

A Core Outcome Measurement Set (COMS) for Research and Clinical Practice in Post COVID-19 Condition (Long COVID) in Adults: An International Delphi Consensus Study

Sarah L. Gorst PhD^{*1}, Nina Seylanova MD^{*2}, Nicola Harman PhD¹, Margaret O'Hara PhD³, Caroline B. Terwee⁴, Paula R. Williamson PhD^{*1}, Dale M. Needham PhD^{*5,6,7}, Daniel Munblit PhD^{*8,9,10}, Timothy R Nicholson PhD^{*11} and the PC-COS Study Group^{**}

1 MRC/NIHR Trials Methodology Research Partnership, Department of Health Data Science, University of Liverpool (a member of Liverpool Health Partners), Liverpool, UK

2 Independent Researcher

3 Long COVID Support charity

5 Amsterdam UMC, The Netherlands

6 Outcomes After Critical Illness and Surgery (OACIS) Research Group, Johns Hopkins University, Baltimore MD

7 Pulmonary and Critical Care Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore MD

8 Department of Physical Medicine and Rehabilitation, Johns Hopkins University School of Medicine, Baltimore MD

9 Department of Paediatrics and Paediatric Infectious Diseases, Institute of Child's Health, Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russia

10 Inflammation, Repair and Development Section, National Heart and Lung Institute, Faculty of Medicine, Imperial College London, London, United Kingdom

11 Research and Clinical Center for Neuropsychiatry, Moscow, Russia

12 Neuropsychiatry Research and Education Group, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

*Authors contributed equally to the study: apart from the two joint first and four joint last authors, authors are listed in alphabetical order.

** Listed at end of manuscript

Corresponding author:

Timothy R Nicholson, MD PhD. Neuropsychiatry Research and Education Group, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK.

Email: timothy.nicholson@kcl.ac.uk

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC), and the World Health Organisation (WHO).

Contributions: DM and TN conceived the idea for the study. PW led the methodological team. DM, TN, PW, DMN, SLG and NS designed the study protocol and were responsible for the day to day running of the project. NS led undertook the literature review, identified outcome measures and categorised them for inclusion in the online Delphi survey. NS also coordinated the data revision process. NH, SLG and NS developed the online Delphi surveys and contributed to the day to day management of the project. DM, TN, CA, DMN, SLG, NS and PW participated in the project methodology discussions throughout the duration of the project. SLG undertook the data analysis. SLG and NH organised the consensus meeting. MO'H

provided and coordinated perspectives of people with lived experience throughout the study design and implementation. SLG and TN drafted the manuscript; all authors reviewed and approved the final manuscript.

The authors alone are responsible for the views expressed in this publication and they do not necessarily represent the decisions, policy or views of the World Health Organization.

Declaration of Interests: DM reports receipt of grants from the British Embassy in Moscow, UK National Institute for Health Research (NIHR) and Russian Foundation for Basic Research. He is also a Co-Chair of International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC) Global Paediatric Long COVID Working Group, member of ISARIC working group on long-term follow-up in adults and co-lead of the PC-COS project. TN reports funding from the UK National Institute for Health and Social Care Research (NIHR). SLG is the COMET Initiative project coordinator. NH reports payment and honoraria for lectures and presentations from Lancaster University. PRW reports that she received funding from the UK National Institute for Health Research (NIHR) and is also a chair of the Core Outcome Measures in Effectiveness Trials (COMET) Management Group. SLG is the COMET project coordinator. Other authors declare that they have no competing interests.

Summary

Background: Agreement is lacking regarding which instruments should be used for measurement of a recently-agreed core outcome set (COS) for clinical practice and research in post COVID-19 condition, also known as Long COVID.

Methods: A rigorous multi-step consensus study was conducted, including a comprehensive literature review informing a three-round online modified Delphi process followed by an online consensus meeting to finalise measurement instruments for this COS, thereby developing a Core Outcome Measurement Set (COMS). 594 participants from 58 countries spanning six continents, participated in the consensus process.

Findings: Consensus was reached for including the 'modified MRC Dyspnea Scale' for measuring the 'Respiratory functioning, symptoms, and conditions' outcome in the COMS. Two outcome measures from the acute COVID-19 COMS were included in this post COVID-19 condition COMS, namely 'Time until death' (for the Survival outcome) and 'Recovery Scale for COVID-19' (for the Recovery outcome). Consensus, based on predefined criteria, was not reached for the other 9 outcomes of the COS. However, based on the study results, suggested measurement instruments are provided for 'Fatigue', 'Post-exertion symptoms'; 'Pain', 'Physical function', 'Work/occupational and study changes' and for 'Cardiovascular', 'Nervous system', 'Cognition' and 'Mental' functioning, symptoms and conditions.

Interpretation: This international COMS provides a framework for assessing post COVID-19 condition in global clinical research and practice settings. Further evaluation is warranted as new data emerges on the performance of measurement instruments for the COS for this new clinical condition.

Box 1: Key messages

Rationale and approach

- Post COVID-19 Condition (Long COVID) encompasses a very wide variety of sequelae that can persist for many months or even years after SARS-CoV-2 infection.
- There has been substantial heterogeneity in the outcomes evaluated in research into, and clinical services for, post COVID-19 condition. We previously undertook a consensus study and developed a core outcome set (COS) — an agreed minimum set of outcome domains that should be measured and reported in clinical research and practice for post-COVID-19 condition in adults worldwide.
- To further advance the field, there is a pressing need for deciding which measurement instruments are the most appropriate for these core outcomes, to optimise data comparability and synthesis.
- This study undertook a rigorous international consensus process of measurement instruments for each of the previously defined core outcome domains to provide a recommended Core Outcome Measurement Set (COMS) .

Findings

- The COMS includes recommended instruments for survival, recovery and respiratory outcomes. No instrument reached consensus to be recommended for the 9 other outcomes in the COS, but the most preferred instruments from this consensus process were identified.

Future Directions and Implications

- Participants at the online consensus meeting agreed that assessment of the relative merits of existing instruments not developed for post COVID-19 condition ('legacy' or generic instruments) versus those developed specifically for post COVID-19 condition (condition-specific measures) is a high priority for future outcome measure research in this condition.
- As new data related to post-COVID-19 condition and related measurement instruments becomes available, the COMS should be re-evaluated.
- The use and reporting of this COMS for adults with post COVID-19 condition is an important step to harmonise and accelerate research, especially the development of evidence-based treatments, and to ensure consistent evaluation and reporting of these core outcomes in both clinical and research settings.

Introduction

Although most people with SARS-CoV-2 infection rapidly recover, many experience persistent symptoms for months or years, known as post COVID-19 condition, Long COVID and other terms.(1) Treatment trials and clinic monitoring require optimal and consistent outcome measurement. However, there has been substantial heterogeneity in selection of relevant outcome domains and associated measurement instruments.(2)

In 2021, a multidisciplinary international group reached consensus on a core outcome set (COS) of outcome domains to be measured in all future clinical studies and care for people with post COVID-19 condition.(3) This COS used established research methods, recommended by the Core Outcome Measures in Effectiveness Trials (COMET) initiative.(4) Consensus on the COS was obtained following a two-round online modified Delphi process delivered in five languages. Stakeholder groups involved in the consensus process included people with post COVID-19 condition and family members; health professionals with post COVID-19 condition; and health professionals and researchers with experience in treating and studying people with post COVID-19 condition. The COS included twelve outcomes: fatigue; pain; post-exertion symptoms; work or occupational and study changes; survival; recovery; and functioning, symptoms, and conditions for each of cardiovascular, respiratory, nervous system, cognitive, mental health and physical outcomes.

After a COS establishes consensus on outcome domains, consensus is then needed on which instrument(s) should be used for each domain by developing a Core Outcome Measurement Set (COMS). When using the term 'instrument,' we are referring to outcome measurement instruments, tools, procedures, etc., used to measure an outcome domain. It is likely that many instruments exist to measure each outcome in a COS; however, the difficulty is reaching consensus on the most appropriate instrument that should be used, as a minimum, by all studies and clinical services evaluating post COVID-19 condition. This is particularly challenging for a new condition where instruments are less likely to have been validated specifically in this clinical population. It is possible that there may be no suitable existing instruments to measure some of the outcome domains in the COS; hence, it is important to identify any potentially appropriate instruments currently under development and evaluation and those from other relevant populations.

Herein, we report on the PC-COS project findings for a COMS for post-COVID-19 condition in adults (≥ 18 years of age) intended for use in clinical research and care.

Methods

This second stage of the PC-COS project followed COSMIN-COMET guidelines on selecting measurement instruments for outcomes included in a COS(5) and was prospectively registered with the COMET (Core Outcome Measures in Effectiveness Trials) initiative database.(6) There was patient involvement throughout the study from funding application development, including two patient partner co-investigators and input and support from the Long COVID SOS and Long COVID Support patient groups, to ongoing development of the protocol and study materials, as well as advising and assisting on recruitment and survey completion strategies. The study protocol was developed a priori, published on the study website(7) and approved, along with all study materials, by the UK Health Research Authority and by the South West Cornwall & Plymouth Research Ethics Committee (REC number 21/SW/0109). Further details regarding study management processes are comprehensively described in the COS paper.(3)

Step 1. Establishing the constructs to consider

Agreeing the construct (i.e. outcome or domain) to be measured in the target population (adults with post COVID-19 condition) was the first step in selecting outcome measurement instruments; outcome domains in the post COVID-19 condition COS(3) were those we aimed to identify and select instruments for. These measures would be the minimum recommended to be used in both clinical research studies and practice across all settings globally, so low as well as high income settings.

Step 2. Finding existing outcome measurement instruments

Detection of instruments currently used in studies

Outcome measurement instruments used in published and ongoing studies for post COVID-19 condition for each outcome domain in the COS were primarily identified from the literature set collected and evaluated in a systematic review conducted during the preceding COS development(3) – see Appendix 1 for full details. Identified instruments were then mapped to domains in the post COVID-19 condition COS and reviewed to remove duplicates and confirm mapping. Instruments that mapped to multiple COS domains were included in a category of multi-domain instruments. Similarly post COVID-19 condition specific instruments were included in their own category. Instruments not mapping to any COS domains were not considered.

The preliminary list of instruments was independently reviewed by three experts (DN, TN and DM) with complementary areas of expertise, who classified each of them as 'include', 'maybe' or 'exclude' for the consensus process. Reasons for exclusion included not relevant to post COVID-19 condition, biological specimen required, cannot be done by phone or post or otherwise not feasible within the scope of the COS (e.g., cannot be undertaken in all resource settings internationally), or other reason (with reason specified). These three assessments were compared and disagreements discussed to reach consensus on a final list of included instruments.

Given the large number of instruments mapping to some COS domains, only the five most frequently used instruments, along with any post COVID-19 condition-specific instruments, were presented in the first round of the modified Delphi consensus process in order to help optimize study feasibility and respondent burden. This inclusion threshold was defined as the five most frequently used instruments for each COS domain in the systematic review. For

domains where less than five instruments were identified, instruments that had only been used once were included. However, if more than five instruments met the domain frequency criteria, only instruments used more than once were included. For instances where more than five instruments met the frequency criteria, and had all been used more than once, the three clinical experts (DN, TN, DM) reviewed the instruments and reached agreement for inclusion based on relevance to post COVID-19 condition and administration feasibility. All instruments not meeting this frequency criterion, other than novel post COVID-19-specific instruments, were excluded at this stage, but could be suggested by participants in Round 1 of the modified Delphi consensus process, along with any other measurement instrument(s) not included for any reason.

Other sources

For domains with less than five instruments identified (as above), we identified instruments that had been recommended to measure outcomes in other COS for relevant related clinical populations, based on the COMET initiative COS international registry. Study group members also suggested relevant PROMIS (Patient-Reported Outcomes Measurement Information System) and Neuro-QoL (Quality of Life in Neurological Disorders) instruments after searching the PROMIS and Neuro-QoL databases.(8,9) Newly developed post COVID-19 condition-specific instruments were also identified via the wider study group and through discussions with other international experts and networks.

Step 3. Assessing quality of outcome measurement instruments

Measurement properties

Given that the measurement properties of many non-COVID specific instruments had not been assessed in a post COVID-19 population, assessment of measurement properties of these instruments was not undertaken. However, for COVID-19 specific instruments we assessed measurement properties using a “mini COSMIN assessment” focused on content validity(10) which involved identifying instrument development papers and additional content validity studies. Two experts (CT and VF) independently rated the instruments according to predefined criteria (see supplementary material, Appendix 2) judging the relevance, comprehensiveness, and comprehensibility of the instruments.

Feasibility aspects

For all instruments feasibility-related data (e.g. time, cost, language/translations) was obtained allowing participants to consider if they could reasonably be used in the intended settings.(11) Feasibility data selection was guided by relevant prior COMS work,(12) to maximise selection of the most important aspects to present to participants for consideration.

Step 4. Reaching agreement on instruments for the COMS

Online modified Delphi consensus process

Instruments were presented in a three round online modified Delphi consensus process using DelphiManager software(13) and grouped into COS outcome domains which were presented in a random order for each participant. Instruments mapping to multiple COS domains, and the post COVID-19 condition-specific measures, were presented in separate categories. Consent for participation was obtained online before participation.

Stakeholder participation

All participants who had participated in the prior COS study were invited to this COMS study. Participants belonged to three stakeholder groups: (1) people with post COVID-19 condition and family members/caregivers; (2) health professionals and researchers with post COVID-19 condition; and (3) health professionals with experience in treating people with post COVID-19 condition and researchers (including psychometricians and PROM experts) with experience in the field of post COVID-19 condition. Additional stakeholders were invited via direct email from the study authors or from relevant patient or professional organisations. We contacted lead investigators of identified study protocols as above. We requested participation via post COVID-19 condition clinics and research groups. To increase global representation, we engaged stakeholders from countries with relatively lower representation in the prior COS project, requesting study dissemination to their networks.

Individuals were also able to express an interest in participating via the PC-COS website (<https://www.pc-cos.org/>) which was promoted via social media. There was no restriction on the number of eligible participants in each stakeholder group.

Materials

Participants were provided with the following for review when completing the surveys:

- List of COS outcome domains with lay definitions of the domain.
- Lists of instruments, mapped to each domain, to be rated.
- 'Instrument Cards' summarising key relevant information about each instrument. These included plain language details, written with input from patient research partners and a web-based link to the instrument itself, when publicly available. See Appendix 3 for an example Instrument Card.
- Definitions of specific measurement properties (e.g., reliability, validity, responsiveness) evaluated on Instrument Cards for COVID-specific instruments.

Scoring

In round 1 participants rated whether instruments should be used to measure the outcome domains using the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) nine-point Likert scale,(14) which is commonly divided into three categories: 'not that important' (1-3), 'important but not critical' (4-6) and 'critically important' (7-9). An option of 'unable to score' was also included if participants didn't feel able to rate specific instruments. Once participants had rated all instruments, prior to submission, they could review their ratings and revise them if they wished. Participants were able to complete the survey over more than one session using a 'save-for-later' option if available time, energy or other health reason required this.

In round 1 participants were also asked to suggest additional relevant instruments not already included in the consensus process. Additional instruments suggested by >1% of participants in each stakeholder group were considered and added to round 2 of the Delphi with an Instrument Card developed for each new instrument. Instruments that reached consensus for exclusion (as defined below) were not retained for round 2.

In round 2, responses for each stakeholder group in round 1 were summarised for each instrument and displayed graphically alongside their own rating and participants were asked to re-rate the instruments, so could choose to change their rating or to keep it the same. At the end of round 2, participants were able to review all their ratings together. Instruments reaching consensus for exclusion were similarly removed..

In round 3, all instruments rated in round 2 were similarly summarised and re-rated, with the addition of being able to reflect on the scores given for the new instruments that were suggested during round 1. COSMIN generally recommend the selection of a single instrument for each outcome in a COMS,(5) therefore in round 3 participants were encouraged to give a single instrument a rating of 7-9 (i.e. 'critically important'), and to rate all other instruments within that domain as 6 or less.

Data analysis

In rounds 1, 2 and 3 of the Delphi, descriptive statistics summarised overall scores of each stakeholder group for the three GRADE categories to determine whether the instruments met the predefined definition of consensus (see below). It was agreed a priori to include responses in the analysis if a participant assessed all instruments in at least one domain. Graphs using R (version 4.0.2) displayed the distribution of ratings for each instrument, stratified by stakeholder group, and shown to participants in rounds 2 and 3.

Following round 3, two additional sensitivity analyses were conducted to identify which instruments were rated most strongly, for each outcome, by each of the three stakeholder groups. Sensitivity analysis 1 involved identifying which instrument(s) within each outcome domain had most frequently been given the overall highest rating, calculated separately for each of the three stakeholder groups. Sensitivity analysis 2 involved calculating which instrument(s) within each outcome domain had most frequently been given a 'critically important' rating (i.e. 7, 8, or 9).

Selection bias between the Delphi process and the subsequent online consensus meeting was assessed by comparing distributions of the mean overall scores from round 3 of the Delphi survey between participants who attended the consensus meeting and those who did not.

Consensus definition

Consensus for an instrument to be excluded from the Delphi process, was defined a priori and pre-specified in the protocol(15) as 50% or less of participants, in **all three** stakeholder groups, rating the instrument as 7-9. Prior to reviewing results of round 2, we modified the exclusion criteria, to reduce participant burden by defining consensus for exclusion as 50% or less participants, in **at least two** stakeholder groups, rating as 7-9.

Consensus for an instrument to be included in the COMS was similarly defined a priori and pre-specified in the protocol(15) as 80% or more participants, in **all three** stakeholder groups, rating as 7-9 **and** <10% participants, in **each** stakeholder group, rating as 1-3.

Online Consensus meeting

At the end of round 3 of the Delphi process, participants were asked to declare conflicts of interest (e.g. related to the development of a specific instrument) and invited to express interest in attending an online consensus meeting. Those with no relevant conflicts of interest were considered, aiming for representation across both stakeholder groups and geographical locations. Prior to the meeting participants received background information about the meeting, their ratings from the Delphi, a summary of the Delphi results and details about the instruments to be discussed at the meeting. People with post COVID-19 condition were invited to attend a pre-meeting to prepare them for the consensus meeting and provide the opportunity to ask questions.

The consensus meeting was held on Zoom, conducted in English and chaired by an experienced independent facilitator (MC). It focused on the results of round 3 and instruments that were rated most highly by all stakeholder groups were prioritised for discussion. For each instrument discussed, participants were invited to provide their arguments in favour of inclusion or exclusion. Following discussion, participants were asked to anonymously vote 'Yes' or 'No' using the Zoom polls, as to whether the proposed instrument should be included in the COMS. Consensus for including an instrument was defined as 80% or more participants in all stakeholder groups agreeing it should be included.

Results

Identification of outcome measurement instruments

The systematic review identified 319 instruments used in post COVID-19 condition studies. Duplicate removal and mapping to core domains reduced this to 293 instruments. Independent review of the instruments reduced this to 91 instruments, which were further reduced to 47 based on frequency criteria (see Methods). An additional five PROMIS instruments and two post COVID-19 condition-specific instruments were added to this list, resulting in 54 outcome measurement instruments presented in round 1 (see supplementary material, Appendix 4) mapped to 12 domains: Cardiovascular functioning, symptoms and conditions, n=1; Fatigue or Exhaustion, n=6; Pain, n=6; Nervous system functioning, symptoms, and conditions, n=3; Cognitive functioning, symptoms, and conditions, n=5; Mental functioning, symptoms, and conditions, n=7; Respiratory functioning, symptoms, and conditions, n=5; Post-exertion symptoms, n=2; Physical functioning, symptoms, and conditions, n=6; Work/occupational changes and study, n=4; Multiple domains, n=5; and Post COVID-19 condition specific, n=4.

Delphi process

Round 1 ran from June 6th-29th 2022, with 711 participants registering; 594 (84%) from 58 countries completed this round. 447 fully completed and 147 partially completed, i.e. rated all instruments for at least one domain. All 594 participants were invited to participate in round 2. Round 2 ran from July 13th to August 8 2022; 394 (66%) participants fully (n=362) or partially (n=32) completed. Round 3 ran from August 12th to September 12th 2022; 359 participants (60%) fully (n=341) or partially (n=18) completed, 324 (82%) of which had completed round 2. Appendix 5 shows response rate breakdowns by stakeholder group across all rounds. Demographic characteristics of participants are presented in Table 1. Further details of participants are in Appendix 6.

Table 1. Demographic characteristics of participants in the Delphi consensus process

	Round 1 (n=594)	Round 2 (n=394)	Round 3 (n=359)
Stakeholder group, n (%)			
People with post COVID-19 condition and family members or caregivers	233 (39)	129 (33)	108 (30)
Health-care professionals and researchers with post COVID-19 condition	65 (11)	45 (11)	40 (11)
Health-care professionals and researchers without post COVID-19 condition	296 (50)	220 (56)	211 (59)
Gender, n (%)			
Female	413 (70)	261 (66)	234 (65)
Age group, n (%)			
18-29	27 (5)	19 (5)	16 (5)
30-39	147 (25)	104 (26)	91 (25)

40-49	203 (34)	127 (32)	116 (32)
50-59	150 (25)	92 (23)	84 (23)
60-69	58 (10)	48 (12)	47 (13)
>=70	9 (2)	4 (1)	5 (1)
Countries, n (%)			
Asia	41 (7)	30 (8)	33 (9)
Africa	16 (3)	10 (3)	11 (3)
Australasia	18 (3)	17 (4)	16 (4)
Europe	359 (60)	229 (58)	203 (57)
North America	138 (23)	96 (24)	83 (23)
South America	22 (4)	12 (3)	13 (4)
Ethnicity, n			
White	438 (74)	292 (74)	264 (74)
South Asian	21 (4)	13 (3)	15 (4)
Hispanic, Latino, Spanish	65 (11)	36 (9)	32 (9)
East Asian, Pacific Islander	21 (4)	15 (4)	16 (4)
Indigenous peoples	2 (<1)	2 (1)	2 (1)
Black	12 (2)	7 (2)	7 (2)
Middle Eastern, North African	13 (2)	13 (3)	8 (2)
Other	22 (4)	16 (4)	15 (4)

After round 1, 13 of the 54 instruments presented met exclusion criteria. One additional instrument was also excluded (Fukuda Criteria for CFS), resulting in 14 instruments excluded for round 2 (see Appendix 7). Suggestions for additional instruments resulted in 139 free-text responses, from which two instruments met a priori inclusion criteria: the WHO Disability Assessment Schedule 2.0 12 item version, and the Nijmegen Questionnaire for respiratory symptoms. An additional two post COVID-19 condition-specific instruments were included: The Long COVID Symptom tool and The Long COVID Impact tool, along with 8 validated post COVID-19 condition-specific subscales (SBQ-Breathing, SBQ-Circulation, SBQ-Fatigue, SBQ-Impact on daily life, SBQ-Memory, thinking & communication, SBQ-Mental health, SBQ-Movement, SBQ-Pain). Two of the post COVID-19 condition-specific subscales were each found to be applicable to two of the outcome domains: SBQ-Fatigue (Fatigue & Exhaustion; Post-exertion symptoms) and SBQ-Impact on daily life (Physical functioning, symptoms, and conditions; Work/occupational changes and study). Thus, these subscales were added into the survey under both domains, resulting in a total of 14 instruments added into round 2.

54 instruments were included in round 2 (see Appendix 8), 12 of which met exclusion criteria (see Appendix 9). Thus, 42 instruments were presented in round 3 (see Appendix 10).

After round 3, none of the 42 instruments met a priori inclusion criteria - see supplementary Appendices 10, 11 and 12 for results of rounds 1, 2 and 3 respectively for all instruments both in total and for each participant group. Results are given as percentages rating instruments as 'not that important' (score 1-3), 'important but not critical' (score 4-6) and 'critically important' (score 7-9) as above. Sensitivity analyses (see Appendix 14) revealed that for two domains a single instrument had been rated most strongly by all three stakeholder groups (Post-exertion symptoms – DePaul Symptom Questionnaire; Physical functioning, symptoms, and conditions: SBQ-LC Impact on daily life scale). Thus, both instruments were taken forward for discussion and voting in the online consensus meeting.

For the pain domain, the Brief Pain Inventory was the only instrument included in round 3 so was also taken forward for discussion and voting in the online consensus meeting. For the Respiratory functioning, symptoms, and conditions domain, the modified Medical Research Council (mMRC) Dyspnoea Scale was rated most strongly by two of the stakeholder groups, so was also included for discussion in the online consensus meeting, alongside the St George's Respiratory Questionnaire, which was rated most strongly by the patients and family members stakeholder group.

For the remaining eight domains, no single instrument was rated most strongly across all stakeholder groups and so were not included for discussion in the online consensus meeting.

Consensus meeting

Ahead of the consensus meeting, a one-hour pre-meeting for people with post COVID-19 condition was held on 27th September 2022 and attended by eight participants. The online consensus meeting was held on 29th September 2022. 37 people attended the three hour consensus meeting, including seven Study Management Group members, four observers, one facilitator and 25 voting participants who had completed the online Delphi survey (10 people with post COVID-19 condition; five healthcare professionals and researchers with post COVID-19 condition; and 10 healthcare professionals and researchers without post COVID-19 condition). Owing to the small number of healthcare professionals and researchers with post COVID-19 condition, participants belonging to this group were asked to select to join one of the other two groups, for the purposes of voting at the meeting, as done during the online meeting for the COS.(3) Voting groups therefore included 12 people with post COVID-19 condition and 13 healthcare professionals and researchers. Some participants were unable to attend the entire meeting or dropped in and out as a result of internet connection. The details of participants who attended the consensus meeting are described in Appendix 15. We found no evidence of selection bias: the average round 3 Delphi scores were similar between participants who did (6.01) and did not (6.31) attend the meeting.

At the beginning of the meeting, attendees were informed about the two measures that were confirmed for inclusion in the COMS. From the outset, it was agreed that the outcome 'Survival' would be measured by 'Time until death' and the outcome 'Recovery' would be measured using the question 'How long it takes to recover, i.e., feel better, no longer having symptoms' for consistency with the acute COVID-19 COS.(16) Following this reminder, the remaining outcome domains and measurement instruments were discussed in the following order: Respiratory functioning, symptoms, and conditions (mMRC Dyspnoea Scale; St George's Respiratory Questionnaire); Pain (Brief Pain Inventory); Post-exertion symptoms (DePaul Symptom Questionnaire); and Physical functioning, symptoms, and conditions (SBQ-LC Impact on daily life scale). Following discussion and voting, the mMRC Dyspnoea Scale was the only instrument meeting the predefined consensus definition for inclusion, with 82% of people with post COVID-19 condition and 92% of healthcare professionals and researchers voting 'Yes' for inclusion so it was added to the COMS (see Table 2).

Table 2. Core Outcome Measurement Set

Core Outcome	Measurement Instrument
Survival	Time until death
Recovery	Recovery Scale for COVID-19
Respiratory functioning, symptoms & conditions	Modified Medical Research Council (mMRC) Dyspnoea Scale

Consensus was not reached in recommending instruments for the remaining nine core outcome domains in COS. Table 3 indicates the instruments with the greatest level of support based on the consensus process. At least one of these instruments can be considered for each of the core outcomes (i.e., where more than one measurement instrument is provided, selection of a single instrument may be appropriate to avoid redundancy and reduce respondent burden).

Table 3. COS outcome measurement instruments for consideration

Core Outcome	Measurement Instrument Options
Fatigue or exhaustion	Fatigue Assessment Scale (FAS)
	Fatigue Severity Scale (FSS)
	Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F)
Post-exertion symptoms	De Paul Symptom Questionnaire
Cardiovascular functioning, symptoms & conditions	The Symptom Burden Questionnaire for Long COVID (SBQ-LC) - Circulation Scale
	New York Heart Association (NYHA) Functional Class
Cognitive functioning, symptoms & conditions	Cognitive Failures Questionnaire (CFQ)
	Montreal Cognitive Assessment 'Blind' version (MoCA-Blind)
Nervous system functioning, symptoms & conditions	Central Sensitization Inventory (CSI)

Mental functioning, symptoms & conditions	Generalized Anxiety Disorder 7 (GAD-7)
	PTSD Checklist for DSM5
Pain	Brief Pain Inventory (BPI)
Physical functioning, symptoms & conditions	The Symptom Burden Questionnaire for Long COVID (SBQ-LC) - Impact on Daily Life Scale
Work/occupational changes and study	Work Ability Index questionnaire (WAI)
	Work Productivity and Activity Impairment (WPAI) questionnaire
	'Your day-to-day work/school?' question item from WHO Global COVID-19 Clinical Platform Case Report Form for Post COVID

Consensus was not reached for any multiple domain outcome measurement instruments, nor any post COVID-19 condition-specific measurement instruments. Table 4 indicates instruments with the highest level of support based on the consensus process (see supplementary material, Appendix 16 for full results).

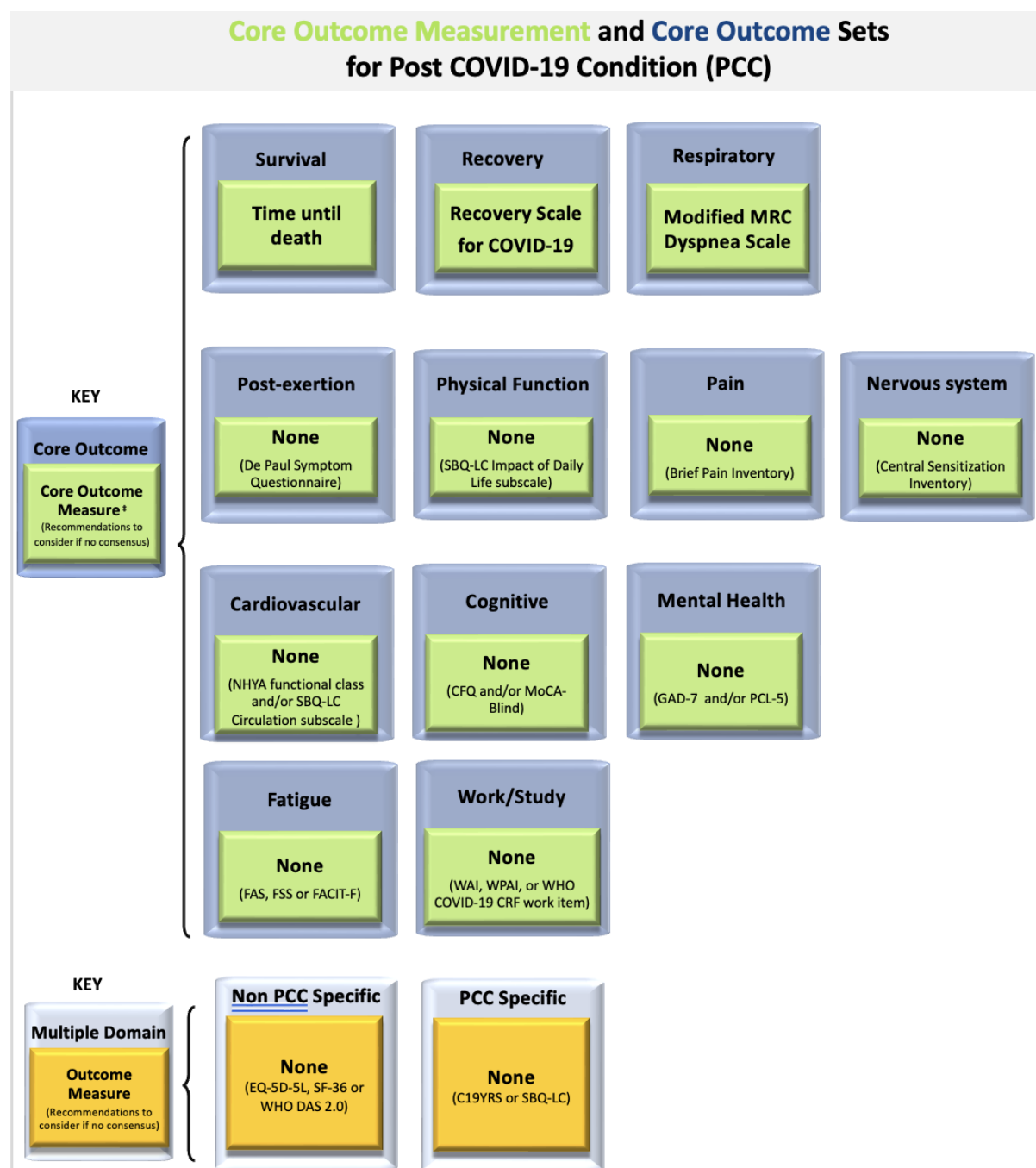
Table 4. Multiple domain and post COVID-19 condition-specific measurement instruments for consideration

Instrument Type	Measurement Instrument Options
Multiple Domain	Euroqol 5-Dimension 5-level (EQ-5D-5L)
	Medical Outcomes Study Short-Form-36 (SF-36)
	WHO Disability Assessment Schedule 2.0 - 12 Item (WHO DAS 2.0)
Post-COVID condition specific	COVID-19 Yorkshire Rehabilitation Screening (C19-YRS) Scale
	Symptom Burden Questionnaire for Long COVID (SBQ-LC)

At the online consensus meeting, there was a very high level of support (90% of people with post COVID-19 condition and 77% of healthcare professionals and researchers) for future research focused on a consensus process regarding use of existing outcome measurement instruments versus post COVID-19 condition-specific instruments versus a combination of both types of instruments for post COVID-19 condition research and clinical practice.

Feedback from participants who attended the consensus meeting was strongly positive and summarised in Appendix 17. A table of available languages for outcome measure instruments can be found in Appendix 18.

Figure 1. Summary of the Core Outcome Measurement Set (COMS) and Core Outcome Set (COS) for post COVID-19 condition



Discussion

This international process achieved consensus for instruments to measure survival, recovery and respiratory outcomes as the recommended post-COVID-19 condition Core Outcome Measure Set (COMS). These instruments are recommended for use and reporting in all clinical research and practice worldwide for post-COVID-19 condition in adults. No measurement instruments achieved consensus for the other 9 core outcome domains for post-COVID 19 condition, although the consensus process reduced the number of potential instruments for measuring the 12 core outcomes from over 300 to 19. The consensus process provided important data regarding which instruments had the greatest level of support despite not reaching consensus. With a goal of reducing the substantial heterogeneity in outcomes measurement for post COVID-19 condition; at least one of these instruments can be considered for each of the 9 core outcomes and should be the focus of future research along with further evaluation of multi-domain and COVID-specific instruments.

Limitations

Many limitations from the first stage of this project, developing the post COVID-19 condition COS for adults,(3) apply to this stage. In brief, the Delphi process and consensus meeting were delivered online, limiting participation to those with sufficient digital skills and access. Geographic and demographic participation is unlikely to be fully representative of post COVID-19 condition experience globally, despite significant efforts to achieve this. The online consensus meeting had insufficient participants with post COVID-19 condition who were also health professionals to constitute their own voting group, limiting the information potentially provided by another group with dual, and potentially unique, perspectives.

There are additional limitations unique to this COMS stage as it was only conducted in English. This decision was based on the extra complexity and substantial time required for translation of the large amount of study materials (e.g, all instrument cards) to multiple languages. This limitation likely impacted on global participation. However, despite this a reasonable global representation was reached.

This stage focused measurement instruments, requiring substantially greater time and detailed evaluations from participants than first stage. This is particularly important for people with post COVID-19 condition where fatigue, low energy levels, and cognitive difficulties are common and often severe and significantly disabling therefore potentially explaining lower completion rates compared to the first stage. To reduce bias from this key issue the order instruments were presented to participants, via the Delphi software, was randomised based on outcome domain. Moreover, major efforts were made, in collaboration with patient partners, to explain that rating all of the instruments for at least one of the outcome domains was the minimum participation required and still would provide valuable data. Across the three rounds, respectively, 75%, 92% and 95% of all participants fully completed the Delphi which reflects remarkable contributions of the participants.

Challenges in achieving consensus

For the nine domains for which consensus was not reached, there are several potential reasons. For example, post COVID-19 condition is a highly heterogeneous disorder potentially influencing instrument preference. Past experience with instruments may have

been highly variable and influenced ratings. As post COVID-19 condition is a new condition there is a lack of high quality data to guide choice between the large number of possible instruments.

It is important to note that not achieving consensus measurement instruments for all core outcome domains is not uncommon, even in other complex multi-domain conditions with decades of research, such as post-intensive care syndrome in survivors of acute respiratory failure.(17) Moreover, post COVID-19 condition is poorly understood mechanistically creating uncertainty regarding use of specific measurement instruments used in other conditions.

Implications for practice

For the three outcomes for which there was consensus on how to measure (Table 2), the recommended instrument should be used in all research and clinical practice. Other measures of these outcomes can be used in addition, if appropriate, since a COMS is the *minimum* set of instruments recommended for all clinical studies and practice to achieve a minimum level of consistency and comparability in measurement and reporting. In particular, this COMS was designed to be appropriate for all settings internationally, irrespective of resources, so additional instruments may be warranted.

For the other outcomes, where there was no consensus on measurement (Table 3), instruments with the greatest level of support can be considered in an effort to reduce the substantial heterogeneity in measurement of post COVID condition in adults.

For the three outcomes where only one instrument is suggested from this study (post exertion, physical function, pain and nervous system) its suitability for the study or setting should be considered. For example, is the instrument available in the appropriate language(s) and is it feasible in terms of patient burden.

If there are two or three instruments available for consideration, these same factors should be weighed up between the instruments. Furthermore, for some outcomes, such as fatigue, it might be measured largely or fully by a single instrument and as such it would be preferable to choose one measure. However, for other outcomes (e.g., mental health or cognition) different measures may cover different aspects of the outcome domain and more than one measure may be required.

Implications for research

An important next step is carefully considering future research regarding clinical outcomes and measurement instruments in the areas of core outcomes without consensus for measurement instruments.

There was high agreement amongst participants at the online consensus meeting that a research priority should be assessing the relative merits of existing instruments not designed for or validated in post COVID-19 condition (e.g. generic or 'legacy' instruments) versus those developed specifically for post COVID-19 condition. Therefore future studies of measurement properties and future consensus projects should focus in this area for advancing research and clinical practice.

It was notable that no instruments from the PROMIS and NeuroQoL modular sets of measures were preferred, despite having advantages related to development using item response theory techniques and extensive validation across multiple conditions. This finding might reflect a lack of familiarity or lack of comparison data (in other relevant conditions) with these measures, or lack of understanding of the potential benefits of their measurement properties.

Based on results of this project, a future research agenda is provided in Box 2.

Box 2: Future research agenda

Overall recommendations

- Study the comparative performance of outcome measurement instruments developed for other conditions versus post COVID-19 specific instruments versus existing generic multi-domain instruments versus Item Response Theory-developed generic measures
- Update the COMS in light of new data in relation to post COVID-19 condition

Outcome domain specific recommendations

Cardiovascular functioning, symptoms and conditions

- Translation and cultural adaptation of NYHA functional class scale

Fatigue or exhaustion

- Study the comparative performance of the three preferred instruments

Nervous system functioning, symptoms and conditions

- Studies phenotyping neurological symptoms in post COVID-19 condition

Cognitive functioning symptoms and conditions

- Studies phenotyping cognitive symptoms in post COVID-19 condition

Mental health functioning symptoms and conditions

- Consider whether a single or multiple measures optimal

Post-exertion symptoms

- Refine the concept and understanding of PEM

Work/occupational changes and study

- Study the comparative performance of the three preferred instruments

Recovery

- Translation and validation of the instrument into languages other than English

Conclusion

A Core Outcome Measure Set (COMS) for post COVID-19 condition in adults is now available to complement the existing Core Outcome Set (COS). This COMS is recommended for immediate use by researchers and clinical services worldwide to improve consistency and comparability of outcome measurement. Within the context of rapidly evolving knowledge regarding post COVID-19 condition, and based on results of this international consensus process, recommendations for future research are provided.

PC-COS Study Group members

Olalekan Lee Aiyegbusi¹, Athena Akrami², Christian Apfelbacher³, Melanie Calvert¹, Jessica Chen⁴, Anastasia Chernyawskaya⁴, Susanna Dodd⁵, Vicky Fallon⁵, Shamil Haroon¹, Sarah E. Hughes¹, Allison Jaure⁶, Alisa Kokorinia⁴, Piero Olliaro⁷, Alexandra Pokrovskaya⁴, Nile Saunders⁴, Jochen Schmitt⁸, Callum Parr⁴, Louise Sigfrid⁹, Frances Simpson¹⁰, Manoj Sivan¹¹

1. Institute of Applied Health Research, University of Birmingham, Birmingham, UK
2. Sainsbury Wellcome Centre, UCL, London, UK
3. Institute of Social Medicine and Health Systems Research, Faculty of Medicine, Otto von Guericke University Magdeburg, Magdeburg, Germany
4. Faculty of Medicine, Imperial College London
5. MRC/NIHR Trials Methodology Research Partnership, Department of Health Data Science, University of Liverpool (a member of Liverpool Health Partners), Liverpool, UK
6. Sydney School of Public Health, The University of Sydney
7. ISARIC Global Support Centre, Nuffield Department of Medicine, University of Oxford, Oxford, UK
8. Center for Evidence-based Healthcare, Medical Faculty Carl Gustav Carus, TU Dresden, Dresden, Germany
9. ISARIC Global Support Centre, Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, UK
10. Coventry University, Coventry, UK
11. University of Leeds, Leeds, UK

Also see supplementary material, Appendix 19

Acknowledgments

We thank the World Health Organization Wouter De Groote, Janet Diaz, Krutika Kuppalli, John Marshall, Srinivas Murthy, Jakobus Preller and Nicoline Schiess and the WHO COVID-19 Clinical research working group for their technical inputs, support in coordination, and dissemination of the survey. We would also like to thank Professor Mike Clarke for his work as an independent facilitator of the online interactive consensus meeting. Finally we would also like to thank Long COVID Support charity members (Christina Barratt, Sophie Evans, Lesley MacNiven, Natalie Rogers and Kirsty Stanley) for providing critical input into the instrument card design.

References

1. Michelen M, Manoharan L, Elkheir N, Cheng V, Dagens A, Hastie C, et al. Characterising long COVID: a living systematic review. *BMJ Glob Health*. 2021 Sep 1;6(9):e005427.
2. Munblit D, Nicholson TR, Needham DM, Seylanova N, Parr C, Chen J, et al. Studying the post-COVID-19 condition: research challenges, strategies, and importance of Core Outcome Set development. *BMC Med*. 2022 Feb 4;20(1):50.
3. Munblit D, Nicholson T, Akrami A, Apfelbacher C, Chen J, De Groote W, et al. A core outcome set for post-COVID-19 condition in adults for use in clinical practice and research: an international Delphi consensus study. *Lancet Respir Med*. 2022 Jul;10(7):715–24.
4. COMET Initiative | Home [Internet]. [cited 2022 Dec 12]. Available from: <https://www.comet-initiative.org/>
5. Prinsen CAC, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a “Core Outcome Set” – a practical guideline. *Trials*. 2016 Sep 13;17(1):449.
6. COMET Initiative | Core Outcome Measures for Post-Covid condition/Long Covid [Internet]. [cited 2022 Nov 3]. Available from: <https://www.comet-initiative.org/Studies/Details/1847>
7. PC-COS study website [Internet]. [cited 2022 Nov 10]. Available from: <https://www.pc-cos.org/publications>
8. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap Cooperative Group During its First Two Years. *Med Care*. 2007 May;45(5):S3.
9. Cella D, Lai JS, Nowinski CJ, Victorson D, Peterman A, Miller D, et al. Neuro-QOL: Brief measures of health-related quality of life for clinical research in neurology. *Neurology*. 2012 Jun 5;78(23):1860–7.
10. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27(5):1159–70.
11. Handbook O. OMERACT Handbook [Internet]. OMERACT Handbook. [cited 2022 Nov 3]. Available from: <https://omeracthandbook.org/handbook>
12. COS Resources – Improving Long-Term Outcomes Research for Acute Respiratory Failure [Internet]. [cited 2022 Nov 3]. Available from: <https://www.improvelto.com/cos-resources/>
13. COMET DelphiManager [Internet]. [cited 2022 Dec 12]. Available from: <https://www.comet-initiative.org/delphimanager/>
14. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol*. 2011 Apr 1;64(4):395–400.
15. PC-COS how protocol_23rd March_clean.pdf [Internet]. Google Docs. [cited 2022 Nov 10]. Available from: https://drive.google.com/file/d/1Scj3CKLKLIO9XgZjBPYkZNXngbUaolJG/view?usp=embed_facebook
16. Tong A, Elliott JH, Azevedo LC, Baumgart A, Bersten A, Cervantes L, et al. Core Outcomes Set for Trials in People With Coronavirus Disease 2019. *Crit Care Med*. 2020 Nov;48(11):1622–35.
17. Needham DM, Sepulveda KA, Dinglas VD, Chessare CM, Friedman LA, Bingham CO, et al. Core Outcome Measures for Clinical Research in Acute Respiratory Failure Survivors. An International Modified Delphi Consensus Study. *Am J Respir Crit Care Med*. 2017 Nov 1;196(9):1122–30.