

Large-scale psychometric assessment and validation of the Modified COVID-19 Yorkshire Rehabilitation Scale (C19-YRSm) patient-reported outcome measure for Long COVID or Post-COVID syndrome

Horton M¹, Smith AB¹, Milne R², Winch D³, Rayner C⁴, Halpin S¹, O'Connor R¹, Rocha Lawrence R⁵, Greenwood DC⁶, Barkley ND^{7,8}, Evans R⁹, Kwon J¹⁰, Dawes H¹¹, Wood C¹², Williams P¹³, Master H¹³, Mansoubi M¹¹, De Kock JH¹⁴, Mullard J¹⁵, Ormerod M¹⁶, Mir G¹⁷, Petrou S¹⁰, O'Connor DB¹⁸, LOCOMOTION Consortium, Sivan M^{1,19,20}.

Affiliations:

¹Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK

²University of Southampton, Southampton, UK

³Kent & Medway ICB Patient Reference Group, Kent, UK

⁴University of Leeds LOCOMOTION Patient Advisory Group, UK

⁵ELAROS 24/7 Ltd., UK

⁶Leeds Institute for Data Analytics, University of Leeds, Leeds, UK.

⁷Northern Care Alliance, UK

⁸Manchester Metropolitan University, Manchester, UK

⁹Department of Respiratory Sciences, University of Leicester, UK

¹⁰Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford,

¹¹Medical School, NIHR Exeter Biomedical Research Centre, Faculty of Health and Life Sciences, University of Exeter, Exeter, UK

¹²Birmingham Community Healthcare NHS Foundation Trust, UK

¹³Hertfordshire Community NHS Trust, UK

¹⁴NHS Highland COVID Recovery Service, North West University, Faculty of Health Sciences, UK

¹⁵Faculty of Medical Sciences, Newcastle University, UK

¹⁶Long Covid Support (UK Registered Charity: 1198938)

¹⁵School of Medicine, University of Leeds, UK

¹⁸School of Psychology, University of Leeds, UK

¹⁹COVID Rehabilitation Service, Leeds Community Healthcare NHS Trust, Leeds, UK.

²⁰National Demonstration Centre of Rehabilitation Medicine, Leeds Teaching Hospitals NHS Trust, Leeds, UK.

Corresponding author:

Dr Adam B. Smith, Ph.D.

Academic Department of Rehabilitation Medicine

Leeds Institute of Rheumatic and Musculoskeletal Medicine

University of Leeds

E-mail address: a.b.smith@leeds.ac.uk

Manuscript word count

3475

Tables

5

Figures

1

Abstract

Background: The C19-YRS was the first condition-specific for Long COVID/Post-COVID syndrome. Although the original C19-YRS evolved to the modified version (C19-YRSm) based on psychometric evidence, clinical content relevance and feedback from patients and healthcare professionals, it has not been validated through Rasch analysis.

Objectives: To psychometrically assess and validate the C19-YRSm using newly collected data from a large-scale, multi-centre study (LOCOMOTION).

Methods: 1278 patients (67% Female; mean age = 48.6, SD 12.7) digitally completed the C19-YRSm. The psychometric properties of the C19-YRSm Symptom Severity (SS) and Functional Disability (FD) subscales were assessed using a Rasch Measurement Theory framework, assessing for individual item model fit, targeting, internal consistency reliability, unidimensionality, local dependency (LD), response category functioning and differential item functioning (DIF) by age group, sex and ethnicity.

Results: Rasch analysis revealed robust psychometric properties of both subscales, with each demonstrating unidimensionality, appropriate response category structuring, no floor or ceiling effects, and minimal LD and DIF. Both subscales also displayed good targeting and reliability (SS: Person Separation Index (PSI)=0.81, Cronbach's alpha=0.82; FD: PSI=0.76, Cronbach's alpha=0.81).

Conclusion: Although some minor anomalies are apparent, the modifications to the original C19-YRS have strengthened its measurement characteristics, and its clinical and conceptual relevance.

Keywords:

Post-COVID Syndrome, Long COVID, C19-YRSm, patient-reported outcome

Introduction

Long Covid (LC) or Post COVID-19 Syndrome (PCS) is a multisystem clinical syndrome where symptoms persist for more than three months after acute infection with SARS-CoV-2.¹ The prevalence of the condition is estimated to be approximately 3.3% (2 million cases) in the UK alone² and up to 36% worldwide.³

Common symptoms include fatigue, shortness of breath, cognitive impairment, muscle and joint pain, chest pain, palpitations, persistent loss of smell and taste, gastro-intestinal upset, and headache.⁴ Symptoms may fluctuate⁵ and approximately 20% of those suffering with LC describe their symptoms as severe.² For some individuals, symptoms may persist for more than 4 years (persistent LC) following the initial COVID-19 infection.⁶ This protracted course of LC leads to a significant compromise on individuals' ability to work and conduct day-to-day tasks and can result in severely reduced health-related quality of life.⁷⁻⁹

Measuring symptom burden, functional disability (or ability) and health-related quality of life (HRQoL) through patient-reported outcome measures (PROMs) is therefore crucial to understand the impacts on health, condition trajectories and the cost-effectiveness of interventions. There have been a number of condition-specific PROMs developed for the condition during or after the pandemic¹⁰⁻¹².

The COVID-19 Yorkshire Rehabilitation Scale (C19-YRS) is the first LC condition-specific PROM reported in the literature¹³. The scale was developed to cover all of the 2001 WHO International Classification of Functioning, Disability and Health (ICF) framework¹⁴ and was designed to capture the symptoms and functional problems associated with LC¹³. The instrument has been widely employed in primary care and community settings,¹⁵⁻¹⁷ rehabilitation interventions,¹⁸ and post-COVID epidemiological studies.^{19,20} Following an initial psychometric analysis of the C19-YRS,²¹ psychometric and clinical evidence, and feedback from patients and healthcare professionals was subsequently integrated, culminating in a modified version of the instrument (C19-YRSm).²² The C19-YRSm has since undergone further classical psychometric validation and has been shown, for instance, to have good internal reliability and convergent validity in a Croatian patient population.²³ More recent validation has demonstrated the C19-YRSm to have good psychometric properties,²⁴ in terms of internal consistency and test-retest reliability, as well as discriminant and convergent validity. Factor analysis supported the instrument's factor structure. Furthermore, an exploratory minimal important difference (MID) and minimal clinically important difference (MCID) were determined for the subscale scores.²⁴

The psychometric properties of the modified C19-YRSm have not yet been assessed using Rasch measurement theory (RMT). In contrast to classical psychometrics, which are focused on the test-level, i.e., the instrument as a whole, RMT allows for item-level analysis. This enables the identification of individual PROM items that may potentially require modification (or removal) in order to improve the measurement properties of a PROM. The aim of this study was therefore to use Rasch measurement methodology to psychometrically assess and further validate the C19-YRSm using data collected from a large-scale, multi-centre study (LOng COvid Multidisciplinary consortium Optimising Treatments and services, LOCOMOTION).²⁵

Materials and Methods

C19-YRSm

The C19-YRSm²² consists of four separate subscales: Symptom Severity (SS), Functional Disability (FD), Other Symptoms (OS), and Overall Health (OH). The OH subscale is a single item, scored on a 0-10 numeric rating subscale, with a score of 0 representing “worst health” and 10 being “best health”. Given the OH is a single-item subscale it cannot be analysed using the Rasch model. The OS subscale consists of a checklist of 25 additional symptoms, where respondents select the symptoms that they have experienced over the last 7 days based on yes/no options. The analysis of the OS subscale is not presented here.

For the remaining subscales, both the SS (26 items condensed to 10 core items, see below) and FD (5 items) are summed individually to form total scores for each subscale. All items on the SS and FD subscales are scored on a 4-point subscale (0=No problem; 1=Mild problem; 2=Moderate problem; 3=Severe problem) where a higher score represents a higher severity of the problem, i.e. worse symptoms or worse functional disability.

Some of the SS subscale core items are grouped within subsets. For these items, the maximum value observed within the subset is selected as the representative value. This scoring step is taken due to the inherent (local) dependency between the items within a subset, as observed during the modification of the original C19-YRS.²⁰ Taking the maximum score from within a set avoids local dependency, whilst maintaining the clinical utility of the individual component items. The sections concerning breathlessness, cough/throat sensitivity, smell/taste, pain/discomfort, cognition, palpitation/dizziness, and anxiety/mood each have a maximum score that is taken from across multiple items in the section.

Data Collection

Data collection was carried out as part of the LOCOMOTION study²⁵ (NIHR Ref: COV-LT2-0016), with the C19-YRSm data collected routinely within 10 participating Long Covid services across the UK between December 2021 and October 2023. Participants with a clinical diagnosis of LC by a qualified healthcare professional were eligible for inclusion in one of the 10 participating centres. Participants had to meet the UK National Institute for Health and Care Excellence (NICE) case definition, i.e., one or more persistent symptoms developed during or post-infection that are consistent with COVID-19 and not explained by alternative diagnoses.²⁶ Consent and clinical data were collected using the ELAROS digital patient-reported outcome measures platform.²⁷ Ethics approval for the LOCOMOTION study was obtained from the Bradford and Leeds Research Ethics Committee on behalf of Health Research Authority and Health and Care Research Wales (reference: 21/YH/0276).

Rasch analysis

Rasch analysis of the data²⁶ data²⁸ was completed with RUMM2030 software,²⁹ and carried out separately for the SS subscale (10 items) and the FD subscale (5 items). Key criteria for the RMT are 1) Unidimensionality - whether the items represent a single factor, 2) Item fit - whether the items fit the Rasch model, 3) Local Dependence - the absence of any further association between items beyond that explained by the underlying trait, 4) Response category functioning (or threshold disordering) - requiring the latent trait to increase monotonically across item response categories and 5) Item invariance (or absence of item bias or differential item functioning (DIF)) - requiring item properties to be invariant to subgroup characteristics (e.g., gender, ethnicity) where latent trait levels are equivalent.

1. Unidimensionality was evaluated by a series of t-tests,³⁰ with multidimensionality indicated if independent subsets of items delivered significantly different person estimates, and the lower bound 95% CI percentage of significantly different t-tests was $> 5\%$.
2. The Rasch analytic process included several standard tests of fit, covering both the overall subscale and item-level fit. All items were assessed individually for fit to the Rasch model (Partial Credit Model)³¹ within the subscale item set to assess whether each item contributes to the underlying construct. Item misfit was indicated where the Bonferroni-adjusted chi-squared p-value was statistically significant for an item and the standardised (z-score) fit-residuals fell outside ± 2.5 .^{32,33}
3. Tests of local dependency (LD) were carried out to determine whether any items in the subscale were more closely related than is explained by the underlying construct; LD was indicated using a residual correlation (Q3 value) criterion cut point of 0.2 above average residual correlation.³⁴
4. Response category functioning was assessed to determine whether the response structure of the items was operating in the intended manner. A functional 0-3 response category structure for each item would be indicated by sequential response thresholds (the crossover points between subsequent response categories) on the underlying (logit) subscale.³⁵
5. Item bias was assessed through uniform and non-uniform differential item functioning (DIF) testing by sex (male/female), age group (16-49; 50+ years) and ethnicity, with significant DIF indicated at a Bonferroni-adjusted ANOVA p-value.

Furthermore, reliability indices were taken as the person separation index (PSI), and the Cronbach's alpha values, and the scale-sample targeting was assessed graphically through the relative distribution of item and person locations, along with the calculation of floor and ceiling effects. When the Rasch model assumptions are satisfied, the sufficiency of the raw score allows for the transformation into a linear, interval-level transformation.³³

Cross-validation

In order to assess the replication of results across independent samples, the complete sample was randomly split into three equally-sized subsamples which were examined separately. This strengthens the analysis through replication, whilst avoiding the over-powering of RUMM fit statistics and misinterpretation that can occur with sample sizes > 500 .^{36,37} The subsamples were used to assess Rasch-based individual item fit and DIF. However, the complete sample was utilised to assess response category functioning, targeting, local dependency and the reliability indices, as these tests operate better with the precision afforded by larger sample sizes. To allow for brevity of reporting, only the results of the first subsample are presented within the manuscript.

Working group/Patient Advisory Group

All empirical results were reported back to a working group made up of clinicians, patients, and social scientists with expertise in PROMs and psychometrics, and additionally to the wider LOCOMOTION team, for sense-checking from both the patient and clinical context. Results were also reported back to a patient advisory group (PAG), and any potential further modifications were discussed within the working group and the PAG, with an emphasis on the practical implications of any potential change in the instrument.

Results

Sample

Data from 1278 patients were included in the study. The mean age was 48.6 (standard deviation, SD: 12.7) years, predominantly female (67%) and White (79%). The mean time since infection was 418 (SD: 268.7) days. Levels of pre-COVID co-morbidity were low with the most commonly reported being mental health (17%) and respiratory conditions (10%). Vaccination status was recorded for only 45% of the sample. The key demographic characteristics of the sample are presented in Table 1. The data for the presented cross-validation subsample were from 423 patients.

Rasch Analysis

Symptom Severity Subscale: Unidimensionality, reliability, floor/ceiling and targeting

A summary of the psychometric properties of both subscales are presented in Table 2, including results for the full sample and for the cross-validation subsample. The results indicated that the SS subscale was unidimensional, although the percentage of significant t-tests fell just outside the lower bound 95% CI, 5% criterion (5.1%) on the full sample. A good level of internal-consistency reliability (0.82) was displayed, with no floor or ceiling effect (see Table 2, and Figure 1).

Symptom Severity Subscale: Local Dependency

The average residual correlation was -0.10 and therefore the criterion value to indicate LD was taken as 0.10 (-0.10 + 0.2). Two local dependencies out of the 45 (4%) pairs were identified (Table 3). The largest LD was observed between the “Fatigue” and “Post-exertional malaise” items (0.17); the other between the “Breathlessness” and “Cough” (0.13). Both of these dependencies appear to be conceptually logical, suggesting that these findings are real, and not just a chance finding.

Symptom Severity Subscale: Individual Item Fit

Most items were within the acceptable fit ranges, although misfit was identified for some items (Table 3). The “Cough” item (covering “cough” and “throat sensitivity”) displayed the largest chi-square and fit residual misfit anomalies, with the high positive fit residual value indicating an underdiscrimination. The “Smell/Taste” and “Cognition” items also demonstrated some misfit although this was borderline in both instances, and inconsistent among the different subsamples.

Symptom Severity Subscale: Category Response structure

The modified 4-response category format mostly displayed an ordered, functional response structure across all items, evidencing a marked improvement from the response functioning of the original C19-YRS. However, the ‘Smell/Taste’ item was consistently disordered among the full sample and all subsamples, with a non-borderline response structure suggesting that a binary response format may be more appropriate. Additionally, the full sample displayed three further items as borderline disordered (‘Fatigue’, ‘PEM’ and ‘Sleep’), with these same items either ordered or borderline disordered among the three smaller subsamples, suggesting that this is not problematic.

Symptom Severity Subscale: Item location ordering (easiest/most difficult items to affirm)

The item location ordering can be observed in Table 3. The “Fatigue” item marked the lowest location on the subscale, meaning that fatigue is observed as the most frequently problematic issue on the SS subscale, i.e., most easily endorsed item by people with LC. Conversely, the “Smell/Taste” item had the highest location on the subscale, meaning that smell/taste is observed as the least frequently problematic issue on the SS subscale.

Symptom Severity Subscale: Differential Item Functioning

No items displayed any consistent significant DIF by sex, age group or ethnicity grouping. However, the limited sample for minority ethnic groups was insufficient to make the finding on ethnicity reliable.

Symptom Severity Subscale: Post hoc analysis addressing the issues found

In order to determine whether amendments could be made to address and thereby potentially resolve the individual item issues that had been identified, the first subsample (N=423) was taken as an experimental dataset and a number of analysis iterations were run.

The first analysis focused on retaining all items in the SS subscale and involved rescoreing the “Smell/Taste” and “Cough” items into a binary response format. This resolved the associated response structure and item fit issues. Furthermore, the “Breathlessness” and “Cough” items were subtested, i.e., added together into a single item, rather than contributing as two separate items, to account for local dependency. This resolved the dependency issues, although the subtested item displayed some misfit (Fit Residual: 3.8).

The second analysis focused on model fit and involved removing “Cough” whilst retaining the rescored (binary) “Smell/Taste” item. This resolved all fit and dependency issues, although the PEM item continued to display a borderline disordered response structure.

Finally, in order to examine the impact of these amendments on person estimates (“scores”), the person estimates from the complete subscale (full sample) were correlated against the full-sample person estimates from both analyses. This indicated that there was strong (Spearman’s) correlation between the complete subscale person estimates and both post-hoc analyses estimates (0.99 and 0.98, respectively), indicating that the subscale amendments have very little effect on person ordering.

Taken together, this suggests that, despite the issues identified, it is perhaps optimal to retain the complete SS subscale in its original format, in order to retain maximum information and allow for continuity of data collection and comparison. Given that Rasch model assumptions have been satisfied, the transformation of the raw ordinal scale scores into interval-level equivalent scores is appropriate, and these transformed scores are available in Table 5. Please note that this transformation is only valid for complete data, where all items have been included in the total score.

Functional Disability subscale: Unidimensionality, reliability, floor/ceiling and targeting

A summary of the psychometric properties for the FD subscale is presented in Table 2. The FD subscale was unidimensional (only 2.6% of unidimensionality t-tests were statistically significant), displayed a good level of internal-consistency reliability (0.82), and had good subscale-sample targeting with no floor or ceiling effect (see Table 2 and Figure 1).

Functional Disability subscale: Local Dependency

The average correlation was -0.23 and therefore the criterion value to indicate LD was taken as -0.03 (-0.23 + 0.2). One local dependency (Table 4) was observed between the “Walking or moving around” and “Personal care” items. Again, there appears to be a conceptual connection between these two items, suggesting that this is a real dependency rather than a chance finding.

Functional Disability subscale: Individual Item Fit

The “Communication” item consistently displayed a fit residual misfit anomaly with the high positive fit residual value indicating an underdiscrimination (Table 4). No other items indicated any evidence of misfit.

Functional Disability subscale: Category Response structure

The response categories displayed an ordered, functional response structure across all items. None of the items were disordered.

Functional Disability subscale: Item location ordering (easiest/most difficult items to affirm)

The “Other activities of daily living” item marked the lowest location on the subscale, i.e., the most frequently problematic (or easily endorsed) issue on the FD subscale. Conversely, the “Personal care” item had the highest location on the subscale, representing the least frequently problematic issue on the FD subscale. The item location ordering can be observed in Table 4.

Functional Disability subscale: Differential Item Functioning

There was no significant DIF by sex, age group or ethnicity grouping indicated for any of the items across any of the samples. As highlighted earlier, however, the ethnic minority sample was insufficiently powered to make the finding on ethnicity reliable. DIF by age group was observed for the “Walking or moving around” item, this was, however, consistent with expectation, as the older group are more likely to report issues with walking.

Functional Disability Subscale: Post hoc analysis addressing the issues found

Although there are relatively few issues found in the FD subscale, the first subsample (N=423) was again taken as an experimental dataset and amendments were made to address the issues that had been identified. This involved sub-testing the “Walking or moving around” and “Personal care” items in order to resolve the dependency issues, although this resulted in a further borderline dependency between the “Other activities of daily living” and “Social role” in the full sample. However, no further sub-testing was carried out, due to the borderline nature of the apparent dependency.

In order to examine the impact of this amendment on person estimates, the person estimates from the complete scale (full sample) were correlated against the full-sample person estimates from the resolved analysis. This indicated a very strong (Spearman’s) correlation of 0.999, indicating that the subscale amendment had very little effect on the ordering of persons. Again, this suggests that the parsimonious retention of the complete FD subscale in its original format would retain maximum information and allow for continuity of data collection and comparison.

Again, given that Rasch model assumptions have been satisfied, the transformation of the raw ordinal scale scores into interval-level equivalent scores is appropriate, and these transformed scores are available in Table 5 (for complete data).

Working group/PAG

A benefit of the C19-YRSm is that it is a relatively short measure with a simple response structure. The feedback from the PAG suggested that it was not burdensome for patients to complete, and that the simplified response format was more appropriate than the previous 11-point numeric rating scale on the original C19-YRS. The PAG suggested that the C19-YRSm was comprehensible, easy to understand, and that the range of symptoms covered across the SS and OS scales (not presented here) was broadly comprehensive, whilst remaining manageable.

Discussion

This study provides the first Rasch analysis of the latest version of the C19-YRSm following the initial development^{13,21} and subsequent psychometric analysis.²² In line with the classical psychometric analysis,²⁴ the results demonstrated evidence of a two-factor structure, comprising unidimensional Symptom Severity and Functional Disability subscales, with both displaying good internal reliability.

Within the subscale analysis, the few item anomalies observed concerned the “Cough” and “Smell/Taste” items, both within the Symptom Severity subscale. It is uncertain why these items were inconsistent with the other subscale items, but there are some potential explanations. For instance, these items also marked the ‘most difficult’ end of the subscale, meaning that they were generally reported to be problematic less frequently than the other items. This positioning means that there is less certainty in respect of the item fit characteristics of the items (given that less sample information is available). Furthermore, it also means that these items are an important demarcation of the upper measurement range of the subscale. The post-hoc removal and amendment of these items had very little effect on the ordering of the person estimates that were generated, therefore the added clinical and measurement information provided by the retention of these items would seem to outweigh any potential improvement in subscale fit.³⁷

Although the cough, throat sensitivity and anosmia items were recognised by the Working Group as common and important symptoms of Covid, the analysis results indicate that their contribution to the impact of Long Covid on a person’s daily life is less clear. It is also possible that a binary (no problem/problem) response format may be more appropriate for these items than the 4-point response structure, especially for the “Smell/Taste” item. Feedback from the Working Group suggested that a binary response would align with the manifestation of these symptoms that present in LC/PCS clinics, especially for the “Smell/Taste” item.

The input from the Working Group / PAG also suggested that the C19-YRSm was not burdensome for patients to complete, and that the simplified response format was more appropriate than the previous 11-point numeric rating subscale. The benefit of the C19-YRSm is that it is a relatively short measure with a simple response structure. If outcome measures are to be repeatedly used in clinical or research settings, they should be relevant, clinically useful, and non-burdensome to patients. Long PROMs may lead to questionnaire response burden, which is recognised as a threat to subscale completion and adherence in trials.³⁸

This study is the first, to the authors knowledge, to investigate item bias or DIF for the C19-YRSm and demonstrated little or no DIF across age groups and gender, thus reinforcing the instrument usability across a wide population of people living with LC. However, one potential limitation to be noted here is that minority ethnic groups were considerably underrepresented within our patient sample. Despite the fact there was no evident item bias, there is still a need for further research that is sufficiently powered to explore any ethnic variation in LC symptoms. Furthermore, although a recent systematic review³⁹ determined that the content validity of the C19-YRSm was sufficient, there is, in general, further research required involving additional input from people with LC as part of the instrument’s ongoing validation process.

Nevertheless, the C19-YRS scored well as one of the suggested LC condition-specific instruments, covering a number of the core outcomes identified in an international consensus study⁴⁰ when evaluated against reporting a core outcome set of 12 outcomes that should be measured in all future clinical studies and in clinical care for people with LC⁴¹. In addition to the current study, previous validation studies^{22,24}

have shown the C19-YRSm to have robust psychometric properties and this is reflected in the growing evidence supporting the clinical utility of the C19-YRSm.^{42,43}

Conclusion

The modified C19-YRSm has been demonstrated to have significant advantages over the original C19-YRS. The content coverage is much improved, including a number of common symptoms that were not included in the original version. This study and preceding validation studies have shown the Symptom Severity and Functional Disability subscales to be far more robust than the original C19-YRS, thereby strengthening the measurement characteristics of the C19-YRSm and enhancing its clinical and conceptual relevance for use in both clinical and community settings. For research purposes, an interval-level score transformation is available, allowing for the calculation of parametric statistics on C19-YRSm scores.

Appendix

The C19-YRSm is available via the University of Leeds licensing platform, here:
<https://licensing.leeds.ac.uk/product/c19-yrs-covid-19-yorkshire-rehabilitation-subscale>

Declarations

Ethics approval and consent to participate

Ethics approval for the LOCOMOTION study was obtained from the Bradford and Leeds Research Ethics Committee on behalf of Health Research Authority and Health and Care Research Wales (reference: 21/YH/0276; Trial registration number NCT05057260, ISRCTN15022307). All participants consented for their data to be used for evaluation and research purposes.

Data availability statement

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Conflicts of interest statement

The authors declare that they have no competing interests.

Author contribution statement

MH, SP, RJOC, HD, RM, SH, GM, CR, DOC, DCG, NDB, MS were responsible for the funding acquisition. MH was additionally responsible for the data analysis and methodology; DCG and ABS additionally contributed to the methodology. All authors contributed to the drafting, review and editing of the manuscript.

Funding statement

The study was funded through a National Institute for Health and Care Research award (COV-LT2-0016). The views expressed in this publication are those of the authors and not necessarily those of NIHR or the Department of Health and Social Care.

Permission to reproduce material from other sources

Not applicable

Clinical trial registration

Not applicable

Acknowledgements

We thank all the participants of this study for their valuable time and feedback on the instrument. The authors are also grateful to all the participating NHS sites and the ELAROS 24/7 digital company for their PROMs platform.

Tables and Figures

Table 1 – Demographic characteristics of sample

	All* (T1)	Non hospitalised	Hospitalised
	1278	1157	120
Female n (%)	856 (67%)	779 (67.3%)	76 (63.3%)
Mean age in years (SD)	48.6 (12.7)	48.2 (12.6)	52.5 (13.2)
Mean weight in kg (SD) (n=249)	82.9 (22.5)	81.5 (22.1) (n=224)	94.9 (24.2) (n=24)
Ethnicity n (%)			
<i>White</i>	1006 (78.7%)	906 (78.3%)	99 (82.5%)
<i>Black</i>	23 (1.8%)	19 (1.6%)	4 (3.3%)
<i>Asian</i>	72 (5.6%)	61 (5.3%)	11 (9.2%)
<i>Mixed/Other</i>	43 (3.3%)	41 (3.5%)	2 (1.7%)
<i>Missing</i>	134 (10.5%)	130 (11.2%)	4 (3.3%)
Smoking status n (%)			
<i>Never smoked</i>	316 (24.7%)	285 (24.7%)	31 (25.9%)
<i>Current smoker</i>	27 (2%)	25 (2.1%)	2 (1.6%)
<i>Ex-smoker</i>	123 (9.6%)	105 (9.1%)	17 (14.2%)
<i>Missing</i>	812 (63.5%)	742 (64.1%)	70 (58.3%)
Admitted to hospital n (%)	120 (9.4%)	0 (0%)	120 (100%)
Mean days since infection (SD)*	418.1 (268.7)	408.3 (268.4)	484.9 (257.7)
Vaccination status n (%)	465 (46%)	418 (45%)	46 (46%)
Pre-COVID co-morbidity n (%)			
<i>Mental health conditions</i>	177 (17%)	157 (17%)	19 (19%)
<i>Respiratory</i>	101 (10%)	88 (10%)	13 (13%)
<i>Cardiovascular</i>	29 (3%)	24 (3%)	5 (5%)
<i>Diabetes</i>	32 (3%)	25 (3%)	7 (7%)

*n = 1021

Table 2 – Overview of psychometric properties of the Symptom Severity (SS) and Functional Disability (FD) subscales

	Symptom Severity (SS)		Functional Disability (FD)		
	Full Sample	Sample A	Full Sample	Sample A	Target Values
total n - (extremes) = valid n	1268 - (11) = 1257	423 - (3) = 420	1268 - (63) = 1205	423 - (17) = 406	
Number of items	10	10	5	5	
Overall Scale Fit (Chi-square p)	p<0.001	p<0.001	p<0.001	p=0.026	p>0.01
Individual Item Fit Residuals	6/10 out of range	2/10 out of range	2/5 out of range	1/5 out of range	within +/-2.5
Individual Item Fit (Chi-square p)	7/10 out of range	4/10 out of range	3/5 out of range	All in range	p>0.05 (Bonferroni adj)
Person Fit Residuals	1.4% outside range	1.2% outside range	1.3% outside range	1.2% outside range	within +/-3.0
Unidimensionality	6.28% (5.1%)	6.43% (4.3%)	2.66%	2.46%	% of significant t-tests <5%
Local Dependency	2/45 pairwise significant	2/45 pairwise significant	1/10 pairwise significant	1/10 pairwise significant	<0.2 above average Q3
Targeting	good alignment	good alignment	good alignment	good alignment	aligned
% sample at floor (min score)	0.0%	0.0%	3.0%	3.3%	
% sample at ceiling (max score)	0.9%	0.7%	2.0%	0.7%	
% sample at floor and ceiling (combined)	0.9%	0.7%	5.0%	4.0%	<15%
Person Separation Index (PSI)	0.81	0.82	0.76	0.77	>0.85
Cronbach's Alpha	0.82	0.84	0.81	0.81	>0.85
Response category threshold ordering	4/10 disordered	2/10 disordered	All ordered	3/5 disordered	All ordered
DIF-by-Sex	No DIF	No DIF	1/5 display DIF (nu)	1/5 display DIF (nu)	p>0.05 (Bonferroni adj)
DIF-by-Age Group	3/10 display DIF	No DIF	3/5 display DIF	No DIF	p>0.05 (Bonferroni adj)
DIF-by-Ethnicity	No DIF	No DIF	No DIF	No DIF	p>0.05 (Bonferroni adj)

Table 3 – Individual item fit and local dependency of the Symptom Severity (SS) subscale

Item	Statement	Item order and fit				Local dependency Q3 correlations									
		Logit Location	SE	Fit Residual	Chi-Square p	SS1	SS2	SS3	SS4	SS5	SS6	SS7	SS8	SS9	SS10
SS1	Breathlessness	-0.16	0.039	1.593	0.286										
SS2	Cough	0.883	0.035	4.961	0.000	0.134									
SS3	Fatigue	-1.054	0.045	-2.217	0.000	-0.051	-0.249								
SS4	Smell & Taste	1.283	0.035	2.468	0.002	-0.108	-0.005	-0.232							
SS5	Pain	-0.359	0.038	-1.51	0.022	-0.084	-0.129	0.013	-0.186						
SS6	Cognition	-0.375	0.039	-2.598	0.002	-0.255	-0.265	0.031	-0.122	-0.066					
SS7	Palpitations & Dizziness	0.497	0.036	0.795	0.098	-0.113	-0.194	-0.118	-0.156	-0.019	-0.036				
SS8	PE Malaise	-0.283	0.035	1.209	0.750	-0.158	-0.235	0.166	-0.254	-0.137	0.044	-0.074			
SS9	Anxiety & Mood	-0.205	0.038	-0.592	0.067	-0.168	-0.162	-0.087	-0.142	-0.081	0.026	-0.118	-0.086		
SS10	Sleep	-0.226	0.036	-0.281	0.359	-0.192	-0.204	-0.057	-0.074	-0.034	-0.051	-0.114	-0.118	-0.024	
		average residual correlation				-0.101									
		average+0.2 (criterion)				0.099									

*Please note that all values are based on the complete sample, except for the Fit Residual and Chi-square p-value, which are based on the first subsample (n=423)
Highlighted items indicate some form of misfit

Table 4 – Individual item fit and local dependency of the Functional Disability (FD) subscale

Item	Statement	Item order and fit				Local dependency Q3 correlations				
		Logit Location	SE	Fit Residual	Chi-Square p	FD1	FD2	FD3	FD4	FD5
FD1	Communication	0.035	0.039	3.54	0.010					
FD2	Walking	-0.045	0.039	0.207	0.572	-0.408				
FD3	Personal Care	1.459	0.044	-0.672	0.599	-0.384	0.017			
FD4	Daily Living	-1.366	0.043	-2.416	0.162	-0.357	-0.134	-0.112		
FD5	Social	-0.083	0.038	-0.786	0.451	-0.133	-0.394	-0.263	-0.112	
		average residual correlation				-0.228				
		average+0.2 (criterion)				-0.028				

*Please note that all values are based on the complete sample, except for the Fit Residual and Chi-square p-value, which are based on the first subsample (n=423)
 Highlighted items indicate some form of misfit

Table 5 - The raw-score to interval-level equivalent transformed score for the Symptom Severity (SS) and Functional Disability (FD) subscales.

*Please note that this conversion is only valid in the case of complete data

Raw score	Interval transformations	
	SS (0-30)	FD (0-15)
0	0.00	0.00
1	2.80	1.84
2	4.74	3.19
3	6.07	4.17
4	7.13	4.97
5	8.03	5.67
6	8.81	6.32
7	9.53	6.96
8	10.19	7.61
9	10.82	8.28
10	11.41	8.98
11	11.99	9.72
12	12.57	10.55
13	13.14	11.56
14	13.71	12.99
15	14.29	15.00
16	14.87	
17	15.47	
18	16.09	
19	16.72	
20	17.36	
21	18.02	
22	18.70	
23	19.40	
24	20.16	
25	20.98	
26	21.92	
27	23.06	
28	24.55	
29	26.75	
30	30.00	

Figure 1 - The relative scale-to-sample targeting of the Symptom Severity (SS) and Functional Disability (FD) subscales.

Key: The (logit) location distribution of the sample is plotted above the x-axis, and the (logit) location distribution of the scale items is plotted below the x-axis. Here, we can see a slight right-skew for the SS scale, and good targeting for the FD scale.

References

1. Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. *Nature Rev Microbiol.* 2023;21(3):133-146.
2. Office of National Statistics (ONS). Self-reported coronavirus (COVID-19) infections and associated symptoms, England and Scotland: 25th April 2024. 2024.
3. Hou Y, Gu T, Ni Z, et al. Global Prevalence of Long COVID, Its Subtypes, and Risk Factors: An Updated Systematic Review and Meta-analysis. *Open Forum Infectious Diseases.* 2025; 12(9): ofaf533.
4. Bowe B, Xie Y, Al-Aly Z. Postacute sequelae of COVID-19 at 2 years. *Nature Med.* 2023;29(9):2347-2357.
5. Greenwood DC, Mansoubi M, Bakerly ND, et al. Physical, cognitive, and social triggers of symptom fluctuations in people living with long COVID: an intensive longitudinal cohort study. *Lancet Regional Health – Europe.* 2024;46.
6. Sivan M, Smith AB, Osborne T, et al. Long COVID clinical severity types based on symptoms and functional disability: a longitudinal evaluation. *J Clin Med.* 2024;13(7):1908.
7. Davis HE, Assaf GS, McCorkell L, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *EClinicalMedicine.* 2021;38
8. Kwon J, Milne R, Rayner C, et al. Impact of Long COVID on productivity and informal caregiving. *Eur J Health Econ.* 2024;25(7):1095-1115.
9. Walker S, Goodfellow H, Pookarnjanamorakot P, et al. Impact of fatigue as the primary determinant of functional limitations among patients with post-COVID-19 syndrome: a cross-sectional observational study. *BMJ Open.* 2023;13(6):e069217.
10. Hughes SE, Haroon S, Subramanian A et al (2022) Development and validation of the symptom burden questionnaire for Long Covid (SBQ-LC): Rasch analysis. *BMJ* 377:e070230 10.1136/bmj-2022-070230.
11. Rofail D, Somersan-Karakaya S, Mylonakis E, Choi JY, Przydzial K, Marquis S, Zhao Y, Hussein M, Norton TD, Podolanczuk AJ, Geba GP. The symptoms evolution of long COVID-19 (SE-LC19): a new patient-reported content valid instrument. *J Patient Rep Outcomes.* 2024 Aug 9;8(1):87.
12. Jandhyala R (2021) Design, validation and implementation of the post-acute (long) COVID-19 quality of life (PAC-19QoL) instrument. *Health Qual Life Outcomes* 19:229.
13. Sivan M, Halpin S, Gee J, et al. The self-report version and digital format of the COVID-19 Yorkshire Rehabilitation Scale (C19-YRS) for Long Covid or Post-COVID syndrome assessment and monitoring. *Adv Clin Neuro Rehab.* 2021;20(3).

14. Patel K, Straudi S, Yee Sien N, Fayed N, Melvin JL, Sivan M. Applying the WHO ICF Framework to the outcome measures used in the evaluation of long-term clinical outcomes in Coronavirus outbreaks. *Int J Environ Res Public Health.* 2020;17(18):6476.
15. López-López L, Calvache-Mateo A, Ortiz-Rubio A, et al. Differences of disabling symptoms between previously hospitalized or non-hospitalized currently working long-COVID survivors one year after infection: a descriptive study. *MDPI;* 2023:2306.
16. Partiprajak S, Krongthaeo S, Piaseu N, Wongsathikun J, Kongsuwan A. The Thai version of the COVID-19 Yorkshire Rehabilitation Scale: a valid instrument for the psychometric assessment of the community members in Bangkok, Thailand. *BMC Pub Health.* 2023;23(1):663.
17. Zhang D, Chung VC-H, Chan DC-C, et al. Determinants of post-COVID-19 symptoms among adults aged 55 or above with chronic conditions in primary care: data from a prospective cohort in Hong Kong. *Frontiers Pub Health.* 2023;11:1138147.
18. Kabir MF, Yin KN, Jeffree MS, et al. Profile of long COVID symptoms needing rehabilitation: a cross-sectional household survey of 12,925 SARS-CoV-2 cases between July and December 2021 in Bangladesh. *Arch Pub Health.* 2023;81(1):132.
19. Ayuso García B, Besteiro Balado Y, Pérez López A, et al. Assessment of post-COVID symptoms using the C19-YRS tool in a cohort of patients from the first pandemic wave in northwestern Spain. *Telemed e-Health.* 2023;29(2):278-283.
20. Ho FF, Xu S, Kwong TMH, et al. Prevalence, patterns, and clinical severity of long COVID among Chinese medicine telemedicine service users: preliminary results from a cross-sectional study. *Int J Environ Res Pub Health.* 2023;20(3):1827.
21. O'Connor RJ, Preston N, Parkin A, et al. The COVID-19 Yorkshire Rehabilitation Scale (C19-YRS): application and psychometric analysis in a post-COVID-19 syndrome cohort. *J M Virol.* 2022;94(3):1027-1034.
22. Sivan M, Preston N, Parkin A, Makower S, Gee J, Ross D, Tarrant R, Davison J, Halpin S, O'Connor RJ, Horton M. The modified COVID-19 Yorkshire Rehabilitation Scale (C19-YRSm) patient-reported outcome measure for Long Covid or Post-COVID-19 syndrome. *J Med Virol.* 2022 Sep;94(9):4253-4264. doi: 10.1002/jmv.27878. Epub 2022 Jun 1. PMID: 35603810; PMCID: PMC9348420.
23. Kustura L, Bobek D, Poljičanin A, et al. Psychometric properties and observational data for COVID-19 Yorkshire Rehabilitation Scale (C19-YRSm) for post-COVID-19 syndrome. *QJM.* 2024;117(1):38-47.
24. Smith A, Greenwood D, Horton M, et al. Psychometric analysis of the modified COVID-19 Yorkshire Rehabilitation Scale (C19-YRSm) in a prospective multicentre study. *BMJ Open Resp Res.* 2024;11(1).
25. Sivan M, Greenhalgh T, Darbyshire JL, et al. LOng COvid Multidisciplinary consortium Optimising Treatments and servIces acrOss the NHS (LOCOMOTION): protocol for a mixed-methods study in the UK. *BMJ Open.* 2022;12(5):e063505.
26. National Institute for Health and Care Excellence (NICE). COVID-19 rapid guideline: managing the long-term effects of COVID-19. NICE guideline [NG188]. 2024.

27. Sivan M, Lawrence RR, O'Brien P. Digital Patient Reported Outcome Measures Platform for Post-COVID-19 Condition and Other Long-Term Conditions: User-Centered Development and Technical Description. *JMIR Human Factors*. 2023;10(1):e48632.
28. Rasch G. Probabilistic models for some intelligence and attainment tests. ERIC; 1993.
29. Andrich D, Sheridan B, Luo G. RUMM 2030. Perth: RUMM Laboratory. 2010.
30. Smith Jr EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J App Measure*. 2002;3(2):205-231.
31. Masters GN. A Rasch model for partial credit scoring. *Psychometrika*. 1982;47(2):149-174.
32. Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative example. *Int J Nurs Stud*. 2009;46(3):380-393.
33. Tennant A, Küçükdeveci AA. Application of the Rasch measurement model in rehabilitation research and practice: early developments, current practice, and future challenges. *Front Rehab Sci*. 2023;4:1208670.
34. Christensen KB, Makransky G, Horton M. Critical values for Yen's Q 3: Identification of local dependence in the Rasch model using residual correlations. *App Psychol Measure*. 2017;41(3):178-194.
35. Andrich D. Rating scales and Rasch measurement. *Exp Review Pharmacoecon Outcomes Res*. 2011;11(5):571-585.
36. Hagell P, Westergren A. Sample size and statistical conclusions from tests of fit to the Rasch model according to the Rasch unidimensional measurement model (RUMM) program in health outcome measurement. *J App Measure*. 2016;17(4):416-431.
37. Hagell P. Testing rating scale unidimensionality using the principal component analysis (PCA)/t-test protocol with the Rasch model: the primacy of theory over statistics. *Open J Stat*. 2014;4(6):456-465.
38. Basch E, Abernethy A, Mullins C, Tiglao M, Tunis S. The development of a guidance for including patient-reported outcomes (PROs) in post-approval clinical trials of oncology drugs for comparative effectiveness research (CER). *Value Health*. 2011;14(3):A10.
39. Baalmann AK, Blome C, Stoletzki N, Donhauser T, Apfelbacher C, Piontek K. Patient-reported outcome measures for post-COVID-19 condition: a systematic review of instruments and measurement properties. *BMJ Open*. 2024;14(12):e084202. doi: 10.1136/bmjopen-2024-084202. PMID: 39806627; PMCID: PMC11667390.
40. Gorst SL, Seylanova N, Dodd SR, Harman NL, O'hara M, Terwee CB, Williamson PR, Needham DM, Munblit D, Nicholson TR, Aiyegbusi OL. Core outcome measurement instruments for use in clinical and research settings for adults with post-COVID-19 condition: an international Delphi consensus study. *Lancet Resp Med*. 2023; 11(12):1101-14.
41. Munblit D, Nicholson T, Akrami A, Apfelbacher C, Chen J, De Groote W, Diaz JV, Gorst SL, Harman N, Kokorina A, Olliari P. A core outcome set for post-COVID-19

condition in adults for use in clinical practice and research: an international Delphi consensus study. *Lancet Resp Med.* 2022;10(7):715-24.

42. D'hoore L, Germonpré P, Rinia B, Caevers L, Stevens N, Balestra C. Effect of normobaric and hyperbaric hyperoxia treatment on symptoms and cognitive capacities in Long COVID patients: a randomised placebo-controlled, prospective, double-blind trial. *Diving Hyperb Med.* 2025. 55(2):104-113. doi: 10.28920/dhm55.2.104-113. PMID: 40544138; PMCID: PMC12267068.
43. Corrado J, Iftekhar N, Halpin S, Li M, Tarrant R, Grimaldi J, Simms A, O'Connor RJ, Casson A, Sivan M. HEART Rate Variability Biofeedback for LOng COVID Dysautonomia (HEARTLOC): Results of a Feasibility Study. *Adv Rehabil Sci Pract.* 2024;13:27536351241227261. doi: 10.1177/27536351241227261. PMID: 38298551; PMCID: PMC10826406.