APTA Cross Sections and Academies Recommendations for COVID-19 Core Outcome Measures

Chris L. Wells, Deb Kegelmeyer, Kirby P. Mayer, Sowmya Kumble, Anne Reilley, Angela Campbell, Rebecca Martin, APTA Cross Academy/Section COVID-19 Core Outcome Measures Task Force

ABSTRACT

Purpose: The novel coronavirus (COVID-19) emerged as a major health concern within the United States in early 2020. Because this is a novel virus, little guidance exists for best practice to evaluate this population within the field of physical therapy.

Methods: An expert task force appointed by the leadership of 9 different academies or sections of the American Physical Therapy Association was formed to develop recommendations for a set of core outcome measures for individuals with or recovering from COVID-19.

Results: This perspective provides guidance on a best practice recommendation to physical therapists and researchers regarding the use of core outcome measures for individuals with or recovering from COVID-19. The process for the selection of core measures for this population is presented and discussed.

Conclusion: Core outcome measures improve the ability to track progress and change across the continuum of care at both the patient and population levels.

The novel coronavirus disease (COVID-19) caused by the respiratory virus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was officially designated a pandemic in the spring of 2020.[1] Since the initial case, the number of confirmed individuals with COVID-19 in the United States has climbed rapidly, reaching over 29 million cases and 500 000 deaths by March 2021.[2] Patients with COVID-19 present with a comprehensive disease spectrum ranging from asymptomatic infection to critical presentations, including severe hypoxic respiratory failure, shock, and multiorgan failure requiring prolonged care in an intensive care unit.[3]

The nature of COVID-19, with the potential for respiratory, cardiovascular, and neurologic compromise, increases the risk of short- and long-term physical disability, especially in the frail and elderly and those with comorbidities such as cancer, hypertension, diabetes mellitus, or stroke.[3,4] Current literature also suggests an increase in hospitalization and death in racial and ethnic minority groups.[4] However, COVID-19 has impacted all ages and all populations ranging from high functioning athletes[5] to patients who are dependent and living in extended-stay facilities.[6] Patients surviving COVID-19 may be at risk for cardiovascular and neurologic events due to direct insult from the virus, with evidence that the virus can cross the blood-brain barrier causing neurologic damage.[7] Additionally, some individuals with COVID-19 develop a severe, systemic, inflammatory response (cytokine storm) that leads to widespread damage and long-term consequences.[8] These sequelae can significantly impact patients' function leading to the need for physical therapy services. Referral to rehabilitation providers at the time of awakening, for the critically ill, and at earliest signs of movement deficits for others can mitigate secondary complications and improve patient-specific outcomes.[9-11]

Chris L. Wells, PT, PhD *University of Maryland Medical Center, Baltimore, Maryland.*

Deb Kegelmeyer, PT, DPT *School of Health and Rehabilitation Sciences, The Ohio State University, Columbus.*

Kirby P. Mayer, DPT, PhD *Department of Physical Therapy, University of Kentucky, Lexington.*

Sowmya Kumble, PT, MPT *Johns Hopkins Hospital, Baltimore, Maryland.*

Anne Reilley, PT, DPT, MS *Onward Physical Therapy, PLLC, Ithaca, New York.*

Angela Campbell, PT, DPT *Department of Physical Therapy, Springfield College, Springfield, Massachusetts.*

Rebecca Martin, PT, DPT, PhD *Department of Physical Therapy, Hanover College, Hanover, Indiana; Department of Physical Therapy, Nova Southeastern University, Fort Lauderdale, Florida. martinr@hanover.edu.*

APTA Cross Academy/Section COVID-19 Core Outcome Measures Task Force was comprised of all authors and the following: Rachel Botkin, PT, MPT (Advanced Competency in Home Health, Representing the Home Health Section, Botkin Rehab Services, LLC); Cristina Faucheux, PT (Representing the Private Practice Section, Moreau Physical Therapy); Tiffany Kendig, PT, DPT, MPH (Certified Lymphedema Therapist, Representing the Academy of Oncologic Physical Therapy, Kessler Rehabilitation Center); Susan Wenker, PT, PhD (Board-Certified Clinical Specialist in Geriatric Physical Therapy—Emeritus, Representing APTA Geriatrics, University of Wisconsin-Madison, Doctor of Physical Therapy Program, Department of Family Medicine and Community Health).

The authors have no conflicts of interest and no source of funding to declare.

Submitted for Publication: January 21, 2021; accepted for publication April 29, 2021; published online September 9, 2021.

DOI: 10.1097/JAT.0000000000000172

Physical therapists in all settings use outcome measures to establish a baseline, aid in clinical decision-making, and track progress across all phases of recovery and all settings of the continuum of care.[12,13] More recently, our profession has begun to recommend using core outcome measures to reduce unwarranted variation in practice and facilitate research initiatives.[14] Core outcomes improve communication between clinicians. Additionally, the use of common outcome measures allows researchers to aggregate the same data of hundreds to thousands of patients to better describe patterns of recovery for patients with a particular disease, such as COVID-19. The purpose of this article is to describe the process used to establish a set of core outcome measures for individuals with or recovering from COVID-19, which serves as a guide for clinical decision-making, reduces unwarranted variation in practice, and facilitates research initiatives.

PROJECT PREMISE: APTA CROSS ACADEMY/ SECTION COVID-19 CORE OUTCOME MEASURE TASK FORCE

On April 15, 2020, leaders from the APTA, including representatives from individual academies and sections, formed the "COVID-19 Response Panel" (Panel), to address the rapidly evolving nature of the novel coronavirus and function as an avenue for leaders to share and work together during the pandemic. This panel met weekly through mid-summer, then as needed through March of 2021. The panel consisted of leaders from 17 of the 18 academies and sections of the APTA. The panel served to reduce redundancies, maximize information and resources being developed, and assist with disseminating succinct and important information to APTA members during the initial months of the pandemic. During a meeting, the panel representatives from multiple sections identified and discussed the need for a set of core outcome measures to be used across settings. Members of the panel and component leaders appointed qualified members of their respective components to the Cross Section/Academies COVID-19 Core Outcome Measures Task Force (Task Force). No component was excluded from this process; however, some components chose not to assign representation to the Task Force. The 12-member Task Force included representatives appointed by APTA Acute Care, Cardiovascular and Pulmonary Section, APTA Geriatrics, Home Health Section, Academy of Neurologic Physical Therapy, APTA Oncology, Academy of Orthopaedic Physical Therapy, and the Private Practice Section. APTA Pediatrics also appointed 3 members to a separate but complementary working group to complete a parallel project for the pediatric population. The Task Force included the chair of the Parkinson Evidence Database to Guide Effectiveness (PD EDGE) Task Force, a member of the Home Health Section Outcome Measures Toolbox Team, members of the APTA Acute Care Outcome Measures Team, and experienced outcome measure researchers and speakers from APTA Acute Care, Cardiovascular and Pulmonary Section, the Academy of Neurologic Physical Therapy, APTA Oncology, and the Academy of Orthopaedic Physical Therapy.

Additionally, the Task Force was comprised of researchers, academicians, and clinicians actively working in various clinical settings. During the Task Force work, 1 member was actively engaged in data collection for research within this population, and several were actively treating patients with COVID-19 in a variety of settings. Consultation with the Core Set of Outcome Measures for Adults with Neurologic Conditions Task Force Chair was an integral part of formulating the Task Force processes.

The panel charged the Task Force to review existing literature and recommend a core set of outcome measures applicable to the greatest number of patients with or recovering from COVID-19 across the continuum of care and all clinical settings. The panel determined that a core set of outcome measures for this population would aid clinicians by improving their ability to track progress across settings, discuss clinical cases with peers, and identify appropriate measures when they are less familiar with the diagnosis. The panel also felt that physical therapy research related to COVID-19 would be expedited by using core outcome measures to allow for data pooling.

The Task Force Chair regularly attended panel meetings to provide updates and solicit feedback as appropriate from all academy/section representatives. Individual Task Force Members and panel members were also integral in soliciting further feedback from members of their respective academies/sections at key times throughout the process. The first task was to devise a process to complete the goals to identify and recommend a core set of outcome measures expeditiously, as shown in Figure 1.

IDENTIFICATION OF THE CONSTRUCTS

The next objective of the Task Force was to identify the constructs that would be of primary importance for individuals with or recovering from COVID-19. The Task Force used various sources, including published literature, World Health Organization and Centers for Disease Control and Prevention reports, information from various professional organizations, and anecdotal reports to define the breadth of the clinical presentation. The clinical presentation was defined across health care settings and by body systems, severity, type of movement dysfunction, and functional limitations of these patients. This search led to adopting the following constructs: cognition, endurance, functional mobility, health-related quality of life, strength/power, and upper extremity function. The Task Force unanimously voted to adopt these constructs. Cognition was included as a construct for multiple reasons. Based on current research, cognition

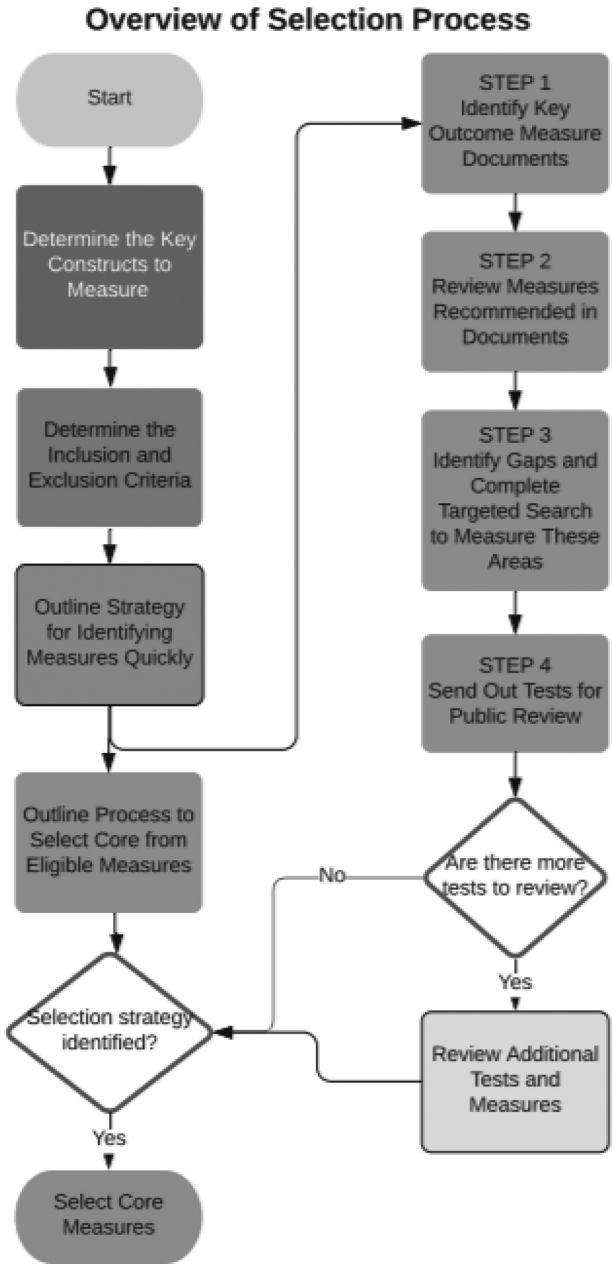

FIGURE 1. Overview of Selection Process.

is often impacted in patients with a moderate to severe presentation of COVID-19.[15] Cognition is often undereval-uated in similar populations.[16,17] To identify the effect of cognitive changes on patients' recovery and to increase the likelihood of identifying neurologic changes, when present, cognitive testing is completed per a recommended schedule as outlined later within this document.

DEVELOPMENT OF INCLUSION AND EXCLUSION CRITERIA

The criteria for the set of core measures needed to have clinical utility for the greatest number of individuals impacted by COVID-19 across the continuum of care and functional level. Perceived and actual barriers to the use of outcomes measures were discussed from the perspectives of all participating section representatives. The Task Force considered the constraints of different practice settings, including telehealth and home health, as well as the effect of isolation precautions when selecting criteria. Measures that required little to no training were prioritized. Cut-off scores related to reliability were included at this stage, but other psychometric properties were considered at a later stage of our process. With this perspective, the inclusion and exclusion criteria were agreed upon and can be found in Table 1.

IDENTIFICATION OF OUTCOME MEASURES

Due to the time-sensitive nature of the translation of our findings, the Task Force unanimously voted for a 4-stage expedited review process as follows:

- *Step 1*. The Task Force identified documents of high levels of evidence recommended or promoted by their respective academy or section, such as clinical practice guidelines, systematic reviews, meta-analysis papers, or measures that were reviewed and recommended by their academy or section. Additionally, Task Force members completed a full review of the sralab.org/rehabilitation-measures and PTNow.org Web sites to identify outcome measures that may be appropriate for inclusion. Specific literature searches were completed in PubMed, MEDLINE, OVID, Cochran, CINAHL, and Google Scholar to identify relevant articles. Search terms include such terms as physical therapy evaluation and assessment, physiotherapy, exercise, rehabilitation, assessment, post-intensive

TABLE 1. Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Interrater reliability >0.75 (administered by tester)	Requires a space larger than a typical hospital room
Test-retest reliability ICC >0.80 (questionnaires)	Materials are not readily available in a telehealth setting
Free to use	Materials cannot be easily cleaned for infection control
Training is free	
<15 min to complete	

ICC, intraclass correlation coefficient.

care syndrome, heart and lung impairment and failure, venous thromboembolism disease, Guillain-Barré, stroke, sepsis, and infection.

- *Step 2*. The Task Force reviewed the documents identified in step 1 and extracted all outcome measures that were recommended after previous rigorous review. These outcome measures were reviewed for inclusion and exclusion criteria and verified the psychometric properties; then, the outcome measures were grouped to identify unrepresented constructs. After review, we determined that all identified measures assessed constructs that are listed earlier, and no other constructs were identified.
- *Step 3*. The Task Force reviewed the measures to add outcome measures to the list for screening based on clinical expertise. The total number of measures identified was 93. Appendix 1 provides details on all considered outcome measures.
- *Step 4*. A public call through APTA Engage asked volunteers to identify any "crucial outcome measures" related to each construct for this population that they believe the Task Force had not yet identified and provide resources as able. The Task Force Chair announced the public call to all academies and sections via the panel. Sections and academies who chose to participate used individualized approaches to make their members aware of the public call. Fifty-two respondents from various clinical practice settings and various specialty areas identified 7 additional measures that fit within the key constructs. These additional measures were screened against the criteria, and those that met inclusion and exclusion criteria were added to the discussion for core outcome measures. The Task Force reviewed all these responses to determine whether the outcome measures addressed the constructs of interest or whether other constructs needed to be included. This process verified that the original 6 constructs covered the spectrum of movement dysfunction that professionals were noting in practice.

The Task Force recognized early that the core set of outcome measures would exhibit floor and ceiling effects in some patients dependent on the patient's functional level at the time of testing. These floor and ceiling effects would need to be addressed through additional, secondary measures. A clinical application algorithm has been developed and published on the APTA Web site, which addresses these and other limitations of the recommended core outcome measure set.[18] Additional outcome measures identified through the 4-step expedited review process that did not meet established criteria were considered for inclusion in the algorithm in areas where no measure met all criteria.

SELECTION OF THE CORE OUTCOME MEASURES

Group consensus of the Task Force was reached to consider the outcome measure characteristics listed in Table 1 when determining a final core set. Additional consideration was given to measures that could capture more than one construct while providing the opportunity to evaluate constructs individually via subscales or that could be related to other measures to span the continuum of care. These recommendations are consistent with those made by other outcome measure task forces.[14,19,20] For example, the Short Physical Performance Battery (SPPB) was selected because it addresses the constructs of both functional mobility and strength/power by utilizing gait speed and a variation of the 5 Times sit-to-stand test, which is recommended as a core outcome measure within the neurologic population,[14] and developed by the National Institute of Aging[21] to assess lower extremity functioning in older adults. Moreover, the SPPB has high clinical utility in a diverse range of clinical populations.[22-26] The 2-minute step test was selected for endurance because of its utility in acutely ill and active populations and because the testing conditions, like step height, could be varied and exercise capacity and metabolic equivalent of task level performance could be calculated to illustrate progress, as functional capacity improved.[20]

As the Task Force compiled and sorted the outcome measures into the constructs, the Task Force members had poor agreement in the inclusion of any upper extremity functional outcome measure. The Task Force put forth additional efforts to identify an appropriate measure for this construct. We determined that the accepted constructs captured global deficits of patients and were sufficient to screen for upper extremity deficits. Further recommendations regarding upper extremity testing can be found in the algorithm.[18]

Finally, all identified measures that met inclusion and exclusion criteria were discussed by construct grouping to identify potential core outcome measures. Candidate measures were compared based on the previously presented criteria. Measures were removed from consideration when a consensus of the Task Force was reached that other measures within the construct category would be more appropriate as a core measure. This process was repeated for each construct until final candidate core outcome measures were identified (see Figure 2). Appendix 1 provides a brief rationale for the removal of each test. Psychometric properties of all final candidate core outcome measures were fully investigated using the original literature (see Table 2).

The measures identified as candidates from each construct were compiled into a list of potential core outcome measures and reviewed for overall time requirements, ability to capture functional levels and be used

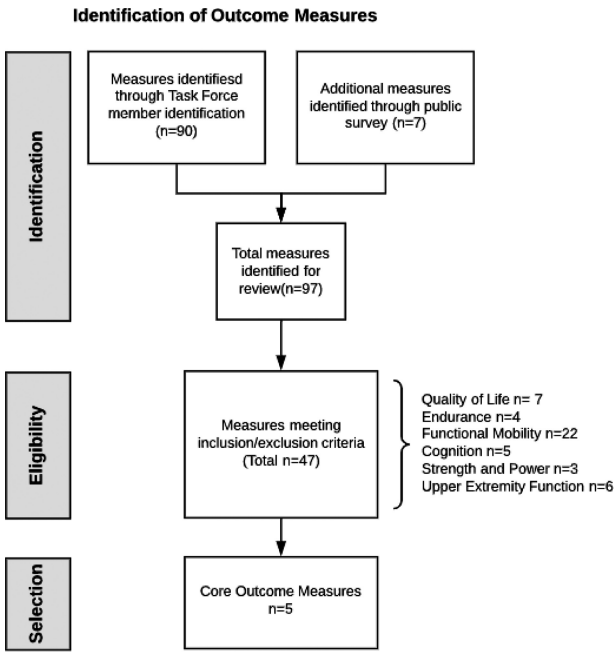

FIGURE 2. Identification of Outcome Measures.

in all settings, possible overlap of constructs, and ability to capture the core standardized tasks recommended in the movement system diagnosis approach.[63] This resulted in identifying a final recommended set of core outcome measures (see Table 3) that was unanimously agreed to on May 29, 2020. A description of domains measured by each measure, scoring, and equipment, space, and time requirements is provided in Table 4. The APTA provided support throughout the process and published the measures on their Web site on June 29, 2020, https://www.apta.org/your-practice/outcomes-measurement/covid-19-core-outcome-measures.[18]

The Task Force recommends using the core set across all clinical settings for patients who have functional impairments secondary to COVID-19 and who have goals to improve in the associated constructs or may have unmonitored cognitive decline that may impact the plan of care. Recommendations for the order of completion of the core measures can be found within the algorithm. A review of these recommendations is encouraged to ensure proper monitoring of cardiovascular and pulmonary tolerance, as the demands of the core outcome measures increase. The algorithm order is progressive in energy demand to decrease fatigue, as patients progress down the pathway.

Clinicians should record the raw data for each measure along with the total score. The raw data allow the clinician and patient to see subtle changes in performance that may be missed when the raw data are converted to an ordinal or nominal score. An example of the use of raw scores on the Medical Research Council-Sum Score (MRC-SS) demonstrates the value of recording raw scores and not just total scores. Consider a patient showing upper extremity strength improvement over the last 6 weeks but is earning a consistent score of 42/60 on the MRC-SS during repeated testing. The raw data illustrate that strength in the upper extremities has been showing continued gains in manual muscle testing. However, an actual loss of muscle performance of the lower extremities has resulted in the same sum score on the MRC-SS. Recording raw scores would be critical to understanding the functional changes of this patient. In addition, the inclusion of raw data for the components of the SPPB, such as recording the gait speed and time for the balance test, may add to the clinical picture for the patient.

We also recommended that clinicians record a score of 0 when a patient is unable to complete a core measure. Having the clinician attempt all the core outcome measures and record a 0 on a test is meaningful information for the next testing cycle and aids in appreciating the trajectory of recovery of the population of patients with and recovering from COVID-19. The importance of 0 can be demonstrated when we consider the SPPB. A patient who is very low functioning due to experiencing the severe effects of COVID-19 may score a 0 on the SPPB for 4 weeks, but on serial testing, the score progresses to a 2/12 at 8 weeks; therefore, the clinician has captured meaningful functional improvements for this patient.[54] If we carry this clinical scenario out further, at 3 months, the patient scores a 10/12 at 12 weeks.

As data become available, recommended measures can be reviewed and adjusted. Waiting for research to be completed before recommending a core set of outcome measures for this population would mean that a wealth of potentially beneficial data may be lost due to a lack of consistent outcome measurement across the continuum of care. The panel and the Task Force's goal is to be proactive, to improve our ability as a profession to demonstrate our value in the care of these patients, and to be able to identify early on best practices for rehabilitating these patients.

TESTING INTERVALS

The Task Force recommends that the core set of measures be completed according to the following prescriptive timeline to understand best the effect of therapy on functional performance and the recovery trajectory.[14] All testing should be done upon initial entry to physical therapy services. The Task Force recognizes the clinician may need 1 to 2 visits to complete the assessment due to patient tolerance. The core set should be done prior to discharge from the current setting and entry to the next setting. Additionally, the core set should be completed when a significant change in clinical presentation occurs. Finally, completing testing at 30 days, 3, 6 and, 12 months post-diagnosis while under physical therapy's care would be extremely helpful and support researchers' efforts to study this novel disease and the response

TABLE 2. Psychometric Properties of Core Measures	
St Louis University Mental Status (SLUMS) examination to assess cognition	
Interrater reliability	Excellent interrater reliability (ICC = 0.990-0.998) and intrarater reliability (ICC = 0.968)[27]
Test-retest reliability	Regression fit: R^2, 0.67; r, 0.82[28]
Validity	When compared to the Mini-Mental Status Examination (MMSE), the SLUMS has similar sensitivities, specificities, and area under the curve in detecting dementia and may be better at detecting mild cognitive decline.[29] The Montreal Cognitive Assessment (MoCA) and the SLUMS have excellent convergent validity (r = 0.91)[30]
Cut-off scores	In individuals with at least a high school education, a score <26 indicates mild cognitive impairment and a score <22 indicates possible dementia. In individuals with less than a high school education, scores <24 indicate mild impairment and <20 indicate possible dementia.[29]
Patient-Reported Outcomes Measurement Information System (PROMIS) Global 10 to assess health-related quality of life	
Interrater reliability	N/A (survey)
Test-retest reliability	Test-retest in stroke was high across all items.[31] Stroke = Responses on all PROMIS Global-10 items were significantly associated with their prior responses (correlation coefficients ≥0.80). Test-retest reliability for PROMIS Global-10 was high (ICC ≥0.85) in patients with lupus.[31]
Validity	Correlates with ED-Q5 at 0.72[32] Correlates with SF-36 = 0.90 (physical function items) = 0.85 (mental health items)[33]
MDC/MCID	MCID of 1 standard deviation from the mean is recommended[34,35]
Cut-off scores	PROMIS measures would use the T-score metric, in which scores have a mean of 50 and a standard deviation of 10 compared with the general population[33]
European-Quality of Life-5D-5L (EQ-5D-5L) to assess health-related quality of life when participating in international studies or when a language barrier prevents the use of PROMIS Global-10	
Interrater reliability	N/A (survey)
Test-retest reliability	For individuals with chronic diseases, the ICC was 0.82.[36]
Validity	Construct validity: for people with hip/knee ostearthritis—high correlation with Western Ontario and McMaster Universities Arthritis Index (WOMAC).[37] No correlation with COPD assessment tool/St George's Respiratory Questionnaire for patents participating in pulmonary rehabilitation,[38] for patients who have had a cerebrovascular accident, poor validity with modified Rankin tool, and Barthel index[39] Content validity: For people with ostearthritis, higher with the 5 levels (5L) rather than 3 levels (3L) especially in domains of mobility, usual activities, and pain/discomfort[37]
MDC/MCID	For individuals with hip/knee ostearthritis MDC = 0.30 in the utility index; MCID = 0.07 pts for improvement, −0.05 patients for becoming worse, 0.32 pts for improvement in patients who did not have symptoms.[40] For patients participating in pulmonary rehabilitation MCID = 0.054 for utility index and 6.99 for the visual analog scale question.[38]
Cut-off scores	No cut-off scores were identified
Medical Research Council-Sum Score (MRC-SS) to assess strength and power	
Interrater reliability	The ICC was 0.94 (95% confidence interval (CI): 0.85-0.98) in patients in the intensive care unit.[41]
Test-retest reliability	Manual muscle testing: For intensive care unit survivors and simulated patients, the ICC was 0.62-100 for the upper extremities and 0.66-1.00 for the lower extremities.[42]
Validity	MRC-SS has moderate to strong correlations with physical performance measures, predictive of 2-y outcomes, and independently predicted delayed extubation in clients with critical illness.[43-45]
MDC/MCID	MCID of 1 standard deviation from the mean is recommended[35]

(continues)

TABLE 2. Psychometric Properties of Core Measures (*Continued*)	
Cut-off scores	<48/60 on MRC-SS for patients surviving a critical illness that required an intensive care unit admission is considered Intensive Care Unit-Acquired Weakness (ICU-AW)[46,47]
Short Physical Performance Battery (SPPB) to assess functional mobility	
Interrater reliability	Interrater reliability is excellent (ICC = 0 .92) in patients with COPD.[48]
Test-retest reliability	Test-retest reliability is excellent (ICC ranged from 0.81 to 0.91) in older adults.[49-51]
Validity	In community-dwelling older adults scores of ≦10 had significantly higher odds of mobility disability at follow-up (odds ratio: 3.38, 95% CI).[52] SPPB scores ≦10 predictive of all-cause mortality in meta-analysis of 17 studies.[53]
MDC/MCID	Range from 0.54 to 2.9 depending on patient population.[49,51,54]
Cut-off scores	Score of <9: poor physical performance and indicative of physical frailty.[55] Score of ≤6: associated with a high fall rate; a score of 7-9: identifies high risk for recurrent falls in women[56]
2-min step test (2MST) to assess endurance	
Interrater reliability	Excellent interrater reliability (ICC = 0.999-1.000, $P < .0001$).[57]
Test-retest reliability	Excellent relative test-retest reliability (ICC = 0.927-0.934, $P < .0001$).[57]
Validity	Convergent = excellent with exercise/activity history or fitness levels. Additional evidence for "relationship between 2MST steps and psycho/cognitive measures, activity performance, training and health status, and age."[58]
MDC/MCID	It is recommended metabolic equivalents (METs) are calculated. [MET = VO_2/3.5 and VO_2 = (0.2 × frequency of stepping rate) + 1.33 × 1.8 (step height in meters)(stepping rate) + 3.5] [20]studies have found that even an increase of 1 MET level can equate to clinically important changes in individuals in cardiac rehabilitation.[20,59-61]
Cut-off scores	Age-related norms for older adults reported by mean (standard deviation): age 60-64: men 101,[21] women 91[24]; age 65-69: men 101,[23], women 90[26]; age 70-74: men 95,[23] women 84[25]; age 75-79: men 91,[27] women 84[24]; age 80-84: men 87,[24] women 75[23]; age 85-89: men 75,[24] women 70[22]; age 90-94: men 69,[26] women 58[21,62]
COPD, chronic obstructive pulmonary disease; ICC, intraclass correlation coefficient; MCID, minimal clinically important difference; MDC, minimal detectable change.	

to interventions thoroughly. The Task Force anticipates a need to consider longer-term testing recommendations such as 2, 3, 5 and 10-year testing intervals based upon longitudinal tracking of other diseases and syndromes.[67,68] However, an alternative frequency schedule is recommended for the St Louis University Mental Status examination due to the possible learning effect of repeated testing. The panel recommends completing cognitive testing if no prior score has been attained, at 1 to 2 months following hospital discharge (when transient

TABLE 3. Core Outcome Measure Recommendations[a]	
Constructs	**Core Outcome Measure**
Cognition	Saint Louis University Mental Status (SLUMS) Examination
Health-related quality of life	PROMIS Global-10 (health-related quality of life measure) or EQ-5D-5L when completing research for an international audience
Muscle strength and power	Medical Research Council-Sum Score (MRC-SS)
Functional mobility	Short Physical Performance Battery (SPPB) • 4-m gait speed subscore • 5 times sit-to-stand subscore • Balance screen
Endurance	2-min step test
[a]Recommendations: include all raw data to examine detailed changes in performance.	

TABLE 4. COVID-19 Recommended Core Outcome Measures Described

COVID Core Outcome Measures	Domains Measured	Scoring	Equipment and Space Needed	Time to Complete
2-min step test[58]	Measures cardiovascular endurance. Highly correlated with the 6-minute walk test.	Number of times the right knee reaches the appropriate height in 2 min	Timing device and tape measure and tape.	5 min
EQ-5D-5L [64]	Measures health-related quality of life in the 5 dimensions of mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.	Describes health status in 5 levels from having no problems to being unable/having extreme problems	Paper and pencil or computer based. Can be self-completed.	<5 min
PROMIS Global-10[65]	Consists of 10 global health items that represent 5 core PROMIS domains (physical function, pain, fatigue, emotional distress, and social health). Four items are used to assess global physical health.	Lower scores indicate more of the construct being measured (eg, more fatigue and more limitations of physical function). Uses a T-score metric in which 50 is the mean of a relevant reference population and 10 is the standard deviation (SD) of that population.	Self-completed either on paper or computer/tablet.	1.8 min
Medical Research Council-Sum Score (MRC-SS)[66]	Muscle strength in supine position for shoulder abduction, elbow flexion, wrist extension, hip flexion, knee extension, and ankle dorsiflexion	The total MRC sum score ranges from 0 (total paralysis) to 60 (normal strength). The score is the sum of the MRC score of 6 muscles (3 at the upper and 3 at the lower limbs) on both sides, each muscle graded from 0 to 5.	Bed, goniometer to measure bed recline	5-10 min
St Louis University Mental Status (SLUMS)[29]	Identify individuals with neurocognitive impairment and is sensitive to change over time. Measures attention, immediate and delayed recall, orientation, numeric calculation and registration, visual spatial relations, executive functioning, and extrapolation.	It is a 30-point, 11-question questionnaire that tests orientation, memory, attention, and executive function. Higher scores indicate better performance.	Paper and pencil. Clinician-administered examination.	7 min
Short Physical Performance Battery (SPPB)[21]	Measures balance in standing in progressively narrowed base of support, 3-m or 4-m walk, and time to rise from a chair 5 times.	Scores range from 0 to 12, with higher scores indicating better performance.	Requires a chair, timer, and 3-m (9.8 ft) area to walk in. May use an assistive device or chair with arms if needed.	10 min

cognitive decline is expected to have resolved),[17] with a decline in condition during postacute care, or at discharge from all physical therapy services when a prior deficit was noted, and the patient is not receiving skilled care from another specialist for cognitive impairments. This frequency schedule has a primary goal to ensure that the therapist is properly adapting the plan of care to the current capabilities of the patient while working to ensure that physical therapists are identifying and referring patients appropriately when cognitive decline may indicate underlying neurologic pathology. Physical therapists should refer the patient when a cognitive decline is initially noted, worsens without explanation, or persists beyond 1 to 2 months after hospital discharge.[17]

LIMITATIONS

The Task Force recognized and discussed several limitations regarding the process for determining the recommendations for the set of core measures. First, the Task Force relied on work that had already been published,

such as clinical practice guidelines, systematic reviews, meta-analysis papers, and Web sites including sralab.org/rehabilitation-measures and PTNow.org, to generate the original outcomes. The Task Force only explored primary articles where gaps existed in the principal sources. The Task Force also relied on selecting measures based upon other diseases and syndromes with a presentation similar to that of patients with or recovering from COVID-19, such as measures recommended for post-ICU syndrome, heart, and lung failure, or Guillain-Barré syndrome. COVID-19 is a novel disease, and the Task Force may have missed key constructs to assess and key outcome measures that may be more sensitive to detect change. Finally, the Task Force recognized and debated the limitations we had imposed on ourselves by our inclusion/exclusion criteria. Measures such as the 6-minute walk test and handgrip strength with well-established value were considered. However, the Task Force appreciated that these measures might be impractical or impossible to administer depending on the limitations of the setting and infection prevention guidelines. We also eliminated some valuable measures because of copyright or license requirements that pose an increased burden to many sites. Thus, many of these common clinical measures were considered, as we developed recommendations for secondary measures within a clinical application algorithm.[18] Finally, the Task Force recognizes the clinician may experience a floor or ceiling effect by only using the core set of outcome measures when assessing a patient with either a low or high level of physical function.

NEXT STEPS

To address the issue of floor and ceiling effects, return to work and sports, and community reintegration, and to provide recommendations for screening and additional tests and measures based on the patient's clinical presentation, the Task Force has designed a clinical application algorithm to guide clinicians using the core measures along with secondary measures to address a large range of functional levels. This can be found at https://www.apta.org/your-practice/outcomes-measurement/covid-19-core-outcome-measures.[18] At the time of submission, a team was completing a second manuscript that demonstrates the utility of the clinical application algorithm through patient case scenarios in different clinical practice settings and severities of movement impairments and dysfunction. A subgroup of the Task Force is also applying a similar approach to reach recommendations for the pediatric population.

CONCLUSION

Due to the novelty of COVID-19, no outcome measure data specific to COVID-19 could be integrated into our process. These recommendations are considered an early guide to provide clinical recommendations based on expert consensus and to align clinical practice to allow for greater compilation of data for research. This should speed the process of ascertaining what outcome measures may be valid and reliable within this population. As such, updates to these core outcome measures will be required as more is learned about COVID-19.

ACKNOWLEDGMENTS

Thank you to Lori Michener and Kirsten Potter, who provided valuable insight at key times within our process.

REFERENCES

1. World Health Organization. Coronavirus Disease (COVID-19) Situation Report–112. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200511-covid-19-sitrep-112.pdf?sfvrsn=813f2669_2.
2. Johns Hopkins Coronavirus Resource Center. New COVID-19 cases worldwide. https://coronavirus.jhu.edu/data/new-cases.
3. Grasselli G, Zangrillo A, Zanella A, et al. Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy region, Italy. *JAMA.* 2020;323(16):1574-1581.
4. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area. *JAMA.* 2020;323(20):2052-2059.
5. Toresdahl BG, Asif IM. Coronavirus disease 2019 (COVID-19): considerations for the competitive athlete. *Sports Health.* 2020;12(3):221-224.
6. McMichael TM, Currie DW, Clark S, et al. Epidemiology of COVID-19 in a long-term care facility in King County, Washington. *N Engl J Med.* 2020;382(21):2005-2011.
7. Paterson RW, Brown RL, Benjamin L, et al. The emerging spectrum of COVID-19 neurology: clinical, radiological and laboratory findings. *Brain.* 2020;143(10):3104-3120.
8. Iannaccone G, Scacciavillani R, Del Buono MG, et al. Weathering the cytokine storm in COVID-19: therapeutic implications. *Cardiorenal Med.* 2020;10(5):277-287.
9. Falvey JR, Ferrante LE. Flattening the disability curve: rehabilitation and recovery after COVID-19 infection. *Heart Lung*. 2020;49(5):440-441.
10. Falvey JR, Krafft C, Kornetti D. The essential role of home- and community-based physical therapists during the COVID-19 pandemic. *Phys Ther*. 2020;100(7):1058-1061.
11. Grabowski DC, Joynt Maddox KE. Postacute care preparedness for COVID-19: thinking ahead. *JAMA.* 2020;323(20):2007-2008.
12. American Physical Therapy Association. *Guide to Physical Therapist Practice 3.0*. Alexandria, VA: American Physical Therapy Association; 2016.
13. Jette DU, Halbert J, Iverson C, Miceli E, Shah P. Use of standardized outcome measures in physical therapist practice: perceptions and applications. *Phys Ther.* 2009;89(2):125-135.
14. Moore JL, Potter K, Blankshain K, Kaplan SL, O'Dwyer LC, Sullivan JE. A core set of outcome measures for adults with neurologic conditions undergoing rehabilitation: a clinical practice guideline. *J Neurol Phys Ther*. 2018;42(3):174-220.
15. Zhou H, Lu S, Chen J, et al. The landscape of cognitive function in recovered COVID-19 patients. *J Psychiatr Res*. 2020;129:98-102.
16. Andrianopoulos V, Gloeckl R, Vogiatzis I, Kenn K. Cognitive impairment in COPD: should cognitive evaluation be part of respiratory assessment? *Breathe.* 2017;13(1):e1-e9.
17. Gordon SM, Jackson JC, Ely EW, Burger C, Hopkins RO. Clinical identification of cognitive impairment in ICU survivors: insights for intensivists. *Intensive Care Med.* 2004;30(11):1997-2008.
18. Martin R, Botkin R, Campbell A, et al. *COVID-19 Core Outcome Measures, APTA Academies and Sections Consensus Statement*. https://www.apta.org/your-practice/outcomes-measurement/covid-19-core-outcome-measures. Published 2020.
19. Spruit M, Jenkins A, Holland A, Force A-hIT. *Report of an Ad-Hoc International Task Force to Develop an Expert-Based Opinion on Early and Short-Term Rehabilitative Interventions After an Acute Hospital Setting in COVID 19 Survivors*. New York, NY: European Respiratory Society; 2020.

20. Riebe D, Ehrmand J, Liguori G, Magal M. *American College of Sports Medicine's Guidelines for Exercise Testing and Prescription*. Philadelphia, PA: Wolters Klumer; 2018.
21. National Institute on Aging, Guralnik J. Short Physical Performance Battery (SPPB): Assessing Physical Performance in the Older Patient. https://www.nia.nih.gov/research/labs/leps/short-physical-performance-battery-sppb.
22. Essam Behiry M, Mogawer S, Yamany A, et al. Ability of the Short Physical Performance Battery Frailty Index to predict mortality and hospital readmission in patients with liver cirrhosis. *Int J Hepatol.* 2019;2019:8092865.
23. Gandotra S, Lovato J, Case D, et al. Physical function trajectories in survivors of acute respiratory failure. *Ann Am Thorac Soc* 2019;16(4):471-477.
24. Mayer KP, Henning AN, Gaines KM, et al. Physical function measured prior to lung transplantation is associated with posttransplant patient outcomes. *Transplant Proc*. 2021;53(1):288-295.
25. Patel MS, Kon SSC, Canavan JL, et al. The Short Physical Performance Battery in COPD. *Am J Respiratory Crit Care Med*. 2013;187:A1366.
26. Verweij NM, Schiphorst AHW, Pronk A, van den Bos F, Hamaker ME. Physical performance measures for predicting outcome in cancer patients: a systematic review. *Acta Oncol (Madr)*. 2016;55(12):1386-1391.
27. Zhang S, Wu YH, Zhang Y, Zhang Y, Cheng Y. Preliminary study of the validity and reliability of the Chinese version of the Saint Louis University Mental Status Examination (SLUMS) in detecting cognitive impairment in patients with traumatic brain injury [published online ahead of print October 24, 2019]. *Appl Neuropsychol Adult*. doi:10.1080/23279095.2019.1680986.
28. Cahn-Hidalgo D, Estes PW, Benabou R. Validity, reliability, and psychometric properties of a computerized, cognitive assessment test (Cognivue®). *World J Psychiatry.* 2020;10(1):1-11.
29. Tariq SH, Tumosa N, Chibnall JT, Perry MH 3rd, Morley JE. Comparison of the Saint Louis University Mental Status examination and the mini-mental state examination for detecting dementia and mild neurocognitive disorder—a pilot study. *Am J Geriatr Psychiatry*. 2006;14(11):900-910.
30. Stewart S, O'Riley A, Edelstein B, Gould C. A preliminary comparison of three cognitive screening instruments in long term care: the MMSE, SLUMS, and MoCA. *Clin Gerontol*. 2012;35:57-75.
31. Katzan IL, Lapin B. PROMIS GH (Patient-Reported Outcomes Measurement Information System Global Health) Scale in stroke: a validation study. *Stroke*. 2018;49(1):147-154.
32. Saad MA, Kassam HF, Suriani RJ Jr, Pan SD, Blaine TA, Kovacevic D. Performance of PROMIS Global-10 compared with legacy instruments in patients with shoulder arthritis. *J Shoulder Elbow Surg*. 2018;27(12):2249-2256.
33. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol*. 2010;63(11):1179-1194.
34. Weaver KE, Forsythe LP, Reeve BB, et al. Mental and physical health-related quality of life among U.S. cancer survivors: population estimates from the 2010 National Health Interview Survey. *Cancer Epidemiol Biomarkers Prev*. 2012;21(11):2108-2117.
35. Rai SK, Yazdany J, Fortin PR, Aviña-Zubieta JA. Approaches for estimating minimal clinically important differences in systemic lupus erythematosus. *Arthritis Res Ther*. 2015;17(1):143.
36. Sakthong P, Sonsa-Ardjit N, Sukarnjanaset P, Munpan W. Psychometric properties of the EQ-5D-5L in Thai patients with chronic diseases. *Qual Life Res*. 2015;24(12):3015-3022.
37. Bilbao A, García-Pérez L, Arenaza JC, et al. Psychometric properties of the EQ-5D-5L in patients with hip or knee osteoarthritis: reliability, validity and responsiveness. *Qual Life Res*. 2018;27(11):2897-2908.
38. Nolan CM, Longworth L, Lord J, et al. The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. *Thorax*. 2016;71(6):493-500.
39. Golicki D, Niewada M, Karlińska A, et al. Comparing responsiveness of the EQ-5D-5L, EQ-5D-3L and EQ VAS in stroke patients. *Qual Life Res.* 2015;24(6):1555-1563.
40. Conner-Spady BL, Marshall DA, Bohm E, et al. Reliability and validity of the EQ-5D-5L compared to the EQ-5D-3L in patients with osteoarthritis referred for hip and knee replacement. *Qual Life Res*. 2015;24(7):1775-1784.
41. Connolly BA, Jones GD, Curtis AA, et al. Clinical predictive value of manual muscle strength testing during critical illness: an observational cohort study. *Crit Care*. 2013;17(5):R229.
42. Fan E, Ciesla ND, Truong AD, Bhoopathi V, Zeger SL, Needham DM. Inter-rater reliability of manual muscle strength testing in ICU survivors and simulated patients. *Intensive Care Med*. 2010;36(6):1038-1043.
43. Mayer K, Thompson Bastin M, Montgomery-Yates A, et al. Acute skeletal muscle wasting and dysfunction predict physical disability at hospital discharge in patients with critical illness. *Critical Care.* 2020;24(1):637.
44. Saccheri C, Morawiec E, Delemazure J, et al. ICU-acquired weakness, diaphragm dysfunction and long-term outcomes of critically ill patients. *Ann Intensive Care*. 2020;10(1):1.
45. Eggmann S, Luder G, Verra ML, Irincheeva I, Bastiaenen CHG, Jakob SM. Functional ability and quality of life in critical illness survivors with intensive care unit acquired weakness: a secondary analysis of a randomised controlled trial. *PLoS One.* 2020;15(3):e0229725.
46. Parry SM, Berney S, Granger CL, et al. A new two-tier strength assessment approach to the diagnosis of weakness in intensive care: an observational study. *Crit Care*. 2015;19(1):52.
47. Jolley SE, Bunnell AE, Hough CL. ICU-acquired weakness. *Chest*. 2016;150:1129-1140.
48. Medina-Mirapeix F, Bernabeu-Mora R, Llamazares-Herrán E, Sánchez-Martínez MP, García-Vidal JA, Escolar-Reina P. Interobserver reliability of peripheral muscle strength tests and Short Physical Performance Battery in patients with chronic obstructive pulmonary disease: a prospective observational study. *Arch Phys Med Rehabil.* 2016;97:2002-2005.
49. Olsen CF, Bergland A. Reliability of the Norwegian version of the Short Physical Performance Battery in older people with and without dementia. *BMC Geriatrics*. 2017;17(1):124.
50. Gómez JF, Curcio CL, Alvarado B, Zunzunegui MV, Guralnik J. Validity and reliability of the Short Physical Performance Battery (SPPB): a pilot study on mobility in the Colombian Andes. *Colomb Med (Cali)*. 2013;44(3):165-171.
51. Mangione KK, Craik RL, McCormick AA, et al. Detectable changes in physical performance measures in elderly African Americans. *Phys Ther*. 2010;90(6):921-927.
52. Vasunilashorn S, Coppin AK, Patel KV, et al. Use of the Short Physical Performance Battery Score to predict loss of ability to walk 400 meters: analysis from the InCHIANTI study. *J Gerontol A Biol Sci Med Sci*. 2009;64(2):223-229.
53. Pavasini R, Guralnik J, Brown JC, et al. Short Physical Performance Battery and all-cause mortality: systematic review and meta-analysis. *BMC Med*. 2016;14(1):215.
54. Perera S, Mody SH, Woodman RC, Studenski SA. Meaningful change and responsiveness in common physical performance measures in older adults. *J Am Geriatr Soc*. 2006;54(5):743-749.
55. da Câmara SM, Alvarado BE, Guralnik JM, Guerra RO, Maciel AC. Using the Short Physical Performance Battery to screen for frailty in young-old adults with distinct socioeconomic conditions. *Geriatr Gerontol Int*. 2013;13:421-428.
56. Veronese N, Bolzetta F, Toffanello ED, et al. Association between Short Physical Performance Battery and falls in older people: the Progetto Veneto Anziani Study. *Rejuvenation Res*. 2014;17(3):276-284.
57. Chow JLJ, Tony S, Chow Y, Tan D, Rand S, Fitzgerald C. The 2-minute step test: a reliable and valid alternative to the 6-minute walk test in measuring aerobic capacity of older adults post coronary revascularisation. *Eur J Preventive Cardiol*. 2017;24:S79.
58. Bohannon RW, Crouch RH. Two-minute step test of exercise capacity: systematic review of procedures, performance, and clinimetric properties. *J Geriatr Phys Ther*. 2019;42(2):105-112.
59. Wicks JR, Oldridge NB, Nielsen LK, Vickers CE. HR index—a simple method for the prediction of oxygen uptake. *Med Sci Sports Exerc*. 2011;43(10):2005-2012.
60. Glass S, GBD. *ACSM's Metabolic Calculations Handbook*. Philadelphia, PA: Lippincott Williams and Wilkins; 2007.
61. Franklin BA, Lavie CJ, Squires RW, Milani RV. Exercise-based cardiac rehabilitation and improvements in cardiorespiratory fitness: implications regarding patient benefit. *Mayo Clin Proc*. 2013;88(5):431-437.
62. Rikli RE, Jones CJ. Functional fitness normative scores for community-residing older adults, ages 60-94. *J Aging Phys Activity*. 1999;7(2):162-181.

63. Hedman LD, Quinn L, Gill-Body K, et al. White paper: movement system diagnoses in neurologic physical therapy. *J Neurol Phys Ther*. 2018;42(2):110-117.
64. American Physical Therapy Association. Quality of Life: EQ-5D-5L (Health-Related Quality of Life Measure). https://www.apta.org/patient-care/evidence-based-practice-resources/test-measures/quality-of-life-eq-5d-5l-health-related-quality-of-life-measure. Published 2020.
65. American Physical Therapy Association. Patient-Reported Outcomes Measurement Information System Global-10, PROMIS Global-10. https://www.apta.org/patient-care/evidence-based-practice-resources/test-measures/patient-reported-outcomes-measurement-information-system-global-10-promis-global-10. Published 2020.
66. Vanpee G, Hermans G, Segers J, Gosselink R. Assessment of limb muscle strength in critically ill patients: a systematic review. *Crit Care Med*. 2014;42(3):701-711.
67. American Physical Therapy Association. Annual Checkup—Why provide an annual checkup. https://www.apta.org/patient-care/interventions/annual-checkup. Published 2018.
68. Østbye T, Yarnall KS, Krause KM, Pollak KI, Gradison M, Michener JL. Is there time for management of patients with chronic diseases in primary care? *Ann Fam Med*. 2005;3(3):209-214.
69. Cummings-Vaughn LA, Chavakula NN, Malmstrom TK, Tumosa N, Morley JE, Cruz-Oliver DM. Veterans Affairs Saint Louis University Mental Status examination compared with the Montreal Cognitive Assessment and the Short Test of Mental Status. *J Am Geriatr Soc*. 2014;62(7):1341-1346.
70. Verghese J, Buschke H, Viola L, et al. Validity of divided attention tasks in predicting falls in older individuals: a preliminary study. *J Am Geriatr Soc*. 2002;50(9):1572-1576.
71. American Physical Therapy Association Home Health Section. The Home Health Section Toolbox of Standardized Tests & Measures. https://aptahhs.memberclicks.net/assets/docs/Home%20Health%20Toolbox.pdf. Published 2013.
72. Graham RC, Smith NM, White CM. The reliability and validity of the physiological cost index in healthy subjects while walking on 2 different tracks. *Arch Phys Med Rehabil*. 2005;86(10):2041-2046.
73. Fritz S, Lusardi M. White paper: "Walking Speed: the Sixth Vital Sign." *J Geriatr Phys Ther*. 2009;32(2):46-49.
74. Renfro M, Bainbridge DB, Smith ML. Validation of evidence-based fall prevention programs for adults with intellectual and/or developmental disorders: a modified Otago exercise program. *Front Public Health*. 2016;4:261.
75. Mikkelsen ME, Still M, Anderson BJ, et al. Society of Critical Care Medicine's International Consensus Conference on Prediction and Identification of Long-Term Impairments After Critical Illness. *Crit Care Med*. 2020;48(11):1670-1679.
76. National Institutes of Health. *P*atient-Reported Outcomes Measurement Information System. https://commonfund.nih.gov/promis/index. Published 2019.
77. Needham DM, Davidson J, Cohen H, et al. Improving long-term outcomes after discharge from intensive care unit: report from a stakeholders' conference. *Crit Care Med*. 2012;40(2):502-509.
78. Needham DM, Sepulveda KA, Dinglas VD, et al. Core Outcome measures for clinical research in acute respiratory failure survivors. an international modified Delphi Consensus Study. *Am J Respir Crit Care Med*. 2017;196(9):1122-1130.
79. Dale LM, Strain-Riggs SR. Comparing responsiveness of the quick disabilities of the arm, shoulder, and hand and the upper limb functional index. *Work*. 2013;46(3):243-250.

APPENDIX 1. Overview of All Considered Outcome Measures		
	Cognition	**Reason for Removal from Consideration**
1.	*Confusion Assessment Method for Delirium	For delirium rather than overall cognition
2.	*Mini-COG *Richmond Agitation Scale	Screening tool rather than outcome measure For agitation and sedation rather than overall cognition
3.	*Short Blesses Test	Screening tool rather than outcome measure
4.	**St Louis University Mental Status Examination*	SELECTED AS A CORE MEASURE. This was the only outcome measure identified meeting inclusion/exclusion criteria to measure the various constructs within cognition that may be affected by COVID-19. Strong interrater reliability. Available in 22 languages. Free training. Comparable to the MoCA in screening for MCI and dementia.[69]
5.	Mini-Mental Status Examination	Not free
6.	Montreal Cognitive Assessment	Not free
7.	Timed Up and Go Cognitive	Reliability unknown
8.	Trail Making Test	Measures limited constructs within cognition
9.	Walking While Talking Examination	Reliability of 0.6[70]
	Endurance	
10	**2-minute step test*	SELECTED AS A CORE MEASURE. Previously recommended by the Home Health Section.[71]
11	*Chester Step Test	Conflicting resources regarding cost to use
12	*Modified Harvard Step Test	Very similar to chosen core measure but not previously recommended by a section to the best of this Task Force's knowledge
13	1-Item Fatigue Questionnaire	Unknown test-retest reliability and measures fatigue rather than endurance
14	2-Item Fatigue Questionnaire	Unknown test-retest reliability and measures fatigue rather than endurance
15	2-minute walk test	Space requirements exceed 10-ft by 12-ft area
16	6-minute arm test	Requires specialized equipment not commonly available in many settings
17	6-minute walk test	Space requirements exceed 10-ft by 12-ft area
18	10-m Shuttle	Space requirements exceed 10-ft by 12-ft area
19	Counting Talk Test	Only found for use within healthy populations
20	Physiological Cost Index	Reliability below 0.75 cut-off[72]
21	Seated Step Test	Reliability below 0.75 cut-off with caveats by the authors indicating it was likely better than the study showed[72]
22	Talk Test	Measures exercise intensity, but not necessarily endurance
	Functional Ability	
23	*3-m Gait Speed Test	Gait speed was a component of the selected core measure, which was able to capture more meaningful data with only a minimal increase in time
24	*4 Square Test	Many individuals are likely to experience a floor effect
25	*Modified 4 Square Test	Many individuals are likely to experience a floor effect
26	*5 times Sit-to-Stand	Very similar to a component of the selected core measure, which was able to capture more meaningful data with only a minimal increase in time
27	*360 Degree Turn Test	Research completed within a limited, nonrelated population
28	*Acute Care Index of Function	Focused on the inpatient setting

(continues)

APPENDIX 1. Overview of All Considered Outcome Measures (*Continued*)		
29	*Barthel Index	Many individuals are likely to experience a ceiling effect
30	*Berg Balance Scale	Required 15-20 minutes to administer, significantly longer than selected measure in this construct
31	*Modified Falls Efficacy Scale	Removed from discussion following the decision that it was more important to use a survey for quality of life and to include a cognitive test and in consideration of survey fatigue
32	*Figure 8 Walking Test	Many individuals are likely to experience a floor effect
33	*Function in Sitting	Many individuals are likely to experience a ceiling effect
34	*Functional Reach Test	May be extremely challenging via telehealth, provides limited information
35	*Functional Status Score for the Intensive Care Unit	Many individuals are likely to experience a ceiling effect
36	Johns-Hopkins Highest Level of Mobility	Not used to assess change. Creates a common language.
37	*Lower Extremity Functional Scale	Many individuals are likely to experience a ceiling effect
38	*Intensive Care Unit Mobility Scale	Many individuals are likely to experience a ceiling effect; only validated in inpatient acute setting
39	*Patient Specific Functional Scale	Difficult to pool data across patients due to differences within the scale for each patient
40	*Physical Function Intensive Care Test	Many individuals are likely to experience a ceiling effect only validated in inpatient acute setting
41	**Short Physical Performance Battery*	SELECTED AS A CORE MEASURE. When recording raw data, this includes gait speed (the sixth vital sign[73]), a balance screen, and a 5 times sit-to-stand (mild variation from Neuro Core Measures recommendation). This allows the physical therapist to capture multiple facets of functional mobility in approximately 10 minutes. Very commonly used in acute care settings.
42	*Single-Limb Stance	Only captures a single construct of functional mobility
43	*Timed Up and Go	Many individuals are likely to experience a floor effect
44	*Trunk Impairment Scale	Many individuals are likely to experience a ceiling effect
45	4-Item Dynamic Gait Index	Reliability unknown
46	4-m gait speed test	Space requirements exceed 10-ft by 12-ft area.
47	4-Stage Balance Test	Reliability below 0.75 cut-off[74]
48	5-m Gait Speed Test	Space requirements exceed 10-ft by 12-ft area.
49	10-m Walk Test	Space requirements exceed 10-ft by 12-ft area.
50	30-second sit-to-stand (modified)	Unknown reliability
51	Activities-Specific Balance Confidence Scale	Not free
52	Activity Measure for Post-Acute Care	Not free
53	Chelsea Critical Care	Requires materials not typically available in all settings.
54	Clinical Test of Sensory Interaction on Balance	Requires materials not typically available in all settings
55	Clinical Test of Sensory Interaction on Balance—modified	Requires materials not typically available in all settings
56	DeMorto Mobility Index	Space requirements exceed 10-ft by 12-ft area.
57	L Test for Mobility	Space requirements exceed 10-ft by 12-ft area.

(continues)

APPENDIX 1. Overview of All Considered Outcome Measures (*Continued*)		
58	Mini Balance Evaluations Systems Test	Space requirements exceed 10-ft by 12-ft area. Requires materials not typically available in all settings
59	Dynamic Gait Index	Space requirements exceed 10-ft by 12-ft area.
60	Functional Gait Assessment	Space requirements exceed 10-ft by 12-ft area.
61	Perme Intensive Care Unit Mobility Score and ICU Mobility Scale	Requires materials not typically available in all settings.
62	Physical Performance Test	Space requirements exceed 10-ft by 12-ft area.
63	Tinetti Performance Oriented Mobility Assessment (POMA)	Space requirements exceed 10-ft by 12-ft area.
	Health-Related Quality Of Life	
64	*Assessment of Quality of Life 8D	Uses a week-long look back period, which was deemed too long to capture changes throughout the continuum of care.
65	***EQ-5D-5L*	RECOMMENDED AS A CORE MEASURE IF COMPLETING RESEARCH WITH AN INTERNATIONAL AUDIENCE. Recommended by the Society of Critical Care Medicine.[75]
66	*Life Satisfaction Questionnaire	10 minutes to complete (double the time of the recommended core measures)
67	*Nottingham Health Profile	10 minutes to complete (double the time of the recommended core measures). Focuses on negative aspects of health rather than quality of life.
68	***PROMIS Global-10*	RECOMMENDED AS A CORE OUTCOME MEASURE. This outcome measure looks at current perceptions regarding quality of life. Recommended by the National Institute of Health.[76] Adapted from the SF-36 and EQ-5D.
69	*Short Form 12	Uses a 4-week look back period, which was deemed too long to capture changes throughout the continuum of care.
70	Short Form 36	Takes longer than 15 minutes to complete.
	Strength and Power	
71	1-minute sit-to-stand	Reliability unknown
72	*5 times sit-to-stand	Similar to a portion of the SPPB selected as a core outcome measure to capture functional ability
73	*30-second sit-to-stand	Similar to a portion of the SPPB selected as a core outcome measure to capture functional ability
74	*Manual Muscle Testing	Less standardized than the very similar selected core outcome measure
75	***Medical Research Council-Sum Score*	SELECTED AS A CORE OUTCOME MEASURE. Applicable across a wide spectrum of functional levels. Recommended in patients surviving acute respiratory failure/critical illness.[77,78] Quick to complete.
76	Hand dynamometry	Requires materials not typically available in all settings
77.	Dynamometry	Requires materials not typically available in all settings
	Upper Extremity Function	
78	9-Hole Peg Test	Requires materials not typically available in all settings
79	Arm Curl Test	Requires materials not typically available in all settings
80	Arm Motor Ability Test	Requires materials not typically available in all settings
81	Action Research Arm Test	Requires materials not typically available in all settings
82	Box and Blocks	Requires materials not typically available in all settings
83	*Disability of Arm, Should, and Hand	May not add to the clinical picture of most individuals with or recovering from COVID-19

(continues)

APPENDIX 1. Overview of All Considered Outcome Measures (*Continued*)		
84	*Finger Tapping Test	Does not provide a comprehensive view of upper extremity function and may not add to the clinical picture of most individuals with or recovering from COVID-19.
85	Hand Dynamometry	Requires materials not typically available in all settings
86	Jebsen Hand Test	Requires materials not typically available in all settings
87	*Katz Index of Independence in Activities of Daily Living	While it provides a picture of upper extremity function, this is more a scale of activities of daily living
88	*Lawton Brody Instrumental Activities of Daily Living Scale	While it provides a picture of upper extremity function, this is more a scale of activities of daily living
89	Purdue Pegboard Test	Requires materials not typically available in all settings
90	*QuickDASH	May not add to the clinical picture of most individuals with or recovering from COVID-19
91	*Upper Extremity Functional Index	Not as responsive as the QuickDASH[79]
92	Upper Extremity Function Test	Requires materials not typically available in all settings
93	Wolf Motor Function Test	Requires materials not typically available in all settings
**Indicates the outcome measure met all inclusion/exclusion criteria following initial review and was included in the construct-specific discussion. The remaining tests were disqualified following initial review of inclusion/exclusion criteria.*		