovisenberg Diaconal University College, Oslo, Norway
13. University Hospital for Psychiatry I, Medical University of Innsbruck, Innsbruck, Austria
14. Department of Otorhinolaryngology, Skåne University Hospital, Lund, Sweden
15. Department of Pharmacy, Sumandeep Vidyapeeth Deemed to be University,Vadodara, India
16. Institute of Community and Public Health, Birzeit University, Birzeit, Occupied Palestinian Territory
17. Faculty of Medicine, University of Oslo, Oslo, Norway

**Correspondent author:**

Dr Cecilie Delphin Amdal, Research support services, Oslo University Hospital, Oslo, Norway.

Postal address: Oslo University Hospital, Sogn Arena, Post Box 4950 Nydalen, NO-0424 Oslo, Norway E-mail: cecia@ous-hf.no. Phone: +47 91517199.

## ORCID-ID:

Cecilie Delphin Amdal (0000-0001-8422-8085); Ragnhild Sørum Falk (0000-0001-8398-3492); Susanne Singer (0000-0001-5784-7964); Dagmara Kulis (0000-0002-2289-2113); Andrew Bottomley; Lambert Tetteh Appiah (0000-0002-8895-2000); Juan Ignacio Arraras (0000-0002-4415-2570); Oliver Bayer ([0000-0003-4648-9481](https://orcid.org/0000-0003-4648-9481)); Eirik Alnes Buanes (0000-0002-1295-6734); Anne Sophie Darlington (0000-0003-4387-7278); Gracia Dekanic Arbanas (0000-0002-6325-1398); Kristin Hofsø (0000-0002-5854-8461); Bernard Holzner (0000-0002-3389-3621); Pernilla Sahlstrand Johnson (0000-0001-5146-3636); Ghansyam Parmar(0000-0002-4614-9797); Madeline Pe; Claire Piccinin (0000-0002-3918-1174); Niveen M Abu Rmeileh (0000-0003-0831-9692); Melanie Schranz (0000-0003-4476-0020); Samantha Sodergren ([0000-0001-8755-146X](https://orcid.org/0000-0001-8755-146X)); Kristin Bjordal (0000-0002-1914-494X).

**Keywords:** COVID-19, quality of life, patient-reported outcome measure, questionnaire, PROM, HRQoL

**Word count: 3998**

**Acknowledgements**

We would like to acknowledge the researchers and clinicians who helped with data collection: Katherine Taylor and Britta Büchler (University Medical Centre Mainz, Germany), Heike Dinklage (Regional Health Department Mainz-Bingen, Germany), Sara Jurdana (Clinical Hospital Centre Rijeka, Croatia), Justine Parer and Marianne Sævik (Haukeland University Hospital Norway), Kojo A. Hutton-Mensah, Jeffery Kudiabor and Samuel Nana Prempeh Agyeman-Gyebi (Komfo Anokye Teaching Hospital, Ghana).

**List of abbreviations and acronyms**

|  |  |
| --- | --- |
| **Abbreviation** | **Term/name** |
| CFI | Comparative Fit Index |
| COSMIN | COnsensus-based Standards for the selection of health Measurement INstruments |
| COVID-19 | coronavirus disease |
| EQ-5D | EuroQoL Questionnaire 5 dimensions |
| EORTC | European Organisation for Research and Treatment of Cancer |
| HRQoL | health-related quality of life |
| ICC | intra-class correlation coefficients |
| Ledidi® | Software company for digital data capturing |
| Long COVID | post-COVID-19 condition |
| PROMs | patient-reported outcome measures |
| PW | provisional weekly |
| QLG | Quality of Life Group |
| QLQ-PW80 | the OSLO COVID-19 QLQ-PW80© |
| QLQ-PW71 | the Oslo Covid-19 QLQ-PW71© |
| QLQ-W61 | the Oslo COVID-19 QLQ-W61© |
| RMSEA | Root Mean Squared Error of Approximation |
| the Oslo Covid-19 QLQ-PW71© | The revised 71-item provisional weekly questionnaire |
| the Oslo COVID-19 QLQ-W61© | The final 61-item weekly questionnaire |
| the OSLO COVID-19 QLQ-PW80© | 80-item provisional weekly questionnaire |
| TLI | Tucker-Lewis Index |
| WHO | World Health Organization |

**Abstract**

**Purpose:** To develop and validate a health-related quality of life (HRQoL) questionnaire for patients with current or previous coronavirus disease (COVID-19) in an international setting.

**Methods:** This multicenter international methodology study followed standardized guidelines for a four-phase questionnaire development. Here we report on the pretesting and validation of our international questionnaire. Adults with current or previous COVID-19, in institutions or at home were eligible. In the pretesting, 54 participants completed the questionnaire followed by interviews to identify administration problems and evaluate content validity. Thereafter, 371 participants completed the revised questionnaire and a debriefing form to allow preliminary psychometric analysis. Validity and reliability were assessed (correlation-based methods, Cronbach’s α and intra-class correlation coefficient).

**Results:** Eleven countries within and outside Europe enrolled patients. From the pretesting, 71 of the 80 original items fulfilled the criteria for item-retention. Most participants (80%) completed the revised 71-item questionnaire within 15 minutes, on paper (n =175) or digitally (n =196). The final questionnaire included 61 items that fulfilled criteria for item-retention or were important to subgroups. Item-scale correlations were >0.7 for all but nine items. Internal consistency (range 0.68-0.92) and test-retest results (all but one scale >0.7) were acceptable. The instrument consists of 15 multi-item scales and six single items.

**Conclusion:** The Oslo COVID-19 QLQ-W61© is an international, stand-alone, multidimensional HRQoL questionnaire that can assess the symptoms, functioning and overall quality of life in COVID-19 patients. It is available for use in research and clinical practice. Further psychometric validation in larger patient samples will be performed.

**Plain Language summary**

Many people have been affected by COVID-19, with an impact on their health-related quality of life. Health-related quality of life includes physical, emotional and social elements. It is important that we can ask the right questions to assess quality of life after having had COVID-19. Our work focuses on testing the questionnaire we have developed, to make sure it includes the right questions, and that patients understand these. We asked patients from eleven countries to complete the questionnaire. A total of 425 patients completed the questionnaire. Sixty-one of 80 questions were kept, and the wording on several items was changed. The final 61-item questionnaire includes the right questions, is complete, and is acceptable to patients. This questionnaire is valid and can be used with patients during and after COVID-19 to measure their health-related quality of life.

## Introduction

For over two years, the SARS-CoV-2 virus has continued its worldwide spread causing a heavy burden for the many affected by symptomatic disease (COVID-19). Vaccination programs have reduced the risk of severe disease, hospitalization, death [1-3], and disease transmission [4]. Fully vaccinated individuals still experience symptoms ranging from mild to serious [5], especially if they are older (≥ 65 years) and have comorbidities [6]. The most frequent symptoms reported in COVID-19 are fever, headache, cough, myalgia, dyspnea, and loss of taste and smell [7-9]. They may also suffer from e.g. sore throat, runny nose, gastrointestinal problems, and chest pain [7-10]which can have a considerable impact on physical, emotional, and social functioning [11, 12]. In April 2020, we conducted a literature review and identified publications related to studies of symptoms and problems with functioning in patients with COVID-19 from a large variety of clinical settings, culture and countries, and from all continents. As the pandemic continued to evolve with new variants of the virus, we updated our literature review twice during the first year [9].

Given the impact of COVID-19 and the potential for persisting problems, it is crucial to evaluate the patient’s perspective using validated patient-reported outcome measures (PROMs). Such measures can be used to assess symptoms and other relevant health-related quality of life (HRQoL) issues over time and in response to therapeutic interventions. They help to ensure that the patient’s perspective remains central and mitigate the effects of underreporting by healthcare professionals [13, 14].

To date, there is no cross-culturally validated COVID-19 specific HRQoL questionnaire available. Although several COVID-19 related clinical outcome assessments have been published [15], they focus on the pandemic’s impact on the general population [16, 17], mental health [18] or describe more general recommendations [19]. Generic HRQoL measures, along with symptom checklists, have been employed to assess issues in patients with COVID-19 [20, 21], but they may not capture the range of potential symptoms and HRQoL issues associated with COVID-19. Moreover, ad hoc non-COVID-specific measures may fail to demonstrate the psychometric properties required for a PROM [22].

In April 2020, our multilingual and multicultural research group, involving clinicians, psychometricians, statisticians and HRQoL specialists, set out to develop a questionnaire to assess HRQoL and symptoms in patients with COVID-19, from diagnosis through active disease and recovery. The first two phases of the questionnaire development have been published [9, 23]. In this paper, we present the next phase of the development process with pretesting of our provisional COVID-19 questionnaire and preliminary psychometric testing of the validity and reliability in an international sample of patients.

**Methods**

The current study is a multicenter international methodology study for a questionnaire development. Guidelines from the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group (QLG) were followed [24]. The four-phase procedure covers general principles for questionnaire development and is supported by the Food and Drug Administration [25]. The first two phases (gathering relevant issues through literature review, interviews with health-care workers and patients in seven countries, and operationalization from issues to items) were performed from April to October 2020 and have been published [9, 23].

To make sure that the questionnaire also covered potential new important issues evolving during the pandemic the literature review was updated in October 2020, with results that was further explored in this phase III study, and in February 2021 without any new issues being reported [9].

The resulting 80-item provisional weekly (PW) questionnaire was named the OSLO COVID-19 QLQ-PW80© (short name QLQ-PW80) and copyrighted by the Oslo University Hospital, Norway, that coordinated the international development. In the current phase III study, this work continued with further pretesting (pilot testing in a small sample) and validation of the provisional questionnaire. The aim was to identify missing or redundant items and ensure comprehensibility (phase IIIA) and perform initial psychometric testing (phase IIIB). All study documents were developed in English. The questionnaire was translated from English into the required languages (n = 9) using a modified forward/backward translation procedure based on international guidelines [26] and reviewed by an experienced translation officer before the patient enrollment started.

**Countries and participants in phase IIIA and phase IIIB**

Countries from all continents were approached through professional networks, including the World Health Organization (WHO) COVID-19 clinical management team network. Partners from 11 countries (Austria, Croatia, Germany, Ghana, India, Norway, Palestine, Spain, Sweden, The Philippines and United Kingdom) enrolled patients ensuring representation from different cultural areas.

The target population was ≥18 years, with verified SARS-CoV-2 infection (according to local/national standards) and active or previous symptomatic COVID-19. Patients in hospitals, nursing homes, at home, or in COVID-19 centers were eligible if they were able to read and comprehend the study documents. Patients in intensive care units could be enrolled after discharge. To ensure content validity of the final questionnaire according to guidelines, pre-specified enrollment matrices were used [24]*.*These matrices were set up to ensure sufficient sample sizes with adequate distribution of participants to represent the target population. In phase IIIA (Appendix 1), we aimed to include 10-15 patients in each cell of the sample matrix, ≥5 patients per country and ≥45 patients in total. In phase IIIB (Appendix 2), we aimed to include 15 patients per cell of the sample matrix and ≥300 patients in total [24], to allow for preliminary evaluation of the psychometric properties, in particular the hypothesized scale structure. In phase IIIB, participants in recovery, expected to have stable disease i.e., unlikely to experience any changes in the physical wellbeing, were asked to complete the questionnaire a second time after 14 days (±2 days) to measure test-retest reliability. The target sample size for this part was 50 patients.

**Phase IIIA: procedure, collection of data**

Data collection in phase IIIA was performed from November 2020 until June 2021. The researchers in the 11 countries approached patients in hospital, nursing homes and patients staying at home. After written informed consent, patients completed the provisional questionnaire, the QLQ-PW80, in their native language and recorded the completion time. This was followed by structured debriefing interviews by the research team documented with field-notes (Appendix 3). Patients’ background information was collected (age, gender, hospitalization, time since diagnosis, disease severity, comorbidity). Whether the patients had experienced each item was assessed considering the whole disease period and reported as *relevance* (1= not at all, 2= a little, 3= quite a bit, 4= very much). If experienced, the extent they were troubled by this was regarded as a *measure of importance* (0= not experienced, 1= not at all to 4= very much). We defined relevant items as those with *relevance* scores 2-4, and important items as those with *importance* scores 3 and 4. Patients were asked if any of the items were difficult to understand or confusing, annoying or upsetting, and if any of the items overlapped with other items. The interviewer also asked about relevance and importance of any other issues not listed, but still experienced by the patient. Six items were assessed in more detail to explore how the patients interpreted/understood these issues (Appendix 4).

**Phase IIIA: decision rules and criteria for retention of items**

A set of six decision rules and criteria for retention of items in phase IIIA was set up as recommended [24]; the first four criteria considered most important (Table 1A). Since COVID-19 is a new disease and a limited number of patients were to be enrolled in phase IIIA, care was taken not to exclude important items simply because of low prevalence. The decision on whether an item should be retained, modified, or removed was made by consensus in the project group, including review of potentially overlapping items and rewording of difficult and annoying items.

**Phase IIIB: procedure, collection of data**

Data collection in phase IIIB was performed over six months from July 2021 to January 2022. An electronic data capture system, Ledidi® was used that fulfilled the necessary national and EU security and privacy policy requirements [27]. After written informed consent, patients in hospital, at home or in nursing home were asked to complete questionnaires on paper or digitally in the Ledidi system®. The set of questionnaires consisted of: 1) background information (as described for phase IIIA, supplemented by vaccination status), 2) the provisional questionnaire resulting from phase IIIA, and 3) a debriefing questionnaire. The debriefing questionnaire documented the time and help needed to complete the questionnaire, whether there were difficult, annoying or overlapping items, or additional issues not covered by the questionnaire. In addition, patients in recovery (more than three months after being infected) were asked to point out any additional issues they had experienced more than three months after being infected.

**Phase IIIB: decision rules and criteria for retention of items**

In phase IIIB, the research group used seven criteria for retention to decide on items that were candidates for removal (Table 1B). The summarized results and consistency across languages were discussed, and decision trails from earlier phases of development consulted. An item was kept if clinically relevant, fulfilling ≥5 of the retention criteria, or found important to subgroups of patients, and not overlapping with other items. An item was modified if >5% of the patients found it difficult or annoying.

**Statistical analyses**

***Descriptive analyses***

Descriptive analyses were presented as frequencies and proportions for categorical data and means, standard deviations, and range for continuous data. For the analysis of content validity, interviews in phase I and pilot testing by the user group in phase II laid the foundations [23]. We investigated this further by exploring feasibility; i.e., patients who needed more than >30 minutes to complete the questionnaire, patients who needed help to understand the items, and whether there were systematic patterns of missing values. We explored possible differences in response patterns between countries by qualitative review of responses and comments in the debriefing questionnaire. New issues raised by the patients were considered if not covered by existing items, not excluded in previous phases, and not clinical parameters or other distinct conditions.

***Psychometric analyses***

The scale structure analyses were based on the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) taxonomy [28]. In addition to the descriptive analyses, we tabulated the results per item along with the number of fulfilled criteria (Appendix 5). The multi-language clinical expert group proposed scales from a clinical point of view. The scale structure was modified based on the following analyses. We assessed the validity by correlation-based methods (e.g., multitrait analysis) using Pearson correlation. Item convergent validity was supported by a correlation of ≥0.40 between an item and its own scale (corrected for overlap). Scales that are conceptually related (e.g., fatigue and malaise) are expected to have a correlation ≥|0.40| while scales that are conceptually different (e.g., worries and temperature) would have a correlation <|0.40|. Scaling errors were calculated as the percentage of items that correlated higher with a different scale than with their own scale, corrected for overlap.

Known group comparisons were performed for pre-defined groups (Appendix 6) based on the patient matrix (Appendix 2), using independent sample t-tests.

To test the reliability, we calculated the internal consistency using Cronbach’s alpha for each multi-item scale. Values >0.70 were considered acceptable for group comparisons. Test-retest reliability was assessed by calculating intra-class correlation coefficients (ICC) for scores at first and second time-points for the multi-item scales.

Confirmatory factor analyses were performed to explore the dimensionality of the questionnaire. We calculated standardized factor loadings for each item with regard to the corresponding scale and considered loadings >0.40 to be sufficient. Model-fit was assessed by the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI), with both indices considered to indicate good fit if >0.95 as well as by the Root Mean Squared Error of Approximation (RMSEA) which is recommended to be <0.06 [29].

The final scale structure was based on the psychometric tests and clinical evaluation and a scoring manual was developed (Appendix 7). All items had responses on a four-point Likert scale ranging from 1 = “not at all” to 4 = “very much”, except the two items in the overall quality of life scale ranging from 1 = “very poor” to 7 = “excellent”. The scores from scales and single items were linearly transformed to scores ranging from 0 to 100. A higher score represented worse symptoms and more problems on functioning scales. For the overall quality of life scale, higher score represented better quality of life.

## Results

**Phase IIIA**

Research teams in 11 countries interviewed 54 patients (Table 2). Adult males and females of all ages, with active ongoing disease or in recovery, admitted to hospital or isolated at home, were included. Twenty-one patients reported various comorbidities. The mean time needed to complete the QLQ-PW80 was 16 minutes (range 4-45). The patient-reported *relevance* and *importance*, and item score distribution with final decisions, are presented in Appendix 8. In summary, 21 items were candidates for removal according to criteria outlined in Table 1A. Patients identified possible overlap for five groups of items. Nine items were removed (Figure 1) due to low importance (n=4, blocked nose, sneezing, shaking hands, heartburn), overlap (n=2, abdominal discomfort, abandoned by family and friends), and an overall consideration with low mean score, floor effect, limited importance, and few patients experiencing them (n=3, constipation, dysuria and hair loss). Five items identified as difficult or annoying were reworded, resulting in the revised provisional 71-item questionnaire named the Oslo Covid-19 QLQ-PW71© (QLQ-PW71).

**Phase IIIB**

A total of 371 patients from 10 countries with ongoing disease or in recovery, admitted to hospital or isolated at home were interviewed (Table 2, Appendix 9). Patients who reported unspecified comorbidity (n=49) commented, e.g., that they had slow metabolism, lumbar prolapse, fibromyalgia, Crohn’s disease, and myalgic encephalomyelitis. The six patients who needed >30 minutes to complete the questionnaire (Table 3) were elderly (67-106 years old) and needed assistance with completion. Forty-four patients needed help to understand the items; mainly from Ghana (n=15), India (n=8), and Germany (n=7), all age groups were represented. Patients in hospital (27/90) and nursing home/other institution (3/17) needed assistance more often compared to patients at home (14/263). Even though they needed assistance, 63% completed the questionnaire within 15 minutes.

Most patients (279, 75%), completed all 71 items. Of the 92 patients with missing values, 68 had only one or two missing. Seven of the 92 patients had completed only the first page of the digital questionnaire (item 1-23), missing the subsequent 48 items. For the others, there were no patterns of missing data. There were no clear differences between the countries regarding number and pattern of missing values.

**Additional patient-reported problems**

Some patients described additional symptoms or problems experienced in the first three months (n=56) and/or more than three months after being diagnosed (n=36). Two of these problems were new: menstrual disturbances (n=2) and word-finding problems or aphasia (n=2). Other proposed issues were either covered by the questionnaire (n=34), excluded previously in phase I or phase IIIA (n=11), or were not patient-reported symptoms but objective signs such as low oxygen saturation and weight loss (n=11) (Appendix 10). The project group agreed that there was not enough evidence to include the two new issues.

**Evaluation of items and (initial) scale structure**

Of the 71 items, 55 items fulfilled ≥5/7 criteria for retention (Appendix 5). Twenty items had a mean score >1.5. Across all items, the whole range of responses was used and patients did not find them difficult or annoying in any language. Compliance was high (>95%) for all but two items (light and heavy housework). All items were reviewed quantitatively and qualitatively. It was decided to remove 10 items that did not fulfill more than 4/7 criteria for retention, and were either not important to subgroups of patients (n=8) or had low compliance and overlapped with other items (n=2). For the initial proposed scales,the internal consistency was satisfactory (Cronbach’s alpha >0.70) for all but one scale (Sensory) (Appendix 5). Nevertheless, the scale was retained because it was considered clinically meaningful.

For the initial proposed scales, the item-scale correlations were satisfactory for all but nine items (drowsy, feeling ill or unwell, headache, chest pain, weakness in hands or feet, appetite loss, carrying a heavy bag upstairs, social activity, worry about financial difficulties). Three were moved to another scale (drowsy, headache, chest pain) while six were kept in the initial scale, as this was considered clinically more meaningful. Three items (palpitation, burning and sore eyes, skin problems) were kept as single items. The ‘Respiratory’ scale was divided into “upper” (throat) and “lower” (chest), as this was clinically more meaningful, and the item-scale correlations were virtually unchanged.

**The validity and reliability of the final questionnaire**

The final 61-item weekly questionnaire resulting from our current phase IIIB development, now consists of 15 multi-item scales and six single items and was named the Oslo COVID-19 QLQ-W61© (QLQ-W61) (Table 4). All scales had acceptable item convergent validity (|≥0.40|) except one item (financial difficulties) in the ‘Worries’ scale. The scaling error was low (0-3.3%). Discrimination across pre-defined groups was observed for most of the scales (Temperature disruption, Sensory, Gastrointestinal, Physical and Social functioning), but not for all (Pain, Worries, Emotional and Cognitive functioning) (Table 5). The ‘Respiratory lower, chest’ scale reached significance for disease status, but not for comorbidity and gender.

Cronbach’s alpha was >0.70 for 14 of the 15 multi-item scales (range 0.68-0.92) (Table 4). The test-retest reliability was good, ICCs between the two time-points were >0.70 except for problems with eyes (ICC 0.44). In the confirmatory factor analysis, all standardized factor loadings exceeded the threshold of 0.40 (range 0.42-0.95), supporting the hypothesized scale structure. The CFI was 0.85, the TLI was 0.83, and the RMSEA was 0.06, somewhat lower than the predefined values of acceptable fit of the model and the data.

## Discussion

This questionnaire was developed to capture HRQoL of patients with COVID-19 from diagnosis through active disease and the recovery period. We have demonstrated that the questionnaire captured relevant and important issues to the international sample of COVID-19 patients, that the final number of items is manageable, and that the instrument shows promising psychometric properties.

 To allow the questionnaire to be feasible and valid in a heterogeneous international setting, we successfully ensured that patients involved represented a good distribution of age, gender, disease phases and severity, comorbidities, and countries. Cross-cultural acceptability was supported by patients from 11 countries in three continents reporting the items to be relevant and important, and discussions among the involved researchers from different cultural areas. Special attention was given to the wording of five items pointed out as difficult to understand to improve cross-cultural acceptance. The word “stigmatised” was difficult to understand in some languages and the group decided to include “or judged negatively” to improve the comprehension of the concept (Appendix 8, item 71). The word “abandoned” was described as too offensive in some languages and, based on earlier input from patients, this was changed to “not receive sufficient attention” (Appendix 8, item 77).

Even though elderly patients might need some assistance, the high compliance, few missing values and the finding that most patients used ≤15 minutes to complete the questionnaire, supported that it is easy to understand and fill in, making it suitable for clinical studies but also for descriptive purposes in clinical practice. Since the digital version of the questionnaire was challenging for some patients, more focused instructions and user-friendly layout are recommended in future studies.

Our results where the participants regarded the functional and overall HRQoL-items as relevant and important, are supported by others studies using e.g. EQ-5D [30]. It is evident that the HRQoL issues relevant to COVID-19 patients are not limited to symptoms alone; how COVID-19 has affected these patients’ functioning and overall quality of life remain important factors in assessing the patients’ experiences.

The preliminary psychometric properties of the QLQ-W61 were robust. All multi-item scales had good internal consistency. Results from *known group comparisons* supported the pre-defined hypotheses for the majority of the scales, but not for all. For ‘Pain’ and ‘Worries’, no differences were found between patients with acute disease and those in recovery; both groups showed elevated levels. This suggests that pain (e.g., muscle aches and pain) and worries (e.g., worries about health) may be issues that need to be flagged for post-COVID-19 condition (Long COVID). Younger and elderly patients had similar results on ‘Cognitive functioning’. This may be a result of selection bias where elderly patients who could respond to the COVID-19 questionnaire (reflecting less problems with cognitive functioning) were selected for this study. For ‘Emotional functioning’, an expected gender difference was not found even though many studies report more emotional issues among females compared to males [31, 32]. This imply that the emotional stress of having COVID-19 may have a similar impact on both genders, but this has to be further explored in larger patient populations. As expected, patients with acute disease had higher mean score on the ‘Respiratory lower, chest’ scale (e.g. shortness of breath, chest pain) than those in recovery. Males and patients with comorbidities are at risk of having more severe COVID-19 [33, 34] and lower respiratory symptoms such as shortness of breath and chest pain are regarded as symptoms of severe disease. However, we were surprised to find that there were no differences regarding gender or comorbidities. The *test-retest reliability* for both multi-item and single item scales were of acceptable levels, except for the eye-item. This shows that the assessment is stable over time unless clinical changes occur in the well-being of the patient. Although results from the *initial confirmatory factor analysis* showed sub-optimal fit (TLI< .90, CFI <.90), this may be explained by the low sample size relative to the complexity of the scale structure for this questionnaire. Furthermore, testing of differences between countries will be performed in the next phase of the questionnaire development involving larger samples of patients from finalized clinical studies in COVID-19 patients.

One limitation might be the questionnaire’s ability to cover issues from new variants of COVID-19, and issues related to Long COVID, although we performed two updates of the literature review to reduce this risk. For example, with the Omicron variant, the most frequent symptoms were runny nose, headache, fatigue, sneezing and sore throat, which were different from the dominant symptoms in the earlier COVID-19 Alpha variant (i.e., fever, cough, and loss of sense of smell or taste) [35]. Also, symptoms of dry and red eyes have been mentioned as more frequent with Omicron than earlier variants of the virus [36]. The QLQ-W61 captures most but not all Omicron symptoms. Runny nose was excluded after phase I, and sneezing after phase IIIA, as patients who had experienced these symptoms did not regard them as important.

In addition, due to the time constrains, there are limitations in the analyses of the questionnaire’s ability to capture all symptoms of Long COVID. Long COVID was unknown at the start of the pandemic and consequently not included in the first two phases of this project. To compensate for this in the current phase, we specifically asked the 181 patients in recovery (more than three months post infection) to point out any additional issues they had experienced more than three months after being infected. The two new issues that were proposed were not included. Menstrual disturbances were more likely to be related to vaccination [37] or secondary to psychological distress [38]. Word-finding problems could be related to general fatigue or be part of a cerebrovascular incidence secondary to COVID-19 related thrombotic complications [39]. In a study on breast cancer patients with Long COVID not yet published (JIA, personal communication), the participants completed the QLQ-PW80 7 or 10 months after being diagnosed with COVID-19. They were asked to describe any new issues, but none were reported. Long COVID has been shown to affect patients in all age groups, with varying background characteristics and levels of disease severity [40].The most commonly seen symptoms are fatigue, reduced physical and cognitive functioning, shortness of breath, and palpitations [41], but also psychological symptoms such as anxiety and depression. Other symptoms reported are loss of taste and smell, muscle pain, headache, skin problems, and hair loss [41]. Almost all symptoms of Long COVID-19 are covered in the QLQ-W61, except less common symptoms such as hair loss (removed in phase IIIA) and conditions such as post-traumatic stress disorder.

Since this process included a broad sample of international patients in various stages of COVID-19, including recovery, we believe the questionnaire fits the needs of various stakeholders. However, a wider distribution of participants would have been optimal in a study of a worldwide pandemic. More countries were approached, but due to ethical approval obstacles, unfortunately, some colleagues were not able to participate and this is a limitation. Some patients involved in this study had comorbidities or pre-existing conditions, and we cannot ascertain that all reported issues are caused by COVID-19. However, since issues were reported as relevant and important by a number of patients, it is ensured that the reported issues are common to at least a subgroup of COVID-19 patients, regardless of their pre-existing conditions.

The Oslo COVID QLQ-W61 is regarded to have the properties needed in clinical trials as a standardized internationally developed tool to evaluate HRQoL of patients during active disease, in the recovery phase, and even in a Long COVID setting. It is a comprehensible instrument that is easy to fill in, and we believe that it could be useful in clinical practice as well. In a clinical setting, the questionnaire could be used to monitor patients, e.g., for symptom control in the clinic and to evaluate whether their HRQoL issues are changing over time. This has to be explored in future studies.

Additional psychometric testing in a large international sample would be preferable, but is considered time consuming and resource-intensive and may be difficult to perform. Therefore, further validation will be based on data from ongoing clinical studies where patients fill in this COVID questionnaire, and data from two such studies (not yet published) are already available. In this study, we tested the questionnaire with a weekly (W) time frame (QLQ-W61), but in future studies, we may test the daily version (QLQ-D61).

**Conclusion:**

The Oslo COVID QLQ-W61© is a stand-alone, multidimensional HRQoL questionnaire that can assess the symptoms, functioning and overall quality of life of COVID-19 patients. The questionnaire is applicable for clinical trials and clinical practice, covers relevant COVID-19 issues and is acceptable to a broad population of COVID-19 patients from many countries. Although the questionnaire still needs to go through a final development phase of international psychometric validation in a large patient sample, this provisional questionnaire is now available for use with nine completed translations.

**Figure Legends**

**Fig. 1** Overview of the development of the international COVID-19 specific health-related quality of life questionnaire

# References

1. Vasileiou E, Simpson CR, Shi T, Kerr S, Agrawal U, Akbari A, et al. Interim findings from first-dose mass COVID-19 vaccination roll-out and COVID-19 hospital admissions in Scotland: a national prospective cohort study. Lancet. 2021;397(10285):1646-57.

2. Lopez Bernal J, Andrews N, Gower C, Robertson C, Stowe J, Tessier E, et al. Effectiveness of the Pfizer-BioNTech and Oxford-AstraZeneca vaccines on covid-19 related symptoms, hospital admissions, and mortality in older adults in England: test negative case-control study. BMJ. 2021;373:n1088.

3. Pritchard E, Matthews PC, Stoesser N, Eyre DW, Gethings O, Vihta KD, et al. Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom. Nat Med. 2021;27(8):1370-8.

4. Levine-Tiefenbrun M, Yelin I, Katz R, Herzel E, Golan Z, Schreiber L, et al. Initial report of decreased SARS-CoV-2 viral load after inoculation with the BNT162b2 vaccine. Nat Med. 2021;27(5):790-2.

5. Bergwerk M, Gonen T, Lustig Y, Amit S, Lipsitch M, Cohen C, et al. Covid-19 Breakthrough Infections in Vaccinated Health Care Workers. N Engl J Med. 2021;385(16):1474-84.

6. Tenforde MW, Self WH, Adams K, Gaglani M, Ginde AA, McNeal T, et al. Association Between mRNA Vaccination and COVID-19 Hospitalization and Disease Severity. JAMA. 2021;326(20):2043-54.

7. Lechien JR, Chiesa-Estomba CM, Place S, Van Laethem Y, Cabaraux P, Mat Q, et al. Clinical and epidemiological characteristics of 1420 European patients with mild-to-moderate coronavirus disease 2019. J Intern Med. 2020;288(3):335-44.

8. da Rosa Mesquita R, Francelino Silva Junior LC, Santos Santana FM, Farias de Oliveira T, Campos Alcântara R, Monteiro Arnozo G, et al. Clinical manifestations of COVID-19 in the general population: systematic review. Wien Klin Wochenschr. 2021;133(7-8):377-82.

9. Amdal CD, Pe M, Falk RS, Piccinin C, Bottomley A, Arraras JI, et al. Health-related quality of life issues, including symptoms, in patients with active COVID-19 or post COVID-19; a systematic literature review. Qual Life Res. 2021:1-15.

10. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020;395(10223):507-13.

11. Evans RA, McAuley H, Harrison EM, Shikotra A, Singapuri A, Sereno M, et al. Physical, cognitive, and mental health impacts of COVID-19 after hospitalisation (PHOSP-COVID): a UK multicentre, prospective cohort study. Lancet Respir Med. 2021;9(11):1275-87.

12. Lamontagne SJ, Winters MF, Pizzagalli DA, Olmstead MC. Post-acute sequelae of COVID-19: Evidence of mood & cognitive impairment. Brain Behav Immun Health. 2021;17:100347.

13. Barr PJ, Berry SA, Gozansky WS, McQuillan DB, Ross C, Carmichael D, et al. No date for the PROM: the association between patient-reported health events and clinical coding in primary care. J Patient Rep Outcomes. 2020;4(1):17.

14. Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. Am J Manag Care. 2008;14(8):530-9.

15. COVID-19 related COAs <https://mapi-trust.org/questionnaires/covid-19-questionnaires/index.xml2022> [cited 2022 11.05.2022]. Available from: https://mapi-trust.org/questionnaires/covid-19-questionnaires/index.xml.

16. Adiukwu F, Orsolini L, Gashi Bytyçi D, El Hayek S, Gonzalez-Diaz JM, Larnaout A, et al. COVID-19 mental health care toolkit: an international collaborative effort by Early Career Psychiatrists section. Gen Psychiatr. 2020;33(5):e100270.

17. Riad A, Huang Y, Zheng L, Elavsky S. The Associations of COVID-19 Induced Anxiety, Related Knowledge and Protective Behavior. Health Psychol Res. 2021;9(1):24768.

18. Repišti S, Jovanović N, Kuzman MR, Medved S, Jerotić S, Ribić E, et al. How to measure the impact of the COVID-19 pandemic on quality of life: COV19-QoL – the development, reliability and validity of a new scale. Global Psychiatry. 2020;3(2):201-9.

19. fda.gov. Assessing COVID-19-Related Symptoms in Outpatient Adult and Adolescent Subjects in Clinical Trials of Drugs and Biological Products for COVID-19 Prevention or Treatment. Guidance for Industry. https://www.fda.gov/regulatory-information/search-fda-guidance-documents.; 2020.

20. Wong AW, Shah AS, Johnston JC, Carlsten C, Ryerson CJ. Patient-reported outcome measures after COVID-19: a prospective cohort study. Eur Respir J. 2020;56(5).

21. Halpin SJ, McIvor C, Whyatt G, Adams A, Harvey O, McLean L, et al. Postdischarge symptoms and rehabilitation needs in survivors of COVID-19 infection: A cross-sectional evaluation. J Med Virol. 2020:1-10.

22. Sivan M, Wright S, Hughes S, Calvert M. Using condition specific patient reported outcome measures for long covid. BMJ. 2022;376:o257.

23. Amdal CD, Taylor K, Kuliś D, Falk RS, Bottomley A, Arraras JI, et al. Health-related quality of life in patients with COVID-19; international development of a patient-reported outcome measure. J Patient Rep Outcomes. 2022;6(1):26.

24. Wheelright SB, Kristin; Bottomley, Andrew; Gilbert, Alexandra; Martinelli, Francesca; Pe, Madeline; Sztankay, Monika; Cocks, Kim; Coens, Corneel; Darlington, Anne-Sophie; Fayers, Peter; Fitzsimmons, Deborah;. EORTC Quality of Life Group Module Development Guidelines 5th ed: EORTC; 2021 2021. 95 p.

25. Administration USDoHaHSFaD. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims In: fda.gov, editor. https://www.fda.gov/media2009. p. https://www.fda.gov/media.

26. Kulis D, Bottomley A, Velikova G, Greimel E, Koller M. EORTC Quality of Life Group Translation Procedure. 2017.

27. Team Lc. Ledidi. Smarter research. Deeper insights. 2021 [Available from: <https://ledidi.com/>.]

28. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.

29. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: a multidisciplinary journal. 1999;6(1):1-55.

30. Tabacof L, Tosto-Mancuso J, Wood J, Cortes M, Kontorovich A, McCarthy D, et al. Post-acute COVID-19 Syndrome Negatively Impacts Physical Function, Cognitive Function, Health-Related Quality of Life, and Participation. Am J Phys Med Rehabil. 2022;101(1):48-52.

31. García-Fernández L, Romero-Ferreiro V, Padilla S, David López-Roldán P, Monzó-García M, Rodriguez-Jimenez R. Gender differences in emotional response to the COVID-19 outbreak in Spain. Brain Behav. 2021;11(1):e01934.

32. Pappa S, Barmparessou Z, Athanasiou N, Sakka E, Eleftheriou K, Patrinos S, et al. Depression, Insomnia and Post-Traumatic Stress Disorder in COVID-19 Survivors: Role of Gender and Impact on Quality of Life. J Pers Med. 2022;12(3).

33. Abate BB, Kassie AM, Kassaw MW, Aragie TG, Masresha SA. Sex difference in coronavirus disease (COVID-19): a systematic review and meta-analysis. BMJ Open. 2020;10(10):e040129.

34. Selvin E, Juraschek SP. Diabetes Epidemiology in the COVID-19 Pandemic. Diabetes Care. 2020;43(8):1690-4.

35. Iacobucci G. Covid-19: Runny nose, headache, and fatigue are commonest symptoms of omicron, early data show. BMJ. 2021;375:n3103.

36. Binotti W, Hamrah P. COVID-19-related Conjunctivitis Review: Clinical Features and Management. Ocul Immunol Inflamm. 2022:1-7.

37. Male V. Menstrual changes after covid-19 vaccination. BMJ. 2021;374:n2211.

38. Demir O, Sal H, Comba C. Triangle of COVID, anxiety and menstrual cycle. J Obstet Gynaecol. 2021;41(8):1257-61.

39. Cheng NM, Chan YC, Cheng SW. COVID-19 related thrombosis: A mini-review. Phlebology. 2022:2683555211052170.

40. Yong SJ. Long COVID or post-COVID-19 syndrome: putative pathophysiology, risk factors, and treatments. Infect Dis (Lond). 2021;53(10):737-54.

41. Walia N, Lat JO, Tariq R, Tyagi S, Qazi AM, Salari SW, et al. Post-acute sequelae of COVID-19 and the mental health implications. Discoveries (Craiova). 2021;9(4):e140.

**Statements and Declarations**

**Funding:** Internal research funds from Oslo University Hospital HF (project nr 41595 and 35444) and research funds from University of Oslo (project nr 102372066). Grant from the Health Department – Government of Navarra (0011-3638-2020-000012) (JI Arraras).

**Competing interests:**

All authors declare no competing interests. Susanne Singer received honoraria for reviewing manuscripts for the Lilly Quality of Life Award, outside of this study. Copyright of the questionnaire belongs to Oslo University Hospital. The questionnaire will be freely available for academic users after registration with Oslo University Hospital.

**Author Contributions:** CDA and KB conceived the idea, wrote the first draft, established the international collaboration and managed the project together with AB. CDA, KB, RSF, SuS, CP, and MP constituted the writing committee for this article. RSF, SuS, MP, DK, CP, ASD gave valuable input to the design and method. LTA, JIA, OB, EAB, GDA, KH, BH, PSJ, GP, NMAR, MS and SaS recruited patients and collected data. CDA, RSF, SuS, MP and KB performed the analyses and interpreted the results. The writing committee wrote the first draft of the publication; all authors were involved in editing successive drafts and approved the final submitted version.

**Ethics approval**: This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Regional Ethical Committee, Norway *(Date 12.10.2020/ ref:****171640)***, Oslo University Hospital Ombudsman, (Date 05.11.2020/ref 20/21865) and in addition, by the appropriate institutional committees and ethical committees in participating countries.

**Consent to participate:** All participants provided written informed consent before they were included in the study.