medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

# Psychometric analysis of the modified Covid-19 Yorkshire Rehabilitation Scale (C19-YRSm) in a prospective multicentre study

**Authors:** Adam B. Smith[1,2], Darren C. Greenwood[1,2], Mike Horton[3], Thomas Osborne[3], Madeline Goodwin[3], Román Rocha Lawrence[3,4], Darren Winch[5], Paul Williams[6], Ruairidh Milne[7], the LOCOMOTION consortium, Manoj Sivan[3,8,9]

**Affiliations:**

[1] School of Medicine, University of Leeds, UK

[2] Leeds Institute for Data Analytics, University of Leeds, UK

[3] Academic Department of Rehabilitation Medicine, University of Leeds, Leeds, UK

[4] ELAROS 24/7

[5] Kent and Medway ICB Long Covid Patient Reference Group

[6] Hertfordshire Community NHS Trust, UK

[7] School of Healthcare Enterprise and Innovation, University of Southampton, Southampton, UK.

[8] Covid Rehabilitation Service, Leeds Community Healthcare NHS Trust, UK

[9] National Demonstration Centre of Rehabilitation Medicine, Leeds Teaching Hospitals NHS Trust, UK

**Corresponding author:**

Adam B. Smith, Ph.D., School of Medicine, University of Leeds

Email: a.b.smith@leeds.ac.uk

**Key words:** Long COVID, C19-YRSm, validity, responsiveness, minimally important difference

**Word count:** 2995

**Tables:** 5

**Figures:** 1

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.


## Abstract

### Background

Long COVID is a novel multisystem clinical syndrome affecting millions of individuals worldwide. The modified COVID-19 Yorkshire Rehabilitation Scale (C19-YRSm) is a condition-specific patient-reported outcome measure designed for assessment and monitoring of people with Long COVID (LC).

### Objectives

To evaluate the psychometric properties of the C19-YRSm in a prospective sample of people with Long COVID.

### Methods

1314 patients attending UK specialist Long COVID clinics completed C19-YRSm and EQ-5D-5L longitudinally. Scale characteristics were derived for C19-YRSm subscales (Symptom Severity, SS; Functional Disability, FD; and Overall Health, OH) and internal consistency (Cronbach's alpha). Convergent validity was assessed using the FACIT-Fatigue scale. Known groups validity was assessed for the Other Symptoms (OS) subscale as tertiles, hospitalisation and intensive care admission. Responsiveness and test-retest reliability was evaluated for C19-YRSm subscales and EQ-5D-5L. The minimal important difference (MID) and minimal clinically important difference (MCID) were estimated. Confirmatory factor analysis was applied to determine the instrument's two-factor structure.

### Results

C19-YRSm demonstrated good scale characteristic properties. Item-total correlations were between 0.37 to 0.65 (for SS and FD), with good internal reliability (Cronbach's alphas >0.8). Item correlations between subscales ranged between 0.46 to 0.72. Convergent validity with FACIT was good (-0.46 to -0.62). The three subscales discriminated between different levels of symptom burden ($p<0.001$), and between patients admitted to hospital and intensive care. There was moderate responsiveness for the three subscales ranging from 0.22 (OH) to 0.50 (SS) and was greater than the EQ-5D-5L. Test-retest reliability was good for both SS 0.86 and FD 0.78. MID was 2 for SS, 2 for FD, and 1 for OH; MCID was 4 for both the SS and FD. The factor analysis supported the two-factor SS and FD structure.

### Conclusions

The C19-YRSm is a condition-specific, reliable, valid, and responsive patient-reported outcome measure for Long COVID.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

**Key messages:**

**What is already known on this topic**

Long Covid or Post-COVID-19 syndrome is a multisystem, fluctuating condition. C19-YRSm is literature's first condition-specific patient reported outcome measure which needed validation in a large population sample.

**What this study adds**

C19-YRSm is a valid, reliable, responsive and easy to administer measure which is able to show clinically meaningful change in the status of the condition in people living with Long Covid.

**How this study might affect research, practice or policy**

C19-YRSm can be used in clinical and research settings to reliably capture the condition trajectory and the effect of interventions and also help inform clinical policy.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

## Introduction

Long Covid (LC) or post-acute sequelae of COVID-19 is a fluctuating, multisystem syndrome [1] with an estimated prevalence of 1.9 million cases in the UK alone [2] and what is estimated to be at least 100 million individuals worldwide [3]. There have been more than 200 symptoms recorded in LC affecting 10 organ systems. The most commonly reported symptoms include fatigue, cognitive problems, pain, sleep problems, and breathlessness [4]. These symptoms may persist for extensive periods following the initial COVID-19 infection [5]. This protracted course of LC leads to a significant negative impact on the individual, in terms of the persistent nature of symptoms and the associated functional disability and adverse health-related quality of life [2,6].

The COVID-19 Yorkshire Rehabilitation Scale (C19-YRS) is a condition-specific patient-reported outcome measure designed to capture the symptoms of LC, as well as assess severity and monitor the persistence of symptoms to inform and guide the rehabilitation of affected patients [7-10]. Since the initial validation of the instrument [7,11], it has been widely used in a variety of LC contexts including symptom evaluation in primary care and community settings [12-14], determining the need for LC rehabilitation interventions [8,9,15], as well as epidemiological assessments of post-COVID symptoms [16,17].

The original 22-item C19-YRS underwent psychometric evaluation including both classical and modern psychometric evaluation methods resulting in a 17-item modified tool, the C19-YRSm [18]. Both the original 22-item and the modified version have undergone a limited degree of subsequent validation [10,19]. The C19-YRS has been shown to have good construct validity but moderate responsiveness [10]. The C19-YRSm has, by contrast, been demonstrated in a Croatian patient population to have good internal reliability and convergent validity [19]. Therefore, the aims of this study were to further validate the C19-YRSm with a longitudinal sample of LC patients, as well as to identify minimally (clinically) important differences to inform use in future randomised controlled trials and clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

## Methods

### *Data*

The data were collated from the LOng COvid Multidisciplinary consortium Optimising Treatments and servIces acrOss the NHS (LOCOMOTION) study [20]. This was a prospective mixed-methods study involving 10 LC services across the UK. 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) [20].

### *Participants*

Eligibility criteria for the study comprised participants: with a diagnosis of LC who were receiving treatment and management of the condition from one of the 10 participating LC services. Participation in the study required participants to be registered on ELAROS: a digital patient-reported outcome measures platform [21]. Informed consent and study data were collected on this platform. Participants were requested to complete the following patient-reported outcome measures (see below) every 3 months after registration. The first patient was registered on the 23 November 2021 and the last on 12 November 2023.

### *Patient and Public Involvement*

Patients have been involved from the outset in the design and implementation of the LOCOMOTION study [20] as part of a 9-member Patient Advisory Group (PAG). The PAG has provided the LOCOMOTION study team with first-hand experience of people living with LC. Two co-authors (RM and DW are members of the LOCOMOTION PAG).

## Instruments

### *Covid-19 Yorkshire Rehabilitation Scale – Modified (C19-YRSm)*

The C19-YRSm is a 17-item instrument [18] designed to capture the key symptoms of LC and its impact on activities of daily living and overall health. The items comprise four subscales: Symptom Severity

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(SS, 10 items), Functional Disability (FD, 5 items), Overall Health (OH, a single item), and Other Symptoms (OS).

The items in the SS subscale comprise the following domains: breathlessness (4 items), cough/throat sensitivity/voice change (2 items), fatigue (one item), smell / taste (2 items), pain / discomfort (five items), cognition (three items), palpitations / dizziness (two items), post-exertional malaise (one item), anxiety / mood (five items), and sleep (one item). FD consists of 5 single items: communication, walking / moving around, personal care, other activities of daily living, and social role.

Responses on the SS and FD subscales are rated on a 0 (no symptom or dysfunction) to 3 (severe life-disturbing symptom or dysfunction) Likert scale. For the SS subscale, the highest value within each of the domains (e.g., breathlessness, pain / discomfort) are added to determine the score for that subscale. Higher scores on both these subscales indicate worse symptomatology and poorer functioning. Responses on the OH subscale are scored on a 0-10 Likert scale (0 being "worst health" and 10 being "best health") with higher scores indicating better health. Other Symptoms (OS) over the last 7 days are also captured from a list of 25 additional symptoms [18].

*Functional Assessment of Chronic Illness Therapy – Fatigue Scale (FACIT-Fatigue)*

The FACIT-Fatigue scale is a 13-item instrument developed to evaluate fatigue and its impact on health-related quality of life and daily activities [22]. Responses are scored on a 5-item Likert scale (from 0 to 4) with a maximum total score of 52 (range 0 to 52). Higher scores indicate better health-related quality of life. Items in the FACIT-Fatigue cover tiredness, fatigue, listlessness, lack of energy, and the impact on daily and social activities. Although originally developed for use with cancer patients, the FACIT-Fatigue scale has been used to evaluate post-COVID fatigue [23].

*The EuroQol 5D-5L (EQ-5D-5L)*

The EuroQol EQ-5D-5L is a preference-based instrument with five domains: Mobility, Usual Activities, Selfcare, Pain / Discomfort, and Anxiety / Depression [24]. It has five response categories ranging from 1 (no problems) to 5 (severe problems). Responses to each domain are collated into a profile score which is converted into a health utility or index score using a country-specific algorithm

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(tariff or value set). Utilities reflect societal preferences for health states and are measured on a metric from 0 (dead) to 1 (perfect health). Utility values less than 0, indicating states worse than dead, are also captured. The EQ-5D-5L scores were mapped onto the EQ-5D-3L using the crosswalk (CW) algorithm to derive UK utility values [25]. The EQ-5D also comprises a visual analogue scale (VAS) measuring self-reported current health on a scale from 0 ("worst health") to 100 ("best health").

**Statistical analysis**

All analyses were performed using R Studio (R version 4.1.1). Descriptive summary statistics (mean, standard deviation, count and percentage) were generated for the following patient demographic and clinical data: age, sex, ethnicity, smoking status, hospital admission, intensive care (ICU) admission. The scale characteristics for the three C19-YRSm subscales (SS, FD, and OH) were derived including: mean (standard deviation, SD), median (inter-quartile range), score range, and skewness (evaluated -0.5 to +0.5). Item characteristics, such as mean item score (SD), missing values, floor and ceiling effects, and item-total correlations were estimated for the SS and FD domains.

The internal reliability of the C19-YRSm SS and FD domains was evaluated using Cronbach's alpha. A Cronbach's alpha >0.7 was considered to be an indicator of adequate internal consistency and >0.8 was considered to be an indicator of good internal consistency.

Convergent validity – the degree to which items or domains on different instruments measure the same constructs – was assessed for the C19-YRSm subscales, SS, FD, and OH using the FACIT-Fatigue. As both SS and FD are negatively scored (a higher score indicates worse symptoms or functioning), negative associations were anticipated between these and the FACIT-Fatigue score. Conversely, positive associations were hypothesised between OH and FACIT-Fatigue. Associations were evaluated using Pearson's product moment.

Known-groups validity was assessed for the C19-YRSm domains using the OS subscale split into tertiles: low number of symptoms (0-3), medium number of symptoms (4-7), and high number of symptoms (7+), as well as hospitalisation and admission to ICU (yes/no). Kruskall-Wallis and Mann-

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Whitney-Wilcoxon rank sum tests were used to evaluate differences in scores across these pre-defined groups.

Responsiveness of the three C19-YRSm subscales (SS, FD, and OH) were evaluated using a subset of patients who had completed the instrument at two timepoints, namely the first assessment and at follow-up 30 days later ($\pm$ 10 days). The responsiveness of the EQ-5D-5L and EQ-5D VAS was also evaluated as a comparator for those patients who had completed both the C19-YRSm and EQ-5D-5L on the same day (at first assessment and 30 days ($\pm$ 10 days)).

Mean change from the first assessment was derived for these domains and an effect size was calculated (standardised mean response, SMR) by dividing this by the standard deviation of the mean change scores. Intra-class correlations and test-retest reliability were also derived to evaluate stability in the instrument subscales over time. Test-retest reliability was evaluated against OH: the reliability coefficient was derived for patients with no change score on the OH between first assessment and day 30 ($\pm$ 10 days).

A half standard deviation of the first assessment domain scores was applied as a putative minimally important difference (MID) [26]. In addition to this, the standard error of measurement (SEM) and reliable change index (RIC) were calculated as follows for the SS and FD domains as indicators of minimal clinically important differences (MCIDs):

SEM = $1.96 \cdot SD_{base} \cdot \sqrt{(1-r)}$ , and r is the test-retest reliability coefficient.

RIC = $1.96 \cdot SD_{base} \cdot \sqrt{[2 \cdot (1-r)]}$, where $SD_{base}$ is the standard deviation at first assessment.

The putative factor structure – a two-dimensional structure encompassing the SS and FD domains – was explored utilising a confirmatory factor analysis (CFA). A number of indices were employed to evaluate the goodness-of-fit of the model: root mean square error of approximation (RMSEA) [27], comparative fit index (CFI) [28], the Tucker-Lewis index (TLI) [29,30], and the standardised root mean squared residual (SRMR) [31]. Various thresholds have been proposed to evaluate model fit. In this

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

study, RMSEA <0.08 [32] was considered to be a reasonable fit; TLI and CFI >0.90 as acceptable fit [29], and SRMR <0.08 as acceptable fit [31]. As no single index provides sufficient evidence alone of model fit, four indices were evaluated in aggregate. The *lavaan* package in R was used for the CFA.

## Results

### *Demographics*

A total of 1314 patients (Table 1) had completed the C19-YRSm on at least one occasion; 263 patients (20%) had completed the instrument at first assessment and at 30 days (± 10 days), and 193 patients had completed the FACIT-Fatigue instrument at least once (15%). The C19-YRSm and EQ-5D-5L had been completed on the same day at both timepoints (first assessment and day 30 (± 10 days)) by 98 patients. The majority (total sample) were Caucasian (76%) females (67%) with an average age of 48 years (SD: 13 years); 10% had been admitted to hospital as a result of COVID-19, and just over 2% had been admitted to ICU.

The mean subscale scores are shown in Table 2. Both means for the SS (18.4, SD: 5.62) and FD (7.1, SD: 3.78) subscales suggested a moderate-to-high level of symptom burden and functional disability. Similarly, OH indicated that patients were at best in moderate health. These three subscales showed little skewness reflecting symmetrical score distributions. The mean number of OS was 5 with positive skew (fewer patients with large numbers of other symptoms).

The item means (Table 3) for the SS and FD subscales ranged approximately between 1 and 2 indicating that patients were on average experiencing at least mild (to moderate) symptom burden and functional disability, although this varied across the items as reflected in the results of the floor and ceiling effects. Missing data was negligible (<3%).

### *Convergent validity*

Table 4 shows the correlation matrix between the domains (see also Supplementary Figure 1). There was a strong positive association between SS and FD. A moderate positive association was

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

determined between OS and SS, and OS and FD. OH was negatively associated with the SS, FD, and OS.

There was a strong negative association between the total FACIT score and the C19-YRSm Fatigue item (r = -0.58, p<0.001, 95%CI: -0.67 to -0.48), and similarly for SS (r = -0.61, p<0.001, 95%CI: -0.69 to -0.51), FD (r = -0.64, p<0.001, 95%CI: -0.72 to -0.55), and OS (r = -0.46, p<0.001, 95%CI: -0.56 to -0.34). The FACIT-Fatigue total was positively associated with OH (r = 0.47, p<0.001, 95%CI: 0.36 to 0.58).

*Known-groups validity*

There was a linear increase in SS score as symptom burden (tertiles of the OS domain) increased in severity from low to high (Table 5a). A similar pattern was observed for FD, whereas OH showed a decrease as symptom burden increased. All these results were statistically significant (p<0.001).

Patients who had been hospitalised for COVID-19 showed higher SS and FD scores (Supplementary Tables 1a and 1b) compared to those who had not been hospitalised (p=0.04 and p=0.008, respectively). No differences between these groups were observed for OH. Statistically significant differences for both SS and FD (but not OH) were also observed between those who had and had not been admitted to ICU (Supplementary Tables 1c).

*Responsiveness*

The mean change over 30 days (± 10 days) was 1.9 (SD: 4.38) for the SS domain. Smaller changes were observed in both the FD and OH domains, 0.7 (SD: 2.53) and 0.3 (SD: 1.67) (Table 5).

The intraclass correlation coefficients for the 3 domains ranged from 0.58 (OH) to 0.76 (FD) (Table 5) suggesting moderate-to-strong content structure over time. A total of 70 patients had stable (unchanged) OH scores over the 30-day evaluation period (± 10 days). The test-retest reliability coefficient for the SS domain was 0.86 and 0.78 for the FD domain, indicating good reliability.

All 3 subscales demonstrated a degree of responsive to change (effect sizes range: 0.22 to 0.50) (Table 5). The responsiveness of the EQ-5D-5L Index was by comparison 0.14 and 0.18 for the VAS.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The 0.5 SD was applied as a metric for the MID. This resulted in the following MIDs: SS = 2; FD = 2; and OH = 1. From Table 6, it may be seen, for instance, that a MID was recorded for SS over the 30-day period following first assessment, but not for either FD or OH. The MCID estimate (based on the SEM) was 4 for both the SS and FD (Table 5).

*Factor Structure*

The results of the confirmatory factor analysis showed a RMSEA of 0.10 (90% confidence intervals, CI: 0.096, 0.107) (Figure 1). The SRMR was 0.066, CFI 0.83, and Tucker-Lewis Index 0.8. Taking all 4 indices together, these indicated reasonable model fit for the two-factor model. These factors were consistent with the interpretation of one factor measuring SS and the other measuring FD.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

## Discussion

The aim of this study was to undertake a further psychometric validation of the C19-YRSm. The results demonstrated good item and scale characteristics. There was good convergent validity with the FACIT-Fatigue scale. Furthermore, the three subscales (SS, FD, and OH) discriminated well between levels of symptom severity, and between patients who had been hospitalised and admitted to ICU. There was also good internal reliability, test-retest reliability, and stability of the subscale scores over time. Furthermore, the convergent correlations were as hypothesised.

The results of the responsiveness analysis showed that the instrument was able to detect changes as patients' symptoms fluctuated and was more sensitive to change than the generic health-related quality of life measure, the EQ-5D-5L (both Index and VAS). This is a potentially important finding for future randomised controlled trials in LC. Although the effect sizes were modest, these must be evaluated in the context of a fluctuating condition [4, 33-34], and it may therefore be that potentially larger effect sizes were being masked by frequent changes in symptoms. The results also suggest some initial metrics for the MID for SS (2), FD (2), and OH (1), as well as the MCID (4 SS and FD). Although these were based on distribution methods, and therefore remain to be confirmed using anchor-based approaches such as patient and clinician global impression of change, these provide useful initial metrics for interpreting meaningful changes in the C19-YRSm scores to aid both clinical interpretation as well as inform sample size considerations for prospective randomised controlled trials.

There was also some support for a 2-factor structure, although the statistics in isolation did not meet the pre-defined thresholds. Although there are no definitive guidelines on what constitutes ideal fit, it is possible that model fit could be improved to a degree that may consequently also positively impact on responsiveness. Previous research has similarly determined moderate responsiveness at the item-level [10]. Further research involving modern psychometric analysis such as Rasch or Item-response theory could explore this issue further and potentially identify individual items that may be removed and / or recalibrated to improve instrument responsiveness.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The results of this study are in line with a previous psychometric validation study of the C19-YRSm, which found both good internal reliability and convergent validity of the instrument [19], providing further evidence for the psychometric properties of the C19-YRSm with meaningful factors or domains, such as SS, FD, and OH. The latter is further bolstered by the large sample size in this study and builds on the earlier development of the instrument [18], supporting its use as one of the few LC-specific patient-reported outcome measures. In addition, it is shorter than other condition-specific instruments such as the Symptom Burden Questionnaire for Long COVID with 131 items [35], thereby minimising patient burden – a particularly important factor in people living with LC, who may present with fatigue and cognitive dysfunction. The instrument's brevity and design lends its use for self-completion by patients, enabling the fluctuating nature of the condition to be monitored on a frequent basis for patients' own awareness, for instance, in determining symptom triggers, as well as by clinicians to evaluate patients' condition over time between clinics appointments.

Given the prevalence of LC, its associated persistence of debilitating symptoms, and the impact of the condition on patients' health-related quality of life, valid and reliable condition-specific patient-reported instruments such as the C19-YRSm are of critical importance in the assessment of LC symptoms, as well as in helping to facilitate appropriate management and rehabilitation of patients suffering with the condition. The evidence presented alongside other studies [19] suggest that C19-YRSm is a condition-specific, reliable, valid, and responsive patient-reported outcome measure for Long COVID.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

**References**

1. Davis HE, McCorkell L, Vogel J, et al. Long COVID: major findings, mechanisms and recommendations. *Nature Rev Microbiol* 2023; (3): 133-46.

2. Office of National Statistics. Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK: 2 February 2023: Office for National Statistics; 2023 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/prevalenceofongoingsymptomsfollowingcoronaviruscovid19infectionintheuk/2february2023]

3. Chen C, Haupert SR, Zimmermann L, et al. Global Prevalence of Post-Coronavirus Disease 2019 (COVID-19) Condition or Long COVID: A Meta-Analysis and Systematic Review. *J Infect Dis* 2022; 226(9): 1593-607.

4. Davis HE, Assaf GS, McCorkell L, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *EClinicalMedicine* 2021: 101019.

5. Callard F, Perego E. How and why patients made Long Covid. *Soc Sci Med* 2021; 268: 113426

6. Sivan M, Rayner C, Delaney B. Fresh evidence of the scale and scope of long covid. *BMJ* 2021; 373: n853.

7. Sivan M, Halpin S, Gee J. Assessing long-term rehabilitation needs in COVID-19 survivors using a telephone screening tool (C19-YRS tool). *Adv Clin Neuro Rehab* 2020; 19 (4): 14-17.

8. Parkin A, Davison J, Tarrant R, et al. A Multidisciplinary NHS COVID-19 Service to Manage Post-COVID-19 Syndrome in the Community. *J Prim Care Community Health* 2021; 12: 21501327211010994.

9. Yilmaz Gokmen G, Durmaz D, Demir C, Yilmaz FN. Determining Post-COVID-19 Symptoms and Rehabilitation Needs in Hospitalized and Nonhospitalized COVID-19 Survivors with Tele-Assessment Methods. *Telemed J E Health* 2023 Sep;29(9):1312-1323. doi: 10.1089/tmj.2022.0349. Epub 2023 Feb 10. PMID: 36757300.

10. Straudi S, Manfredini F, Baroni A, et al. Construct Validity and Responsiveness of the COVID-19 Yorkshire Rehabilitation Scale (C19-YRS) in a Cohort of Italian Hospitalized COVID-19 Patients. *Int J Environ Res Public Health* 2022; 19(11): 6696. doi: 10.3390/ijerph19116696. PMID: 35682280; PMCID: PMC9180312.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

11. 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 Neurosci Rehab* 2021; 20(3): 1-5.

12. 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. *Healthcare*. 2023; 11(16): 2306. doi: 10.3390/healthcare11162306. PMID: 37628505; PMCID: PMC10454028.

13. Zhang D, Chung VC, Chan DC, 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. *Front Public Health* 2023;11:1138147. doi: 10.3389/fpubh.2023.1138147. PMID: 37213637; PMCID: PMC10196359.

14. Partiprajak S, Krongthaeo S, Piaseu N, et al. 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 Public Health* 2023;23(1):663. doi: 10.1186/s12889-023-15566-2. PMID: 37041552; PMCID: PMC10088103.

15. 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 Public Health* 2023; 81(1):132. doi: 10.1186/s13690-023-01140-0. PMID: 37461092; PMCID: PMC10351147.

16. 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 J E Health* 2023; 29(2):278-283. doi: 10.1089/tmj.2022.0085. Epub 2022 Jun 21. PMID: 35727139.

17. 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 Public Health* 2023; 20(3):1827. doi: 10.3390/ijerph20031827. PMID: 36767195; PMCID: PMC9914360.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

18. Sivan M, Preston N, Parkin A, et al. 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; 94(9): 4253-64.

19. 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* 2023: hcad224. doi: 10.1093/qjmed/hcad224. Epub ahead of print. PMID: 37788123.

20. 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.

21. Sivan M, Rocha Lawrence R, 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 Hum Factors* 2023; 10:e48632. doi: 10.2196/48632. PMID: 37665334; PMCID: PMC10592725.

22. Yellen SB, Cella DF, Webster K, et al. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. *J Pain Symptom Manage*. 1997; 13(2):63-74. doi: 10.1016/s0885-3924(96)00274-6. PMID: 9095563.

23. Hartung TJ, Neumann C, Bahmer T, et al. Fatigue and cognitive impairment after COVID-19: A prospective multicentre study. *EClinicalMedicine* 2022; 53:101651. doi: 10.1016/j.eclinm.2022.101651. PMID: 36133318; PMCID: PMC9482331.

24. Herdman M, Gudex C, Lloyd A, et al. *Qual Life Res* 2011; 20(10):1727-36. doi: 10.1007/s11136-011-9903-x. Epub 2011 Apr 9. PMID: 21479777; PMCID: PMC3220807.

25. van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*. 2012; 15(5): 708-15.

26. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003; 41(5):582-92. doi: 10.1097/01.MLR.0000062554.74615.4C. PMID: 12719681.

27. Steiger JH. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behav Res* 1990; 25: 173–180.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

28. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull* 1990; 107: 238–246.

29. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. *Psychol Bull* 1980; 88: 588–606.

30. Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika* 1973; 38: 1–10.

31. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 1999; 6: 1–55.

32. Browne MW, Cudeck R. Alternative ways of assessing model fit. 1993. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp. 136–162). Newbury Park, CA: Sage.

33. National Institute for Health and Care Research (NIHR). Living with Covid19 – Second review. 16/03/2021 ed, 2021.

34. Ziauddeen N, Gurdasani D, O'Hara ME, et al. Characteristics and impact of Long Covid: Findings from an online survey. *PLOS ONE* 2022; 17(3):e0264331. doi: 10.1371/journal.pone.0264331.

35. Hughes SE, Haroon S, Subramanian A, et al. Development and validation of the symptom burden questionnaire for long covid (SBQ-LC): Rasch analysis. *BMJ*. 2022 Apr 27;377:e070230. doi: 10.1136/bmj-2022-070230. PMID: 35477524; PMCID: PMC9043395.

**The C19-YRSm is available at:** https://c19-yrs.com/the-c19-yrs/

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

## Tables and Figures

### Table 1. Basic sociodemographic details

| **Characteristic** | **N = 1,314**[1] |
|---|---|
| **Sex** | |
| Female | 882 (67%) |
| Male | 432 (33%) |
| **Age (years)** | 47.9 (13.0) |
| **Ethnicity** | |
| Asian (includes any Asian background, for example, Bangladeshi, Chinese, Indian, Pakistani) | 84 (6.4%) |
| Black, African, black British or Caribbean (includes any black background) | 29 (2.2%) |
| Mixed or multiple ethnic groups (includes any mixed background) | 20 (1.9%) |
| White (includes any white background) | 1003 (76%) |
| Another ethnic group (includes any other ethnic group, for example, Arab) | 18 (1.4%) |
| Not recorded | 152 (12%) |
| **Smoking status** | |
| Current occasional smoker | 21 (1.6%) |
| Current regular smoker | 11 (0.8%) |
| Ex-smoker | 105 (8%) |
| Never smoked | 271 (21%) |
| Not recorded | 906 (69%) |
| **Hospital admission** | |
| No | 1179 (90%) |
| Yes | 134 (10%) |
| **ICU admission** | |
| No | 1283 (98%) |
| Yes | 31 (2.0%) |

[1] N (%)

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

**Table 2. C19-YRSm domain characteristics**

| Subscale (scale range) | Valid scores (N) | Mean (SD) | Median (IQR) | Score range | Skewness |
|---|---|---|---|---|---|
| **Symptom Severity (0-30)** | 1314 | 18.4 (5.62) | 18 (14.5 - 22) | 0 - 30 | -0.17 |
| **Functional Disability (0-15)** | 1314 | 7.1 (3.80) | 7.0 (4.0 - 10.0) | 0 - 15 | 0.13 |
| **Overall Health (0-10)** | 1314 | 4.5 (1.90) | 5.0 (3.0 - 6.0) | 0 - 10 | 0.13 |
| **Other Symptoms (0-25)** | 1314 | 5.6 (4.3) | 5.0 (2.0 – 8.0) | 0 – 23 | 0.78 |

*****N, number; SD, standard deviation; IQR, interquartile range

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

**Table 3. Scaling assumptions, targeting, and internal reliability**

| **Scaling assumptions (range)** | **Symptom Severity** | **Functional Disability** |
|---|---|---|
| Item means | 0.9 - 2.4 | 0.7 - 2.1 |
| Item SD | 0.7 - 1.1 | 0.9 - 1.1 |
| Item-total correlation | 0.37 - 0.61 | 0.46 - 0.65 |
| **Targeting** | | |
| Missing data (%): range | 0.8 - 2.3 | 0.8 - 1.1 |
| Floor effects (%): range | 11.0 - 48.0 | 6.0 - 40.0 |
| Ceiling effects (%): range | 2.0 - 50.0 | 6.0 - 61.0 |
| **Internal reliability** | | |
| Cronbach's alpha | 0.82 | 0.81 |

*SD, standard deviation.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

**Table 4. Correlation matrix C19-YRSm domains**

| **Domains** | **Symptom Severity** | **Functional Disability** | **Overall Health** | **Other Symptoms** |
|---|---|---|---|---|
| **Symptom Severity** | 1.00 | 0.73 | -0.47 | 0.63 |
| **Functional Disability** | 0.73 | 1.00 | -0.52 | 0.55 |
| **Overall Health** | -0.47 | -0.52 | 1.00 | -0.34 |
| **Other Symptoms** | 0.63 | 0.55 | -0.34 | 1.00 |

*****N=1314; $p<0.001$

**Table 5. Responsiveness C19-YRSm domains**

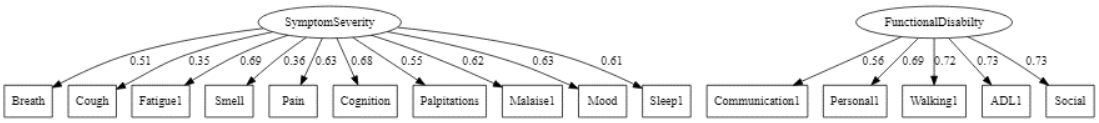
| Subscale (scale range) / Change from 1st assessment | Mean (SD) | Median | Min, Max | ICC | TRR | ES | MID | MCID (SEM) | MCID (RIC) |
|---|---|---|---|---|---|---|---|---|---|
| **Symptom Severity (0-30)** | 2.0 (4.0) | 2 (0 to 3) | (-19.0, 23.0) | 0.68 | 0.86 | 0.50 | 2 | 4 | 6 |
| **Functional Disability (0-15)** | 0.8 (2.4) | 1 (-1 to 2) | (-5.0, 10.0) | 0.76 | 0.78 | 0.33 | 2 | 4 | 5 |
| **Overall Health (0-10)** | 0.3 (1.67) | 0.4 (-1 to 1) | (-7.0, 7.0) | 0.58 | - | 0.22 | 1 | - | - |

*****SD, standard deviation; min, minimum; max, maximum; ICC, intraclass correlation; TRR, test-retest reliability; ES, effect size; MID, minimally important difference; MCID, minimal clinically important difference; SEM, standard error of measurement; RIC, reliable change index

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

**Figure 1. Factor structure of the C19-YRSm Symptom Severity and Functional Disability Domains**



medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

**Supplementary Table 1a. C19-YRSm domain mean domain scores by symptom burden ("Other Symptoms")**

| Domain | **Low**, N = 507[1] | **Medium**, N = 441[1] | **High**, N = 396[1] | **p-value**[2] |
|---|---|---|---|---|
| **Symptom Severity (0-30)** | 14.6 (4.8) | 18.8 (4.5) | 22.8 (4.2) | <0.001 |
| **Functional Disability (0-15)** | 5.1 (3.1) | 7.0 (3.3) | 9.9 (3.3) | <0.001 |
| **Overall Health (0-10)** | 5.1 (1.8) | 4.4 (1.7) | 3.8 (1.8) | <0.001 |

[1]Mean (Standard Deviation); [2]Kruskal-Wallis rank sum test; Low, Medium, High based on tertiles of the "Other Symptom" domain

**Supplementary Table 1b. C19-YRSm domain mean domain scores by hospital admission**

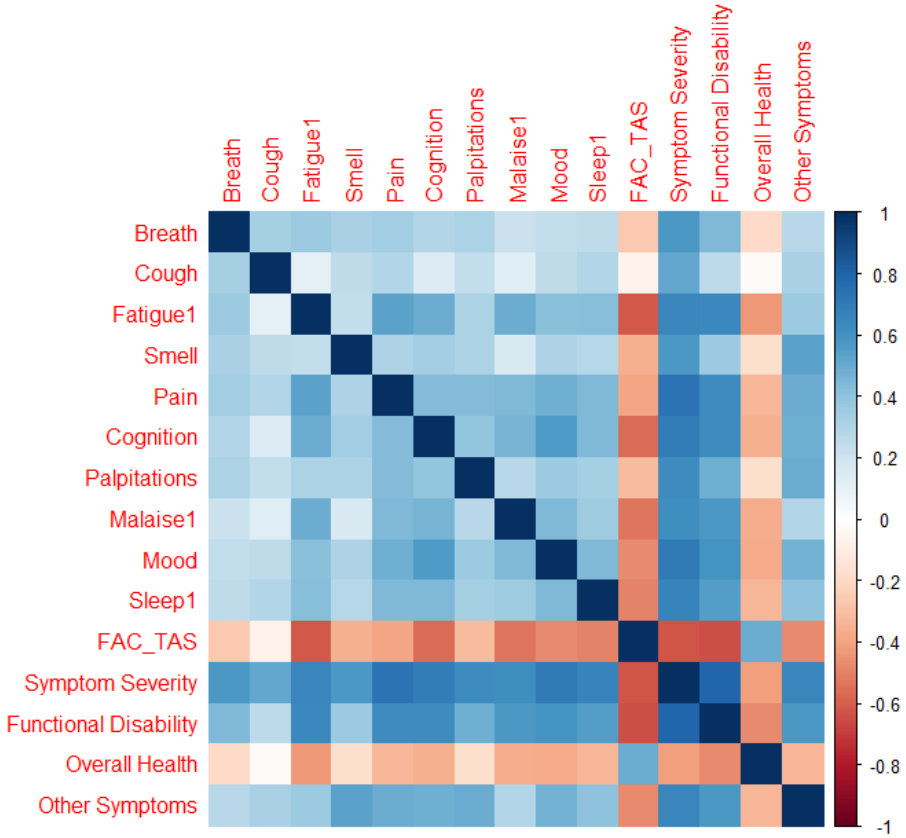
| Hospital Admission | **No**, N = 1179[1] | **Yes**, N = 134[1] | **p-value**[2] |
|---|---|---|---|
| **Symptom Severity (0-30)** | 18.3 (5.6) | 19.3 (6.0) | 0.07 |
| **Functional Disability (0-15)** | 7.0 (3.7) | 8.0 (4.2) | 0.03 |
| **Overall Health (0-10)** | 5.0 (1.9) | 4.5 (1.9) | 0.8 |

[1] Mean (Standard Deviation); [2] Mann-Whitney-Wilcoxon rank sum test; Missing values, N=8

**Supplementary Table 1c. C19-YRSm domain mean domain scores by ICU admission**

| Characteristic | **No**, N = 1,283[1] | **Yes**, N = 31[1] | **p-value**[2] |
|---|---|---|---|
| **Symptom Severity (0-30)** | 18.3 (5.6) | 20.2 (6.0) | 0.07 |
| **Functional Disability (0-15)** | 7.1 (3.8) | 8.8 (4.5) | 0.03 |
| **Overall Health (0-10)** | 4.5 (1.9) | 4.3 (2.2) | 0.9 |

[1] Mean (Standard Deviation); [2] Mann-Whitney-Wilcoxon rank sum test; Missing values, N=8**Figure 1.**

medRxiv preprint doi: https://doi.org/10.1101/2023.12.22.23300424; this version posted December 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

**Supplementary Figure 1.Correlation matrix C19-YRSm and the FACIT total**