Appendix 3

Table of Contents

1. Outcome measure instrument selection methods

The outcome measurement instruments were selected from those used in published and ongoing studies and research protocols for post-COVID-19 condition in children for each outcome domain. The literature set was collected and evaluated in a systematic review conducted by a comprehensive search of Medline, Embase, the WHO COVID-19 Research Database (from inception until December 29, 2021). Clinical trial protocols were identified from two clinical trial registries (ICTRP database and ClinicalTrials.gov). Additional search was performed on June 1, 2023 to screen for recent evidence. All articles and protocols were evaluated independently by two researchers (NS, AC, AM, ND, AA, LX, PB, PR, KA). After data extraction outcome measurement instruments were categorised by the core group into 3 types: scales/questionnaires, laboratory tests and clinical assessment tools. Instruments requiring trained personnel, additional software, clinical facilities, or not pertaining to "core outcomes" were excluded by core group pre-Delphi.

A list of remaining instruments was anonymously reviewed by a group of independent international experts - 11 healthcare professionals and researchers. They provided feedback on each instrument and suggested potential additions, which were assessed for feasibility and applicability by the core group. Approved new instruments were presented in the second round for further review. Experts were reminded to evaluate each instrument's feasibility and suitability, specifically for diverse settings and the paediatric population. In the second round, each expert received an anonymised feedback-incorporated list of instruments. After reviewing the comments from the first round, they had the liberty to modify their initial selection or retain it. Each expert indicated their preference for each instrument's inclusion in the consensus workshop. Instruments that garnered "include" or "maybe" responses from more than half of the experts were forwarded to the consensus workshop.

2. Example instrument card*

THE POST-COVID CORE OUTCOME SET (PC-COS)
FOR CHILDREN AND YOUNG PEOPLE
PC-COS CHILDREN

INSTRUMENT CARD

Outcome 1: Cardiovascular functioning, symptoms, and conditions

Description: New onset or worsening of problems affecting the heart (e.g. irregular heartbeat, palpitations, pounding or racing heartbeat, resting heartbeat changes, pericarditis/myocarditis
(heart inflammation)); problems with the blood vessels (i.e., veins or arteries), changes in blood pressure.

GENERAL INSTRUCTIONS

As a first step of this project **seven outcomes** were selected as the most critical for children and young people with Long Covid, forming a Core Outcome Set.

Now we need to decide on the most appropriate instruments to be used for the assessment of each of these outcomes.

Please thoroughly review the list of prioritised instruments provided. For each instrument, you will find a summary and expert feedback regarding its appropriateness for assessing Long Covid in children and young people.

Prioritise: As you review each instrument, consider which one you believe is the most suitable for assessing each outcome in children and young people with Long Covid. This selection should be made considering the instrument's feasibility (i.e., can be used in all settings) and suitability for the paediatric population.

Consideration of Expert Feedback: Read through the feedback provided by a group of eleven international experts. However, please note that it is entirely up to you to decide which of the instruments you prioritise over the others, and we will have a chance to discuss this at the meeting.

Keep the balance: Strive to strike a balance between an instrument's reliability and feasibility for research and clinical practice. The most effective tools will both provide reliable outcomes and be practical to use in a variety of settings.

Voting at the workshop: During the workshop, you will have the opportunity to discuss and vote for the most appropriate instrument for each outcome ranking them. The goal of this process is to reach a consensus on the best tools for assessing long Covid outcomes in children and young people.

Contents

- Instrument sample: Malmo POTS score (MAPS)

Full instrument samples were provided for every instrument for consensus workshop participants' review

Instruments Summary Information

Outcome 1: Cardiovascular functioning, symptoms, and conditions

Description: New onset or worsening of problems affecting the heart (e.g. irregular heartbeat, palpitations, pounding or racing heartbeat, resting heartbeat changes, pericarditis/myocarditis (heart inflammation)); problems with the blood vessels (i.e., veins or arteries), changes in blood pressure.

Instrument	Link	Time to complete	N of items	Age group	Validation in children	Languages	Cost
PedsQL™ Cardiac Module	https://drive.google.com/file/d/1hjmQtxVmC42mg4d11668W_ROjTPUk-n_/view	3-5 minutes	Toddlers (age 2-4): 23 items Young Children (ages 5-7): 25 items Children, Teens, Young Adults and Adults: 27 items	Self reported and Parent-reported: Toddlers (2-4 years) Young Child (5-7 years Child (8-12 years) Adolescent (13-18 years) Young Adult (18-25 years) Self-reported only: Adults (>26 years)	Yes	Available in multiple languages (100+)	Free and commercial licence available
Symptom Burden Questionnaire for Long COVID (Circulation scale)	https://drive.google.com/file/d/1pKHaHjqDW9NqyfIIIJEE_khAkJbEfL5Z/view	1-2 minutes (Circulation scale)	4 items (Circulation scale)	Adults 18+	No	US English Chinese, Arabic, Japanese	Free and commercial licence available
Malmo POTS score (MAPS)	https://drive.google.com/file/d/1W02IQPmeFdscWUEWyDw8hYASwU_Es44R/view	5 minutes	12 items	Adults 18+	No	English, Swedish	Free

The information provided in the table is accurate to the best of our knowledge

Experts review *(Experts were asked to select instruments that should be discussed at the meeting from a long list of instruments. Only instruments that will be discussed at the meeting are presented)*

Measurement instruments	Decision after two rounds of revision										
	1	2	3	4	5	6	7	8	9	10	11
PedsQL™ Cardiac Module	Include	Include	Maybe	Include	Include	Include	Include	Include	Include	Include	Include
Symptom Burden Questionnaire for Long COVID (Circulation scale)	*Unvoted*	Maybe	Include	Maybe	Include	Maybe	Maybe	Include	Include	Include	*Unvoted*
Postural orthostatic tachycardia syndrome (POTS)	*Unvoted*	Exclude	Maybe	Exclude	Maybe	Include	Maybe	Exclude	Maybe	Include	Include

Summary of additional comments from the experts

PedsQL™ Cardiac Module	Some experts express difficulties in understanding the module and indicate potential issues with its accessibility. Yet, others appreciate the PedsQL, a well-validated and widely used questionnaire set, often favoured in most studies due to its generic quality of life assessment, which may be more appropriate than other cardiac-specific measures. Experts also acknowledge the scale's beneficial features, like ability to use in a paper format, existence of age-specific questions, and the inclusion of cognitive scores. Despite this, some criticise its relevance to specific outcomes, suggesting that several questions may not pertain to the interest outcomes, and others may make assumptions such as “past surgery”. Its applicability for younger children was also raised as a concern, with its current format may require in-person interactions for accurate rating.
Symptom Burden Questionnaire for Long COVID (Circulation scale)	Experts have expressed mixed views about the given scale. The scale was originally developed for adults, and while some believe it's adaptable for children, others note that it would require modification and validation for paediatric populations. The clarity in defining degrees of severity, such as mild, moderate, and severe, was considered not easy to implement. Despite being in development, some experts appreciate the scale's design, finding it comprehensive and potentially superior to other outcome measures if certain sections were removed. However, they caution that it's not fully validated yet, and its reliance on a 7-day recall period might be insufficient given the fluctuating nature of many symptoms. It's also viewed as a feasible tool that captures relevant aspects of Long COVID, yet it notably lacks a focus on chest pain. Adaptation of this tool for younger people is currently underway.

Postural orthostatic tachycardia syndrome (POTS)	Some experts believe that this instrument is not ideal for children, particularly younger ones, implying that it may be more suitable for older children or adults. Others point out that it also incorporates questions for several non-cardiac issues, suggesting it may be too broad in scope. There is a consensus that some of the questions are too specific to POTS or that they are replicated in other questionnaires, making it less unique or potentially redundant. Despite these criticisms, some experts found the questionnaire straightforward, and believe that the content is appropriate and relevant, though there are reservations regarding the psychometrics of the scale.

3. List of unique outcome measures for COS outcomes

COS outcome	Outcome Measure	Result
Cardiovascular functioning, symptoms and conditions	PedsQL™ Cardiac Module	Excluded following discussions at consensus meeting
	Symptom Burden Questionnaire for Long COVID (Circulation scale)	Excluded following discussions at consensus meeting
	Malmo POTS score (MAPS)	Excluded following discussions at consensus meeting
	ADHD Cardiac screening questionnaire	Excluded following expert Delphi process*
	Paediatric Sudden Cardiac Arrest Signal questions	Excluded following expert Delphi process*
	SCL-90 Scale	Excluded by core group prior to expert Delphi (not feasible)
	ISARIC COVID-19 Health and Wellbeing Follow-Up Survey for Children	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Questionnaire (adapted for children from the adult WHO CRF for post-COVID-19 conditions) by Vanesa Seery et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9892252/bin/mmc2.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Telephone follow-up using standardised clinical proforma by Cara J Bossley et al. (https://assets.researchsquare.com/files/rs-1001103/v1/1c14f9553af8d1d272de0e35.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Telephone interview using original questionnaire by Ali A Asadi-Pooya et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8414448/bin/12519_2021_457_MOESM2_ESM.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Self-reported data through a mobile application by Erika Molteni et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8443448/bin/mmc1.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Ellinor Sterky et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8444740/bin/APA-110-2578-s001.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original survey for paediatricians by Giuseppe Fabio Parisi et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8467017/table/childre	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)

	n-08-00769-t001/?report=objectonly)	
	Original questionnaire by Ieva Roge et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8586002/bin/Data_Sheet_2.PDF)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original online survey for the children's parent/guardian by Maria Zavala et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8767867/bin/ciab991_suppl_Supplementary_Data.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Roxane Dumont et al. (https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-022-34616-8/MediaObjects/41467_2022_34616_MOESM1_ESM.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original online questionnaire by Adriana Prato et al. (12887_2023_4035_MOESM1_ESM.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Limor Adler et al. (https://bmjopen.bmj.com/content/bmjopen/suppl/2023/02/21/bmjopen-2022-064155.DC1/bmjopen-2022-064155supp001_data_supplement.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	B-type natriuretic peptide	Excluded by core group prior to expert Delphi (laboratory investigation)
	Blood tests - Troponin	Excluded by core group prior to expert Delphi (laboratory investigation)
	Pro-BNP	Excluded by core group prior to expert Delphi (laboratory investigation)
	Troponin I	Excluded by core group prior to expert Delphi (laboratory investigation)
	Troponin T	Excluded by core group prior to expert Delphi (laboratory investigation)
	12-lead electrocardiogram	Excluded by core group prior to expert Delphi (clinical investigation)
	24 hours ambulatory ECG recording	Excluded by core group prior to expert Delphi (clinical investigation)
	24hr ECG	Excluded by core group prior to expert Delphi (clinical investigation)
	6MWT	Excluded by core group prior to expert Delphi (clinical investigation)
	Angiogram	Excluded by core group prior to expert Delphi (clinical investigation)
	Blood pressure	Excluded by core group prior to expert Delphi (clinical investigation)
	Cardiac examination	Excluded by core group prior to expert Delphi (clinical investigation)

	Cardiac MRI	Excluded by core group prior to expert Delphi (clinical investigation)
	Cardiac ultrasound	Excluded by core group prior to expert Delphi (clinical investigation)
	Cardiopulmonary exercise test	Excluded by core group prior to expert Delphi (clinical investigation)
	CT-pulmonary angiograms	Excluded by core group prior to expert Delphi (clinical investigation)
	Detailed echocardiography	Excluded by core group prior to expert Delphi (clinical investigation)
	Doppler Ultrasound (Baseline blood flow measurements in the brachial artery)	Excluded by core group prior to expert Delphi (clinical investigation)
	Doppler Ultrasound (Flow-mediated vasodilation (VMF) in the brachial artery)	Excluded by core group prior to expert Delphi (clinical investigation)
	ECG	Excluded by core group prior to expert Delphi (clinical investigation)
	Echocardiogram	Excluded by core group prior to expert Delphi (clinical investigation)
	Echocardiographical M mode - LateraL E/E‘ ratio	Excluded by core group prior to expert Delphi (clinical investigation)
	Echocardiographical M mode - Left atrial to aortic ratio	Excluded by core group prior to expert Delphi (clinical investigation)
	Echocardiographical M mode - Left ventricular ejection fraction	Excluded by core group prior to expert Delphi (clinical investigation)
	Echocardiographical M mode - Left ventricular end diastolic diameter	Excluded by core group prior to expert Delphi (clinical investigation)
	Echocardiographical M mode - Left ventricular posterior wall diameter	Excluded by core group prior to expert Delphi (clinical investigation)
	Echocardiographical M mode - Mitral septal E/E‘, M/S ratio	Excluded by core group prior to expert Delphi (clinical investigation)
	Electrocardiogram Conduction block	Excluded by core group prior to expert Delphi (clinical investigation)
	Electrocardiogram including arrhythmia	Excluded by core group prior to expert Delphi (clinical investigation)
	Electrocardiogram ST-T change	Excluded by core group prior to expert Delphi (clinical investigation)
	Exercise stress test	Excluded by core group prior to expert Delphi (clinical investigation)
	Heart rate	Excluded by core group prior to expert Delphi (clinical investigation)
	Holter	Excluded by core group prior to expert Delphi (clinical

		investigation)
	Medical imaging of the heart	Excluded by core group prior to expert Delphi (clinical investigation)
	Non-contrast cardiac MRI	Excluded by core group prior to expert Delphi (clinical investigation)
	Oscillometric BP device	Excluded by core group prior to expert Delphi (clinical investigation)
	Peripheral vascular examination	Excluded by core group prior to expert Delphi (clinical investigation)
	Stress test using treadmill ergometry	Excluded by core group prior to expert Delphi (clinical investigation)
	Tissue Doppler	Excluded by core group prior to expert Delphi (clinical investigation)
Gastrointestinal functioning, symptoms, and conditions	PedsQL™ Gastrointestinal Symptoms Scales	Included in the COMS as a measurement instrument for "Gastrointestinal functioning, symptoms, and conditions"
	Questionnaire on Pediatric Gastrointestinal Symptoms (QPGS)	Excluded following discussions at consensus meeting
	Symptom Burden Questionnaire for Long COVID (Stomach and Digestion Scale)	Excluded following discussions at consensus meeting
	EAT-10 score	Excluded following expert Delphi process*
	Section within the SCL-90 scale	Excluded following expert Delphi process*
	Original questionnaire by Mostafa M. Khodeir et al. (http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0260259.s002)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Telephone follow-up using standardised clinical proforma by Cara J Bossley et al. (https://assets.researchsquare.com/files/rs-1001103/v1/1c14f9553af8d1d272de0e35.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Luise Borch et al. (https://static-content.springer.com/esm/art%3A10.1007%2Fs00431-021-04345-z/MediaObjects/431_2021_4345_MOESM1_ESM.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Telephone interview using original questionnaire by Ali A Asadi-Pooya et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8414448/bin/12519_2021_457_MOESM2_ESM.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Self-reported data through a mobile application by Erika Molteni et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8443448/bin/mmc1.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original survey for paediatricians by Giuseppe Fabio Parisi et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8467017/table/children-08-00769-t001/?report=objectonly)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Ieva Roge et al.	Excluded by core group prior to expert Delphi (non-validated

	(https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8586002/bin/Data_Sheet_2.PDF)	questionnaire/CRF)
	Original online survey for the children's parent/guardian by Maria Zavala et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8767867/bin/ciab991_suppl_Supplementary_Data.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	CLoCk Questionnaire by Terence Stephenson et al. (https://www.thelancet.com/cms/10.1016/S2352-4642(22)00022-0/attachment/15f4036a-7343-461f-9399-85fcb36b5042/mmc1.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Roxane Dumont et al. (https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-022-34616-8/MediaObjects/41467_2022_34616_MOESM1_ESM.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original online questionnaire by Adriana Prato et al. (12887_2023_4035_MOESM1_ESM.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Questionnaire (adapted for children from the adult WHO CRF for post-COVID-19 conditions) by Vanesa Seery et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9892252/bin/mmc2.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Measuring stomach reflux symptom by Visual Analog Score (VAS)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	ISARIC COVID-19 Health and Wellbeing Follow-Up Survey for Children	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Blood analysis	Excluded by core group prior to expert Delphi (laboratory investigation)
	Albumin	Excluded by core group prior to expert Delphi (laboratory investigation)
	Triglycerides	Excluded by core group prior to expert Delphi (laboratory investigation)
	Total cholesterol	Excluded by core group prior to expert Delphi (laboratory investigation)
	Total Bilirubin	Excluded by core group prior to expert Delphi (laboratory investigation)
	Stool Sample (faeces or rectal swab)	Excluded by core group prior to expert Delphi (laboratory investigation)
	Alkaline phosphatase (ALP)	Excluded by core group prior to expert Delphi (laboratory investigation)
	Metagenomic sequencing on rectal swabs/stools	Excluded by core group prior to expert Delphi (laboratory investigation)
	Amilase	Excluded by core group prior to expert Delphi (laboratory investigation)
	Alanine Aminotransferase (ALT)	Excluded by core group prior to expert Delphi (laboratory investigation)

	Aspartate Aminotransferase (AST)	Excluded by core group prior to expert Delphi (laboratory investigation)
	Bilirubin	Excluded by core group prior to expert Delphi (laboratory investigation)
	Gamma-glutamyl Transferase (GGT)	Excluded by core group prior to expert Delphi (laboratory investigation)
	Lactate Dehydrogenase (LDH)	Excluded by core group prior to expert Delphi (laboratory investigation)
	Faecal routine test	Excluded by core group prior to expert Delphi (laboratory investigation)
	Lipase	Excluded by core group prior to expert Delphi (laboratory investigation)
	Volume-Viscosity Swallowing Test (V-VST)	Excluded by core group prior to expert Delphi (clinical investigation)
	Abdominal examination	Excluded by core group prior to expert Delphi (clinical investigation)
	Abdominal ultrasound	Excluded by core group prior to expert Delphi (clinical investigation)
Fatigue or Exhaustion	PedsQL™ Multidimensional Fatigue Scale	Included in the COMS as a measurement instrument for "Fatigue or Exhaustion"
	Chalder fatigue questionnaire	Excluded following discussions at consensus meeting
	PROMIS Paediatric Fatigue	Excluded following discussions at consensus meeting
	Symptom Burden Questionnaire for Long COVID (Fatigue scale)	Excluded following discussions at consensus meeting
	Multidimensional Fatigue Inventory, MFI-20	Excluded following expert Delphi process*
	Fried Frailty phenotype	Excluded following expert Delphi process*
	Bell's Functionality Score	Excluded following expert Delphi process*
	ISARIC COVID-19 Health and Wellbeing Follow-Up Survey for Children	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Modified Rankin scale (mRS)	Excluded by core group prior to expert Delphi (not suitable for the purpose of this COS)
	pedsFACIT-F	Excluded by core group prior to expert Delphi (not suitable for the purpose of this COS)
	Question verbally on the phone, "In the last month, have you felt tired for a great part of the day?"	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Limor Adler et al. (https://bmjopen.bmj.com/content/bmjopen/suppl/2023/02/21/bmjopen-2022-064155.DC1/bmjopen-2022-064155supp001_data_supplement.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original online questionnaire by Adriana Prato et al.	Excluded by core group prior to expert Delphi (non-validated

	(12887_2023_4035_MOESM1_ESM.docx)	questionnaire/CRF)
	CLoCk Questionnaire by Terence Stephenson et al. (https://www.thelancet.com/cms/10.1016/S2352-4642(22)00022-0/attachment/15f4036a-7343-461f-9399-85fcb36b5042/mmc1.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Mostafa M. Khodeir et al. (http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0260259.s002)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Telephone follow-up using standardised clinical proforma by Cara J Bossley et al. (https://assets.researchsquare.com/files/rs-1001103/v1/1c14f9553af8d1d272de0e35.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Luise Borch et al. (https://static-content.springer.com/esm/art%3A10.1007%2Fs00431-021-04345-z/MediaObjects/431_2021_4345_MOESM1_ESM.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Standardised clinic proforma by Daniela Say et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8057863/bin/mmc1.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Telephone interview using original questionnaire by Ali A Asadi-Pooya et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8414448/bin/12519_2021_457_MOESM2_ESM.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Self-reported data through a mobile application by Erika Molteni et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8443448/bin/mmc1.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	An original questionnaire by Ellinor Sterky et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8444740/bin/APA-110-2578-s001.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original survey for paediatricians by Giuseppe Fabio Parisi et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8467017/table/children-08-00769-t001/?report=objectonly)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Ieva Roge et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8586002/bin/Data_Sheet_2.PDF)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Questionnaire (adapted for children from the adult WHO CRF for post-COVID-19 conditions) by Vanesa Seery et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9892252/bin/mmc2.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Roxane Dumont et al. (https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-022-34616-8/MediaObjects/41467_2022_34616_MOESM1_ESM.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
Post-exertion symptoms	CDC symptom inventory for CFS	Excluded following discussions at consensus meeting
	PEM items from DePaul Symptom Questionnaire	Excluded following discussions at consensus meeting

	Symptom Burden Questionnaire for Long COVID (Fatigue scale)	Excluded following discussions at consensus meeting
	ISARIC COVID-19 Health and Wellbeing Follow-Up Survey for Children	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Roxane Dumont et al. (https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-022-34616-8/MediaObjects/41467_2022_34616_MOESM1_ESM.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
Neuro-cognitive system functioning, symptoms, and condition	Peds QL Cognitive Functioning Scale	Included in the COMS as a measurement instrument for "Neuro-cognitive system functioning, symptoms, and conditions"
	PROMIS Pediatric Cognitive Function - Short Form 7a	Excluded following discussions at consensus meeting
	Symptom Burden Questionnaire for Long COVID (Memory, Thinking & Communication scale, movement scale, muscles and joints, pain scales)	Excluded following discussions at consensus meeting
	Addenbrooke's Cognitive Examination (ACE-III)	Excluded following expert Delphi process*
	ASQ assessment (for infants born >29 weeks gestation)	Excluded following expert Delphi process*
	Bayley-IV neurological examination	Excluded following expert Delphi process*
	Chalder fatigue scale	Excluded following expert Delphi process*
	From Body Vigilance Scale (BVS)	Excluded following expert Delphi process*
	Functional Independence measure (FIM)	Excluded following expert Delphi process*
	IQCODE	Excluded following expert Delphi process*
	Short Blessed Test	Excluded following expert Delphi process*
	Vanderbilt ADHD assessment	Excluded following expert Delphi process*
	NIH Toolbox	Excluded following expert Delphi process*
	ISARIC COVID-19 Health and Wellbeing Follow-Up Survey for Children	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	MentalPlus® (a scale of assessment and cognitive rehabilitation)	Excluded by core group prior to expert Delphi (not applicable in low-resource settings)
	Modified Rankin scale (mRS)	Excluded by core group prior to expert Delphi (not suitable for the purpose of this COS)
	Montreal Cognitive Assessment	Excluded by core group prior to expert Delphi (not suitable for the purpose of this COS)
	SCL-90 scale	Excluded by core group prior to expert Delphi (not feasible)
	SDQ (Hyperactivity scale)	Excluded by core group prior to expert Delphi (not suitable for the purpose of this COS)
	Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWS)	Excluded by core group prior to expert Delphi (not suitable for the purpose of this COS)
	Original questionnaire by Limor Adler et al. (https://bmjopen.bmj.com/content/bmjopen/suppl/2023/02/21/bmjop	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)

	en-2022-064155.DC1/bmjopen-2022-064155supp001_data_supplement.pdf)	
	Original questionnaire by Roxane Dumont et al. (https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-022-34616-8/MediaObjects/41467_2022_34616_MOESM1_ESM.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Mostafa M. Khodeir et al. (http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0260259.s002)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Telephone follow-up using standardised clinical proforma by Cara J Bossley et al. (https://assets.researchsquare.com/files/rs-1001103/v1/1c14f9553af8d1d272de0e35.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Luise Borch et al. (https://static-content.springer.com/esm/art%3A10.1007%2Fs00431-021-04345-z/MediaObjects/431_2021_4345_MOESM1_ESM.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Telephone interview using original questionnaire by Ali A Asadi-Pooya et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8414448/bin/12519_2021_457_MOESM2_ESM.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Self-reported data through a mobile application by Erika Molteni et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8443448/bin/mmc1.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	An original questionnaire by Ellinor Sterky et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8444740/bin/APA-110-2578-s001.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Ieva Roge et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8586002/bin/Data_Sheet_2.PDF)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original online survey for the children's parent/guardian by Maria Zavala et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8767867/bin/ciab991_suppl_Supplementary_Data.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	CLoCk Questionnaire by Terence Stephenson et al. (https://www.thelancet.com/cms/10.1016/S2352-4642(22)00022-0/attachment/15f4036a-7343-461f-9399-85fcb36b5042/mmc1.pdf)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original online questionnaire by Adriana Prato et al. (12887_2023_4035_MOESM1_ESM.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Questionnaire (adapted for children from the adult WHO CRF for post-COVID-19 conditions) by Vanesa Seery et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9892252/bin/mmc2.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	NSE, S100B, neurofilament proteins in blood	Excluded by core group prior to expert Delphi (laboratory

		investigation)
	Attention Bias test of automatic biases towards disease-associated words	Excluded by core group prior to expert Delphi (requires trained personnel)
	Function Acquisition Speed Test	Excluded by core group prior to expert Delphi (requires trained personnel)
	Hopkins Verbal Learning Test-Revised (HVLT-R)	Excluded by core group prior to expert Delphi (requires trained personnel)
	Behavior Rating Inventory of Executive Function 2nd Edition (BRIEF-2)	Excluded by core group prior to expert Delphi (requires trained personnel)
	California Verbal Learning Test Children's Version (CVLT-C)	Excluded by core group prior to expert Delphi (requires trained personnel)
	Child and Adolescent Memory Profile List (ChAMP) List	Excluded by core group prior to expert Delphi (requires trained personnel)
	Conners Comprehensive Behavior Rating Scale (Conners CBRS)	Excluded by core group prior to expert Delphi (not feasible)
	Conners Early Childhood (Conners EC)	Excluded by core group prior to expert Delphi (not feasible)
	Delis Kaplan Executive Functioning System Verbal Fluency (D-KEFS)	Excluded by core group prior to expert Delphi (requires trained personnel)
	MVP Verbal Subtest and Reliable Digit Span	Excluded by core group prior to expert Delphi (requires trained personnel)
	NEPSY-II Auditory Attention	Excluded by core group prior to expert Delphi (requires trained personnel)
	Oral Symbol Digits Modalities Test (SDMT)	Excluded by core group prior to expert Delphi (not suitable for the purpose of this COS)
	Test of Everyday Attention of Children Score (TEA-Ch Score)	Excluded by core group prior to expert Delphi (requires trained personnel)
	Wechsler Intelligence Scale for Children 5th Edition Digit Span (WISC-V)	Excluded by core group prior to expert Delphi (not suitable for the purpose of this COS)
	Digit Span forward and backward test	Excluded by core group prior to expert Delphi (requires trained personnel)
	The Babinski reflex	Excluded by core group prior to expert Delphi (clinical investigation)
	Hoffman's sign	Excluded by core group prior to expert Delphi (clinical investigation)
	Brain fMRI during resting state and a fatigue-provoking test	Excluded by core group prior to expert Delphi (clinical investigation)
	Neurological examination	Excluded by core group prior to expert Delphi (clinical investigation)
Physical	EQ5D (family of instruments)	Included in the COMS as a measurement instrument for "Physical functioning, symptoms, and conditions"

functioning, symptoms, and conditions	PROMIS Early Childhood Parent Report Physical Activity 7a	Excluded following discussions at consensus meeting
	PROMIS Pediatric Physical Activity – Short Form 8a	Excluded following discussions at consensus meeting
	Symptom Burden Questionnaire for Long COVID (Impact on Daily Life Scale)	Excluded following discussions at consensus meeting
	Barthel Index	Excluded following expert Delphi process*
	Basic Activity of Daily Living (BADL)	Excluded following expert Delphi process*
	Clinical Frailty Scale (CFS)	Excluded following expert Delphi process*
	Duke Activity Status Index (DASI)	Excluded following expert Delphi process*
	Fried Frailty phenotype	Excluded following expert Delphi process*
	Functional Independence Measure (WeeFIM or FIM)	Excluded following expert Delphi process*
	International Physical Activity Questionnaires Short Form (IPAQ-SF)	Excluded following expert Delphi process*
	Post COVID-19 Functional Status Scale (Scale 0-64 points)	Excluded following expert Delphi process*
	PROMIS Pediatric Physical Activity – Short Form 4a	Excluded following expert Delphi process*
	Bell's Functionality Score	Excluded following expert Delphi process*
	The motor skills module activity questionnaire (MOMO) (Available in German only)	Excluded following expert Delphi process*
	Global Physical Activity Questionnaire (GPAQ)	Excluded by core group prior to expert Delphi (not suitable for the purpose of this COS)
	Growth indices	Excluded by core group prior to expert Delphi (not suitable for the purpose of this COS)
	ISARIC COVID-19 Health and Wellbeing Follow-Up Survey for Children	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Functional Status Scale (FSS)	Excluded by core group prior to expert Delphi (requires trained personnel)
	Medical Outcome Study Short Form (MOS SF)-36 Score	Excluded by core group prior to expert Delphi (not suitable for the purpose of this COS)
	Short Physical Performance Battery (SPPB)	Excluded by core group prior to expert Delphi (clinical investigation)
	Telephone follow-up using standardised clinical proforma by Cara J Bossley et al. (https://assets.researchsquare.com/files/rs-1001103/v1/1c14f9553af8d1d272de0e35.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Telephone interview using original questionnaire by Ali A Asadi-Pooya et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8414448/bin/12519_2021_457_MOESM2_ESM.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	An original questionnaire by Ellinor Sterky et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8444740/bin/APA-	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)

	110-2578-s001.docx)	
	Original questionnaire by Ieva Roge et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8586002/bin/Data_Sheet_2.PDF)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original online questionnaire by Adriana Prato et al. (12887_2023_4035_MOESM1_ESM.docx)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Pulse oximetry (SpO2) at rest, before 6-minute walk test	Excluded by core group prior to expert Delphi (clinical investigation)
	Pulse oximetry (SpO2) during exercise, at the end of 6-minute walk test	Excluded by core group prior to expert Delphi (clinical investigation)
	Incremental Cardiopulmonary exercise test (Dyspnea during exercise, 10-point categorical Borg scale)	Excluded by core group prior to expert Delphi (clinical investigation)
	Incremental Cardiopulmonary exercise test (Inspiratory capacity during exercise, L and % of predicted)	Excluded by core group prior to expert Delphi (clinical investigation)
	Incremental Cardiopulmonary exercise test (Minute-ventilation/carbon dioxide output during exercise (L/L))	Excluded by core group prior to expert Delphi (clinical investigation)
	Incremental Cardiopulmonary exercise test (Oxygen uptake at peak exercise (% of predicted))	Excluded by core group prior to expert Delphi (clinical investigation)
	Incremental Shuttle Walk Test (SWT)	Excluded by core group prior to expert Delphi (clinical investigation)
	Berg Balance Test	Excluded by core group prior to expert Delphi (clinical investigation)
	Standardised stadiometer (calculating standard deviation, growth curves, and growth speed)	Excluded by core group prior to expert Delphi (clinical investigation)
	Actigraph (3D accelerometer) model G-Walk during the 10 metre gait test	Excluded by core group prior to expert Delphi (clinical investigation)
	Actigraph (3D accelerometer) model G-Walk during the 6-minute walk test	Excluded by core group prior to expert Delphi (clinical investigation)
	Actigraph (3D accelerometer) model G-Walk used during the "timed up and go" test	Excluded by core group prior to expert Delphi (clinical investigation)
	Musculoskeletal ultrasound	Excluded by core group prior to expert Delphi (clinical investigation)
	ActivPAL	Excluded by core group prior to expert Delphi (clinical investigation)
	Six-minute walk test	Excluded by core group prior to expert Delphi (clinical investigation)
Work/ occupational and study changes	Symptom Burden Questionnaire for Long COVID (Impact on Daily Life Scale)	Excluded following discussions at consensus meeting
	WHO DAS 2 Children and Youth 36-Item Version	Excluded following discussions at consensus meeting
	Work Productivity and Activity Impairment Questionnaire: General Health V2.0 (WPAI:GH)	Excluded following expert Delphi process*

	ISARIC COVID-19 Health and Wellbeing Follow-Up Survey for Children	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Original questionnaire by Ieva Roge et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8586002/bin/Data_Sheet_2.PDF)	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)
	Number of absent days from school/work due to illness	Excluded by core group prior to expert Delphi (non-validated questionnaire/CRF)

*- Did not meet a priori predefined criteria ("include" or "maybe" responses from more than half of the experts)

4. Full details of expert Delphi participants

Name	Surname	Gender	Institution	Country	Stakeholder group (HCP/Researcher)	Primary expertise related to Long COVID
Ali Akbar	Asadi-Pooya	Male	Epilepsy Research Center, Shiraz University of Medical Sciences; Jefferson comprehensive epilepsy center, Thomas Jefferson University	Iran/USA	Health professional/Researcher	Neurological/cognitive problems in long COVID
Dr Anbarasu Theodore	Anbu	Male	Alder Hey Children's NHS Foundation Trust	United Kingdom	Health professional, Paediatrician	Lead for CYP Long Covid and ME/CFS service at Alder Hey Children's NHS Foundation Trust Hub
Carlos R.	Oliveira	Male	Yale University School of Medicine	USA	Health professional/Researcher	Diagnosis and treatment of paediatric Long COVID patients.
Danilo	Buonsenso	Male	Fondazione Policlinico Universitario A. Gemelli IRCCS	Italy	Health professional/Researcher	Paediatric infectious diseases
Sarah	Hughes	Female	University of Birmingham	United Kingdom	Researcher	Outcome measure development (patient-reported outcomes
Laura	Malone	Female	Kennedy Krieger Institute & Johns Hopkins	USA	Health professional/Researcher	Paediatric long COVID
Liat	Ashkenazi-Hoffnung	Female	Schneider Children's Medical Center	Israel	Health professional	Pediatric Infectious Diseases

Olalekan Lee	Aiyegbusi	Male	University of Birmingham	United Kingdom	Researcher	Patient and public involvement lead for the NIHR-funded TLC Study. Conducted reviews of long COVID literature and was involved in the development of the SBQ a PRO measure for assessing symptoms of long COVID
Daniele	Dona'	Male	Department for Women's and Children's Health, University of Padua	Italy	Researcher	Pediatric Infectious Diseases Consultant, co-leader of the Clinical Working Group of the VERDI project (101045989), which is funded by the European Union.
Claire	Thorne	Female	Population, Policy and Practice Dept, University College London GOS Institute of Child Health	United Kingdom	Researcher	Infectious diseases epidemiology
Terry	Segal	Female	University College London Hospitals NHS Foundation Trust	United Kingdom	Health professional/Researcher	Adolescence, paediatric endocrinology (growth and puberty), chronic fatigue syndrome, obesity, anorexia nervosa (medical aspects), chronic medically unexplained symptoms

5. Results following expert Delphi

Outcome 1: Cardiovascular functioning, symptoms, and conditions

Measurement instruments	Round of expert Delphi	Expert voting										
		1	2	3	4	5	6	7	8	9	10	11
PedsQL™ Cardiac Module	Round 1	Maybe	Include	Exclude	Maybe	Include	Include	Include	Include	Maybe	Include	Include

	Round 2	Include	Include	Maybe	Include	Include	Include	Include	Include	Include	Include	Include
ADHD Cardiac screening questionnaire	Round 1	Exclude	Exclude	Maybe	Maybe	Exclude	Exclude	Exclude	Exclude	Maybe	Exclude	Exclude
	Round 2	Exclude	Exclude	Maybe	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude
Paediatric Sudden Cardiac Arrest Signal questions	Round 1	Exclude	Exclude	Maybe	Exclude	Maybe	Maybe	Maybe	Exclude	Exclude	Exclude	Maybe
	Round 2	Exclude	Exclude	Maybe	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Maybe
Symptom Burden Questionnaire for Long COVID (Circulation scale)	Round 1	NEWLY SUGGESTED										
	Round 2	*Unvoted*	Maybe	Include	Maybe	Include	Maybe	Maybe	Include	Include	Include	*Unvoted*
Malmo POTS score (MAPS)	Round 1	NEWLY SUGGESTED										
	Round 2	*Unvoted*	Exclude	Maybe	Exclude	Maybe	Include	Maybe	Exclude	Maybe	Include	Include

Measurement instruments	**Summary of additional comments from the experts in rounds 1/2**
PedsQL™ Cardiac Module	Some experts express difficulties in understanding the module and indicate potential issues with its accessibility. Yet, others appreciate the PedsQL, a well-validated and widely used questionnaire set, often favoured in most studies due to its generic quality of life assessment, which may be more appropriate than other cardiac-specific measures. Experts also acknowledge the scale's beneficial features, like ability to use in a paper format, existence of age-specific questions, and the inclusion of cognitive scores. Despite this, some criticise its relevance to specific outcomes, suggesting that several questions may not pertain to the interest outcomes, and others may make assumptions such as “past surgery”. Its applicability for younger children was also raised as a concern, with its current format may require in-person interactions for accurate rating.

ADHD Cardiac screening questionnaire	Overall, experts have mixed opinions on the ADHD Cardiac screening questionnaire. Some believe it is not ideal for follow-up visits and has questions that are unrelated to Long COVID. However, others find the first questions on intolerance, ECG, and fainting to be good. The relevance of the questions for family history is disputed. The questionnaire is considered short and simple, but there are concerns about its applicability to the paediatric population and the potential distress caused by some of the sensitive questions. Additionally, experts mention that the questionnaire focuses more on congenital and sudden death screenings rather than long COVID.
Paediatric Sudden Cardiac Arrest Signal questions	There are mixed opinions among experts regarding the appropriateness of the Paediatric Sudden Cardiac Arrest Signal questions for follow-up visits. Some experts believe that only the first part of the questionnaire is suitable. The focus on family history is a concern for several experts, as it may increase anxiety without providing any useful information. Overall, experts suggest modifying the questionnaire to include only the first five questions and rewording them for improved clarity.
Symptom Burden Questionnaire for Long COVID (Circulation scale)	Experts have expressed mixed views about the given scale. The scale was originally developed for adults, and while some believe it's adaptable for children, others note that it would require modification and validation for paediatric populations. The clarity in defining degrees of severity, such as mild, moderate, and severe, was considered not easy to implement. Despite being in development, some experts appreciate the scale's design, finding it comprehensive and potentially superior to other outcome measures if certain sections were removed. However, they caution that it's not fully validated yet, and its reliance on a 7-day recall period might be insufficient given the fluctuating nature of many symptoms. It's also viewed as a feasible tool that captures relevant aspects of Long COVID, yet it notably lacks a focus on chest pain. Adaptation of this tool for younger people is currently underway.
Malmo POTS score (MAPS)	Some experts believe that this instrument is not ideal for children, particularly younger ones, implying that it may be more suitable for older children or adults. Others point out that it also incorporates questions for several non-cardiac issues, suggesting it may be too broad in scope. There is a consensus that some of the questions are too specific to POTS or that they are replicated in other questionnaires, making it less unique or potentially redundant. Despite these criticisms, some experts found the questionnaire straightforward, and believe that the content is appropriate and relevant, though there are reservations regarding the psychometrics of the scale.

Outcome 2: Gastrointestinal functioning, symptoms, and conditions

Measurement instruments	Round of expert Delphi	Expert voting										
		1	2	3	4	5	6	7	8	9	10	11
EAT-10 score	Round 1	Exclude	Include	Exclude	Maybe	Include	Include	Exclude	Include	Exclude	Exclude	Include

	Round 2	Exclude	Exclude	Exclude	Exclude	Maybe	Include	Exclude	Exclude	Exclude	Exclude	*Unvoted*
Section within the SCL-90 scale	Round 1	Exclude	Exclude	Exclude	Exclude	Include	*Unvoted*	Maybe	Exclude	Maybe	Maybe	Exclude
	Round 2	Exclude	Exclude	Exclude	Exclude	Maybe	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude
PedsQL™ Gastrointestinal Symptoms Scales	Round 1	Maybe	Maybe	Include	Maybe	Maybe	Include	Include	Include	Include	Include	Include
	Round 2	Include	Include	Include	Include	Include	Include	Include	Include	Maybe	Include	Include
Questionnaire on Pediatric Gastrointestinal Symptoms (QPGS)	Round 1	Exclude	Include	Maybe	Maybe	Maybe	Include	Include	Exclude	Exclude	Include	Include
	Round 2	Exclude	Include	Include	Maybe	Maybe	Include	Exclude	Exclude	Maybe	Include	Include
Symptom Burden Questionnaire for Long COVID (Stomach and Digestion Scale)	Round 1	NEWLY SUGGESTED										
	Round 2	*Unvoted*	Maybe	Include	Exclude	Include	Exclude	Maybe	Include	Include	Maybe	*Unvoted*

Measurement instruments	**Summary of additional comments from the experts in rounds 1/2**
EAT-10 score	Overall, experts have varying opinions on the EAT-10 score questionnaire. Some believe that there are too many questions, many of which are irrelevant or unrelated to swallowing. Others find the questionnaire easy to complete and relevant, especially for monitoring purposes. There is also a recognition that difficulty swallowing is an important symptom to assess and that it is not covered by other tools. However, there is a consensus among experts that the questionnaire may be too long and that it may not accurately capture the symptoms commonly seen in paediatric patients. Overall, the relevance and usefulness of the EAT-10 score questionnaire seem to depend on the specific focus on swallowing difficulties and the individual needs of the patient population

	being assessed.
Section within the SCL-90 scale	Experts have differing opinions on the usefulness and appropriateness of the Section within the SCL-90 scale. Some experts feel that the section has too many questions and is too long, making it potentially burdensome for respondents, especially younger children. They also note that the section includes items that are not relevant to the specific outcomes of interest. Some experts point out that the scale was originally designed for psychiatric patients, which may reduce compliance and limit its applicability to other groups. Additionally, extracting individual items from the scale is seen as problematic, as the psychometric properties pertain to the scale as a whole, rather than individual items.
PedsQL™ Gastrointestinal Symptoms Scales	Some experts believe that it may lack specific relevance to the gastrointestinal (GI) implications of COVID and propose the development of a new scale, others praise its broad coverage of GI symptoms, especially in a paediatric setting, and its use of validated questionnaires, making it a go-to for most studies. The scale's length and accessibility of its questions are points of contention, with critics citing it as potentially too long or unclear. Despite these criticisms, the scale's comprehensive range of questions and its previous validation across a variety of paediatric GI conditions are applauded. Some suggest that the scale may even be redundant if a COVID symptom questionnaire is available, while others see its generality and age specificity as strengths. It is also recognised for covering symptoms included in COS and its development and validation within a paediatric population.
Questionnaire on Pediatric Gastrointestinal Symptoms (QPGS)	Mixed views were expressed, some experts believe that it is a comprehensive tool, specifically designed for functional gastrointestinal disorders, covering a broad range of symptoms and suitable for all ages, including follow-up visits. It is particularly noted for its potential applicability to populations experiencing Functional GI symptoms as seen in Chronic Fatigue Syndrome and Long Covid. Some experts appreciate its detailed nature, despite its length, and believe it could be feasibly completed periodically for continued monitoring. However, concerns are raised about its length - 83 questions - and redundancy, particularly when compared with the PedsQL GIS questionnaire that covers similar questions anyway. Critics also highlight weak temporal stability in items evaluating the impact of symptoms on school and social/family activities. A significant limitation flagged is that it has not been fully validated in children yet. Therefore, there is debate about its usage versus other tools like the PedsQL which is validated in children.
Symptom Burden Questionnaire for Long COVID (Stomach and Digestion Scale)	The scale has received mixed expert feedback. While it is recognised for its strengths, particularly its suitability for assessing gastrointestinal issues and its comprehensive coverage of major symptoms, it also raises concerns. Primary among them is the lack of clarity in defining symptom severity levels like mild, moderate, and severe. Also, experts pointed out the need for validation and adaptation for use in children and adolescents, as its primary development was for adults. Adaptation for younger age groups (11-17 years) is just under development, some professionals still express reservations about its usage in paediatrics, suggesting reliance on previously used, validated tools.

Outcomes 3 and 4: Fatigue or Exhaustion AND Post-exertion symptoms

Measurement instruments	Round of expert Delphi	Expert voting										
		1	2	3	4	5	6	7	8	9	10	11
Chalder fatigue questionnaire	Round 1	Maybe	Maybe	Maybe	Exclude	Maybe	Exclude	Exclude	Include	Include	Include	Maybe
	Round 2	Maybe	Maybe	Maybe	Exclude	Exclude	Exclude	Exclude	Include	Include	Include	Exclude
Fried Frailty phenotype	Round 1	Exclude	Exclude	Maybe	Maybe	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude
CDC symptom inventory for CFS	Round 1	Include	Maybe	Exclude	Maybe	Maybe	Maybe	Maybe	Include	Maybe	Include	Maybe
	Round 2	Maybe	Maybe	Exclude	Maybe	Exclude	Exclude	Exclude	Include	Maybe	Include	Maybe
PEM items from DePaul Symptom Questionnaire	Round 1	Maybe	Exclude	Include	Maybe	Exclude	Include	Include	Exclude	Include	Maybe	Exclude
	Round 2	Maybe	Exclude	Include	Maybe	Maybe	Maybe	Exclude	Exclude	Include	Exclude	Maybe
PROMIS Paediatric Fatigue	Round 1	Maybe	Include	Include	Maybe	Include	Maybe	Maybe	Include	Maybe	Exclude	*Unvoted*
	Round 2	Include	Include	Include	Maybe	Maybe	Maybe	Include	Include	Maybe	Exclude	Include
PedsQL™ Multidimensional Fatigue Scale	Round 1	Exclude	Include	*Unvoted*	Maybe	Maybe	Include	Include	Include	Include	Maybe	*Unvoted*

	Round 2	Maybe	Include	Exclude	Include	Maybe	Include	Include	Include	Include	Include	Maybe
Multidimensional Fatigue Inventory, MFI-20	Round 1	NEWLY SUGGESTED										
	Round 2	Maybe	Exclude	Include	Exclude	Exclude	Exclude	Maybe	Exclude	Maybe	Exclude	Maybe
Symptom Burden Questionnaire for Long COVID (Fatigue scale)	Round 1	NEWLY SUGGESTED										
	Round 2	Exclude	Maybe	Include	Exclude	Include	Include	Maybe	Include	Include	Maybe	*Unvoted*
Bell's Functionality Score	Round 1	NEWLY SUGGESTED										
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Include	Maybe	Include	Maybe

Measurement instruments	**Summary of additional comments from the experts in rounds 1/2**
Chalder fatigue questionnaire	The Chalder Fatigue Questionnaire has elicited varied opinions among experts. Some highlight the presence of unrelated questions, notably on memory, while others question its applicability in children, given its predominant use in adult populations. Additionally, it appears to elicit cognitive issues, with responses ranging from appreciation for its succinctness to critique for its potential controversy. The scale's validation in paediatric populations also raises questions. Further, some experts cast doubt on the scale due to controversy surrounding Chalder's work, particularly in relation to cognitive behavioural therapy for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome, which may not adequately capture long COVID population. However, others appreciate its simplicity, shortness, and its established use in conditions like Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. They argue that controversies related to the authors should not discount its use. It is also noted that the scale's age range is 18-65, questioning its content validity and highlighting the lack of validation in paediatric populations.

Fried Frailty phenotype	Experts agree that the Fried Frailty phenotype is suitable for adults, specifically focused on assessing frailty in older adults. However, it is not appropriate for children. The Fried Frailty phenotype was initially developed for cardiac patients and is mostly focused on the adult population. It requires tests and equipment that may not be feasible or suitable for use in children and adolescents. It is suggested that alternative measures, such as the Physical Activity Questionnaire for Children (PAQ-C) and Adolescents (PAQ-A), may be more useful in assessing frailty in this population.
CDC symptom inventory for CFS	Instrument has been viewed as a robust but relatively complicated tool. Experts appreciate its good internal consistency, excellent convergent validity, and its specificity for Chronic Fatigue Syndrome symptoms, finding it somewhat comparable to the Long Covid scale. However, they have raised concerns about its lengthy and intricate scoring process, which they perceive to be a challenge in quantifying symptoms. Additionally, the scale is criticised for its breadth, spanning multiple domains rather than being narrowly focused. It was also noted that the instrument has not been developed specifically for paediatrics, limiting its applicability in younger populations. It has also not been validated for Primary Care Clinics, which can pose questions about its reliability and validity in these settings. Lastly, if only specific components, such as Fatigue and Exhaustion data, are to be extracted from the general inventory, it could increase the administration burden and potentially affect its reliability and validity.
PEM items from DePaul Symptom Questionnaire	The key concerns highlighted include the length and complexity of the DSQ, with 91 sections deemed time-consuming, and possibly leading to low compliance due to the high burden on patients. Some experts also noted its potential limitations when applying it to children. However, many expressed appreciation for the DSQ's extensive coverage of symptoms beyond fatigue, and its detailed assessment of frequency and severity over a longer period (3 months) than other instruments. Experts seem interested in the potential utility of the DSQ's paediatric version (DSQ-Ped), which is currently being validated. While it's noted that the questionnaire might be better adapted for a wider age range, the DSQ could be a suitable patient-reported outcome (PRO) tool, particularly if a paediatric version becomes available. There's also the possibility of extracting individual items from the questionnaire, though there are concerns about how this might impact scoring, reliability, and validity. The consensus seems to be that the DSQ's use requires further discussion, and decisions should be made based on the similarity of the assessed factors.
PROMIS Paediatric Fatigue	Most experts appreciate its combination of information acquisition and feasibility, citing it as a validated measure that's short, specific to paediatrics, and widely used in practice. The scale's specific focus on fatigue was recognized as advantageous by some, given that it doesn't encompass aspects outside its targeted domain. However, several experts expressed concern about the scale's lack of attention to cognitive fatigue, key components in the broader concept of fatigue. These limitations suggest that the PROMIS Paediatric Fatigue scale might best be used in conjunction with other instruments to ensure comprehensive fatigue assessment.
PedsQL™ Multidimensional Fatigue Scale	Experts found this instrument as a generally valid and useful tool, especially for assessing fatigue in children over 8 years old. Although the scale, featuring a relatively large set of 45 questions, may appear extensive, it provides a detailed analysis of various aspects of fatigue, including cognitive fatigue. This factor is crucial when evaluating children in certain populations. While the quality of life (QOL) aspect is not directly linked with the Core Outcome Set, it is still considered feasible and beneficial, as it is regularly used in clinical practice. Overall, the scale is recognised for its focus on fatigue but needs to be understood within its limitations.
Multidimensional Fatigue Inventory, MFI-20	Experts have varying opinions on the Multidimensional Fatigue Inventory (MFI-20) in relation to its applicability to children. It is generally agreed that the MFI-20 is not completely suitable for children and young people, as it has only been validated in adults. Some experts find the item wording to be clear and easy to understand, while others believe the questions are too broad and may reflect symptoms other than fatigue.

Symptom Burden Questionnaire for Long COVID (Fatigue scale)	Experts suggest that the focus of the current version on adults raises issues concerning its applicability to younger populations, although it is noted to be adaptable to children in the future. There is ambiguity on how to define the intensity levels, such as mild, moderate, or severe. Yet, concerns are raised regarding its "7 day" time frame as an outcome measure given the fluctuating nature of Long COVID symptoms. Its lack of focus on function and Activities of Daily Living is another point of criticism.
Bell's Functionality Score	Experts do not consider Bell's Functionality Score to be appropriate for children and adolescents due to its complexity and difficulty for children to understand. They suggest that it might be more suitable for assessing physical functioning. The inclusion of work-related questions and irrelevant item wording also makes it unsuitable for the paediatric population. Overall, experts agree that Bell's Functionality Score is only suitable for adults and not children.

Outcome 5: Neuro-cognitive system functioning, symptoms, and conditions

Measurement instruments	Round of expert Delphi	Expert voting										
		1	2	3	4	5	6	7	8	9	10	11
Addenbrooke's Cognitive Examination (ACE-III)	Round 1	Maybe	Exclude	Maybe	Maybe	Exclude	Exclude	Exclude	Exclude	Exclude	Maybe	Exclude
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude
ASQ assessment (for infants born >29 weeks gestation)	Round 1	Exclude	Exclude	Exclude	Include	Include	Include	Maybe	Exclude	Exclude	Exclude	Include
	Round 2	Exclude	Exclude	Exclude	Exclude	Maybe	Maybe	Exclude	Exclude	Exclude	Exclude	Maybe
Bayley-IV neurological examination	Round 1	Maybe	Include	Exclude	Maybe	Exclude	Exclude	Include	Exclude	Exclude	Exclude	Include
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Maybe	Exclude	Exclude	Exclude	Maybe
Chalder fatigue scale	Round 1	Maybe	Maybe	Exclude	Exclude	Maybe	Maybe	Exclude	Maybe	Include	Include	Exclude
	Round 2	Maybe	Maybe	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Include	Include	Exclude
From Body Vigilance Scale (BVS)	Round 1	Exclude	Exclude	Exclude	Exclude	Exclude	Maybe	Exclude	Exclude	Maybe	Maybe	Maybe
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Maybe	Exclude	Exclude

Functional Independence measure (FIM)	Round 1	Exclude	Exclude	Maybe	Exclude	Exclude	Include	Exclude	Exclude	Include	Maybe	Maybe
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude
IQCODE	Round 1	Exclude	Exclude	Exclude	Exclude	Exclude	Maybe	Exclude	Exclude	Exclude	Exclude	Exclude
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude
Short Blessed Test	Round 1	Exclude	Exclude	Exclude	Exclude	Exclude	Include	Maybe	Exclude	Exclude	Exclude	Exclude
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude
PROMIS Pediatric Cognitive Function - Short Form 7a	Round 1	Include	Include	Include	Include	Include	Include	Include	Include	Maybe	Maybe	Include
	Round 2	Include	Include	Include	Include	Include	Include	Include	Include	Include	Include	Include
Vanderbilt ADHD assessment	Round 1	NEWLY SUGGESTED										
	Round 2	Maybe	Maybe	*Unvoted*	Exclude	Exclude	Include	Exclude	Exclude	Maybe	Exclude	*Unvoted*
Peds QL Cognitive Functioning Scale	Round 1	NEWLY SUGGESTED										
	Round 2	Maybe	Include	*Unvoted*	Maybe	Include	Include	Include	Maybe	Include	Maybe	*Unvoted*

Symptom Burden Questionnaire for Long COVID (Memory, Thinking & Communication scale, movement scale, muscles and joints, pain scales)	Round 1	NEWLY SUGGESTED										
	Round 2	Exclude	Maybe	Include	Exclude	Include	Exclude	Maybe	Include	Include	Maybe	*Unvoted*
NIH Toolbox	Round 1	NEWLY SUGGESTED										
	Round 2	Exclude	Include	*Unvoted*	Exclude	Maybe	Exclude	Exclude	Include	*Unvoted*	Exclude	*Unvoted*

Measurement instruments	**Summary of additional comments from the experts in rounds 1/2**
Addenbrooke's Cognitive Examination (ACE-III)	The experts generally agree that the Addenbrooke's Cognitive Examination (ACE-III) is a good instrument for assessing cognitive abilities in adults. However, they also highlight that it is not suitable for use with children, as it contains questions that are not developmentally appropriate. The length of the examination is also seen as a drawback. Another point of agreement among the experts is that the ACE-III is not feasible for use in a post or online format. Additionally, they believe that it may not be relevant for assessing cognitive abilities in long COVID patients.
ASQ assessment (for infants born >29 weeks gestation)	Experts have provided mixed opinions on the ASQ assessment for infants born >29 weeks gestation. Some experts feel that the assessment is very age-specific and should be completed at 2 years of age. They also find it unclear and not suitable for older children, and that it may not be fully related to long covid. However, other experts believe that the ASQ assessment is standard, well-validated, simple, and easy to use. They suggest using other instruments for older children and adolescents. Additionally, as the ASQ was developed specifically for children with prematurity, some experts feel it may be relevant to all paediatric populations.
Bayley-IV neurological examination	Experts have differing opinions on this instrument. Some find it to be a good tool, but note that it can be very lengthy. It is important to note that the BSID-IV is mostly representative of the U.S. population, which may impact its applicability in other countries. Additionally, the BSID-IV requires specific equipment and must be conducted by a healthcare professional with specific training. This may make it less feasible for certain settings or individuals. There are also concerns about the suitability of the BSID-IV for long COVID patients, as it is primarily focused on developmental achievements and may not be relevant to their specific needs.
Chalder fatigue scale	Overall, experts have mixed opinions on the Chalder fatigue scale. It is seen as more applicable in adults and less applicable in children. It is mainly used to assess fatigue and may not be as suitable for assessing neuro-cognitive abilities or sleep. There is some controversy surrounding its validation, particularly in children. Some experts suggest that it may be more relevant for assessing fatigue than neuro-cognitive complaints but problematic in young children. It is considered a simple and easy-to-complete scale, but it may not be suitable for younger children who are still developing language skills. Additionally, word finding difficulties should not be confused with pre-existing language disorders or other developmental difficulties.

From Body Vigilance Scale (BVS)	Experts have mixed opinions on the From Body Vigilance Scale (BVS). Some feel that the scale has too many questions and is too specific, making it less applicable in children. They also express doubts about its relevance in measuring sensitivity and awareness of internal sensations in young people and children. Furthermore, some experts question the feasibility and validation of the scale in children. However, others believe that the tool is complex and requires further exploration to understand its effectiveness. Additionally, experts warn about the importance of considering developmental differences when using the BVS.
Functional Independence measure (FIM)	Experts have mixed opinions on the Functional Independence Measure (FIM) as a quantitative tool in paediatric rehabilitation. While some experts believe it is only appropriate for adults and the elderly, others feel it is too specific to gastrointestinal issues and not appropriate for younger children or long COVID patients. It appears to be a clinician-reported measure and may not reflect a change in performance. However, some experts find it useful and applicable to the most severe patients, and suggest assessing if the questions are age appropriate.
IQCODE	The experts' opinions on the IQCODE suggest that it is not suitable for use with children. They believe it is more applicable for severe neurocognitive problems typically found in elderly individuals with dementia. The questionnaire's focus on comparing the current condition with that of 10 years ago is not considered appropriate for paediatric use.
Short Blessed Test	Overall, experts tend to agree that the Short Blessed Test may not be suitable for children and adolescents. It is not considered appropriate for individuals with cognitive impairments, intellectual disabilities, or severe cognitive impairment or language difficulties. It may also not capture the full range of cognitive abilities in children and adolescents. Some experts also mention that the test is primarily designed for assessing dementia in adults and may not be appropriate for children, especially when considering developmental considerations. Additionally, it is noted that the test cannot be done by post or online.
PROMIS Pediatric Cognitive Function - Short Form 7a	Experts largely hold a positive view on the Pediatric Cognitive Function - Short Form 7a scale. They appreciate its design, emphasising its appropriateness for paediatric patients, especially in identifying symptoms commonly reported. Its brevity is highly commended, making it a manageable tool for children to complete, although there is a noted limitation for its applicability primarily to older children. Overall, the consensus among professionals suggests that it is a short, precise, and appropriate tool for assessing cognitive function in children.
Vanderbilt ADHD assessment	The experts have varying opinions on the Vanderbilt ADHD assessment for assessing symptoms such as brain fog. Some find the inattention questions relevant and helpful, while others feel that the instrument is not relevant or necessary for all patients. It is noted that the assessment is ADHD-specific and not useful for exploring other possible symptoms or conditions. Additionally, some experts express concerns about the length and potential worry it may cause for patients and parents. Overall, the assessment is seen as more suitable for parents of children aged 8-12 years and may not be relevant for Long COVID patients.
Peds QL Cognitive Functioning Scale	Experts were generally positive in their views on the PedsQL Cognitive Functioning Scale. Some view it as overly lengthy, while others see it as a relatively concise and validated tool. These contrasting perspectives could stem from difficulty in accessing the entirety of the questions, an issue noted by a few of the experts. Despite this, some experts regard it as a potentially better option than PROMIS, acknowledging its routine use and age-appropriate design. The scale's appropriateness for long COVID symptoms has been mentioned as well, highlighting its potential application in ongoing pandemic-related research. Despite these differing opinions, the common thread seems to be an appreciation for the scale's validation and frequent use in practice.

Symptom Burden Questionnaire for Long COVID (Memory, Thinking & Communication scale, movement scale, muscles and joints, pain scales)	The questionnaire is yet to be validated for use in paediatric populations. The simplicity, feasibility, and relevance of the questionnaire to the Long COVID population have been noted positively, although questions about its validation persist.
NIH Toolbox	The purpose and usage of the NIH Toolbox are considered unclear by some experts. However, it is regarded as a comprehensive and beneficial tool for adults and children aged three and above. It offers normative data for children as young as three years old, yet some of the tests and instruments necessitate assessment by an examiner, which poses a limitation. Concerns also arise regarding the lack of validation for the youngest children and the potential resource and access issues associated with acquiring electronic versions, which can be expensive. Nonetheless, it is widely utilised and has been validated in diverse conditions.

Outcome 6: Physical functioning, symptoms, and conditions

Measurement instruments	Round of expert Delphi	Expert voting										
		1	2	3	4	5	6	7	8	9	10	11
Barthel Index	Round 1	Include	Maybe	Exclude	Maybe	Exclude	Include	Include	Include	Maybe	Exclude	Exclude
	Round 2	Maybe	Exclude	Exclude	Exclude	Exclude	Maybe	Exclude	Include	Exclude	Exclude	Exclude
Basic Activity of Daily Living (BADL)	Round 1	Exclude	Maybe	Exclude	Exclude	Exclude	Include	Exclude	Include	Exclude	*Unvoted*	Maybe
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Maybe	Exclude	Exclude
Clinical Frailty Scale (CFS)	Round 1	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Maybe	Maybe	Exclude	Exclude	Exclude
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude
Duke Activity Status Index (DASI)	Round 1	Exclude	Exclude	Exclude	Exclude	Exclude	Include	Maybe	Include	Maybe	Exclude	Exclude
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Maybe	Exclude	Exclude
EQ5DY instrument	Round 1	Exclude	Include	*Unvoted*	Include	Include	Include	Include	Exclude	Include	Include	Include
	Round 2	Maybe	Include	*Unvoted*	Include	Include	Include	Include	Exclude	Include	Include	Include

Fried Frailty phenotype	Round 1	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Maybe
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude
Functional Independence Measure (WeeFIM or FIM)	Round 1	Exclude	Exclude	Exclude	Exclude	Exclude	Include	Maybe	Exclude	Maybe	Maybe	Include
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Maybe
International Physical Activity Questionnaires Short Form (IPAQ-SF)	Round 1	Exclude	Exclude	Exclude	Exclude	Maybe	Include	Include	Maybe	Maybe	Maybe	Maybe
	Round 2	Exclude	Exclude	Exclude	Exclude	Maybe	Maybe	Maybe	Maybe	Exclude	Exclude	Maybe
Post COVID-19 Functional Status Scale (Scale 0-64 points)	Round 1	Exclude	Maybe	Include	Exclude	Exclude	Include	Maybe	Include	Maybe	Exclude	Exclude
	Round 2	Exclude	Exclude	Maybe	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude
PROMIS Early Childhood Parent Report Physical Activity 7a	Round 1	Exclude	Include	Maybe	Maybe	Include	Exclude	Include	Include	*Unvoted*	Exclude	Include
	Round 2	Maybe	Include	Maybe	Maybe	Include	Exclude	Include	Include	Maybe	Exclude	Include
PROMIS Pediatric Physical Activity – Short Form 8a	Round 1	Exclude	Include	Exclude	Maybe	Include	Maybe	Include	Include	*Unvoted*	Exclude	Include
	Round 2	Maybe	Include	Exclude	Maybe	Maybe	Include	Include	Include	Exclude	Maybe	Include

PROMIS Pediatric Physical Activity – Short Form 4a	Round 1	Exclude	Exclude	Exclude	Maybe	Maybe	Exclude	Include	Include	*Unvoted*	Exclude	Include
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Exclude	Include	Exclude	Exclude	Include
Bell's Functionality Score	Round 1	NEWLY SUGGESTED										
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Include	Exclude	Include	Maybe	Include	Maybe
The motor skills module activity questionnaire (MOMO) (Available in German only)	Round 1	NEWLY SUGGESTED										
	Round 2	Exclude	Maybe	Exclude	Exclude	*Unvoted*	Exclude	Exclude	Exclude	Exclude	Exclude	*Unvoted*
Symptom Burden Questionnaire for Long COVID (Impact on Daily Life Scale)	Round 1	NEWLY SUGGESTED										
	Round 2	Exclude	Maybe	Maybe	Exclude	Include	Exclude	Maybe	Exclude	Exclude	Maybe	*Unvoted*

Measurement instruments	**Summary of additional comments from the experts in rounds 1/2**
Barthel Index	The Barthel Index has received varied opinions from experts. While some experts believe that it can be adapted for use in children and appreciate its brevity and comprehensiveness, others argue that it is more suitable for older adults with severe dementia. It is particularly challenging to interpret the scale in younger children, but it might be possible for a family member to complete it on their behalf. Modifications may be necessary to make it more applicable for young children, including those experiencing Post COVID Condition (Long COVID). Additionally, the index primarily revolves around adults and assumes independence, which may not align with the developmental needs of younger children.
Basic Activity of Daily Living (BADL)	There are varying opinions among experts regarding the use of the Basic Activity of Daily Living (BADL) scale. Certain experts believe that it could be adapted for use with children, while others argue that the scale is more suitable for older adults. Furthermore, some experts express doubt about the practicality of using the BADL scale with children due to the need for direct observation.

Clinical Frailty Scale (CFS)	The Clinical Frailty Scale (CFS) is a tool that is primarily used in geriatrics and focused on older adults, particularly those with dementia. However, it is not appropriate for use with children or individuals with stable long-term disabilities or learning disabilities. Some experts suggest that the CFS may need tailoring for use with paediatric populations, as its rating system appears to be more focused on terminal illness.
Duke Activity Status Index (DASI)	Experts agree that the Duke Activity Status Index (DASI) is not fully suitable for the paediatric population as it is designed for adults and includes questions about sexual activity and work, which are not developmentally appropriate for children. However, some experts note that the DASI is short and may be used in the older population, but it has limited questions relevant to child daily functioning and may not capture all aspects of frailty in children.
EQ5DY instrument	The EQ5DY instrument is largely praised by experts as an effective, simple, and focused tool designed for the paediatric population. It is regarded as sufficient by itself, highlighted by its popularity and broad application in children's health economic analyses. The EQ5DY is noted for its validity in assessing children's health, with a specific proxy version available for young children. Its range of assessment is not limited to physical activity but extends to various facets like mobility, self-care, usual activities, and psychological states, such as feeling worried, sad, or unhappy. The tool's practicality and user-friendliness, particularly for children, are appreciated. A child-friendly version of the EQ-5D further underscores its suitability and adaptability for this demographic.
Fried Frailty phenotype	Experts agree that the Fried Frailty phenotype is suitable for adults, specifically focused on assessing frailty in older adults. However, it is not appropriate for children. The Fried Frailty phenotype was initially developed for cardiac patients and is mostly focused on the adult population. It requires tests and equipment that may not be feasible or suitable for use in children and adolescents. It is suggested that alternative measures, such as the Physical Activity Questionnaire for Children (PAQ-C) and Adolescents (PAQ-A), may be more useful in assessing frailty in this population.
Functional Independence Measure (WeeFIM or FIM)	Experts have expressed mixed opinions on the use of the Functional Independence Measure (WeeFIM or FIM) as a quantitative tool in paediatric rehabilitation. Some experts feel that the tool is too long and complicated, making it difficult to use in online or postal settings. They also believe that it may not be suitable for assessing self-care abilities in younger children and that the FIM may be too complicated for parents to understand and accurately complete.
International Physical Activity Questionnaires Short Form (IPAQ-SF)	Experts have varying opinions on the International Physical Activity Questionnaires Short Form (IPAQ-SF). Some experts feel that it may be too complex or detailed for younger children to accurately recall their activity levels. The consensus is that it is most applicable for older children and adolescents, with some experts suggesting it is suitable for older teens and potentially younger teens as well. However, there is concern that parents may struggle to estimate activity levels for younger children based on the wording of the questionnaire.
Post COVID-19 Functional Status Scale (Scale 0-64 points)	The experts had mixed opinions on the appropriateness of the Post COVID-19 Functional Status Scale for children. Some felt that it could be adapted for children and that it was the easiest and most appropriate option available. Others felt that the questions were too specific to adults and not suitable for children. Some experts mentioned that the scale was not developmentally appropriate for dependent children and that the responses were difficult to differentiate. Overall, the experts were unsure about its suitability for children.
PROMIS Early Childhood Parent Report Physical Activity 7a	General feedback is that an instrument has been explicitly designed for assessing physical activity in very young children. Experts note its potential utility, acknowledging it as a promising, albeit not flawless, tool for the younger age groups. Some have reservations, unsure if they would implement all three sections of the tool. Others advocate for its use in combination with additional assessments for older children. Overall, its focus on intense physical activity and its perceived suitability for the target population have garnered positive responses from the expert community.

PROMIS Pediatric Physical Activity – Short Form 8a	This tool has garnered a mix of opinions from experts. There is concern about the reference to strenuous exercises, as these could in themselves cause the symptoms being referred to, creating potential ambiguity. While there is some favour for the PROMIS scales, which offer a broad range of measurement, experts are hesitant to include all three due to potential interpretational issues among different populations. There is also concern about the 7-day recall period, as this might not capture the fluctuating nature of long COVID symptoms adequately, making it hard to observe systematic changes over time. There's a viewpoint that the scale may be more suited to older children. Some experts see the scale as a great tool for quantification, but there are also reservations regarding its relevance, with criticism that it might confuse strenuous physical activity with symptoms like sweating, which could cross into autonomic territory or yield false positives. Lastly, there's a note that the scale focuses predominantly on hard activity, which could be a limitation.
PROMIS Pediatric Physical Activity – Short Form 4a	The experts felt that the Pediatric Physical Activity - Short Form 4a is too short and less informative compared to other scales mentioned. They mentioned that it focuses on hard activity and does not consider post-exertional malaise. They also mentioned that for older children, a longer scale might be more appropriate.
Bell's Functionality Score	The experts generally agreed that Bell's Functionality Score is not suitable for children as it is primarily designed for adults. They also noted that the item wording is not relevant to the paediatric population. Some experts suggested that it could be adapted for children by replacing the reference to work with school. Overall, the experts felt that the scale is more appropriate for adults and would require modifications to be applicable to children.
The motor skills module activity questionnaire (MOMO) (Available in German only)	The experts had mixed opinions on The Motor Skills Module Activity Questionnaire (MOMO). Some found it to be an interesting tool that offers good coverage of both adults and children. However, others noted that it is only available in German, making it difficult for non-German speakers to evaluate. Additionally, some experts found it to be too long and challenging to complete, which may impact its feasibility in research studies or clinical settings. Overall, the lack of availability in additional languages was seen as a limitation of the questionnaire.
Symptom Burden Questionnaire for Long COVID (Impact on Daily Life Scale)	Experts have expressed mixed views about the given scale. The scale was originally developed for adults, and while some believe it's adaptable for children, others note that it would require modification and validation for paediatric populations. The clarity in defining degrees of severity, such as mild, moderate, and severe, was considered not easy to implement. Despite being in development, some experts appreciate the scale's design, finding it comprehensive and potentially superior to other outcome measures if certain sections were removed. However, they caution that it's not fully validated yet, and its reliance on a 7-day recall period might be insufficient given the fluctuating nature of many symptoms. It's also viewed as a feasible tool that captures relevant aspects of Long COVID, yet it notably lacks a focus on chest pain. Adaptation of this tool for younger people is currently underway.

Outcome 7: Work/occupational and study changes

Measurement instruments	Round of expert Delphi	Expert voting										
		1	2	3	4	5	6	7	8	9	10	11
None	Round 1	No scales/instruments reported in the reviewed evidence										

Work Productivity and Activity Impairment Questionnaire: General Health V2.0 (WPAI:GH)	Round 1	NEWLY SUGGESTED										
	Round 2	Exclude	Exclude	Exclude	Exclude	Exclude	Include	Maybe	Include	Exclude	Exclude	Maybe
Symptom Burden Questionnaire for Long COVID (Impact on Daily Life Scale)	Round 1	NEWLY SUGGESTED										
	Round 2	Exclude	Maybe	Include	Exclude	Include	Exclude	Maybe	Exclude	*Unvoted*	Maybe	*Unvoted*

Measurement instruments	**Summary of additional comments from the experts in rounds 1/2**
Work Productivity and Activity Impairment Questionnaire: General Health V2.0 (WPAI:GH)	Experts indicate that WPAI:GH instrument is focused on adults' activities and may not be appropriate for children. Some experts believe that the questions should be adapted to include school-related activities for paediatric use. However, others feel that with mild adaptations, the questionnaire can be used for older children. The ability of younger children to respond to some items requesting time estimates is questioned, indicating a need for more suitable questions for this age group.
Symptom Burden Questionnaire for Long COVID (Impact on Daily Life Scale)	Experts have expressed mixed views about the given scale. The scale was originally developed for adults, and while some believe it's adaptable for children, others note that it would require modification and validation for paediatric populations. The clarity in defining degrees of severity, such as mild, moderate, and severe, was considered not easy to implement. Despite being in development, some experts appreciate the scale's design, finding it comprehensive and potentially superior to other outcome measures if certain sections were removed. However, they caution that it's not fully validated yet, and its reliance on a 7-day recall period might be insufficient given the fluctuating nature of many symptoms. It's also viewed as a feasible tool that captures relevant aspects of Long COVID, yet it notably lacks a focus on chest pain. Adaptation of this tool for younger people is currently underway.

6. Consensus workshop participants

	Total number (%)[3]	**Voting participants (%)**
Healthcare professionals/Researchers	**29 (100)**	**22 (100)**
Delphi stakeholder group:		
- Health professional (including those who also do research)[1]	16 (55)	11 (50)
- Researcher (without any clinical patient care	13 (45)	11 (50)

duties)[2]		
Country of residence		
Australia	2 (7)	1 (4·5)
Chile	2 (7)	1 (4·5)
Germany	1 (3·4)	1 (4·5)
Israel	1 (3·4)	1 (4·5)
Italy	1 (3·4)	1 (4·5)
Lithuania	1 (3·4)	0 (0)
Latvia	1 (3·4)	1 (4·5)
Malaysia	1 (3·4)	1 (4·5)
Netherlands	1 (3·4)	1 (4·5)
Poland	1 (3·4)	1 (4·5)
Romania	2 (7)	2 (9)
Switzerland	1 (3·4)	0 (0)
UK	9 (31)	7 (32)
USA	5 (17)	4 (18)
Children and young people (≤18 years old) with Long COVID and their family and carers	**9 (100)**	**8 (100)**
Delphi stakeholder Group:		
- Family/caregivers of CYP with Long COVID	9 (100)	8 (100)
Country of residence		
Ireland	1 (11)	1 (13)
Netherlands	1 (11)	1 (13)
UK	6 (66·6)	5 (63)
USA	1 (11)	1 (13)

[1] Health professionals who care for people with Long COVID/post COVID-19 condition
[2] Researchers who undertake research in Long COVID/post COVID-19 condition
[3] One observer did not provide information on their stakeholder group and country of residence

7. Consensus workshop voting results

COS outcome	Outcome Measure	N (%) participants voting to INCLUDE in consensus meeting	Result

Cardiovascular functioning, symptoms and conditions	PedsQL Cardiac Module	16/28 (57)	Not included in the COMS
	Symptom Burden Questionnaire for Long COVID (Circulation scale)	7/27 (25)	Not included in the COMS
	Malmo POTS score (MAPS)	18/27 (64)	Not included in the COMS
Gastrointestinal functioning, symptoms, and conditions	PedsQL Gastrointestinal Symptoms Scales	23/26 (88)	Included in the COMS
	Questionnaire on Pediatric Gastrointestinal Symptoms (QPGS)	2/26 (8)	Not included in the COMS
	Symptom Burden Questionnaire for Long COVID (Stomach and Digestion Scale)	6/26 (23)	Not included in the COMS
Fatigue or Exhaustion	Chalder fatigue questionnaire	3/26 (12)	Not included in the COMS
	PROMIS Paediatric Fatigue	3/26 (12)	Not included in the COMS
	PedsQL Multidimensional Fatigue Scale	26/26 (100)	Included in the COMS
	Symptom Burden Questionnaire for Long COVID (Fatigue scale)	3/26 (12)	Not included in the COMS
Post-exertion symptoms	CDC symptom inventory for CFS	5/26 (19)	Not included in the COMS
	PEM items from DePaul Symptom Questionnaire	10/26 (38)	Not included in the COMS
	Symptom Burden Questionnaire for Long COVID (Fatigue scale)	6/26 (23)	Not included in the COMS
Neuro-cognitive system functioning, symptoms,	PROMIS Pediatric Cognitive Function - Short Form 7a	9/24 (36)	Not included in the COMS

and conditions	PedsQL Cognitive Functioning Scale	21/25 (84)	Included in the COMS
	Symptom Burden Questionnaire for Long COVID (Memory, Thinking & Communication scale, movement scale, muscles and joints, pain scales)	4/24 (16)	Not included in the COMS
Physical functioning, symptoms, and conditions	EQ5DY instrument	24/25 (96)	Included in the COMS
	PROMIS Physical Activity	2/25 (8)	Not included in the COMS
	Symptom Burden Questionnaire for Long COVID (Impact on Daily Life Scale)	3/25 (12)	Not included in the COMS
Work/occupational and study changes	Symptom Burden Questionnaire for Long COVID (Impact on Daily Life Scale)	5/22 (23)	Not included in the COMS
	WHO DAS 2 Children and Youth 36-Item Version	7/23 (30)	Not included in the COMS