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], Janet V. Diaz MD[3], Nicola Harman PhD[1], Krutika Kupali MD[3], Margaret O'Hara[4], Caroline B. Terwee[5], Paula R. Williamson PhD*[1], Dale M. Needham PhD*[6,7,8], Daniel Munblit PhD*[9,10,11], Timothy R Nicholson PhD*[12] 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 Clinical Management Team, World Health Organization, Geneva, Switzerland
4 Birmingham Hospitals
5 Amsterdam
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, who contributed equally, 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 NSe designed the study protocol and were responsible for the day to day running of the project. NSe, CP, AK and JC undertook the literature review, identified outcome measures and categorised them for inclusion in the online Delphi survey. NSe coordinated the data revision process. NH, SLG and NSe developed the online Delphi surveys and contributed to the day to day management of the project. DM, TN, JP, PO, CA, AT, JS, DMN, SLG, NSe 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. KK and JVD led the WHO administrative aspects of the study. MO'H provided and coordinated invaluable perspectives of people with lived experience throughout the study into its 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). FS reports input as PPE SLG is the COMET Initiative project coordinator. Nicola Harman reports payment or honoraria for lectures, presentations from the Lancaster University. CA reports grants or contracts from Dr Wolff Group, Bionorica and The European Cooperation in Science and Technology (COST). He also acknowledges consulting fees from the Dr Wolff Group, Bionorica, Sanofi-Aventis Germany and honoraria from AstraZeneca. He serves as a Co-Chair Harmonising Outcome Measures for Eczema (HOME) initiative and Co-Chair Hand Eczema Core Outcome Set (HECOS) initiative. JP was employed by WHO in the Case Management team, HQ, WHE at the time of the manuscript writing. He is also a chair of the e-Learning committee and member of the Council at the European Society of Intensive Care Medicine. JVD is the lead of the clinical management response pillar for COVID-19 and in that capacity convene the WHO Clinical Characterization and Management Research working group. The Post COVID-19 COS steering committee was a sub working group of this bigger group. PRW reports that she received funding from the UK National Institute for Health Research (NIHR). She 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

It is not currently agreed which instruments should be used to measure a recently agreed core outcome set (COS) for clinical practice and research in post COVID-19 condition, also known as Long COVID. A rigorous multi-step modified Delphi 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 the measurement instruments for this COS, thereby developing a Core Outcome Measurement Set (COMS). 594 participants from 58 countries, representing six continents, participated in the Delphi survey. Consensus was reached for including the modified MRC Dyspnea Scale for measurement of the outcome of 'Respiratory functioning, symptoms, and conditions' in the COMS. Two outcome measures from the acute COVID-19 COMS were included in this COMS for consistency; namely 'Survival' (Time until death) and 'Recovery' (Recovery Scale for COVID-19).

Consensus, based on predefined criteria, was not reached for the other 9 outcome domains of the COS, but preferences were expressed for a variable number of measurement instruments 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. This international consensus-based COMS provides a framework for assessing post COVID-19 condition in global clinical research and practice settings which can be reappraised as new data emerges on the performance of different measures.

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 .
- Research and clinical care focused on post COVID-19 condition have substantial heterogeneity in the outcomes evaluated. We previously undertook a consensus study and developed a core outcome set (COS) — an agreed minimum of critical outcomes that should be measured and reported in post-COVID-19 condition in adults for use in clinical research and practice worldwide.
- However, there is a pressing need for defining which outcome measures are the most appropriate to be used to measure critical outcomes, to ensure data comparability and allow for the data synthesis.
- This study sought to provide a consensus on measurement instruments to be used for each of previously defined “core” outcomes” using gold-standard methodology.

Findings

- Consensus was reached with regards to instruments to be used for survival, recovery and respiratory outcomes. No single instrument reached a consensus to be recommended for use with the other outcomes but the most preferred instruments were identified.

Future Directions and Implications

- An important next step is considering the publication of new information regarding the suitability of existing outcome measurement instruments for post-COVID-19 condition, or the publication of new instruments for the core outcomes, for which there is not currently consensus.
- There was agreement amongst participants at the consensus meeting that assessment of the relative merits of existing instruments not developed for post COVID-19 condition (‘legacy’ instruments) versus those developed specifically for Long COVID is one of the primary priorities of post COVID-19 condition research.
- The use and reporting of this COS for adults with post COVID-19 condition is an important step to optimise and accelerate research, especially the development of evidence-based treatments, and to ensure consistent evaluation of these important outcomes in clinical settings.

Introduction

Although the majority of people previously infected with SARS-CoV-2 infection rapidly recover, a substantial number of individuals experience persistent symptoms for months or even years. Results of a recent large multinational study,(1) which included data for 1.2 million individuals from 22 countries, estimated global proportions of individuals with persistent fatigue, cognitive, and respiratory symptom clusters following symptomatic COVID-19 at 6.2% (95% uncertainty interval [UI], 2.4%-13.3%). Taking into account the number of individuals who had asymptomatic COVID-19, the real prevalence of post COVID-19 condition can be even higher. With an increased recognition of the problem multiple clinical trials were launched to tackle the absence of appropriate management options. However, selection of valid outcomes as a measurement of intervention effectiveness quickly turned into one of the main methodological challenges of post COVID-19 condition research.(2)

In 2021, a multidisciplinary international group of experts defined the core set of outcome (COS) domains to be used in all future clinical studies and care for people with post COVID-19 condition.{Updating} This COS used agreed research methods, recommended by the Core Outcome Measures in Effectiveness Trials (COMET) initiative. Consensus on outcomes to be included in the COS was reached 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. Agreement was reached on the inclusion of twelve outcomes in the COS for post COVID-19 condition clinical research and care.(3) Outcomes included fatigue; pain; post-exertion symptoms; work or occupational and study changes; survival; and functioning, symptoms, and conditions for each of cardiovascular, respiratory, nervous system, cognitive, mental health, physical outcomes and recovery.

Now that a COS for post COVID-19 condition is identified at the outcome domain level (i.e. WHAT to measure), it is important to achieve consensus on HOW these domains should be measured, i.e., which outcome measurement instruments should be used. When using the term 'instrument,' we are referring to any outcome measurement instruments, tools, procedures, etc., that are used to measure an outcome domain. It is likely that many measurement instruments exist to measure each of the domains included in the post COVID-19 condition COS; however, the difficulty will be identifying which measurement instruments are valid, reliable, and feasible for use in this condition. This issue is particularly challenging in post COVID-19 condition, as instruments are unlikely to be validated specifically in this population. It is possible that there may be no suitable existing instruments to measure some of the outcome domains included in the post COVID-19 condition COS; hence, it will be important to identify any relevant instruments that are currently under development and evaluation to inform a research agenda. Once we have identified, evaluated and reached consensus on a set of instruments, the objective will be to promote their implementation in future post COVID-19 condition clinical research and care.

Here, we report on the PC-COS project findings, which have led to development of a COMS for post-COVID-19 condition in adults (≥18 years of age) that is intended for use in clinical research and care.

Methods

This second stage of the PC-COS project followed the COSMIN-COMET guideline on the selection of outcome measurement instruments for outcomes included in a COS. This guidance, The PC-COS study was prospectively registered with the COMET Initiative (Core Outcome Measures in Effectiveness Trials).(4)] The study protocol was developed a priori and was approved by the UK Research Ethics Committee. Details regarding study management process are comprehensively described in the COS paper.(3)
.

Step 1. Establishing the aspects to consider

Agreeing the construct (i.e. outcome or domain) to be measured in the target population was the first step in the selection of outcome measurement instruments. The target population included adults with post COVID-19 condition. The selected instruments would be the minimum recommended measures to be used in both clinical research studies and practice across all settings (e.g., high and low income) throughout the world. The outcome domains included in the post COVID-19 condition COS are those for which we were aiming to identify and select instruments.

Step 2. Finding existing outcome measurement instruments

Systematic review of instruments currently used in studies
A systematic review was undertaken to identify all outcome measurement instruments used in published and ongoing studies for post COVID-19 condition for each outcome domain in the COS. This review used the literature set collected and evaluated for the COS.(3) Thus, included studies published between 1st January 2020 up to 17th March 2021, which were identified in the ISARIC (International Severe Acute Respiratory and emerging Infection Consortium) living systematic review.(5) Clinical trial protocols were identified from two clinical trial registries (ICTRP database and ClinicalTrials.gov), which were searched on the 19th May 2021. Additional major review articles and clinical trials that were not captured in the above searches, but considered critical and were included. Search dates ranged from 25th March 2021 up to 25th May 2021. All articles and protocols were evaluated independently by two researchers (three pairs from the extended team of NS, AC, JC, CP, AP, NS). All team members had experience with systematic reviews and/or patient-reported outcome measures.

Once the instruments were collected, they were mapped to one of the domains in the post COVID-19 condition COS. The list was then reviewed by the Study Management Group to remove double entries and other errors. Instruments that mapped to multiple COS domains, including the post COVID-19 condition-specific measures, were also included in a category of multi-domain instruments or post COVID-19 condition specific instrument, respectively. All other instruments that did map to any of the COS domains were not considered.

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

Due to the extensive nature of the list of domain mapped instruments, it was decided that only the most frequently used, along with any post COVID-19 condition-specific instruments, would go through to Step 3 in order to reduce the respondent burden in the modified Delphi consensus process. Frequency was defined according to the top five most used instruments, relating to each COS domain, in the articles and protocols identified in the above systematic review. For those domains, where less than five instruments were identified for a domain, instruments that had only been used once were included. However, if more than five instruments met the frequency criteria for a domain, only instruments that had been 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 expert members of the Study Management Group (DN, TN, DM) reviewed the instruments and reached agreement on the top five to be included based primarily on consideration of their relevance to post COVID-19 condition and feasibility of administration. All instruments that did not meet this frequency criterion, other than the novel post COVID-19-specific instruments, were excluded at this stage.

Other sources

For core outcomes with less than five instruments identified based on the prior process, we identified instruments that had been recommended to measure outcomes in other COS in relevant populations. based on existing records of records of the COMET Initiative which maintains an international registry of COS projects. We also identified relevant PROMIS (Patient-Reported Outcomes Measurement Information System) and Neuro-QoL (Quality of Life in Neurological Disorders) instruments via a search of the PROMIS and Neuro-QoL databases.(6,7) Newly developed post COVID-19 condition-specific instruments were also identified through discussion with international experts.

Step 3. Assessing quality of outcome measurement instruments

Measurement properties

The Study Management Group considered it unlikely that there would be any existing studies that had assessed the measurement properties of any of the non-COVID instruments for a post COVID-19 population. Therefore, owing to the nature of this condition and in the interests of speed and feasibility, a revised quality assessment approach was adopted, which did not involve a COSMIN quality assessment. In Step 2, we endeavoured to identify frequently used generic instruments, for which there would likely be existing studies on measurement properties in other populations.

We undertook a rapid mini COSMIN assessment to assess the quality of these post COVID-19 condition-specific instruments. which involved identifying the patient-reported outcome measure (PROM) development papers and additional content validity studies if available. Two PROM experts (CT and VF) independently rated the instruments according to predefined COSMIN criteria (see Appendix 1), judging the relevance, comprehensiveness, and comprehensibility of the instruments.

Feasibility aspects

In addition to assessing the measurement properties of the newly developed post COVID-19 condition-specific instruments, feasibility aspects were considered for all instruments to determine whether the instrument could be easily applied in the intended settings, accounting for other constraints, i.e. time, cost, interpretability.(8) Guided by previous studies that have selected instruments for outcomes included in COS,(9) we selected the most important feasibility aspects to be considered. This included consideration about the languages and translations available for each instrument.

Step 4. Reaching agreement on the instruments to be included in the COS

Online Delphi survey

The list of instruments was prioritised in a three round online Delphi consensus process delivered using DelphiManager software. The instruments,grouped according to the COS outcome domains, were presented in a random order for each participant. Instruments that mapped to multiple COS domains, and the post COVID-19 condition-specific measures, were presented in separate categories. Consent for participation in the online survey was sought online prior to accessing the online consensus process.

Stakeholder Involvement

All participants who participated in the COS were invited to participate. These participants belonged to the following 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 of working in the field of post COVID-19 condition. Additional stakeholders were invited to participate via direct email from the study authors or from relevant patient or professional organisations. We contacted the lead investigators of the study protocols, which were registered in the clinical trial registries following the systematic review search date. We also used current Long COVID clinics and research groups. To increase global representation, we engaged with stakeholders from countries for which there was a relatively lower representation in the prior COS project, asking them to disseminate details about this study to their networks. Individuals were also able to express an interest in participating in the survey via the PC-COS website (https://www.pc-cos.org/). There was no restriction on the number of eligible participants in each stakeholder group completing the Delphi survey.

Materials

Participants were provided access to the following materials which they were asked to review when completing the Delphi survey:

- The list of outcome domains that were agreed for inclusion in the COS, along with lay definition of the outcome).
- The instruments, which had been mapped to the COS outcome domains, to be rated in the Delphi.
- Instrument Cards summarising the feasibility and the measurement properties of the post COVID-19 condition-specific instruments.
- Definitions of the specific measurement properties (e.g., reliability, validity, responsiveness) that were evaluated on the Instrument Cards.

Standardised “Instrument Cards” were produced for all instruments included in the Delphi survey as previously done.(9) The Instrument Cards also included plain language details, written with input from the patient research partners (see Appendix 2).

Scoring

In round 1 of the online Delphi survey, participants were asked to rate whether they thought the instruments presented should be used to measure the outcome domains included in the COS. Participants rated all instruments using the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) nine-point Likert scale,(10) 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 for participants to use if they did not feel able to rate a specific instrument. Once participants had rated all instruments, prior to submission, they could review their ratings and change if they wished. Participants also had the option to partially complete the survey using the ‘Save for later’ option.

As part of round 1, participants were asked to suggest any additional relevant instruments that was not already included in the consensus process. Any additional instruments suggested by >1% of participants, in each stakeholder group, was considered and added to round 2 of the Delphi survey with"Instrument Cards" developed.

In the second round of the Delphi, responses for each stakeholder group were summarised for each instrument and displayed graphically. Any instruments that reached consensus for exclusion (as defined below) in round 1 were not retained for round 2. For all instruments included in round 2, participants were able to view the responses by stakeholder group alongside their own rating from round 1 and asked to re-rate the instrument. Participants 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.

All instruments rated in round 2 were summarised and those reaching consensus for exclusion (as defined below) were not retained for round 3. The process and scoring followed the same format as round 2, with the addition of participants being able to reflect on the scores given for the new instruments that were suggested during round 1. COSMIN recommend the selection of a single instrument for each outcome in a core outcome measurement set (COMS),(11) therefore in round 3, participants were encouraged to give a single instrument a rating of 7-9 (i.e. 'critically important') within most of the domains, and to rate all other instruments within that domain as 6 or less.

Data analysis

In rounds 1, 2 and 3 of the Delphi, we used descriptive statistics to summarise the 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 that responses would be included in the analysis if a participant assessed all instruments in at least one domain. We produced graphs using R (version 4.0.2) to display the distribution of ratings for each instrument, stratified by stakeholder group, which were shown to participants in rounds 2 and 3 of the Delphi survey.

Following the round 3 instructions, where participants were encouraged to give a single instrument a rating of 7-9 within most domains, additional sensitivity analyses were conducted to identify which instruments had been rated most strongly by each of the three stakeholder groups, within each of the domains. Two sensitivity analyses were undertaken. Sensitivity analysis 1 involved identifying which instrument(s) within each 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 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 the distribution of the mean overall scores from the third round 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 subsequent rounds of the Delphi process, was defined as 50% or less participants, in **each** stakeholder group, rating as 7-9. Prior to reviewing results from round 2, we modified the exclusion criteria for this round, to reduce participant burden and consensus for an instrument to be excluded from round 3 of the Delphi, was defined 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 defined as 80% or more participants, in **each** stakeholder group, rating as 7-9 **and** <10% participants, in **each** stakeholder group, rating as 1-3.

Online Consensus meeting

At the end of the final round of the Delphi survey, participants were invited to express their interest in attending an online consensus meeting. Expressions of interest were considered, allowing for representation across stakeholder groups and geographical locations. Prior to attending the meeting, participants received background information, a copy of their own ratings from the Delphi survey, a summary of the results from the Delphi survey and details about the instruments to be discussed at the meeting. In advance of the consensus meeting, people with post COVID-19 condition were invited to attend a pre-meeting to inform them about what would happen at the consensus meeting and provide the opportunity to ask any questions.

The consensus meeting, which was held via Zoom, was conducted in English and chaired by an experienced independent facilitator (MC). The meeting focused on the results of the final round of the Delphi survey and instruments that were rated most highly by all stakeholder groups were prioritised for discussion. For each instrument being 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, within their predefined stakeholder groups, as to whether the proposed instrument should be included in the COMS. Consensus for an instrument to be included in the COMS was defined as 80% or more participants in both stakeholder groups agreeing it should be included.

Results

Identification of outcome measurement instruments

The systematic review identified a total of 319 instruments used in post COVID-19 condition studies. Following removal of duplicates and mapping of instruments to the core domains, this list was reduced to 293 instruments. independent review of the instruments (see Methods) resulted in 91 instruments, which was 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 in the Delphi round 1 (see Appendix 3). These 54 instruments were mapped under 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). The outcome domains were randomised in the online Delphi process and thus displayed in a random order for each participant.

Delphi process

Round 1 of the online Delphi was open from June 6-29, 2022, with 711 individuals registering to participate. Of these 711, 594 participants (84%) from 58 countries completed the first round, with 447 fully completing and the remaining 147 partially completing (i.e. rating all instruments for at least one, but not all, domains). All 594 participants were invited to participate in round 2. Round 2 was open from July 13 to Aug 8, 2022, during which time 394 (66%) participants fully (n=362) or partially (n=32) completed. Round 3 was open from Aug 12 to Sep 12, 2022, with 359 participants (60%) fully (n=341) or partially (n=18) completing it. Of these 359 participants, 324 (82%) had completed round 2. Appendix 4 shows the breakdown of response rate, by stakeholder group, across Delphi rounds. Demographic characteristics of the Delphi participants are presented in Table 2. Further details of Delphi participants are presented in Appendix 5.

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

	Delphi round 1 (n=594)	**Delphi round 2 (n=394)**	**Delphi 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)

At the end of round 1, 13 instruments met the criteria for exclusion. One additional instrument was also excluded (Fukuda Criteria for CFS), resulting in 14 instruments not being retained for round 2 (see Appendix 6). suggestions for additional instruments resulted in139 free-text responses from which two instruments met a prior criteria for inclusion: WHO Disability Assessment Schedule 2.0 – 12 Item, and Nijmegen Questionnaire. 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 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 being added into round 2.

Fifty-four instruments were included in round 2 (see Appendix 7), 12 of which met exclusion criteria (see Appendix 8). Thus, 42 instruments were retained for inclusion in round 3 (see Appendix 9).

At the end of round 3, none of the 42 instruments met the a priori criteria for inclusion. The results of the Delphi survey are provided in Appendices 10-12. The sensitivity analyses revealed that for two of the 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 consensus meeting. For the pain domain, the Brief Pain Inventory was the only instrument included for rating in round 3 and so was also taken forward for discussion and voting in the 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, this instrument was also recommended for inclusion in the COS for acute COVID-19,(3) and thus was also included for discussion in the 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 deemed necessary for discussion in the consensus meeting. The results of the sensitivity analyses are provided in Appendix 13.

Consensus meeting

In advance of the consensus meeting, a one-hour pre-meeting for people with post COVID-19 condition was held on the 27th September 2022. This session, which was attended by eight participants, provided information on what would happen at the consensus meeting and offered an opportunity to ask questions.

The online consensus meeting was held on the 29th September 2022. Thirty-seven people attended the 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, the 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 happened previously. Therefore, the voting groups included 12 people with post COVID-19 condition and 13 healthcare professionals and researchers. Some participants were unable to attend for 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 14. We found no evidence of selection bias: the average round 3 Delphi scores were similar between participants who attended the consensus meeting (6.01) and those who did not attend the meeting (6.31).

At the beginning of the meeting, the attendees were informed about the two measures that were confirmed for inclusion in the COMS. From the outset, it was agreed that the COS outcome 'Survival' would be measured by *'Time until death'*, as consistent with previous COS. It was also agreed that the COS outcome 'Recovery' would be measured using the question recommended in the COS for acute COVID-19 - *'How long it takes to recover, i.e., feel better, no longer having symptom'*.[ref] Following this reminder of the a priori Survival and Recovery measurement decisions, 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' to include it in the COMS. The mMRC Dyspnoea Scale was therefore added to the COMS, as the recommended instrument for 'Respiratory functioning, symptoms, and conditions', alongside the two previously agreed instruments for 'Survival' and 'Recovery' (see Table 3).

Table 2. Core Outcome Measurement Set

Core Outcome	**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 outcome measurement instruments for the remaining nine core outcome domains in the post COVID-19 condition COS. Table 4 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 domain-specific 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 also not reached for any multiple domain outcome measurement instruments, nor any post COVID-19 condition-specific measurement instruments. Table 5 indicates instruments with the highest level of support based on the consensus process. The results from the Delphi survey and the consensus meeting are included in Appendix 15.

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

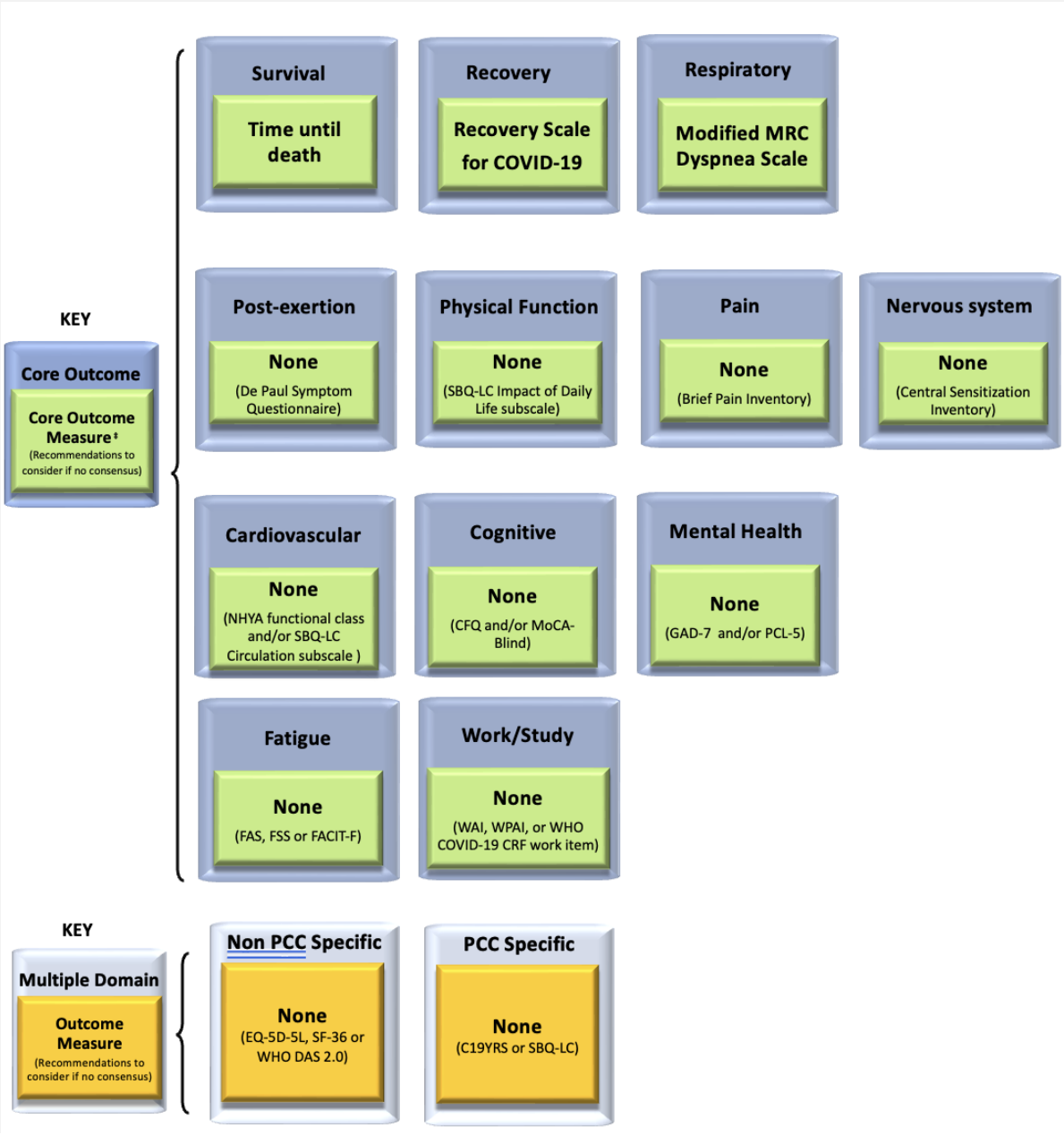
Instrument Type	**Measurement Instrument Options**
Multiple Domain	Euroqol 5-Dimension 5-level (EQ-5D-5L)
	Medical Outcomes Study Short-Form-36 (SF-36)
	WHO Disability Assessment Schedule 2.0 - 12 Item (WHO DAS 2.0)
Post-COVID condition specific	COVID-19 Yorkshire Rehabilitation Screening (C19-YRS) Scale
	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 (see Appendix 16).

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

Core Outcome Measurement and Core Outcome Sets for Post COVID-19 Condition (PCC)



Discussion

This consensus process resulted in three measures meeting the criterion for inclusion in the Core Outcome Measure Set (COMS) - as such these should be measured and reported in post-COVID-19 condition in adults for use in clinical research and practice worldwide. Consensus was reached for instruments to measure survival, recovery and respiratory outcomes. No measurement instruments achieved consensus for the other 9 core outcome domains, although the consensus process did reduce the number of potential instruments measuring the 12 core outcomes from over 300 to 19.

Limitations

There are several key limitations of this study to note, many of which have been described in detail in the publication presenting results of the first stage of the project (3). In brief, the Delphi survey 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. In the consensus meeting there was an insufficient number of people with post COVID-19 condition who were also health professionals to constitute their own voting group, limiting the information that could have been provided by another group with dual, and potentially unique, perspectives.

There are additional limitations in this 'how to measure' stage. The study was only conducted in English. This decision was taken based on the extra complexity associated with potential translation of the process to multiple languages, which would include needing to translate all information materials as well as all instruments assessed, some of which were available only in English. This will have resulted in significant resource requirements and slowing down of the study progress. This limitation is likely to have impacted on global participation, especially from ethnic and social groups less likely to speak English. However, despite this limitation a reasonable global representation was reached.

This 'how to measure' stage required substantially greater involvement from the participants and was more demanding than the 'what to measure' stage, particularly for people with post COVID-19 condition where fatigue, low energy levels, and cognitive difficulties are common and often severe and significantly disabling. This likely explains the lower completion rates and higher attrition through the three Delphi rounds when compared to the "what to measure" stage. In order to mitigate these limitations, the order in which the instruments were presented was randomised based on outcome domain, minimising the potential for bias due to survey fatigue. Major efforts were made, in collaboration with patient partners, to explain that rating all of the instruments for just one of the outcome domains was the minimum participation required and would provide highly valuable data. Participants were encouraged to rate as many measurements as they could, even if they were unable to complete all ratings.

Challenges in achieving consensus

For the nine outcome domains for which consensus was not reached there are several potential reasons why this occurred. For example, post COVID-19 condition is a highly heterogeneous disorder and different combinations of symptoms, and even variable experience of specific symptoms, can occur potentially influencing instrument preference.

For clinicians and researchers, their past experience with measurement instruments may have been highly variable, influencing their ratings.

As this is a new condition there is understandably a lack of high quality data to guide choices between both existing instruments used in post COVID-19 condition studies and those designed specifically for the condition. It is therefore arguable that consensus would be harder to achieve and the threshold for consensus applied here may have been too high for such a situation.

It is important to note that not reaching consensus for how to measure a number of core outcomes is not uncommon and has been previously reported in projects developing COS for other disorders. For example, even for a relatively well defined and understood condition such as post ICU syndrome, consensus was not reached despite previous extensive research and therefore significant experience and data to guide preference (ref Dale's COMS and perhaps one other well conducted COMS). It is important to bear in mind that post COVID-19 condition is not only a new condition with high heterogeneity but it is still very poorly understood at a mechanistic level. As such, so is the relationship and potential similarity with other disorders and therefore the appropriateness, or otherwise, of measures designed for, and validated in other conditions.

Implications for practice

For the three outcomes for which there was consensus on how to measure them (Table 2), the instrument should be used in all research and clinical practice. Other measures of these outcomes can be used in addition if it is felt they provide extra value without increasing the burden unduly in a given setting.

For the other outcomes, where there was no consensus on how to measure them (Table 3) these instruments should be considered for use after factoring in the specific setting in which they are to be used.

For outcomes where just one instrument is suggested (post exertion, physical function, pain and nervous system) a simple consideration of this instrument's suitability for the study or setting is required. For example, is the instrument available in the appropriate languages, is it feasible in terms of patient burden, is it affordable if there is a fee for use.

If there are two or three instruments the relative merits of these same factors should be weighed up between the options. Furthermore some outcomes, such as fatigue, might be measured largely or fully by most instruments and as such it would generally be preferable to choose one measure, whereas for other outcomes it might be that different measures cover different aspects of the outcome domain and it might be that more than one measure is required in a given setting.

Implications for research

An important next step is considering the publication of any new information regarding the suitability of existing outcome measurement instruments for Long COVID, or the publication of new instruments for the core outcomes where there is not currently consensus on how they should be measured.

There was agreement amongst participants at the 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 ('legacy' instruments) versus those developed specifically for post COVID-19 condition. Therefore studies should focus on comparing legacy instruments to post COVID-19 condition-specific instruments and combinations of both types of instruments both in research and clinical practice.

It was notable that no instruments from the PROMIS and NeuroQoL modular sets of measures were preferred, despite having been extensively validated across multiple conditions. This might reflect a lack of experience and familiarity with these measures, or lack of understanding of their potential benefits, particularly for use in new disorders where a more 'generic' instrument, that makes no assumptions about mechanism or similarity to other disorders could be beneficial. These measures have also been developed with item response theory (IRT) so can potentially reduce the number questions required to produce equivalent measurement and therefore the burden for patients and researchers.

There were some important points raised about measurement instruments during the Delphi studies and consensus meeting that warrant further research. A summary of a proposed future research agenda can be seen in Box 2.

Box 2: Future research agenda

Overall recommendations

- Update the COMS when sufficient new data becomes available on existing measures used in post COVID-19 condition, and/or new measures are developed
- Study the comparative performance of existing ('legacy') versus post COVID-19 specific versus Item Response Theory measures

Outcome domain specific recommendations

Cardiovascular

- Translation and validation of NYHA into languages other than English

Fatigue

- 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

- Study the comparative performance of the three preferred instruments

Recovery

- Translation and validation into languages other than English

Conclusion

A Core Outcome Measure Set (COMS) for post COVID-19 condition is now available to compliment the Core Outcome Set (COS). It is hoped researchers and clinical services worldwide rapidly adopt this to optimise clinical monitoring as well as data collation and comparison to accelerate evidence generation to better understand and treat this new condition with a major global health and socioeconomic impact. More work is needed to develop better understanding of the optimal outcome measures for this condition and reaching consensus on how to best to measure more outcomes and further develop a COMS as new data updates knowledge and the views of key stakeholders. This will require a significant and coordinated effort by the international research community in close partnership with patients with lived experience of the condition and other experts and key stakeholders.

PC-COS Study Group members

Olalekan Lee Aiyegbusi[1], Athena Akrami[2], Christian Apfelbacher[3], Melanie Calvert[1], Jessica Chen[4], Anastasia Chernyawskaya[4], Wouter De Groote[5], Susanna Dodd[6], Vicky Fallon[6], Shamil Haroon[1], Sarah E. Hughes[1], Alisa Kokorinia[4], John Marshall[7], Srinivas Murthy[8], Piero Olliaro[9], Alexandra Pokrovskaya[4], Jakobus Preller[10], Nile Saundres[4], Nicoline Schiess[11], Jochen Schmitt[12], Malcolm "Calum" G. Semple[4], Louise Sigfrid[13], Frances Simpson[14], Manoj Sivan[15], Allison Tong[16]

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. NCD Department, Rehabilitation Programme, World Health Organization, Geneva, Switzerland
6. MRC/NIHR Trials Methodology Research Partnership, Department of Health Data Science, University of Liverpool (a member of Liverpool Health Partners), Liverpool, UK
7. University of Toronto, Toronto, Canada
8. University of British Columbia, Canada
9. ISARIC Global Support Centre, Nuffield Department of Medicine, University of Oxford, Oxford, UK
10. Health Care Readiness Unit, Health Emergencies Unit, World Health Organization, Switzerland
11. Brain Health Unit, World Health Organization, Geneva, Switzerland
12. Center for Evidence-based Healthcare, Medical Faculty Carl Gustav Carus, TU Dresden, Dresden, Germany
13. ISARIC Global Support Centre, Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, UK
14. Coventry University, Coventry, UK
15. University of Leeds, Leeds, UK
16. Sydney School of Public Health, The University of Sydney

Acknowledgments

We would like to thank Professor Mike Clarke for his work as an independent facilitator of the online interactive consensus meeting. We would also like to thank Long COVID Support charity (<<List names of those that helped>>) for providing input into the Instrument card design.

References

1. Global Burden of Disease Long COVID Collaborators. Estimated Global Proportions of Individuals With Persistent Fatigue, Cognitive, and Respiratory Symptom Clusters Following Symptomatic COVID-19 in 2020 and 2021. JAMA. 2022 Oct 25;328(16):1604–15.
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 | Core Outcome Measures for Post-Covid condition/Long Covid [Internet]. [cited 2022 Nov 3]. Available from: https://www.comet-initiative.org/Studies/Details/1847
5. 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.
6. 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.
7. 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.
8. Handbook O. OMERACT Handbook [Internet]. OMERACT Handbook. [cited 2022 Nov 3]. Available from: https://omeracthandbook.org/handbook
9. COS Resources – Improving Long-Term Outcomes Research for Acute Respiratory Failure [Internet]. [cited 2022 Nov 3]. Available from: https://www.improvelto.com/cos-resources/
10. 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.
11. 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.
12. Munblit D, Buonsenso D, Sigfrid L, Vijverberg SJH, Brackel CLH. Post-COVID-19 condition in children: a COS is urgently needed. Lancet Respir Med. 2022 Jul 1;10(7):628–9.