Core outcome measurement instruments for use in clinical and research settings for adults with post-COVID-19 condition: an international Delphi consensus study

Sarah L Gorst, Nina Seylanova*, Susanna R Dodd, Nicola L Harman, Margaret O'Hara, Caroline B Terwee, Paula R Williamson†, Dale M Needham†, Daniel Munblit†, Timothy R Nicholson†, and the PC-COS study group‡*

Post-COVID-19 condition (also known as long COVID) is a new, complex, and poorly understood disorder. A core outcome set (COS) for post-COVID-19 condition in adults has been developed and agreement is now required on the most appropriate measurement instruments for these core outcomes. We conducted an international consensus study involving multidisciplinary experts and people with lived experience of long COVID. The study comprised a literature review to identify measurement instruments for the core outcomes, a three-round online modified Delphi process, and an online consensus meeting to generate a core outcome measurement set (COMS). 594 individuals from 58 countries participated. The number of potential instruments for the 12 core outcomes was reduced from 319 to 19. Consensus was reached for inclusion of the modified Medical Research Council Dyspnoea Scale for respiratory outcomes. Measures for two relevant outcomes from a previously published COS for acute COVID-19 were also included: time until death, for survival, and the Recovery Scale for COVID-19, for recovery. Instruments were suggested for consideration for the remaining nine core outcomes: fatigue or exhaustion, pain, post-exertion symptoms, work or occupational and study changes, and cardiovascular, nervous system, cognitive, mental health, and physical outcomes; however, consensus was not achieved for instruments for these outcomes. The recommended COMS and instruments for consideration provide a foundation for the evaluation of post-COVID-19 condition in adults, which should help to optimise clinical care and accelerate research worldwide. Further assessment of this COMS is warranted as new data emerge on existing and novel measurement instruments.

Lancet Respir Med 2023

Published Online November 2, 2023 https://doi.org/10.1016/S2213-2600(23)00370-3

*Joint first authors

†Joint last authors

‡Study group members listed in the appendix (p 2)

MRC-NIHR Trials Methodology Research Partnership, Department of Health Data Science, University of Liverpool, Liverpool, UK (S L Gorst PhD, N L Harman PhD, S R Dodd PhD, Prof P R Williamson PhD); **Independent researcher, London, UK** (N Seylanova MD); **Long Covid Support, London, UK** (M O'Hara PhD); **Methodology Program, Amsterdam Public Health Research Institute, and Department of Epidemiology and Data Science, Amsterdam UMC, Amsterdam, Netherlands** (Prof C B Terwee PhD); **Outcomes After Critical Illness and Surgery Research Group, Pulmonary and Critical Care Medicine, Department of Medicine, and Department of Physical Medicine and Rehabilitation, Johns Hopkins University School of Medicine, Baltimore, MD, USA** (Prof D M Needham PhD); **Department of Paediatrics and Paediatric Infectious Diseases, Institute of Child's Health, Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russia** (D Munblit PhD); **Research and Clinical Center for Neuropsychiatry, Moscow, Russia** (D Munblit); **Care for Long Term Conditions Division, Florence Nightingale Faculty of Nursing, Midwifery and Palliative Care** (D Munblit) **and Neuropsychiatry Research and Education Group, Institute of Psychiatry, Psychology and Neuroscience** (T R Nicholson PhD), **King's College London, London, UK**

Introduction

Although many people fully recover from COVID-19 within a few weeks to months of SARS-CoV-2 infection, an estimated 45% continue to have one or more symptoms beyond the acute phase, regardless of hospitalisation status.[1] Such symptoms are collectively known as post-COVID-19 condition or long COVID, although other terms are also used.[2] WHO has defined post-COVID-19 condition as occurring in individuals with a history of probable or confirmed SARS-CoV-2 infection, usually 3 months from onset, with symptoms lasting for at least 2 months that cannot be explained by an alternative diagnosis. Common symptoms include fatigue, shortness of breath, and cognitive dysfunction, which can impact everyday functioning. Symptoms can be new onset following initial recovery from an acute COVID-19 episode or persist from the initial illness and might fluctuate or relapse over time.

Clinical monitoring and trials of interventions for post-COVID-19 condition will not only benefit from the identification and use of optimum (eg, the most valid and reliable) measurement instruments but will also benefit from consistent use of these instruments across services and studies, which will facilitate comparisons and collation of data, thus accelerating evidence synthesis. However, there is substantial heterogeneity in the evaluation and reporting of relevant outcomes and associated measurement instruments.[3] In 2021, a multidisciplinary international group reached consensus on a core outcome set (COS) of 12 outcomes that should be measured in all future clinical studies and in clinical care for people with post-COVID-19 condition.[4] Development of this COS involved the use of established research methods, recommended by the Core Outcome Measures in Effectiveness Trials (COMET) Initiative. Consensus on the COS was obtained following a two-round online modified Delphi process delivered in five languages. Stakeholder groups in the consensus process included people with post-COVID-19 condition and their family members or caregivers, health-care professionals and researchers with post-COVID-19 condition, and health-care professionals and researchers with experience in treating and studying people with post-COVID-19 condition. The COS included twelve outcomes: fatigue or exhaustion; pain; post-exertion symptoms; work or occupational and study changes; survival; recovery; and functioning, symptoms, and conditions for each of cardiovascular, nervous system, cognitive, mental health, respiratory, and physical outcomes.

After a COS has been established, consensus is needed on the instrument(s) that should be used for each core outcome through the development of a core outcome measurement set (COMS). When using the term instrument, we are referring to outcome measurement instruments, tools, and procedures used to measure an outcome. Many instruments are likely to exist to measure each outcome in a COS; however, the challenge is to achieve consensus on the most appropriate instrument that should be used, as a minimum, by all studies and clinical services. This issue is particularly challenging for post-COVID-19 condition because it is a new condition

Key messages

Rationale and approach

- Post-COVID-19 condition (also known as long COVID) is a complex disorder that comprises a wide array of signs and symptoms that can last for many months or years after SARS-CoV-2 infection; substantial heterogeneity in the evaluation and reporting of outcomes, and associated measurement instruments, for post-COVID-19 condition, even within symptom domains, has hampered progress in research and clinical services
- We previously conducted a consensus study that led to the development of a 12-item core outcome set (COS)—an agreed minimum set of outcomes 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 agreement on the most appropriate measurement instruments for these core outcomes to advance research and clinical services
- We aimed to develop a core outcome measurement set (COMS) for the 12 outcomes in the post-COVID-19 COS via an international consensus study that involved a literature review, a three-round online modified Delphi process (with 594 participants from 58 countries, 50% of whom were people with post-COVID-19 condition and their family members or caregivers, rating 54 different instruments), and an online consensus meeting

Findings

- The number of potential instruments for measuring the 12 core outcomes was reduced from 319 to 19 during the consensus process
- Measurement instruments for three outcomes—survival, recovery, and respiratory functioning, symptoms, and conditions—were included in the final COMS, which we recommend for use in clinical and research settings for adults with post-COVID-19 condition; for the remaining nine core outcomes, although consensus was not reached for any single instrument, those with the greatest level of support, based on the consensus process, were identified and should be considered for use in clinical practice and research

Future directions and implications

- Use of this COMS for adults with post-COVID-19 condition will facilitate the consistent measurement and reporting of core outcomes in both clinical and research settings and, through optimised comparisons and synthesis of data, help to accelerate research, particularly the development of much-needed evidence-based interventions
- As new data pertaining to post-COVID-19 condition and related measurement instruments becomes available, the COMS should be subjected to periodic reassessment, and further methodological research investigating and comparing suggested outcome measures (that did not reach consensus) should be undertaken
- Participants in the consensus meeting agreed that assessment of the relative merits of existing instruments that were not developed for post-COVID-19 condition versus those developed specifically for post-COVID-19 condition should be a high priority for future outcome measurement research

Correspondence to:
Dr Timothy R Nicholson, Neuropsychiatry Research and Education Group, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, UK.
timothy.nicholson@kcl.ac.uk

See Online for appendix

and instruments are less likely to have been validated specifically in this clinical population. It is possible that no suitable instrument exists to measure some outcomes in the COS; hence, it is important to identify any potentially suitable instruments that are currently under development and evaluation as well as instruments that are used in other relevant populations.

To address the need for a consensus on measurement instruments for the outcomes identified in the COS for post-COVID-19 condition in adults,[4] the second stage of the Post-COVID-19 Core Outcome Set (PC-COS) project was undertaken by an international group of multidisciplinary experts from, and in collaboration with, the COMET Initiative, the WHO COVID-19 Clinical Research working group, the International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC), and people with post-COVID-19 condition. Further input from people with post-COVID-19 condition was provided throughout the project by representatives of patient organisations and charities (Long Covid Support and Long Covid SOS) and patient partners in the study team. Here, we report on the results of this second stage of the PC-COS project, which led to the development of a COMS for post-COVID-19 condition in adults (≥18 years of age) that is intended to be the minimum set of instruments recommended for use in clinical practice and research globally, including in low-income and high-income settings.

Methods

This second COMS stage of the PC-COS project followed Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN)–COMET guidelines on selecting measurement instruments for outcomes included in a COS.[5] The study comprised three phases: (1) a literature review to identify measurement instruments for consideration by the participants; (2) a three-round online modified Delphi process to rate the importance of the selected instruments for a COMS; and (3) an online interactive consensus meeting to review and agree upon the final COMS. The study protocol[6] was developed a priori and was approved, along with all study materials, by the UK Health Research Authority and by the South West–Cornwall and Plymouth Research Ethics Committee (REC number 21/SW/0109). The project was prospectively registered on the COMET database.[7]

Study groups and participants

The core study team (SLG, NS, PRW, DMN, DM, and TRN) designed the study protocol, oversaw the second phase of the PC-COS project, identified and invited individuals with relevant expertise to form the core author group, and were responsible for the day-to-day running of the project. The core author group had expertise in methodology, various fields of clinical medicine and clinical research, and public and patient engagement. A methods team (SLG, NS, CBT, PRW, DMN, DM, and TRN), led by PRW, was established to develop and oversee the project methods, with additional methodological input from NLH and SRD. A PC-COS study group was established by the core study team. Study group participants were identified and invited through expert networks including ISARIC and the COMET Initiative, and support groups for people with lived experience of long COVID. The PC-COS study group comprised 22 members from four countries, including health-care professionals, researchers representing a range of

medical fields, methodologists, and people with post-COVID-19 condition and their carers, and was actively involved in the design and conduct of this project (see appendix p 2 for further details of the study group members). The PC-COS study group was kept informed about and provided feedback on study progress throughout the project.

MO'H was the patient partner lead and also advised and assisted on recruitment and survey completion strategies. Two patient partners in the PC-COS study group (AA and FS) and the Long Covid SOS and Long Covid Support patient groups had involvement throughout the study, including input into and support with funding application development and ongoing development of the protocol and study materials. Further details regarding study management processes are the same as, and comprehensively described in, the COS study.[4]

For the Delphi process, all individuals who had participated in the COS study[4] were invited to participate in this COMS study (see elsewhere[4] for details of participant recruitment). Participants were classified into the following three stakeholder groups: people with post-COVID-19 condition and their family members or caregivers; health-care professionals and researchers with post-COVID-19 condition; and health-care professionals and researchers without post-COVID-19 condition. Health-care professionals and researchers (including psychometricians and experts in patient-reported outcome measures) must have had experience of treating people with post-COVID-19 condition and research in the field of post-COVID-19 condition, respectively.

Additional individuals in the above stakeholder groups who had not participated in the COS study[4] were identified through their contact with the PC-COS study website (which was promoted via Twitter and Facebook) and were invited to take part by direct email from the core study team. We also invited lead investigators of identified study protocols (see below) to participate, and research groups and patients attending long COVID clinics were invited via personal contacts, group email lists, and announcements at meetings. To increase global representation, we engaged stakeholders (eg, from clinical, research, and patient groups) from countries with relatively lower representation in the COS project,[4] requesting study dissemination to their networks. Relevant criteria for participation and contact details were provided on the PC-COS study website; as before, individuals who responded to calls for participants were screened for eligibility before being registered for the survey. There was no restriction on the number of eligible participants in each stakeholder group. Consent for participation was obtained online before the start of the study.

At the end of the third round of the Delphi process, participants were asked to declare any conflicts of interest (eg, related to the development of a specific instrument) and to express their interest in participating in the online consensus meeting. Those with no relevant conflicts of interest were considered for inclusion, with the aim of achieving representation across stakeholder groups and geographical locations.

Literature review

A list of potential outcome measurement instruments used in published and ongoing studies of post-COVID-19 condition for each of the ten outcomes in the COS without instruments defined a priori was generated to inform the COMS consensus process. Outcome measures were identified primarily from the results of the literature review conducted during the preceding COS development exercise[4] (see appendix p 3 for full details). The literature review included data from a living systematic review on long COVID,[2] which was based on searches of MEDLINE, CINAHL (EBSCO), Global Health (Ovid), the WHO Global Research Database on COVID-19, and LitCovid for articles published in English from Jan 1, 2020, to March 17, 2021. Additionally, ClinicalTrials.gov and the International Clinical Trials Registry Platform were searched on May 19, 2021, to extract clinical trial protocol data. Between March 25, 2021, and May 25, 2021, we manually searched for additional review articles and clinical trials that had not been captured in the above searches, but that were considered important to include. All articles and protocols were evaluated by two of seven independent reviewers (NS from the core study team and JC, AC, AK, CP, AP, and NS from the PC-COS study group) who have experience with systematic reviews, patient-reported outcome measures, or both. Further details of the search strategy used for the literature review, including the living systematic review,[2] are presented elsewhere.[2,4]

An updated literature search was conducted to evaluate the frequency of use of the measurement instruments for the core outcomes and to identify any reports of new measurement instruments or new data on existing instruments published since our original searches. We reviewed all measurement instruments included in the most recent update of the living systematic review on long COVID,[2] which had a search date of Oct 19, 2022, and in a Cochrane systematic review of post-COVID-19 condition, which had a search date of May 11, 2022.[8] All articles in these reviews were evaluated by SRD, who has extensive experience with systematic reviews and patient-reported outcome measures.

Instruments identified from the literature searches were mapped to outcomes in the post-COVID-19 condition COS[4] and reviewed to remove duplicates and confirm mapping. Instruments that mapped to multiple COS outcomes were included in a category of multidomain instruments. Similarly, post-COVID-19 condition-specific instruments were included in their own category. Instruments that did not map to any COS outcomes were not considered.

For more on the **COMET Initiative** see https://www.comet-initiative.org/

For more on the **PC-COS project** see https://www.pc-cos.org/publications

For more on **ISARIC** see https://isaric.org/

For more on **Long Covid Support** see https://www.longcovid.org/

For more on **Long Covid SOS** see https://www.longcovidsos.org/

The preliminary list of instruments was independently reviewed by three clinical experts (DMN, DM, and TRN) with complementary specialist areas, who classified each instrument into one of three categories—include, maybe, or exclude—for the consensus process. Instruments were excluded for the following reasons: not being relevant to post-COVID-19 condition; a biological specimen would be required; measurement could not be done by phone or post or is otherwise not feasible within the scope of the COS (eg, cannot be undertaken in all settings internationally); or other (with reason specified). The assessments from each of the three reviewers were compared and disagreements were resolved with discussion to reach consensus on a final list of instruments to be included in the Delphi process.

Given the large number of instruments that mapped to some COS outcomes, only the five most frequently used instruments identified in the literature review for each outcome, along with any post-COVID-19 condition-specific instruments, were selected for presentation in the first round of the modified Delphi consensus process to help to optimise study feasibility and reduce respondent burden. For outcomes for which fewer than five instruments were identified, instruments that had been used only once were included; however, if more than five instruments were identified, only instruments that had been used more than once were included. For outcomes for which more than five instruments were identified, all of which had been used more than once, the three clinical experts (DMN, DM, and TRN) reviewed the instruments and reached agreement for inclusion on the basis of relevance to post-COVID-19 condition and administration feasibility. All instruments not meeting these frequency criteria, other than novel post-COVID-19-specific instruments, were excluded at this stage, but could be suggested by participants in the first round of the modified Delphi consensus process, along with suggestions for any other measurement instrument(s).

For outcomes for which fewer than five instruments were reported (as above), we identified instruments that had been recommended to measure outcomes in other COMS studies of relevant related clinical populations, based on data from the COMET Initiative international database of COS studies. PC-COS study group members also suggested relevant instruments based on searches of the Patient-Reported Outcomes Measurement Information System (PROMIS)[9] and Quality of Life in Neurological Disorders (Neuro-QoL)[10] databases. Newly developed post-COVID-19 condition-specific instruments were also identified by the study group and through discussions with other international experts and networks such as COMET, WHO, ISARIC, and national and international post-COVID-19 condition clinical and research networks.

From the outset, it was agreed by the PC-COS study group that measures for two outcomes that were included in a previously published COS for acute COVID-19[11] were relevant to post-COVID-19 condition, and should therefore be automatically included in the final post-COVID-19 COMS—ie, survival would be measured by time until death and recovery would be measured using the Recovery Scale for COVID-19.[12]

Assessment of instrument quality

Given that the measurement properties of many non-COVID-19-specific instruments had not been assessed in a post-COVID-19 population, evaluation of the measurement properties of these instruments was not undertaken. However, for post-COVID-19-specific instruments, we assessed measurement properties, focusing on content validity,[13] which involved identifying reports of instrument development and additional content validity studies. Two experts (CBT from the core author group and VF from the PC-COS study group) independently rated the post-COVID-19-specific instruments according to predefined criteria using a mini COSMIN assessment[13] (appendix p 4) to judge the relevance, comprehensiveness, and comprehensibility of the instruments, and the results of these assessments were presented to participants (see below).

For all instruments, feasibility-related data (eg, time requirements, costs, and languages of instruments or availability of translations) were obtained, allowing participants to consider whether they could reasonably be used in the intended settings.[13] Selection of feasibility data was guided by previous relevant COMS initiatives[14,15] to maximise selection of the most important aspects to present to participants for consideration.

Delphi process

The consensus process included a three-round online modified Delphi process[16] to rate the instruments for each outcome. In the first round, survey participants were asked whether the instruments should be used to measure the outcomes by rating each instrument anonymously using the Grading of Recommendation, Assessment, Development, and Evaluation (GRADE) nine-point Likert scale,[17] which is commonly divided into three categories: not important (1–3), important but not critical (4–6), and critically important (7–9). An option of "unable to score" was included in case participants did not feel able to rate specific instruments. A free-text option was also included in the first round for suggestions of additional relevant instruments not already included in the consensus process. Additional instruments suggested by more than 1% of participants in each stakeholder group were considered by the core study team for inclusion in the second round, with an instrument card (see below) developed for each new instrument. Once participants had rated all instruments, they could review their ratings and revise them before submission if they wished. Participants could complete the survey over more than one session using a "save-for-later" option. Instruments that reached consensus for

exclusion (as defined below) were not retained for rating in subsequent rounds; all other instruments were included in subsequent rounds.

In the second round of the Delphi process, responses of the three stakeholder groups for all instruments rated in the first round that did not meet exclusion criteria were summarised and displayed graphically alongside participants' own ratings. Participants were asked to re-rate the instruments using the same GRADE scale, with the option of changing or retaining their previous ratings. At the end of the second round, participants were able to review their ratings.

In the third round, responses for all instruments rated in the second round, but not excluded, were summarised and re-rated, as above, with the provision of additional data from stakeholder group summary scores given for new instruments suggested during the first round. The COSMIN guideline generally recommends the selection of a single instrument for each outcome in a COMS;[13] therefore, in the third round, participants were strongly encouraged to give a single instrument a rating of 7–9 (ie, critically important) for each outcome and to give all other instruments for that outcome ratings of 6 or lower.

Consensus for an instrument to be excluded from the Delphi process was prespecified in the protocol[6] as 50% or less of participants, in all three stakeholder groups, rating the instrument as 7–9. Before reviewing the results of the second round, we modified the exclusion criteria to reduce participant burden by defining consensus for exclusion as 50% or less of participants, in at least two stakeholder groups, rating the instrument as 7–9. Consensus for an instrument to be included in the COMS was prespecified in the protocol[6] as 80% or more of participants, in all three stakeholder groups, rating an instrument as 7–9, and less than 10% of participants, in each stakeholder group, rating an instrument as 1–3. A level of agreement of 70% or 80% is a generally accepted threshold for inclusion in a COS;[16] 80% was chosen for this study with the aim of identifying a single measurement instrument for each outcome.

Participants were provided with the following items for review when completing the surveys: (1) a list of COS outcomes with lay definitions of each outcome; (2) a list of instruments, mapped to each outcome, for rating; (3) instrument cards summarising key relevant information about each instrument, including plain language details written with input from patient research partners and a web-based link to the instrument itself when publicly available (see appendix p 5 for an example instrument card); and (4) definitions and evaluations of specific measurement properties (eg, reliability, validity, and responsiveness), and the results of the mini COSMIN assessments, on instrument cards for post-COVID-19 condition-specific instruments. The order in which instruments were presented to participants was randomised by COS outcome categories. Instruments mapping to multiple outcomes and the post-COVID-19 condition-specific measures were presented in separate categories. The surveys and all materials were presented in English. The Delphi survey was delivered using DelphiManager software.[18]

Direct communication during the Delphi process was by email, supplemented by public announcements on private long COVID patient-support group pages on social media (Twitter and Facebook), to encourage group members who were participating in the Delphi process to complete their ratings. Furthermore, a video interview about the importance of completing the Delphi process, conducted by a patient representative (M'OH) with the study lead (TRN), was made publicly available on the study website and circulated through patient-support channels on social media.

Consensus meeting

Before the online consensus meeting, participants received background information about the meeting, their ratings from the Delphi process, a summary of the Delphi results, and details about the instruments to be discussed at the meeting. People with post-COVID-19 condition and their family members or caregivers 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. It focused on the results of the third round of the Delphi process, with instruments rated most highly by all stakeholder groups prioritised for discussion. For each instrument discussed, participants were invited to provide arguments in favour of inclusion or exclusion. Following discussion, participants were asked to vote anonymously "yes" or "no", using Zoom polls, for whether each instrument should be included in the COMS. Consensus for inclusion of an instrument in the COMS was defined as 80% or more of participants in all stakeholder groups voting that it should be included.

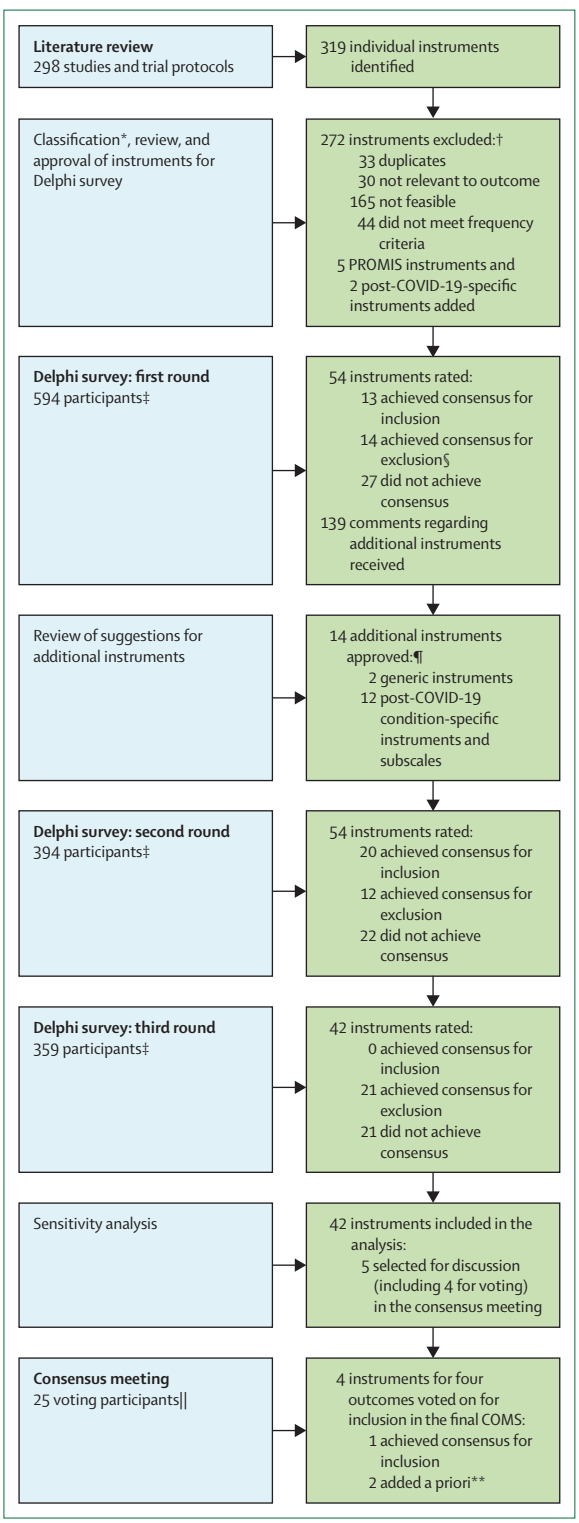
Data analysis

In the three rounds of the Delphi process, descriptive statistics were used 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. It was agreed a priori to include responses in the analysis if a participant assessed all instruments for at least one outcome. Graphs displaying the distribution of ratings for each instrument, stratified by stakeholder group, were produced using R (version 4.0.2) and shown to participants in the second and third Delphi rounds.

After the third Delphi round, two sensitivity analyses were conducted to identify the instruments that were most highly rated for each outcome by each of the three stakeholder groups to guide inclusion in the

consensus meeting. The first sensitivity analysis involved identifying which instrument(s) for each outcome had most frequently been given the overall highest rating, calculated separately for each of the three stakeholder groups. The second sensitivity analysis involved identifying which instrument(s) for each outcome had most frequently been rated as critically important (ie, score 7–9).

Selection bias between the Delphi process and the subsequent online consensus meeting was assessed by comparing distributions 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.



Results

Literature review

The literature review resulted in the identification of 298 studies or trial protocols on post-COVID-19 condition reported up to May 25, 2021, that were eligible for inclusion; 319 individual instruments were reported in these studies and trial protocols. After removal of duplicates, mapping to core outcomes, independent review of the instruments, and exclusion of instruments that did not meet the frequency criteria, 47 instruments were approved for consideration in the first round of the Delphi process. A further five relevant PROMIS instruments and two post-COVID-19 condition-specific instruments identified by the PC-COS study group were added, resulting in 54 outcome measurement instruments presented in the first round of the Delphi process (appendix pp 6–8) that mapped to 12 outcomes or categories: cardiovascular functioning, symptoms, and conditions (one instrument); fatigue or exhaustion (six instruments); pain (six instruments); nervous system functioning, symptoms, and conditions (three instruments); cognitive functioning, symptoms, and conditions (five instruments); mental health functioning, symptoms, and conditions (seven instruments); respiratory functioning, symptoms, and conditions (five instruments); post-exertion symptoms (two instruments); physical functioning, symptoms, and conditions (six instruments); work or occupational and study changes (four instruments); multidomain (five instruments); and post-COVID-19 condition specific (four instruments). The COMS development steps are summarised in figure 1.

Figure 1: **Overview of the COMS development process**
For the Delphi survey, instruments that reached consensus for exclusion were not retained for rating in the subsequent rounds. COMS=core outcome measurement set. COS=core outcome set. *Instruments were classified by three reviewers into the following categories: include, maybe, and exclude. †Instruments were mapped to the 12 outcomes in the COS.[4] ‡Participants were classified into three stakeholder groups: people with post-COVID-19 condition and family members or caregivers, health-care professionals and researchers with post-COVID-19 condition, and health-care professionals and researchers without post-COVID-19 condition. §Includes one item that was excluded after the consensus process. ¶Includes instruments suggested by Delphi participants and by the core study team. ||Participants were classified into two stakeholder groups: people with post-COVID-19 condition, and health-care professionals and researchers. **Additional instruments were included for two outcomes that were part of a previously published COS for acute COVID-19.[11]

Our updated literature searches for outcome measurement instruments published since our original search—reported in the living systematic review[2] (up to Oct 19, 2022) and the Cochrane systematic review[8] (up to May 11, 2022)—resulted in the identification of 120 new studies of post-COVID-19 condition. However, no new eligible measurement instruments were identified for the following core outcomes: cardiovascular functioning, symptoms, and conditions; pain; nervous system functioning, symptoms, and conditions; post-exertion symptoms; and survival. For the remaining core outcomes, although new instruments were identified, none met the frequency criteria and therefore would not have been considered for inclusion in the Delphi process. Furthermore, the new data on existing instruments did not alter the frequencies relative to those of the original search.

Delphi process

The first round of the online Delphi process took place from June 6 to June 29, 2022. 711 individuals registered to take part in the study and 594 participants (84%) from 58 countries took part in the first round; 447 participants fully completed the survey and 147 partially completed the survey (ie, rated all instruments for at least one outcome). Of the 594 participants who were invited to participate in the second round, 394 (66%) took part; 362 participants fully completed the survey and 32 partially completed the survey. 359 of the 594 invited participants (60%) took part in the third round, 324 (82%) of whom had completed the second round; 341 participants fully completed the survey and 18 partially completed the survey. See appendix (p 9) for response rates, stratified by stakeholder groups, across all rounds. Demographic characteristics, by Delphi round, are presented in table 1. Further details of the Delphi participants are presented in the appendix (pp 10–12).

At the end of the first round of the Delphi process, 13 of the 54 instruments met prespecified criteria for exclusion. One additional instrument was excluded (the Fukuda Criteria for chronic fatigue syndrome) because it was identified as a diagnostic, rather than a measurement, instrument (appendix p 13). Of the remaining 40 instruments, 13 met criteria for inclusion and 27 did not meet inclusion or exclusion criteria. Thus, these 40 instruments were taken forward to the second round.

A total of 139 free-text responses regarding suggestions for additional instruments were received and reviewed, two of which met criteria for inclusion in the second round of the Delphi process: the WHO Disability Assessment Schedule 2.0 12-item version and the Nijmegen Questionnaire for respiratory symptoms, each suggested by four participants. Two additional post-COVID-19 condition-specific instruments identified by the core study team were included—the Long COVID Symptom Tool and the Long COVID Impact Tool[19]—as well as eight validated post-COVID-19 condition-specific subscales of the Symptom Burden Questionnaire for Long Covid (SBQ-LC): breathing; circulation; fatigue; impact on daily life; memory, thinking and communication; mental health; movement; and pain. Two of these subscales were each found to be applicable to two outcomes: SBQ-LC–fatigue (fatigue or exhaustion; post-exertion symptoms) and SBQ-LC–impact on daily life (physical functioning, symptoms, and conditions; work or occupational and study changes). Thus, these subscales were added to the survey for both outcomes, resulting in the addition of 14 instruments for rating in the second round.

	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)
Male	174 (29)	126 (32)	120 (33)
Non-binary	5 (1)	5 (1)	3 (1)
Prefer not to answer	1 (<1)	1 (<1)	1 (<1)
Age group, n (%)			
18–29 years	27 (5)	19 (5)	16 (5)
30–39 years	147 (25)	104 (26)	91 (25)
40–49 years	203 (34)	127 (32)	116 (32)
50–59 years	150 (25)	92 (23)	84 (23)
60–69 years	58 (10)	48 (12)	47 (13)
≥70 years	9 (2)	4 (1)	5 (1)
Geographical area, 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)
Ethnic origin, 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)

Not all percentages add up to 100% owing to rounding. *One participant in each Delphi round did not specify their gender. †Countries are listed in the appendix (pp 10–12).

Table 1: **Characteristics of participants in the Delphi consensus process**

The second round of the Delphi process took place from July 13 to Aug 8, 2022. Following rating by 394 participants, 12 of the 54 instruments included in the second round (appendix pp 14–16) met prespecified criteria for exclusion (appendix p 17). 20 met criteria for inclusion and 22 did not meet inclusion or exclusion criteria. Thus, 42 instruments were taken forward to the third round (appendix pp 18–19).

The third round of the Delphi process took place from Aug 12 to Sept 12, 2022. Following rating by 359 participants, 21 of the 42 instruments met prespecified criteria for exclusion. None met criteria for inclusion and 21 did not meet inclusion or exclusion criteria. See appendix pp 20–34 for results from the three Delphi rounds as percentages of participants, by stakeholder group, rating each instrument by GRADE categories of importance.

Sensitivity analyses (appendix pp 35–39) showed that for two outcomes, a single instrument had been most highly rated by all three stakeholder groups: the DePaul Symptom Questionnaire for post-exertion symptoms, and the SBQ-LC–impact on daily life subscale for physical functioning, symptoms, and conditions. Thus, both instruments were taken forward for discussion and voting in the online consensus meeting.

For the pain outcome, the Brief Pain Inventory was the only instrument included in the third round, and was therefore taken forward for discussion and voting in the online consensus meeting. For the respiratory functioning, symptoms, and conditions outcome, the modified Medical Research Council (mMRC) Dyspnoea Scale was most highly rated by two stakeholder groups, and was therefore included for discussion and voting in the online consensus meeting. The St George's Respiratory Questionnaire (SGRQ) was most highly rated by the group of patients and their family members or caregivers, and therefore it was included for discussion in the consensus meeting as a potential measure for consideration alongside the mMRC Dyspnoea Scale; however, there was no voting on whether the SGRQ should be included in the COMS because it was not highly rated in the Delphi process by either of the health-care professional and researcher stakeholder groups.

For the remaining eight outcomes or categories, no single instrument was most highly rated across all stakeholder groups and instruments for these outcomes were therefore not included for discussion and voting in the online consensus meeting. Consensus was also not reached for any multidomain outcome measurement instruments, nor for any post-COVID-19 condition-specific measurement instruments. Thus, five instruments for four outcomes were taken forward for discussion, including four instruments for voting, in the consensus meeting.

Consensus meeting

The online consensus meeting was held on Zoom on Sept 29, 2022. 37 people attended this 3-h meeting, including seven members of the core author group, four observers, one independent facilitator, and 25 voting participants who had completed the online Delphi surveys: ten people with post-COVID-19 condition; five health-care professionals and researchers with post-COVID-19 condition; and ten health-care professionals and researchers without post-COVID-19 condition. Owing to the small number of attendees in the group of health-care professionals and researchers with post-COVID-19 condition, participants in this group were asked to choose one of the other two groups to join for voting at the meeting, as done for the COS study.[4] Voting groups therefore included 12 people with post-COVID-19 condition and 13 health-care professionals and researchers. Three participants (two from the group of people with post-COVID-19 condition and one from the group of health-care professionals and researchers) were unable to continuously attend the entire meeting but their votes were still counted if they returned to the meeting. See appendix (p 40) for details of participants who attended the consensus meeting. We found no evidence of selection bias: the average Delphi round three scores were similar between participants who attended the consensus meeting (6·01) and those who did not attend the meeting (6·31).

At the start of the online meeting, attendees were reminded about the two measures that were confirmed for inclusion in the COMS: time until death, to measure survival, and the Recovery Scale for COVID-19, to measure recovery, for consistency with the pre-existing COS for acute COVID-19.[11] The remaining outcomes 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 subscale). After discussion and voting, the mMRC Dyspnoea Scale was the only instrument that met the predefined consensus definition for inclusion, with nine (82%) people with

Panel 1: **Core outcome measurement set for adults with post-COVID-19 condition**

Survival*
Time until death

Recovery*
Recovery Scale for COVID-19[12]

Respiratory functioning, symptoms, and conditions
Modified Medical Research Council Dyspnoea Scale[20]

Links to instrument cards, containing instrument-specific information, including licensing information and details of how to access the instruments, are available on the PC-COS study website.[21] *Measures for two outcomes that were included in a previously published core outcome set for COVID-19[11] were automatically included in the final COMS owing to their relevance to post-COVID-19 condition.

post-COVID-19 condition and 11 (92%) health-care professionals and researchers voting for inclusion; thus, this scale was added to the COMS, making a total of three measurement instruments (panel 1).[21]

Consensus was not reached on inclusion of instruments in the COMS for the remaining nine core outcomes in the COS. Table 2 indicates the instruments for the COS outcomes with the greatest level of support based on the consensus process.[21] At least one of these instruments can be considered for measurement of each of the core outcomes (ie, where more than one instrument is indicated, selection of a single instrument might be appropriate to avoid redundancy and reduce respondent burden). Table 3 indicates the multidomain and post-COVID-19 condition-specific instruments[21] with the greatest level of support based on the consensus process (see appendix p 41 for full results). See appendix (p 42) for a table of available languages for the 19 recommended and suggested outcome measurement instruments. Figure 2 provides an overview of the process and results for the COS and COMS development exercises.

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

	Delphi round 3: % of participants giving a GRADE rating of 7–9			Consensus meeting: % of participants voting to include instrument in the COMS	
	People with post-COVID-19 condition and family members or caregivers	Health-care professionals and researchers with post-COVID-19 condition	Health-care professionals and researchers without post-COVID-19 condition	People with post-COVID-19 condition	Health-care professionals and researchers
Cardiovascular functioning, symptoms, and conditions					
Symptom Burden Questionnaire for Long COVID–circulation subscale	63	62	46	..	..
New York Heart Association Functional Class scale	60	60	58	..	..
Fatigue or exhaustion					
Fatigue Assessment Scale	59	59	54	..	..
Fatigue Severity Scale	61	50	56	..	..
Functional Assessment of Chronic Illness Therapy–fatigue subscale	71	47	53	..	..
Pain					
Brief Pain Inventory	60	56	45	55	67
Nervous system functioning, symptoms, and conditions					
Central Sensitisation Inventory	78	44	34	..	..
Cognitive functioning, symptoms, and conditions					
Cognitive Failures Questionnaire	66	53	22	..	..
Montreal Cognitive Assessment–Blind version	60	45	59	..	..
Mental health functioning, symptoms, and conditions					
GAD-7 questionnaire	48	23	64	..	..
PTSD Checklist for DSM-5	59	56	51	..	..
Post-exertion symptoms					
DePaul Symptom Questionnaire	81	68	53	33	8
Physical functioning, symptoms, and conditions					
Symptom Burden Questionnaire for Long COVID–impact on daily life subscale	70	59	49	75	54
Work or occupational and study changes					
Work Ability Index	61	29	35	..	..
Work Productivity and Activity Impairment questionnaire	53	50	58	..	..
WHO post-COVID-19 Case Report Form occupational status item	59	53	37	..	..

Links to instrument cards, containing instrument-specific information, including licensing information and details of how to access the instruments, are available on the PC-COS study website.[21] Owing to the relatively small numbers of health-care professionals and researchers with post-COVID-19 condition who took part in the consensus meeting, participants from this group were classified into two stakeholder groups: people with post-COVID-19 condition, and health-care professionals and researchers. DSM-5=Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. GAD-7=Generalized Anxiety Disorder 7. GRADE=Grading of Recommendation, Assessment, Development, and Evaluation. PTSD=post-traumatic stress disorder.

Table 2: **Core outcome measurement instruments for consideration based on key results from the Delphi survey and consensus meeting**

	People with post-COVID-19 condition and family members or caregivers (%)	Health-care professionals and researchers with post-COVID-19 condition (%)	Health-care professionals and researchers without post-COVID-19 condition (%)
Multidomain instruments			
EuroQol 5-Dimension 5-Level survey	45	45	67
Short Form (36) Health Survey	64	49	46
WHO Disability Assessment Schedule 2.0 12-item version	58	59	40
Post-COVID-19 condition-specific instruments			
COVID-19 Yorkshire Rehabilitation Screening Scale	64	65	54
Symptom Burden Questionnaire for Long COVID	66	73	54

Links to instrument cards, containing instrument-specific information, including licensing information and details of how to access the instruments, are available on the PC-COS study website.[21] COMS=core outcome measurement set. GRADE=Grading of Recommendation, Assessment, Development, and Evaluation.

Table 3: **Multidomain and post-COVID-19 condition-specific measurement instruments for consideration based on GRADE ratings of 7–9 in round 3 of the Delphi survey**

instruments for post-COVID-19 condition research and clinical practice. Feedback on how the consensus meeting was conducted from participants who attended the meeting was strongly positive (summarised in appendix p 43).

Discussion

With the second phase of the PC-COS project, a large international consensus study, we aimed to address the pressing need for a COMS for post-COVID-19 condition in adults. The final COMS comprised instruments to measure survival (time until death), recovery (the Recovery Scale for COVID-19), and respiratory outcomes (the mMRC Dyspnoea Scale). These instruments are recommended for use and reporting in all clinical research and practice settings worldwide involving adults (≥18 years of age) with post-COVID-19 condition. The consensus process also provided important data on the instruments with the greatest level of support, and through this process, the number of potential instruments for measuring the 12 core outcomes was reduced from 319 to 19, although no single measurement instrument reached consensus for nine of the 12 core outcomes. With the goal of reducing the substantial heterogeneity in outcome measurements for post-COVID-19 condition, at least one of these instruments, with high levels of support but without consensus agreement, can be considered for use with each of the nine remaining core outcomes (table 2) and should be the focus of future research, in addition to further evaluation of multidomain and post-COVID-19 condition-specific instruments.

Participants from the first stage of the PC-COS project, focused on the set of outcomes that should be measured and reported,[4] were invited to take part in this second stage, to identify the instruments that are most appropriate to measure those outcomes. The number of participants in the online Delphi consensus process was smaller for the second stage compared with the first stage (table 1),[4] but still included good representation from all stakeholder groups, with a similar distribution across regions of the world. There was a slightly higher proportion of health-care professionals and researchers in this second stage compared with the first stage (table 1);[4] nevertheless, the proportion of participants with post-COVID-19 condition and their family members or caregivers was still high at 50% in the first round and 41% in the third round—a strength of this study, particularly given that only 28% of previous COMS studies have included the views of people with lived experience.[22]

Considering the number of instruments to be rated over three Delphi rounds and the short response time permitted (with the goal of expediting availability of the COMS results), the participation and retention rates were high: 60% of participants from round one took part in round three. Participation and retention during the Delphi process were facilitated by a communication strategy designed with patient partners, involving reminders through email and social media (the primary mode of communication among patients with long COVID), as well as a video interview informing patients about the importance of completing the Delphi process that was disseminated via social media.

Limitations from the COS development study,[4] the first stage of the PC-COS project, apply to this second stage. Briefly, the Delphi process and consensus meeting were delivered online, limiting participation to people with sufficient digital skills and access, although potentially increasing numbers beyond those who could have participated in person. Participation across geographical and demographic groups is unlikely to have been fully representative of the post-COVID-19 condition experience globally, despite substantial efforts to achieve broad representation. The online consensus meeting had insufficient participants with post-COVID-19 condition who were also health-care professionals or researchers to constitute their own voting group, thus limiting the information provided by this group with dual, and potentially unique, perspectives.

There were additional limitations that were unique to this COMS study. For example, the survey and consensus meeting were conducted only in English. This decision was based on the extra complexity and substantial time required for translation of the large amount of study materials (eg, all instrument cards) into multiple languages. This limitation probably also affected the geographical and demographic diversity of participants; nevertheless, a reasonable global representation was achieved (table 1).

This second stage of the project, focused on measurement instruments, required substantially greater time and detailed evaluations from participants than did

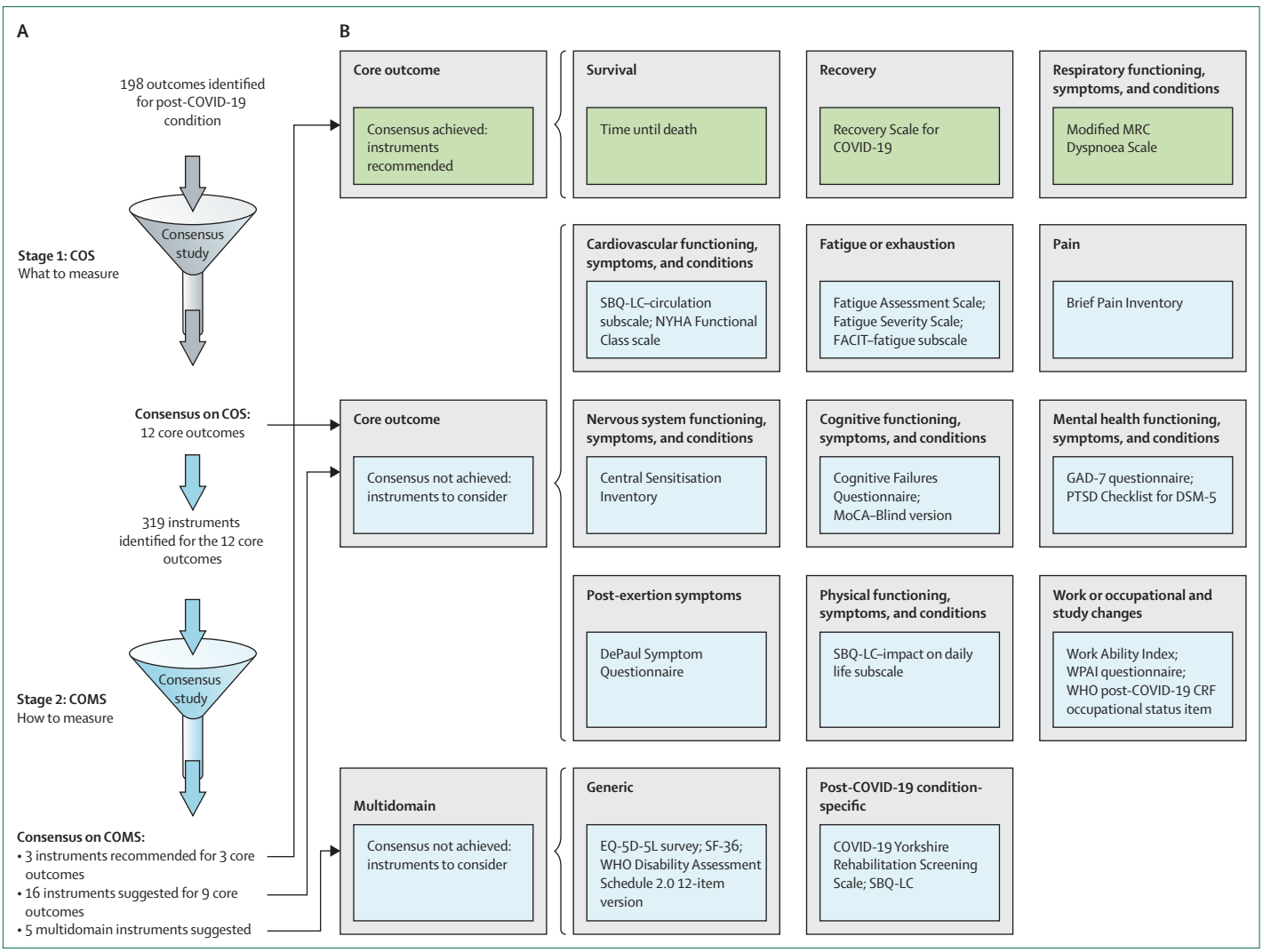


Figure 2: **Summary of the development process and results for the COS and COMS for post-COVID-19 condition in adults**
(A) Development process for the COS (what to measure) and COMS (how to measure). (B) Results for the COS and COMS. COMS=core outcome measurement set. COS=core outcome set. CRF=Case Report Form. DSM-5=Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. EQ-5D-5L=EuroQol 5-Dimension 5-Level. FACIT=Functional Assessment of Chronic Illness Therapy. GAD-7=Generalized Anxiety Disorder 7. MoCA=Montreal Cognitive Assessment. MRC=Medical Research Council. NYHA=New York Heart Association. PTSD=post-traumatic stress disorder. SBQ-LC=Symptom Burden Questionnaire for Long COVID. SF-36=Short Form (36) Health Survey. WPAI=Work Productivity and Activity Impairment.

the COS development stage. This issue is particularly important for people with long COVID, in whom fatigue, low energy levels, and cognitive difficulties are common and often significantly disabling, which might explain the lower completion rates compared with those in the first stage. To reduce bias, the order of presentation of the instruments was randomised on the basis of outcomes. Moreover, major efforts were made, in collaboration with patient partners, to explain that rating of all instruments for at least one of the outcomes was the minimum participation required and would still provide valuable data. Across the three rounds, 75%, 92%, and 95% of all participants, respectively, fully completed the Delphi surveys, which reflects remarkable contributions of the participants.

A consensus-based approach to define a set of measurement instruments has been used in the majority of COMS projects.[22] Participants in the Delphi consensus process had access to the instrument cards (appendix p 5), which included a link to the measurement instrument, to provide an opportunity to assess face validity. For the post-COVID-19-specific instruments, an additional full assessment against each of the ten predefined COSMIN criteria[13] was undertaken, results of which were included on the instrument cards. As this was intended to be a global COMS, aspects of feasibility were also included on the instrument cards. In this context, it was not possible to control (or be aware of) the aspects of the instruments that were considered to be most important by participants during their ratings.

Because of the rapid expansion in post-COVID-19 condition research and publications during the course of this study, we conducted further literature searches after the study was completed to identify any reports of relevant new outcome measurement instruments or new data on existing instruments that had been published since our original search in October 2021. Although we found 120 new studies through these updated searches, reassuringly, we identified no new outcome measurement instruments that met eligibility criteria for inclusion in the Delphi consensus process.

***Panel 2*: Recommendations for future research**

Overall recommendations

- Comparison of the performance of measurement instruments developed for other conditions, post-COVID-19 condition-specific instruments, generic multidomain instruments, and generic measures developed using item response theory techniques
- Reassessment and updating of the core outcome measurement set as new data emerge on post-COVID-19 condition and related measurement instruments

Outcome-specific recommendations

Cardiovascular functioning, symptoms, and conditions

- Translation and cultural adaptation of the New York Heart Association Functional Class scale

Fatigue or exhaustion

- Comparison of the performance of the three preferred instruments: the Fatigue Assessment Scale, the Fatigue Severity Scale, and the Functional Assessment of Chronic Illness Therapy fatigue subscale

Nervous system functioning, symptoms, and conditions

- Phenotyping studies of neurological symptoms in post-COVID-19 condition

Cognitive functioning, symptoms, and conditions

- Phenotyping studies of cognitive symptoms in post-COVID-19 condition

Mental health functioning, symptoms, and conditions

- Investigation of whether a single measure or multiple measures are optimal

Post-exertion symptoms

- Refinement of the concept and understanding of post-exertion symptoms in post-COVID-19 condition; comparison of the performance of measurement instruments developed for chronic fatigue syndrome and post-COVID-19 condition-specific instruments

Work or occupational and study changes

- Comparison of the performance of the three preferred instruments: the Work Ability Index questionnaire, the Work Productivity and Activity Impairment questionnaire, and the WHO post-COVID-19 Case Report Form occupational status item

Recovery

- Translation and cultural adaptation of the Recovery Scale for COVID-19

Regarding the absence of consensus on instruments for nine of the outcomes, there are several potential reasons. For example, post-COVID-19 condition is a highly heterogeneous disorder that potentially influences instrument preference. Past experience with instruments might have been highly variable, which could have 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. Moreover, post-COVID-19 condition is poorly understood mechanistically, creating uncertainty regarding the use of specific measurement instruments that were developed for other conditions.

It is important to note that consensus on measurement instruments for all core outcomes is often not achieved, even in other complex multidomain conditions with decades of research, such as post-intensive care syndrome in survivors of acute respiratory failure.[23] In a review of 118 studies that have aimed to develop a COMS for a COS, a single instrument for each core outcome was recommended in only 11 (9%) studies.[22] Of the 80 studies for which a consensus procedure was used, a rigorous Delphi process was undertaken in eight and the remainder involved only a meeting or were not clearly reported. Optimal methods for determining consensus on outcome measurement instruments is the subject of ongoing research.

For the three outcomes for which measurement instruments have been included in the COMS (panel 1), the recommended instruments should be used in all post-COVID-19 research and clinical practice settings. Additional measures of these outcomes can also be used, if appropriate, because a COMS is the minimum set of instruments recommended for all clinical studies and services 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, and therefore additional instruments might be warranted.

For the other outcomes, for which there was no consensus on measurement instruments (table 2), the instruments with the greatest level of support should be considered for use in an effort to reduce the substantial heterogeneity in the measurement of post-COVID-19 condition in adults. For the four outcomes for which only one instrument is suggested from this study (pain; nervous system functioning, symptoms, and conditions; post-exertion symptoms; and physical functioning, symptoms, and conditions), suitability for the study or setting—for example, the availability of the instrument in the appropriate language(s) and its feasibility in terms of patient burden—should be considered carefully. Post-exertion symptoms are under-researched and consequently poorly understood despite being commonly identified as having particular importance for patients;

hence, more detailed research-based recommendations are needed for these symptoms. If two or three instruments are available, the above factors should also be considered in instrument selection. Furthermore, for some outcomes (eg, cognitive functioning, symptoms, and conditions; and mental health functioning, symptoms, and conditions), different measures might cover different aspects of the outcome and more than one measure might be appropriate.

Further research regarding clinical outcomes and measurement instruments for the core outcomes without consensus for measurement instruments will be an important next step. There was a high level of agreement among participants at the consensus meeting that a research priority should be the assessment of the relative merits of existing instruments that were not designed for or validated in post-COVID-19 condition (ie, generic or legacy instruments) versus those that have been developed specifically for post-COVID-19 condition. For example, how does an instrument designed for post-exertion symptoms in chronic fatigue syndrome (also known as myalgic encephalomyelitis) perform compared with instruments that have been developed for post-COVID-19 condition? Future studies of measurement properties and future consensus projects should focus on these questions to advance research and clinical practice.

It was notable that no instruments from the PROMIS or Neuro-QoL modular sets of outcome measures were preferred, despite the advantages of having been developed using item response theory techniques and having 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 a lack of understanding of the potential benefits of their measurement properties. In the context of rapidly evolving knowledge regarding post-COVID-19 condition, and based on the results of this international consensus process, recommendations for future research are provided in panel 2.

With the vast numbers of people affected by post-COVID-19 condition worldwide, internationally agreed outcome measurements are essential to advance research and clinical care. The COMS and suggested measurement instruments for post-COVID-19 condition in adults presented here complement the existing COS,[4] and are relevant to people with lived experience of long COVID. This COMS is recommended for immediate use by researchers and clinical services worldwide to improve the quality, consistency, and comparability of outcome measurement. Its uptake should thus optimise clinical care and harmonisation of research data, which should, in turn, accelerate understanding of post-COVID-19 condition and how to treat it.

Contributors

DM and TRN conceived the idea for the study. SLG, NS, PRW, DMN, DM, and TRN designed the study protocol and were responsible for the day-to-day running of the project as members of the core study team. PRW led the methods team. NS undertook the literature review, identified outcome measures, and categorised them for inclusion in the online Delphi survey. NS also coordinated the data-revision process. SLG, NS, and NLH developed the online Delphi surveys and contributed to the day-to-day management of the project. As members of the methods team, SLG, NS, CBT, PRW, DMN, DM, and TRN participated in discussions about project methods, led by PRW, throughout the duration of the project, with additional methodological input from SRD and NLH. SLG undertook the data analysis. SLG and NLH organised the consensus meeting. MO'H provided and coordinated the perspectives of people with lived experience throughout the design and implementation of the study. SRD conducted and integrated an updated search for outcome measurement instruments. SLG and TRN drafted the manuscript; all authors reviewed and approved the final manuscript. SLG, NS, and TRN have accessed and verified the data. All authors have full access to all the data reported in the study and accept responsibility to submit for publication.

Declaration of interests

SLG is the project coordinator of the Core Outcome Measures in Effectiveness Trials (COMET) Initiative. NLH has received payment and honoraria for lectures and presentations from Lancaster University. PRW is chair of the COMET Management Group. DM is a co-chair of the International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC) Global Paediatric Long COVID working group and a member of the ISARIC working group on long-term follow-up in adults. All other authors declare no competing interests.

Acknowledgments

This study was funded by the UK's National Institute for Health and Care Research (NIHR; grant COV-LT2-0072). The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. TRN and PRW have received funding from the NIHR. DM has received grants from the British Embassy in Moscow, the NIHR, and the Russian Foundation for Basic Research. NLH has received grants from Lyme Disease Action and the Foundation for Sarcoidosis Research. We thank all participants of the consensus development process, and particularly the people with lived experience of post-COVID-19 condition and their carers; Wouter De Groote, Janet Diaz, Krutika Kuppalli, John Marshall, Srinivas Murthy, Jakobus Preller, and Nicoline Schiess of WHO and the WHO COVID-19 Clinical Research working group for their technical input and support in coordination and dissemination of the survey; Mike Clarke for his work as an independent facilitator of the online interactive consensus meeting; and Christina Barratt, Sophie Evans, Lesley MacNiven, Natalie Rogers, and Kirsty Stanley, members of Long Covid Support, for providing critical input into the design of the instrument cards. The views expressed in this publication are those of the authors and do not necessarily represent the views of the International Severe Acute Respiratory and Emerging Infection Consortium, the NIHR, or WHO.

References

1 O'Mahoney LL, Routen A, Gillies C, et al. The prevalence and long-term health effects of long Covid among hospitalised and non-hospitalised populations: a systematic review and meta-analysis. *EClinicalMedicine* 2022; **55:** 101762.

2 Michelen M, Manoharan L, Elkheir N, et al. Characterising long COVID: a living systematic review. *BMJ Glob Health* 2021; **6:** e005427.

3 Munblit D, Nicholson TR, Needham DM, et al. Studying the post-COVID-19 condition: research challenges, strategies, and importance of core outcome set development. *BMC Med* 2022; **20:** 50.

4 Munblit D, Nicholson T, Akrami A, 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; **10:** 715–24.

5 Prinsen CAC, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "core outcome set" – a practical guideline. *Trials* 2016; **17:** 449.

6 PC-COS Core study team. Post COVID condition (long COVID) core outcome set study: PC-COS protocol. King's College London. March 23, 2022. https://drive.google.com/file/d/1Scj3CKLKLIO9XgZjBPYkZNXngbUaolJG/view (accessed Oct 6, 2023).

7 Munblit D, Nicholson T, Williamson P, et al. Core outcome measures for post-Covid condition/long Covid. https://www.comet-initiative.org/Studies/Details/1847 (accessed Oct 6, 2023).

8 Saif-Ur-Rahman KM, Kothari K, Sadlier C, et al. Effect of pharmacological interventions for the treatment of people with post-COVID-19 condition: a rapid review. *Cochrane Evid Synth Methods*. 2023; **1:** e12001.

9 Cella D, Yount S, Rothrock N, 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; **45** (suppl 1)**:** S3–11.

10 Cella D, Lai JS, Nowinski CJ, et al. Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. *Neurology* 2012; **78:** 1860–67.

11 Tong A, Elliott JH, Azevedo LC, et al. Core outcomes set for trials in people with coronavirus disease 2019. *Crit Care Med* 2020; **48:** 1622–35.

12 Tong A, Baumgart A, Evangelidis N, et al. Core outcome measures for trials in people with coronavirus disease 2019: respiratory failure, multiorgan failure, shortness of breath, and recovery. *Crit Care Med* 2021; **49:** 503–16.

13 Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018; **27:** 1159–70.

14 Needham DM. COS resources – improving long-term outcomes research for acute respiratory failure. 2023. https://www.improvelto.com/cos-resources/ (accessed Oct 6, 2023).

15 Beaton D, Maxwell L, Grosskleg S, Shea B, Tugwell P (eds). The OMERACT Handbook Version 2.1. OMERACT, April 2021. https://omeract.org/handbook/ (accessed Oct 6, 2023).

16 Williamson PR, Altman DG, Bagley H, et al. The COMET handbook: version 1.0. *Trials* 2017; **18** (suppl 3)**:** 280 (section 2.7.6.1).

17 Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol* 2011; **64:** 395–400.

18 COMET. DelphiManager. https://www.comet-initiative.org/delphimanager/ (accessed Oct 6, 2023).

19 Tran VT, Riveros C, Clepier B, et al. Development and validation of the long coronavirus disease (COVID) symptom and impact tools: a set of patient-reported instruments constructed from patients' lived experience. *Clin Infect Dis* 2022; **74:** 278–87.

20 Primary Care Respiratory Society. MRC Dyspnoea Scale. https://www.pcrs-uk.org/mrc-dyspnoea-scale (accessed Oct 6, 2023).

21 PC-COS Project. Outcome measurement instruments. 2022. https://www.pc-cos.org/instruments (accessed Oct 6, 2023).

22 Gorst SL, Prinsen CAC, Salcher-Konrad M, Matvienko-Sikar K, Williamson PR, Terwee CB. Methods used in the selection of instruments for outcomes included in core outcome sets have improved since the publication of the COSMIN/COMET guideline. *J Clin Epidemiol* 2020; **125:** 64–75.

23 Needham DM, Sepulveda KA, Dinglas VD, 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; **196:** 1122–30.