RESEARCH ARTICLE

Validation of a markerless motion capture app for automated scoring of sit-to-stand, timed up and go, and short physical performance battery tests in adults with chronic disease

Jennifer K. Bertrand[1], Margaret L. McNeely[2], Jack Bates[1], Joshua Joy[2], Jenil Kanani[3], Victor E. Ezeugwu[2], Puneeta Tandon [1]*

1 Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Canada, 2 Department of Physical Therapy, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, Canada, 3 Department of Computing Science, Faculty of Science, University of Alberta, Edmonton, Canada

* ptandon@ualberta.ca

Abstract

Physical performance tests such as the 30-second Sit-to-Stand (30s-STS), Timed Up and Go (TUG), and Short Physical Performance Battery (SPPB) are widely used to assess physical function in older adults and are predictive of key health outcomes. However, their routine use in clinical practice is limited by time, resource, and personnel constraints. This study aimed to validate the automated scoring of physical performance assessments using a mobile, markerless motion capture (MMC) app compared to scoring by a certified exercise physiologist (CEP), and to quantify the rate and reasons for technology-related data loss. 228 adults (mean age = 61.6 ± 11.9 years) with at least one chronic medical condition were enrolled. Participants completed seven performance assessments: 30s-STS, TUG, and all components of the SPPB (Side-by-Side, Semi-Tandem and Tandem balance stands, 5-times Sit-to-Stand (5xSTS), and Gait Speed). All tests were scored simultaneously by a CEP and the MMC app using a Light Detection and Ranging (LiDAR)-enabled iPad. Agreement was assessed using intraclass correlation coefficients (ICCs) and weighted Cohen's kappa. Agreement between the MMC app and CEP was good to excellent for all assessments. ICCs ranged from 0.812 (Tandem Stand) to 0.995 (5xSTS). The overall SPPB score showed almost perfect agreement (κ = 0.808). Perfect agreement with no variability was observed for the Side-by-Side and Semi-Tandem balance tests. The overall tech-related data loss rate was low (3.1%), with the most common issue being poor motion tracking quality (1.3%). Automated scoring of physical performance tests using a self-contained MMC app demonstrated high agreement with expert scoring and low data loss in a cohort of participants with a range of chronic medical conditions. These findings support the use of MMC-enabled mobile applications for scalable, accessible, and objective assessment of physical function in clinical settings, with future potential for remote and asynchronous use.

Author summary

Tests of physical function—like standing up from a chair or walking a short distance—are powerful tools for understanding mobility, frailty, and future health risks in older adults. However, these tests are rarely used in routine care due to time, space, and staff constraints. In our study, we evaluated a mobile application that uses the iPad's built-in camera to automatically measure how a person performs common movement-based tests. We tested the app with 228 older adults living with chronic illness and compared its automated scores to those from a trained health professional. We found strong agreement between the two, suggesting that accurate scoring can be achieved without a human rater. The app also performed reliably, with very few technical failures or lost data. By reducing the need for trained observers or specialized equipment, our tool could help make mobility assessments more accessible—both in clinics and at home. In future work, we aim to expand the technology for use with standard smartphone cameras and support fully self-directed testing. Ultimately, our goal is to enable broader, earlier detection of functional decline, and to support more personalized care through digital health tools that are scalable, affordable, and easy to use.

1. Introduction

Physical performance assessments, such as the 30-second Sit-to-Stand (STS) test [1], the Short Physical Performance Battery (SPPB) [2], and the Timed Up and Go (TUG) test [3], are essential tools for evaluating physical function in older adults, including those with chronic diseases [4]. These tests were designed to improve upon self-reported measures by providing objective and reliable insights into a patient's physical capabilities including aerobic fitness, strength, and balance. In clinical or research settings, trained clinicians typically guide patients through these assessments and carry out direct observation and scoring. Large normative datasets support established cutoffs and performance thresholds for frailty, fall risk, and other clinical indicators [3,5–8]. These assessments are not only valuable for quantifying change with interventions (e.g., exercise) [9–11], but they are also highly predictive of outcomes such as hospitalization, disability, reduced quality of life, post-operative complications, increased mortality, and healthcare costs [2,7,12–18]. Consequently, physical performance tests are valuable not only for patients but also for clinicians and health care systems, enabling informed care and proactive decision-making [13,19].

Advances in digital technology have created new opportunities to modernize physical performance assessments, enhancing their automation, sensitivity, accessibility, and standardization. For decades, biomechanical research has leveraged digital tools for measuring physical movement, including optical tracking systems and sensors like ground force plates and inertial measurement units (IMUs) [20]. More recently, these

tools have been applied to physical performance assessments, with in-depth biomechanical analyses of gait [21–23], STS [24–26], and TUG [27,28]. Most of these biomechanical studies have been conducted in controlled laboratory environments using specialized, research-grade equipment (e.g., a computerized walkway with embedded pressure sensors (GAITRite, CIR Systems, USA) [22], or a 5-sensor wearable IMU system (LegSys+, BioSensics, USA) [24]), which are costly, lab-intensive, and not well-suited for routine clinical use due to challenges in data extraction, patient burden, and workflow integration [29–32]. However, with a recognized need to use more readily available devices such as smartphone cameras to facilitate proactive health monitoring from a patient's home [29,31], there has been an emerging intersection of biomechanical knowledge and consumer-grade technology including markerless motion capture (MMC technology).

Markerless motion capture (MMC) technology offers a low-cost and convenient solution for tracking movement during physical performance assessments. Unlike traditional marker-based motion capture systems, which require multiple near-infrared cameras and reflective markers placed on specific body landmarks [29,33], MMC leverages consumer camera technology, such as those found in smartphones and tablets, paired with advancements in computer vision and deep learning algorithms for pose estimation [34,35] These advancements have brought MMC technology closer to the accuracy levels of marker-based systems [29,33,36,37]. The move to MMC provides significant benefits, especially by reducing the time and cost associated with marker-based approaches [29,35,38], which have often been impractical for clinical settings. Beyond its cost-effectiveness and efficiency, the portability of MMC systems makes it well-suited for telehealth applications, and its rich kinematic data output creates opportunities for automated assessments and more sensitive indicators of performance.

While MMC technology has shown promise as a tool for physical performance assessments, its application in clinical research remains relatively narrow and underexplored. A 2023 review by Lam et al. found that although 65 studies applied MMC technology for clinical measurements in rehabilitation, more than half specifically focused on movement disorders such as Parkinson's Disease and Cerebral Palsy [29]. Only a small subset of these studies assessed the validity of MMC systems for specific physical performance tests, such as shoulder range of motion [39,40]. Similarly, a systematic review on MMC technology's use in neurodegenerative contexts revealed that most studies (24 out of 26) concentrated on gait analysis [32]. There have been efforts to automate evaluations of specific tests such as the TUG using the Xbox Kinect [41] and conventional video cameras [42–45], as well as the STS [46–50]. Despite these developments, few studies have explored the application of MMC systems for a comprehensive range of physical performance assessments or validated its automated scoring against the clinical reference standard of an expert clinician evaluation in adults with chronic illnesses.

The current study aims to validate the automated scoring of physical movement assessments using an app-based MMC system by comparing it to the clinical reference standard of scoring by a certified exercise physiologist (CEP) in a large sample of older adults with chronic medical conditions. A secondary aim is to explore the rate and reasons for MMC system-related data loss across this diverse sample.

2. Methods

2.1. Study design and population

This study is part of a larger research project investigating multidimensional contributors to the development of frailty by way of movement assessments, nutritional analysis, daily physical activity, and self-reported survey data. For this paper, only movement assessment data, self-reported survey data (e.g., demographic information), and data obtained through medical chart abstraction (e.g., Charlson Comorbidity Index) are included.

The study population consisted of community-dwelling adults recruited through local research clinics, including the Cancer Rehabilitation Clinic, the Edmonton Cirrhosis Care Clinic, and targeted social media advertisements. Inclusion criteria were: (i) a history of at least one chronic medical condition associated with increased frailty risk (condition verified by medical chart review); (ii) ≥18 years of age; (iii) ability to perform functional movement assessments, wear a smartwatch

for one week, and log dietary intake for two days; and (iv) capacity to provide informed consent. All research activities were approved by the University of Alberta's Health Research Ethics Board (Pro00117501), and all participants provided written informed consent prior to participation.

2.2. Procedure

The study was conducted at the University of Alberta in Edmonton, Alberta, Canada, between December 2023 and August 2024. Movement assessments were conducted in a clinical laboratory setting. Initially, the first 32 participants completed the test battery over two separate laboratory visits spaced 6–14 days apart, with approximately half of the assessments performed at each visit. Following protocol refinement to simplify logistics and minimize participant burden, all subsequent participants completed the full test battery in a single laboratory visit, with a scheduled mid-session rest period to reduce fatigue and ensure participant comfort.

Prior to the clinic visit, participants completed an online consent form, a demographic survey, and the Physical Activity Readiness Questionnaire (PAR-Q). Upon arrival, their PAR-Q responses were reviewed by a certified exercise physiologist (CEP) experienced in working with individuals with chronic medical conditions. The review included discussions about limitations affecting their ability to complete the movement assessments, review of blood pressure and heart rate, and participants were invited to opt out of any tests for which they felt they would experience discomfort. Prior to beginning the movement assessments, clearance for physical activity was determined by the CEP, with as needed, but infrequent physician consultation.

2.2.1. Movement assessments. During their visit, participants performed a series of movement assessments under the instruction of the CEP. The assessments included the 30-second Sit-to-Stand (30s-STS) test [1], the Timed Up and Go (TUG) test [3], and the 5 tests of the Short Physical Performance Battery (SPPB) [2]: the Side-by-Side Stand, the Semi-Tandem Stand, the Tandem Stand, the 5-times Sit-to-Stand (5xSTS) test (corresponding to the 'Repeated Chair Stand Test' described in the original SPPB validation [2]), and the Gait Speed test. These assessments were selected for their widespread use in clinical practice and research, with large normative datasets that establish population-based reference values for each test [2,51,52], systematic reviews documenting their psychometric properties [53–59], and consistent inclusion in geriatric assessment guidelines [60–62].

The tested movement assessments are further detailed in Table 1. The SPPB total score was calculated following the SPPB scoring protocol [2], with the five constituent test scores summed to yield a total score between 0 (lowest function) and 12 (highest function). To ensure participant safety, for the Side-by-Side, Semi-Tandem, and Tandem Stands, participants were offered the option of having a chair placed within arm's reach for support if balance was lost.

Because the objective of this study was to compare scoring agreement between a MMC system and clinician measurement, and not to assess change or performance differences across conditions or sessions, randomization of assessment order was not required. A fixed order of assessments was implemented to align with clinical workflows and optimize participant comfort, with a 15-minute break (used for instructions on other study elements) scheduled halfway through the tests.

2.2.2. The Ameya MMC App. A custom-built iOS application, the Ameya MMC app (Fig 1), was developed for this study to support multidimensional frailty assessment, including automated scoring of physical performance tests (Tables 1 and 2) using MMC technology. The app captures 3D joint position data from 34 anatomical landmarks at approximately 30 Hz using the Light Detection and Ranging (LiDAR) depth-sensing camera built into certain iPad Pro and iPhone Pro models.

The app's MMC component is built using a skeletal tracking module which supports both RGB and RGB-D camera input. In this study, real-time pose estimation was performed using the LiDAR-based depth camera on the iPad Pro (Fifth/Sixth generation, Apple Inc., US). The Ameya MMC app incorporated this tracking framework to enable automated, test-specific scoring procedures for the investigated physical performance assessments. Participants were assigned individual profiles within the app, which allowed trained research assistants (RAs) to initiate the appropriate assessments using the clinician-facing interface.

Table 1. Description of movement assessments and MMC system setup.

Assessment Name	Assessment Description	Clinician Measurement Protocol	Measurement Type	MMC System Setup (iPad Placement)
30-second Sit-to-Stand (30s-STS) [1]	Repeated sit-to-stand movements from a chair over 30 seconds	Count number of full upright stands completed within 30 seconds	Continuous (0–X reps complete)	Camera placed at front-right angle to participant and chair, ~2.5 m distance
Timed Up and Go (TUG) [3]	Stand from chair, walk 3 meters, turn, return, and sit down again	Time to complete full sequence	Continuous (s)	Camera placed in front of participant, ~5 m from starting position/ chair
SPPB: Side-by-Side Stand [2]	Standing with feet together for up to 10 seconds	Time held (up to 10 seconds)	Continuous (0–10s)	Camera placed directly in front of participant, ~2.5 m distance
SPPB: Semi-Tandem Stand [2]	Standing with one heel of one foot next to the instep of the other for up to 10 seconds	Time held (up to 10 seconds)	Continuous (0–10s)	Camera placed to side of participant, ~2.5 m away, with back foot closest to camera
SPPB: Tandem Stand [2]	Standing with one foot directly in front of the other for up to 10 seconds	Time held (up to 10 seconds)	Continuous (0–10s)	Camera placed to side of participant, ~2.5 m away
SPPB: 5-times Sit-to-Stand (5xSTS) [2]	Rising from a seated position to standing and sitting again, 5 times, as quickly as possible	Time to complete 5 stands	Continuous (s)	Camera placed at front-right angle to participant and chair, ~2.5 m distance
SPPB: Gait Speed [2]	Walking 3 meters at a usual pace	Time to walk 3 meters, divided by 3 meters	Continuous (s)	Camera placed in front of participant, ~5 m from start line
SPPB: Total Score [2]	Composite score summing results from balance tests, Gait Speed, and 5xSTS	Calculated with established scoring protocol (range: 0–12), sum of 5 component test scores	Ordinal (0–12)	Derived from video recordings of individual tasks; MMC algorithms apply scoring logic per test

https://doi.org/10.1371/journal.pdig.0001172.t001
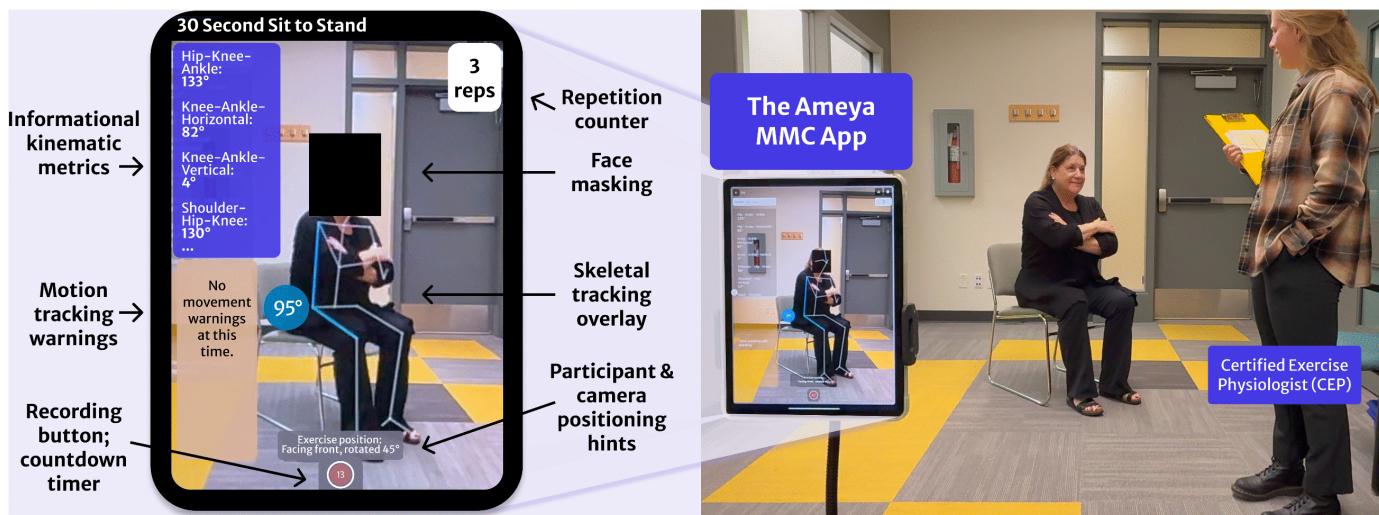
Fig 1. Ameya MMC App and in-clinic setup with the CEP.

https://doi.org/10.1371/journal.pdig.0001172.g001

A consistent setup protocol was used during data collection (iPad placed in a tripod approximately 2.5 m away from the participant, with the camera lens approximately 0.5 m off the ground; see Table 1 and Fig 1), however, assessments were conducted in multiple clinical rooms with variation in flooring, lighting, and background. The tracking system operated robustly under these conditions without additional configuration or environmental controls.

Table 2. Summary of test-specific algorithm logic and output metrics.

Assessment Name	Input Signal	Core Detection Logic	Output Metric(s) Displayed in App
30-second Sit-to-Stand (30s-STS) [1]	Vertical body motion	Detect repeated sit-to-stand transitions over a fixed 30-second period; apply movement thresholds and filters to avoid false counts.	Number of full stands completed; data quality flag.
Timed Up and Go (TUG) [3]	Body-height change during sit-to-stand and return-to-sit	Detect stand-up and sit-down events from body-height transitions; compute time between leaving and returning to the seated position.	Total time (seconds); data-quality flag.
SPPB Balances (Side-by-Side, Semi-Tandem, Tandem) [2]	Hand and foot/ankle motion relative to start position	Detect loss of balance using hand or foot movement.	Duration held (up to 10 seconds); pass/fail balance score; data-quality flag.
SPPB: 5-times Sit-to-Stand (5xSTS) [2]	Vertical body motion	Identify five complete sit-to-stand cycles; measure total time between the first and fifth stands.	Time to complete five stands (seconds); data-quality flag.
SPPB: Gait Speed [2]	Forward body (core) displacement	Estimate forward walking speed from body motion over 3 meters; verify consistency across partial-distance segments	3-meter gait duration (seconds) and derived score; data-quality flag.

https://doi.org/10.1371/journal.pdig.0001172.t002

Test-specific scoring algorithms were executed in real time during each assessment (see Table 2). Each algorithm was tailored to the corresponding movement test, where movement signals like vertical displacement or forward motion were used to identify key events like stand transitions, walking phases, or balance loss. These scoring algorithms included repetition detection and counting for the 30s-STS test and the 5xSTS test, time estimation for the TUG test and the Gait Speed test, and duration thresholds for balance tests. Output metrics were calculated automatically and displayed immediately on the app interface after each test. MMC app recordings were stored locally on the iPad and automatically synchronized with a secure cloud server, and included in-app face masking (see Fig 1) for enhanced participant privacy.

The application also included error detection and interface elements to support use by non-technical staff. These included countdown timers, real-time movement counters, and visual indicators when pose estimation was compromised (e.g., due to occluded joints or camera misalignment).

During pilot testing, scoring algorithms were tested and refined to accommodate varied movement speeds, body sizes, and compensatory postures observed in the target population. Subsequent minor adjustments to the scoring algorithms were performed iteratively during data collection in response to observed edge cases, optimizing smoothing parameters for variable joint trajectories such as irregular repetitions, and increasing tolerance to partial occlusions when limbs were temporarily obscured. These refinements did not alter the underlying scoring logic but improved the robustness of detection across participants. All recordings were reprocessed at the conclusion of the study's data collection period using the finalized algorithm set (one per movement test) to ensure scoring consistency across the full dataset.

Altogether, the Ameya MMC app's architecture includes real-time pose tracking, task-specific scoring logic, automated data handling, and built-in quality control mechanisms. These components were developed to support standardized and scalable assessment of physical performance in structured clinical research settings.

2.2.3. Additional measurements. Outside of the in-person session, participants completed additional questionnaires via REDCap, including a demographic survey, the revised Edmonton Symptom Assessment Scale [63], the Edmonton Frail Scale [64], the EQ-5L-5D [65], the Duke Activity Status Index (DASI) [66], and a fall history survey. For participants who gave explicit consent (n = 226), medical chart abstraction was performed to obtain clinical data (e.g., history of chronic disease or cancer) and calculate the Charlson Comorbidity Index (CCI) [67]. The CCI is a measure of comorbidity burden, assigning weighted scores to chronic conditions, with higher scores indicating increased comorbidity and mortality risk.

2.3. Outcomes

The primary outcome was the level of agreement between the Ameya MMC app and a clinician (CEP), evaluated separately for each of the seven movement assessments, as well as for the overall SPPB score. Agreement was assessed with Intraclass Correlation Coefficients (ICCs) for the following metrics: (i) TUG duration (seconds); (ii) number of repetitions in the 30s-STS test; (iii-v) balance duration (up to 10 seconds) for the Side-by-Side, Semi-Tandem, and Tandem Stands; (vi) time (seconds) to complete five sit-to-stand repetitions; and (vii) walking speed during the Gait Speed test. Agreement for the SPPB total score (range: 0–12) was also assessed.

The rate and reasons for tech-related data loss per movement assessment and overall were tracked. Tech-related data loss was defined as the percentage of MMC app recordings that were not analyzable, omitting participant or protocol-related exclusions.

Four types of tech-related issues were identified: (i) app crashes (any instance where the MMC app failed after recording had begun), (ii) incomplete data (occurred where earlier versions of the real-time algorithm prematurely terminated a test before the completion criteria was met), (iii) poor motion tracking quality, determined through visual inspection (e.g., skeletal overlay misaligned with participant joints), field notes (e.g., tracking of background objects was observed by the iPad operator), or detection of implausible joint movements (e.g., multi-joint velocity spikes = faster than 5 m/s, a conservative cutoff well above expected speeds for these clinical tasks [24,27,68–72]), and (iv) stand count discrepancies with the 5xSTS test (the CEP counted five repetitions but the MMC system did not, leading to the manual stopping of recordings that prevented the MMC app from calculating a duration).

2.4. Statistical analysis

Descriptive statistics were calculated using appropriate summary measures for continuous and ordinal outcomes. Agreement between the CEP and MMC app on continuous test outcomes was assessed using ICCs with 95% confidence intervals, calculated using single-rating, absolute-agreement, two-way mixed-effects models (ICC(2,1)) following guidelines from Koo and Li [73]. ICC values were interpreted using the following thresholds: values below 0.5 were considered to reflect poor agreement, values between 0.5 and 0.75 indicated moderate agreement, values between 0.75 and 0.9 indicated good agreement, and values greater than 0.9 indicated excellent agreement [73]. ICCs were undefined when no between-subject variability was present.

Agreement on the discrete, ordinal SPPB total score (range: 0–12) was evaluated using a weighted Cohen's kappa with quadratic weights. This method accounts for agreement by chance and penalizes larger differences more heavily. Kappa values were interpreted based on the classification system by Landis and Koch [74]: values below 0.00 were considered poor agreement; values between 0.00 and 0.20 slight; 0.21 to 0.40 fair; 0.41 to 0.60 moderate; 0.61 to 0.80 substantial; and 0.81 to 1.00 almost perfect agreement. All analyses were conducted using Jamovi (version 2.3) and the 'seolmatrix' module [75] based on the 'irr' package [76] in R.

3. Results

3.1. Characteristics of the participants

A total of 228 adults with at least one chronic medical condition or a history of cancer were enrolled. Demographic characteristics are summarized in Table 3. Of the total sample, 62% were female, 79% identified as white, and mean age was 61.6 ± 11.9 years. The cohort demonstrated a high burden of chronic disease and comorbidity, with a mean Charlson Comorbidity Index (CCI) score of 4.52 ± 2.37. The most commonly reported chronic conditions were a history of cancer (37%) and cirrhosis (25%). Additional diagnoses (grouped into the Other category in Table 3 Disease History) spanned a broad range, including neurological conditions (e.g., Parkinson's disease, multiple sclerosis), post-transplant status, inflammatory arthritis, and diabetes.

Table 3. Participant characteristics.

Participant Characteristics	n (%) or M±SD
Age (years)	61.6±11.9
Sex	
Female	142 (62.3)
Male	86 (37.7)
Gender (n=227)	
Woman	139 (61.2)
Man	87 (38.3)
Other	1 (<1)
Race (n=224)	
White	177 (79.0)
East Asian	12 (5.4)
South Asian	10 (4.5)
Aboriginal	6 (2.7)
Latino	4 (1.8)
Middle Eastern	2 (0.9)
Black	1 (<1)
Other	4 (1.8)
Mixed race	8 (3.6)
BMI (kg/m²) (n=226)	29.2±6.6
Disease History* (n=226)	
Cancer	84 (37.2)
Cirrhosis	57 (25.2)
Other	103 (45.6)
Charlson Comorbidity Index (n=226)	4.52±2.37
Gait Aid Used (n=218)	16 (7.3)
Falls 1 year prior (n=218)	
None	135 (61.9)
One	45 (20.6)
Two	20 (9.2)
More than two	18 (8.3)
Duke Activity Status Index score (n=212)	42.7±12.3
EQ-5D-5L VAS rating (n=218)	68.9±18.4
Edmonton Symptom Assessment Scale (n=210)	
Pain	2.49±2.19
Tiredness (lack of energy)	3.72±2.68
Drowsiness (feeling sleepy)	2.70±2.53
Nausea	0.50±1.22
Appetite	3.28±2.76
Shortness of breath	1.50±2.26
Depression	1.64±2.33
Anxiety	1.90±2.29
Wellbeing (overall feeling)	3.69±2.70
Edmonton Frail Scale score (n=218)	
Fit (score 0–3)	102 (46.8)
Vulnerable (score 4–5)	53 (24.3)
Mild Frailty (score 6–7)	40 (18.3)

(Continued)

Table 3. (Continued)

Participant Characteristics	n (%) or M±SD
Moderate Frailty (score 8–9)	16 (7.3)
Severe Frailty (score 10+)	7 (3.2)

N = 228 unless otherwise indicated. Sample sizes vary due to missing data. *Categories are not mutually exclusive; participants may have multiple conditions.

Participants reported moderate functional capacity in daily life, with a mean DASI score of 42.7 ± 12.3 (n = 212). Overall health status, as rated on the EQ-5D-5L visual analog scale, averaged 68.9% ± 18.4% (n = 218). A gait aid was regularly used by 7% of the participants, and 38% reported at least one fall in the previous year. Symptom burden, as measured by the Edmonton Symptom Assessment Scale (n = 210), was generally low, with the highest mean scores observed for tiredness (3.72 ± 2.68), overall wellbeing (3.69 ± 2.70), and appetite concerns (3.28 ± 2.76). Based on the Edmonton Frail Scale (n = 218), 46.8% were categorized as not frail, while 24.3% were classified as vulnerable, and 25.6% with mild to moderate frailty. Seven participants (3.2%) were classified as severely frail.

3.2. Agreement between CEP and the MMC app

Agreement between the outcomes from the CEP and the MMC app (Table 4) was excellent for the TUG [ICC = 0.96, 95% CI: 0.93–0.98], the 30s-STS [ICC = 0.93, 95% CI: 0.90–0.95], and the 5xSTS test [ICC = 0.995, 95% CI: 0.98–0.99]. Agreement on Gait Speed was also high [ICC = 0.87, 95% CI: 0.80–0.91].

For both the Side-by-Side and Semi-Tandem Stands, the MMC app recorded perfect agreement with the CEP. However, ICCs were undefined due to a ceiling effect; all participants completed the full 10-second duration, resulting in no between-subject variability.

Of all tests, the Tandem Stand showed the lowest agreement between the CEP and the MMC app [ICC = 0.81, 95% CI: 0.76–0.86], though the value remained in the good range. Further inspection revealed six instances in which participants lost balance and used a nearby chair for support. While the CEP recorded this as the end of the test, the MMC app was unable to detect the chair support as it occurred behind the patient and out of the field of view of the camera. When these six cases were excluded, agreement improved to excellent [ICC = 0.98, 95% CI: 0.97–0.98].

Table 4. Agreement between MMC and clinician (CEP).

Movement Assessment	Outcome	n	ICC (95% CI)	Cohen's kappa
Timed Up and Go (TUG)	Duration (s)	226	**0.962** (0.927, 0.977)	
30-second Sit-to-Stand (30s-STS)	Count (reps)	222	**0.928** (0.895, 0.949)	
SPPB: Complete Battery	Score (0–12)	178		**0.808**
Side-by-Side Stand	Duration (0–10s)	220	perfect agreement (*ICC undefinable**)	
Semi-Tandem Stand	Duration (0–10s)	220	perfect agreement (*ICC undefinable**)	
Tandem Stand	Duration (0–10s)	200	**0.812** (0.758, 0.855)	
Tandem Stand (*hand support excl.***)	Duration (0–10s)	194	**0.979** (0.972, 0.984)	
SPPB: 5-times Sit-to-Stand (5xSTS)	Duration (s)	213	**0.995** (0.983, 0.998)	
SPPB: Gait Speed	Speed (m/s)	226	**0.868** (0.796, 0.910)	

*There was perfect agreement and no variability within participants for the Side-by-Side and Tandem Stands, so an ICC value could not be computed for these two movement assessments.

**Additional exclusions on Tandem Stand are the removal of 6 cases where participants braced themselves with the back of a chair that was placed beside them for safety.

For the total SPPB score, calculated as the sum of the individual SPPB components, agreement between the CEP and the MMC app was high, with a weighted Cohen's kappa of 0.808, indicating almost perfect agreement.

3.3. Tech-related data loss

Tech-related data loss was tracked across the seven movement assessments and is summarized in Table 5. After removing non-technical exclusions (e.g., participant opt-outs, protocol deviations), 1576 total MMC app recordings remained. Of these, 1527 were analyzable, resulting in a tech-related data loss rate of 3.1% (49 recordings).

Of the small number of excluded recordings (49 of 1576 total recordings), the most common reason for exclusion was poor motion tracking quality (21 recordings, or 1.3% of all MMC app recordings). Other exclusion reasons included incomplete data due to errors in early algorithm versions (13 recordings) and app crashes (10 recordings), though each type of exclusion amounted to less than 1% of data loss.

Three assessments—Gait Speed, 30s-STS, and TUG—had minimal tech-related data loss (<1%). The 5xSTS test had a specific issue related to stand count discrepancies. In five cases, MMC detected only four repetitions, while the CEP identified five. Because the MMC app's stopping condition was not met, the duration could not be calculated and agreement could not be assessed.

The Tandem Stand had the highest tech-related data loss rate, with 20 unusable MMC app recordings (9% of recordings for that test). Most of these were due to poor motion tracking (12 recordings). The algorithm struggled with depth cues and differentiating foot placement when participants wore bulky clothing (e.g., long shorts, sweatpants) or dark-colored attire. These issues and potential solutions are elaborated on in the Discussion.

4. Discussion

This study validated the use of the Ameya MMC app for automated scoring of physical performance tests in a sample of 228 adults with chronic disease. Agreement between gold-standard clinician-determined scores and app-based MMC system scores was assessed for the 30-second STS test, the TUG test, and all 5 tests of the SPPB. Good to excellent agreement was observed across all assessments, with the TUG, 30s-STS and 5xSTS tests yielding ICC values ranging from 0.928 to 0.995. Clinician and MMC app agreement on the overall SPPB score (composite score of 5 summed scores) was found to be almost-perfect (weighted Cohen's kappa of 0.808). Across all seven physical performance tests, only two tests, the Tandem Stand and Gait Speed, were below the threshold for excellent agreement, with ICCs of 0.812

Table 5. Participant exclusions and MMC app data loss.

	Side-by-Side Stand	Semi-Tandem Stand	Tandem Stand	Gait Speed	5xSTS	30s-STS	TUG	All Tests Combined
N Total Participants	228	228	228	228	228	228	228	1596
Non-Tech Exclusions								
Participant Opt-Out			2		7	2		11
Protocol Deviation		1	6			2		9
N After Non-Tech Exclusions	228	227	220	228	221	224	228	1576
Tech Exclusions								
App Crash	2	2	3			2	1	10
Incomplete Data (Early Algorithm)	4		5		3		1	13
Poor Motion Tracking Quality	2	5	12	2				21
Stand Count Discrepancy					5			5
N After Tech Exclusions	220	220	200	226	213	222	226	1527
% Tech-Related Data Loss	3.5	3.1	9.1	0.9	3.6	0.9	0.9	3.1

https://doi.org/10.1371/journal.pdig.0001172.t005

and 0.868, respectively. Overall, these results demonstrate that with the use of carefully developed algorithms, app-based scoring of in-clinic physical performance assessments using the Ameya MMC app provides good to excellent agreement with gold-standard clinician-determined scoring.

Despite strong evidence linking physical performance tests to clinical outcomes such as hospitalization, disability, and mortality [12–17,77,78], gold-standard, clinician-determined assessments remain underutilized in routine care, including in rehabilitation and chronic disease management [79–81]. Across care settings, clinicians cite time constraints, lack of trained or certified personnel, and other practical limitations as barriers to administering these standardized and objective tests [82–85]. The current study begins to address these challenges by demonstrating that physical performance assessments can be accurately scored using a self-contained app running on an iPad, eliminating the need for external sensors, computing infrastructure, or post-testing analysis. This easy-to-use setup lowers the barrier to in-clinic use and suggests the potential of future remote, asynchronous use, expanding access for patients and scalability for health systems.

Others have explored the supplementary use of consumer-grade technology and MMC technology with physical performance assessments, using the Microsoft Kinect or other simple RGB cameras for motion tracking, however, these approaches typically require additional computing infrastructure, technical expertise, and are often limited to a single assessment type [29]. In contrast, the current study is, to our knowledge, the first to demonstrate expert-level scoring across a suite of standardized physical performance tests using a standalone, real-time mobile app. This includes seven validated tests administered during a single clinical session across a large and clinically diverse population of individuals with at least one chronic condition predisposing to frailty. Although previous studies have tested large samples and/or multiple performance tasks [86,87], none have combined portability, real-time scoring, clinical usability, and a broad range of validated assessments, within a single, plug-and-play app compatible with any LiDAR-enabled device (e.g., iPhone Pro or iPad Pro models). This tool may help address practical barriers to physical performance assessment in clinical settings by offering an accessible and standardized measurement approach.

Building on this unique approach, these findings can be situated alongside recent efforts to automate physical performance assessments using non-laboratory, MMC technology-based tools. In a series of studies [88–90], an electronic SPPB kiosk demonstrated excellent agreement with manually scored assessments (ICC = 0.92 for the total score [90]). While the kiosk also used LiDAR sensor technology, it incorporated additional hardware, including a load cell array (weight-sensitive mat) for the three balance tests and the 5xSTS [90]. Despite this added hardware, its scoring precision was comparable to or lower than that of the current study. For instance, the balance ICC (0.89, 95% CI: 0.81–0.93) and 5xSTS ICC (0.84, 95% CI: 0.75–0.90) [90] were similar to or lower than those observed here (0.81 for Tandem Stand before exclusions, and 0.99 for 5xSTS). Their data loss rate (3.3%, defined as the proportion of participants excluded due to incomplete data capture [90]) was also similar to that observed in the present study.

In another large-scale study, Liu et al. [86] tested 665 older adults using a Microsoft Kinect, custom software, a Mini PC and a custom TV-based interface. Their system incorporated seven functional assessments, including STS, TUG, and Gait Speed, and produced correlation coefficients between automated and clinician scores. Most correlations were strong and consistent with the current study findings, except for Gait Speed, where their correlation (r = 0.493) showed much weaker agreement than our measure of agreement (ICC = 0.868). Data loss was not reported. In a much smaller sample of 30 older adults, Kaewakaen et al. [91] achieved excellent agreement with an expert rater using a Kinect for the 5STS, further supporting the feasibility of automated scoring in this domain. Together, these comparisons illustrate the promise of prior MMC systems but also underscore the benefits of a streamlined, app-based approach.

These findings support the use of the Ameya MMC app as a practical tool to facilitate accurate, objective physical performance testing in clinical settings. Rather than replacing clinician expertise, the app is designed to complement it, reducing scoring burden while maintaining accuracy. Advantages include automated scoring, improved efficiency, and the potential for non-specialized staff to conduct assessments, which may enable more frequent monitoring of physical function [31,32].

The tool may be particularly relevant in resource-limited settings [29,92], such as rural clinics or long-term care facilities, where access to trained assessors is often constrained [31]. In these contexts, it could support earlier identification of deficits that may be modifiable with intervention. In research contexts, the automated evaluation through the app could also streamline data collection and make it more feasible to include physical performance testing as an objective outcome measure across larger or more distributed studies. By facilitating accurate, scalable, and accessible assessment [29,31,32], this approach supports more equitable, patient-centered models of care [93,94] and serves as a promising adjunct to traditional in-person assessments.

4.1. Limitations

While the automated MMC scoring approach demonstrated strong overall agreement, several limitations and design considerations emerged. One key challenge was validating this MMC app approach against clinician judgment, which includes a degree of subjectivity. For example, CEPs may have counted a borderline STS repetition that occurred just after the timer ended or have allowed for slight deviations in form (such as not standing fully upright), whereas the MMC app's rigid nature with fixed thresholds did not accommodate these nuances. This occasionally led to discrepancies, such as the MMC app scoring only four repetitions when the CEP counted five, highlighting a trade-off between algorithmic precision and assessor flexibility. CEPs also relied on the MMC app's auditory countdown to synchronize their stopwatch, likely introducing small timing inconsistencies due to reaction time, which may have added noise to the reported agreement measures.

The Tandem Stand posed the greatest challenge in terms of both scoring accuracy and data quality. For safety, the test requires a nearby chair for support, and the test is stopped if the participant uses the chair. While CEPs could directly observe whether support of the chair was used, the MMC app could not detect this, as the chair was often occluded from the camera's view. Modifications to the algorithm that account for this scenario are underway. Exclusion of these six cases substantially improved agreement. Beyond this challenge, the static nature and sagittal orientation of the Tandem Stand reduced the MMC app's ability to detect and track positions of the key body joints, especially in the lower body, leading to poor tracking in some participants. These issues were most common when participants wore dark or loose-fitting clothing or footwear that obscured joint landmarks. Future improvements may improve participant positioning and algorithmic handling of common occlusion scenarios.

Gait Speed measurement also presented unique challenges, as it required participants to move over a longer distance than the other tests. Unlike CEPs, who timed a marked 3-meter section, the MMC app does not track floor markers and instead relies on depth measurements that may not be captured over an identical time window to the CEP. This approach introduced some variability, particularly given the LiDAR sensor's reduced depth accuracy beyond 5 meters—an issue that occasionally affected tracking quality when participants were furthest from the iPad (at the start of their Gait Speed test).

Because both the MMC app and the CEP were scoring the same test instances, participant instructions were delivered solely by the CEP, and the app included only a countdown timer. Observing the instructional flow and challenges across a large sample has helped inform the design of automated instructions to be integrated in future app versions, similar to the approach taken by Liu et al. [86]

Beyond these procedural considerations, sample composition also presents limitations in this study. Approximately 80% of participants identify as White, and the cohort is largely composed of older adults. This limits generalizability to younger populations with chronic disease and to individuals from diverse racial and ethnic backgrounds. Prior work has shown that computer vision systems may perform less accurately across different skin tones and body types [95,96], and broader digital health research warns that inadequate diversity in training and validation can perpetuate health inequities [97–99].

Finally, these findings are currently limited to supervised, in-clinic use and are not yet generalizable to remote or unsupervised settings. The current approach also depends on LiDAR-enabled devices, which are typically newer, higher-end consumer models—potentially limiting accessibility for some users. Taken together, these limitations illustrate

important design considerations that will guide further app refinement to improve usability, performance, reliability, and generalizability.

4.2. Future directions

Building on the insights from this study, next steps will focus on expanding the app's utility beyond supervised, in-clinic use. This includes integrating automated instructions to guide participants through assessments without a clinician present, refining test algorithms to better handle edge cases, especially those stemming from safety considerations (e.g., detecting compensatory arm movements during balance tests), and adapting assessments so they could be used with 2D/non-LiDAR smartphone cameras (e.g., with Android phones or tablets, and non-Pro iOS models). Real-time tracking quality checks will be explored to flag cases where tracking is compromised—such as due to dark or loose clothing—allowing participants to correct the issue before data is recorded (and thus reducing the rate of data loss to poor motion tracking).

The MMC app is being adapted for remote and asynchronous use, allowing users to perform assessments on their own at home. Doing so requires thoughtful updates, not just to the technology (e.g., adjustments to enable non-LiDAR, 2D smartphone camera use), but also to safety protocols and user interface design, especially to support older adults or individuals with limited mobility. We envision a future where a simple 30s-STS task, captured with a tablet in a clinic or home setting, becomes a routine part of health monitoring, just like blood pressure, weight, or height [12,100,101]. Just as automated blood pressure cuffs have enabled non-clinicians to reliably capture vital signs, advances in MMC have the potential to bring the same level of objectivity and convenience to the assessment of physical function.

In parallel, the next steps will explore how the rich kinematic data captured during app-based MMC assessments could yield deeper clinical insight—without placing any extra burden on patients or clinicians. These continuous movement signals, already collected by the MMC app, may offer a more nuanced and objective view of how these tasks are performed, beyond the final score. For example, two individuals might complete 12 repetitions in a 30-second STS test, but one performs with fluid, controlled motion while the other shows erratic pacing or compensatory strategies that could signal instability, muscular deficits, disease, or fall risk [24,26,70,102–104]. Reducing the performance metrics into a single repetition count or duration obscures meaningful differences in movement quality [32,105,106]. By utilizing the full potential of kinematic data from MMC technology, we can enhance clinical interpretation, identifying subtle impairments that traditional scoring methods may miss. This represents an important step to maximizing the value of digital assessments—moving from simple tools for quantifying gross test performance to a platform that provides deeper insights into functional ability. These next steps are intended to advance the integration of physical function monitoring into both research and routine care in a more consistent and accessible manner.

5. Conclusion

This study validated a self-contained, app-based MMC system for scoring physical performance assessments in a large and clinically diverse sample of adults with chronic medical conditions. Good to excellent agreement was found between the app's automated scores and those of a trained clinician (CEP) across multiple functional tests, supporting its accuracy and feasibility for in-clinic use. By enabling automated scoring within a portable app, this approach may help to reduce barriers to standardized physical function assessment. With continued development in this field, there is potential to support more scalable, accessible, and patient-centered approaches to frailty and fall-risk screening, prehabilitation and tracking rehabilitation progress, as well as proactive monitoring of physical, functional health across the lifespan.

Supporting information

S1 Table. Average scores recorded by CEP and MMC app for each movement assessment.
(PDF)

S1 Dataset. Per subject data for each physical performance test.
(ZIP)

Acknowledgments

Author contributions

Conceptualization: Margaret L. McNeely, Victor E. Ezeugwu, Puneeta Tandon.

Data curation: Jennifer K. Bertrand.

Formal analysis: Jennifer K. Bertrand, Jenil Kanani.

Funding acquisition: Margaret L. McNeely, Puneeta Tandon.

Investigation: Jennifer K. Bertrand, Jack Bates, Joshua Joy.

Methodology: Jennifer K. Bertrand, Margaret L. McNeely, Puneeta Tandon.

Project administration: Jennifer K. Bertrand, Margaret L. McNeely, Puneeta Tandon.

Resources: Margaret L. McNeely, Puneeta Tandon.

Supervision: Margaret L. McNeely, Victor E. Ezeugwu, Puneeta Tandon.

Visualization: Jennifer K. Bertrand.

Writing – original draft: Jennifer K. Bertrand.

Writing – review & editing: Jennifer K. Bertrand, Margaret L. McNeely, Victor E. Ezeugwu, Puneeta Tandon.

References

1. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. Res Q Exerc Sport. 1999;70(2):113–9. https://doi.org/10.1080/02701367.1999.10608028 PMID: 10380242

2. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol. 1994;49(2):M85-94. https://doi.org/10.1093/geronj/49.2.m85 PMID: 8126356

3. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991;39(2):142–8. https://doi.org/10.1111/j.1532-5415.1991.tb01616.x PMID: 1991946

4. Beaudart C, Rolland Y, Cruz-Jentoft AJ, Bauer JM, Sieber C, Cooper C, et al. Assessment of muscle function and physical performance in daily clinical practice : a position paper endorsed by the European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO). Calcif Tissue Int. 2019;105(1):1–14. https://doi.org/10.1007/s00223-019-00545-w PMID: 30972475

5. Rikli RE, Jones CJ. Development and validation of a functional fitness test for community-residing older adults. J Aging Phys Activity. 1999;7(2):129–61. https://doi.org/10.1123/japa.7.2.129

6. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci. 2001;56(3):M146-56. https://doi.org/10.1093/gerona/56.3.m146 PMID: 11253156

7. Studenski S, Perera S, Patel K, Rosano C, Faulkner K, Inzitari M, et al. Gait speed and survival in older adults. JAMA. 2011;305(1):50–8. https://doi.org/10.1001/jama.2010.1923 PMID: 21205966

8. Bohannon RW, Wang Y-C. Four-meter gait speed: normative values and reliability determined for adults participating in the NIH toolbox study. Arch Phys Med Rehabil. 2019;100(3):509–13. https://doi.org/10.1016/j.apmr.2018.06.031 PMID: 30092204

9. 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):743–9. https://doi.org/10.1111/j.1532-5415.2006.00701.x PMID: 16696738

10. Rodriguez-Mañas L, Laosa O, Vellas B, Paolisso G, Topinkova E, Oliva-Moreno J, et al. Effectiveness of a multimodal intervention in functionally impaired older people with type 2 diabetes mellitus. J Cachexia Sarcopenia Muscle. 2019;10(4):721–33. https://doi.org/10.1002/jcsm.12432 PMID: 31016897

11. Giné-Garriga M, Roqué-Fíguls M, Coll-Planas L, Sitjà-Rabert M, Salvà A. Physical exercise interventions for improving performance-based measures of physical function in community-dwelling, frail older adults: a systematic review and meta-analysis. Arch Phys Med Rehabil. 2014;95(4):753–769.e3. https://doi.org/10.1016/j.apmr.2013.11.007 PMID: 24291597

12. Fritz S, Lusardi M. White paper: "walking speed: the sixth vital sign". J Geriatr Phys Ther. 2009;32(2):46–9. https://doi.org/10.1519/00139143-200932020-00002 PMID: 20039582

13. Studenski S, Perera S, Wallace D, Chandler JM, Duncan PW, Rooney E, et al. Physical performance measures in the clinical setting. J Am Geriatr Soc. 2003;51(3):314–22. https://doi.org/10.1046/j.1532-5415.2003.51104.x PMID: 12588574

14. de Fátima Ribeiro Silva C, Ohara DG, Matos AP, Pinto ACPN, Pegorari MS. Short physical performance battery as a measure of physical performance and mortality predictor in older adults: a comprehensive literature review. Int J Environ Res Public Health. 2021;18(20):10612. https://doi.org/10.3390/ijerph182010612 PMID: 34682359

15. Volpato S, Cavalieri M, Sioulis F, Guerra G, Maraldi C, Zuliani G, et al. Predictive value of the short physical performance battery following hospitalization in older patients. J Gerontol A Biol Sci Med Sci. 2011;66(1):89–96. https://doi.org/10.1093/gerona/glq167 PMID: 20861145

16. Cesari M, Kritchevsky SB, Newman AB, Simonsick EM, Harris TB, Penninx BW, et al. Added value of physical performance measures in predicting adverse health-related events: results from the Health, Aging And Body Composition Study. J Am Geriatr Soc. 2009;57(2):251–9. https://doi.org/10.1111/j.1532-5415.2008.02126.x PMID: 19207142

17. Hanada M, Yamauchi K, Miyazaki S, Oyama Y, Yanagita Y, Sato S, et al. Short-Physical Performance Battery (SPPB) score is associated with postoperative pulmonary complications in elderly patients undergoing lung resection surgery: a prospective multicenter cohort study. Chron Respir Dis. 2020;17. https://doi.org/10.1177/1479973120961846 PMID: 32985263

18. Guralnik JM, Ferrucci L, Pieper CF, Leveille SG, Markides KS, Ostir GV, et al. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. J Gerontol A Biol Sci Med Sci. 2000;55(4):M221–31. https://doi.org/10.1093/gerona/55.4.m221 PMID: 10811152

19. Ramirez CT, Legault NJ, Sketeris KA. Using mobility tools within multidimensional assessments to increase prognostic awareness and assist in shared decision-making to enhance goal concordant care. Semin Oncol Nurs. 2024;40(4):151675. https://doi.org/10.1016/j.soncn.2024.151675 PMID: 38902181

20. Robertson DGE, Caldwell GE, Hamill J, Kamen G, Whittlesey S. Research Methods in Biomechanics. Human Kinetics; 2013.

21. Mobbs RJ, Perring J, Raj SM, Maharaj M, Yoong NKM, Sy LW, et al. Gait metrics analysis utilizing single-point inertial measurement units: a systematic review. Mhealth. 2022;8:9. https://doi.org/10.21037/mhealth-21-17 PMID: 35178440

22. Verghese J, Holtzer R, Lipton RB, Wang C. Quantitative gait markers and incident fall risk in older adults. J Gerontol A Biol Sci Med Sci. 2009;64(8):896–901. https://doi.org/10.1093/gerona/glp033 PMID: 19349593

23. Warmerdam E, Hansen C, Romijnders R, Hobert MA, Welzel J, Maetzler W. Full-body mobility data to validate inertial measurement unit algorithms in healthy and neurological cohorts. Data. 2022;7(10):136. https://doi.org/10.3390/data7100136

24. Park C, Sharafkhaneh A, Bryant MS, Nguyen C, Torres I, Najafi B. Toward remote assessment of physical frailty using sensor-based sit-to-stand test. J Surg Res. 2021;263:130–9. https://doi.org/10.1016/j.jss.2021.01.023 PMID: 33652175

25. Tateoka K, Tsuji T, Shoji T, Tokunaga S, Okura T. Relationship between acceleration in a sit-to-stand movement and physical function in older adults. Geriatrics (Basel). 2023;8(6):123. https://doi.org/10.3390/geriatrics8060123 PMID: 38132494

26. Atrsaei A, Paraschiv-Ionescu A, Krief H, Henchoz Y, Santos-Eggimann B, Büla C, et al. Instrumented 5-time sit-to-stand test: parameters predicting serious falls beyond the duration of the test. Gerontology. 2022;68(5):587–600. https://doi.org/10.1159/000518389 PMID: 34535599

27. Ansai JH, Farche ACS, Rossi PG, de Andrade LP, Nakagawa TH, Takahashi ACDM. Performance of different timed up and go subtasks in frailty syndrome. J Geriatr Phys Ther. 2019;42(4):287–93. https://doi.org/10.1519/JPT.0000000000000162 PMID: 29210935

28. Ortega-Bastidas P, Gómez B, Aqueveque P, Luarte-Martínez S, Cano-de-la-Cuerda R. Instrumented timed up and go test (iTUG)—more than assessing time to predict falls: a systematic review. Sensors. 2023;23(7):3426. https://doi.org/10.3390/s23073426

29. Lam WWT, Tang YM, Fong KNK. A systematic review of the applications of markerless motion capture (MMC) technology for clinical measurement in rehabilitation. J Neuroeng Rehabil. 2023;20(1):57. https://doi.org/10.1186/s12984-023-01186-9 PMID: 37131238

30. Hafer JF, Vitali R, Gurchiek R, Curtze C, Shull P, Cain SM. Challenges and advances in the use of wearable sensors for lower extremity biomechanics. J Biomech. 2023;157:111714. https://doi.org/10.1016/j.jbiomech.2023.111714 PMID: 37423120

31. Scott B, Seyres M, Philp F, Chadwick EK, Blana D. Healthcare applications of single camera markerless motion capture: a scoping review. PeerJ. 2022;10:e13517. https://doi.org/10.7717/peerj.13517 PMID: 35642200

32. Jeyasingh-Jacob J, Crook-Rumsey M, Shah H, Joseph T, Abulikemu S, Daniels S, et al. Markerless motion capture to quantify functional performance in neurodegeneration: systematic review. JMIR Aging. 2024;7:e52582. https://doi.org/10.2196/52582 PMID: 39106477

33. Song K, Hullfish TJ, Scattone Silva R, Silbernagel KG, Baxter JR. Markerless motion capture estimates of lower extremity kinematics and kinetics are comparable to marker-based across 8 movements. J Biomech. 2023;157:111751. https://doi.org/10.1016/j.jbiomech.2023.111751 PMID: 37552921

34. Zhu X, Boukhennoufa I, Liew B, Gao C, Yu W, McDonald-Maier KD, et al. Monocular 3D human pose markerless systems for gait assessment. Bioengineering (Basel). 2023;10(6):653. https://doi.org/10.3390/bioengineering10060653 PMID: 37370583

35. Colyer SL, Evans M, Cosker DP, Salo AIT. A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system. Sports Med Open. 2018;4(1):24. https://doi.org/10.1186/s40798-018-0139-y PMID: 29869300

36. Ceriola L, Taborri J, Donati M, Rossi S, Patanè F, Mileti I. Comparative analysis of markerless motion capture systems for measuring human kinematics. IEEE Sensors J. 2024;24(17):28135–44. https://doi.org/10.1109/jsen.2024.3431873

37. Scataglini S, Abts E, Van Bocxlaer C, Van den Bussche M, Meletani S, Truijen S. Accuracy, validity, and reliability of markerless camera-based 3D motion capture systems versus marker-based 3D motion capture systems in gait analysis: a systematic review and meta-analysis. Sensors (Basel). 2024;24(11):3686. https://doi.org/10.3390/s24113686 PMID: 38894476

38. Johnson PK, Kincaid CJ, Lush NL, Jackson SJ, Eggett D, Feland JB, et al. Administering movement assessments via markerless motion capture provides new normative values over clinical tests. Res Square. 2024. https://doi.org/10.21203/rs.3.rs-3955493/v1

39. Lee SH, Yoon C, Chung SG, Kim HC, Kwak Y, Park H-W, et al. Measurement of shoulder range of motion in patients with adhesive capsulitis using a kinect. PLoS One. 2015;10(6):e0129398. https://doi.org/10.1371/journal.pone.0129398 PMID: 26107943

40. Matsen FA 3rd, Lauder A, Rector K, Keeling P, Cherones AL. Measurement of active shoulder motion using the Kinect, a commercially available infrared position detection system. J Shoulder Elbow Surg. 2016;25(2):216–23. https://doi.org/10.1016/j.jse.2015.07.011 PMID: 26341024

41. Kampel M, Doppelbauer S, Planinc R. Automated Timed Up & Go Test for functional decline assessment of older adults. Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. New York, NY, USA: Association for Computing Machinery; 2018. pp. 208–16. https://doi.org/10.1145/3240925.3240960

42. Kamnardsiri T, Thawinchai N, Parameyong A, Pholjaroen P, Wonglangka K, Prupetkaew P, et al. Conventional video-based system for measuring the subtask speed of the Timed Up and Go Test in older adults: validity and reliability study. PLoS One. 2023;18(6):e0286574. https://doi.org/10.1371/journal.pone.0286574 PMID: 37267315

43. Choi Y, Bae Y, Cha B, Ryu J. Deep Learning-based subtask segmentation of timed up-and-go test using RGB-D Cameras. Sensors (Basel). 2022;22(17):6323. https://doi.org/10.3390/s22176323 PMID: 36080782

44. Herath HMKKMB, Jayasekara AGBP, Madhusanka BGDA, Karunasena GMKB. Non-Invasive Tools for Early Detection and Monitoring of Sarcopenia in Older Individuals. 2023 Moratuwa Engineering Research Conference (MERCon). 2023. pp. 219–24. https://doi.org/10.1109/MERCon60487.2023.10355475

45. Li T, Chen J, Hu C, Ma Y, Wu Z, Wan W, et al. Automatic timed up-and-go sub-task segmentation for Parkinson's disease patients using video-based activity classification. IEEE Trans Neural Syst Rehabil Eng. 2018;26(11):2189–99. https://doi.org/10.1109/TNSRE.2018.2875738 PMID: 30334764

46. Ejupi A, Brodie M, Gschwind YJ, Lord SR, Zagler WL, Delbaere K. Kinect-based five-times-sit-to-stand test for clinical and in-home assessment of fall risk in older people. Gerontology. 2015;62(1):118–24. https://doi.org/10.1159/000381804 PMID: 26021781

47. Maudsley-Barton S, Yap MH. Objective falls risk assessment using markerless motion capture and representational machine learning. Sensors (Basel). 2024;24(14):4593. https://doi.org/10.3390/s24144593 PMID: 39065991

48. Morinan G, Peng Y, Rupprechter S, Weil RS, Leyland L-A, Foltynie T, et al. Computer-vision based method for quantifying rising from chair in Parkinson's disease patients. Intell-Based Med. 2022;6:100046. https://doi.org/10.1016/j.ibmed.2021.100046

49. Shin JH, Yu R, Ong JN, Lee CY, Jeon SH, Park H, et al. Quantitative gait analysis using a pose-estimation algorithm with a single 2D-video of Parkinson's disease patients. J Parkinsons Dis. 2021;11(3):1271–83. https://doi.org/10.3233/JPD-212544 PMID: 33935106

50. Soltaninejad S, Rosales-Castellanos A, Ba F, Ibarra-Manzano MA, Cheng I. Body Movement Monitoring for Parkinson's Disease Patients Using A Smart Sensor Based Non-Invasive Technique. 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom). 2018. pp. 1–6. https://doi.org/10.1109/HealthCom.2018.8531197

51. Rikli RE, Jones CJ. Functional fitness normative scores for community-residing older adults, ages 60-94. J Aging Phys Activity. 1999;7(2):162–81. https://doi.org/10.1123/japa.7.2.162

52. Mayhew AJ, So HY, Ma J, Beauchamp MK, Griffith LE, Kuspinar A, et al. Normative values for grip strength, gait speed, timed up and go, single leg balance, and chair rise derived from the Canadian longitudinal study on ageing. Age Ageing. 2023;52(4):afad054. https://doi.org/10.1093/ageing/afad054 PMID: 37078755

53. Silva PFS, Quintino LF, Franco J, Faria CDCM. Measurement properties and feasibility of clinical tests to assess sit-to-stand/stand-to-sit tasks in subjects with neurological disease: a systematic review. Braz J Phys Ther. 2014;18(2):99–110. https://doi.org/10.1590/s1413-35552012005000155 PMID: 24839043

54. Vaidya T, Chambellan A, de Bisschop C. Sit-to-stand tests for COPD: a literature review. Respir Med. 2017;128:70–7. https://doi.org/10.1016/j.rmed.2017.05.003 PMID: 28610673

55. Cardoso NL, de Sá JF, do Nascimento LFE, Mendes LA, Bruno S, Torres-Castro R, et al. Psychometric properties of the sit-to-stand test for patients with pulmonary hypertension: a systematic review. Braz J Phys Ther. 2025;29(4):101223. https://doi.org/10.1016/j.bjpt.2025.101223 PMID: 40347595

56. Rydwik E, Bergland A, Forsén L, Frändin K. Psychometric properties of timed up and go in elderly people: a systematic review. Phys Occup Ther Geriat. 2011;29(2):102–25. https://doi.org/10.3109/02703181.2011.564725

57. Exter SH, Koenders N, Wees P, Berg MGA. A systematic review of the psychometric properties of physical performance tests for sarcopenia in community-dwelling older adults. Age Ageing. 2024;53(6):afae113. https://doi.org/10.1093/ageing/afae113 PMID: 38851214

58. Christopher A, Kraft E, Olenick H, Kiesling R, Doty A. The reliability and validity of the Timed Up and Go as a clinical tool in individuals with and without disabilities across a lifespan: a systematic review. Disabil Rehabil. 2021;43(13):1799–813. https://doi.org/10.1080/09638288.2019.1682066 PMID: 31656104

59. Kameniar K, Mackintosh S, Van Kessel G, Kumar S. The psychometric properties of the short physical performance battery to assess physical performance in older adults: a systematic review. J Geriatr Phys Ther. 2024;47(1):43–54. https://doi.org/10.1519/JPT.0000000000000337 PMID: 35442231

60. Stevens JA, Phelan EA. Development of STEADI: a fall prevention resource for health care providers. Health Promot Pract. 2013;14(5):706–14. https://doi.org/10.1177/1524839912463576 PMID: 23159993

61. Montero-Odasso M, van der Velde N, Martin FC, Petrovic M, Tan MP, Ryg J, et al. World guidelines for falls prevention and management for older adults: a global initiative. Age Ageing. 2022;51(9):afac205. https://doi.org/10.1093/ageing/afac205 PMID: 36178003

62. Eckstrom E, Vincenzo JL, Casey CM, Gray S, Cosley K, Caulley J, et al. American Geriatrics Society response to the World Falls Guidelines. J Am Geriatr Soc. 2024;72(6):1669–86. https://doi.org/10.1111/jgs.18734 PMID: 38131656

63. Watanabe SM, Nekolaichuk C, Beaumont C, Johnson L, Myers J, Strasser F. A multicenter study comparing two numerical versions of the Edmonton Symptom Assessment System in palliative care patients. J Pain Symptom Manage. 2011;41(2):456–68. https://doi.org/10.1016/j.jpainsymman.2010.04.020 PMID: 20832987

64. Rolfson DB, Majumdar SR, Tsuyuki RT, Tahir A, Rockwood K. Validity and reliability of the Edmonton Frail Scale. Age Ageing. 2006;35(5):526–9. https://doi.org/10.1093/ageing/afl041 PMID: 16757522

65. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727–36. https://doi.org/10.1007/s11136-011-9903-x PMID: 21479777

66. Hlatky MA, Boineau RE, Higginbotham MB, Lee KL, Mark DB, Califf RM, et al. A brief self-administered questionnaire to determine functional capacity (the Duke Activity Status Index). Am J Cardiol. 1989;64(10):651–4. https://doi.org/10.1016/0002-9149(89)90496-7 PMID: 2782256

67. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373–83. https://doi.org/10.1016/0021-9681(87)90171-8 PMID: 3558716

68. Steffen TM, Hacker TA, Mollinger L. Age- and gender-related test performance in community-dwelling elderly people: six-minute walk test, berg balance scale, timed up & go test, and gait speeds. Physical Therapy. 2002;82(2):128–37. https://doi.org/10.1093/ptj/82.2.128

69. Zakaria NA, Kuwae Y, Tamura T, Minato K, Kanaya S. Quantitative analysis of fall risk using TUG test. Comput Methods Biomech Biomed Engin. 2015;18(4):426–37. https://doi.org/10.1080/10255842.2013.805211 PMID: 23964848

70. Millor N, Lecumberri P, Gómez M, Martínez-Ramírez A, Izquierdo M. An evaluation of the 30-s chair stand test in older adults: frailty detection based on kinematic parameters from a single inertial unit. J Neuroeng Rehabil. 2013;10:86. https://doi.org/10.1186/1743-0003-10-86 PMID: 24059755

71. Millor N, Cadore EL, Gómez M, Martínez A, Lecumberri P, Martirikorena J, et al. High density muscle size and muscle power are associated with both gait and sit-to-stand kinematic parameters in frail nonagenarians. J Biomech. 2020;105:109766. https://doi.org/10.1016/j.jbiomech.2020.109766

72. Kumar V, Yoshiike T, Shibata T. Predicting sit-to-stand adaptations due to muscle strength deficits and assistance trajectories to complement them. Front Bioeng Biotechnol. 2022;10:799836. https://doi.org/10.3389/fbioe.2022.799836 PMID: 35372315

73. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–63. https://doi.org/10.1016/j.jcm.2016.02.012 PMID: 27330520

74. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159. https://doi.org/10.2307/2529310

75. Seol H. seolmatrix: Correlations suite for jamovi (Version 3.7.1). 2023. Available from: https://github.com/hyunsooseol/seolmatrix

76. Gamer M, Lemon J, Fellows I, Singh P. Package "irr." 2012. Available from: https://www.r-project.org/

77. Pavasini R, Guralnik J, Brown JC, di Bari M, Cesari M, Landi F, et al. Short physical performance battery and all-cause mortality: systematic review and meta-analysis. BMC Med. 2016;14(1):215. https://doi.org/10.1186/s12916-016-0763-7 PMID: 28003033

78. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. N Engl J Med. 1995;332(9):556–61. https://doi.org/10.1056/NEJM199503023320902 PMID: 7838189

79. Ooi H, Welch C. Obstacles to the early diagnosis and management of sarcopenia: current perspectives. Clin Interv Aging. 2024;19:323–32. https://doi.org/10.2147/CIA.S438144 PMID: 38404480

80. Crane HM, Miller ME, Pierce J, Willig AL, Case ML, Wilkin AM, et al. Physical Functioning Among Patients Aging With Human Immunodeficiency Virus (HIV) Versus HIV uninfected: feasibility of using the short physical performance battery in clinical care of people living with HIV Aged 50 or Older. Open Forum Infect Dis. 2019;6(3):ofz038. https://doi.org/10.1093/ofid/ofz038 PMID: 30882010

81. Gauthier ML, Unverzagt CA, Mendonça LDM, Seitz AL. Missing the forest for the trees: a lack of upper extremity physical performance testing in sports physical therapy. Int J Sports Phys Ther. 2023;18(2):419–30. https://doi.org/10.26603/001c.73791 PMID: 37020447

82. McDonnell B, Stillwell S, Hart S, Davis RB. Breaking down barriers to the utilization of standardized tests and outcome measures in acute care physical therapist practice: an observational longitudinal study. Phys Ther. 2018;98(6):528–38. https://doi.org/10.1093/ptj/pzy032 PMID: 29471539

83. 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):125–35. https://doi.org/10.2522/ptj.20080234 PMID: 19074618

84. Reijnierse EM, de van der Schueren MAE, Trappenburg MC, Doves M, Meskers CGM, Maier AB. Lack of knowledge and availability of diagnostic equipment could hinder the diagnosis of sarcopenia and its management. PLoS One. 2017;12(10):e0185837. https://doi.org/10.1371/journal.pone.0185837 PMID: 28968456

85. Swinkels RAHM, van Peppen RPS, Wittink H, Custers JWH, Beurskens AJHM. Current use and barriers and facilitators for implementation of standardised measures in physical therapy in the Netherlands. BMC Musculoskelet Disord. 2011;12:106. https://doi.org/10.1186/1471-2474-12-106 PMID: 21600045

86. Liu X-T, Nikkhoo M, Wang L, Chen CP, Chen H-B, Chen C-J, et al. Feasibility of a kinect-based system in assessing physical function of the elderly for home-based care. BMC Geriatr. 2023;23(1):495. https://doi.org/10.1186/s12877-023-04179-4 PMID: 37587451

87. Hsiao M-Y, Li C-M, Lu I-S, Lin Y-H, Wang T-G, Han D-S. An investigation of the use of the Kinect system as a measure of dynamic balance and forward reach in the elderly. Clin Rehabil. 2018;32(4):473–82. https://doi.org/10.1177/0269215517730117 PMID: 28879781

88. Jung H-W, Roh H, Cho Y, Jeong J, Shin Y-S, Lim J-Y, et al. Validation of a Multi-Sensor-Based Kiosk for Short Physical Performance Battery. J Am Geriatr Soc. 2019;67(12):2605–9. https://doi.org/10.1111/jgs.16135 PMID: 31441514

89. Hernandez HHC, Ong EH, Heyzer L, Tan CN, Ghazali F, Yang DZ, et al. Validation of a multi-sensor-based kiosk in the use of the short physical performance battery in older adults attending a fall and balance clinic. Ann Geriatr Med Res. 2022;26(2):125–33. https://doi.org/10.4235/agmr.22.0022

90. Yang DZ, Rodrigues EE, Hernandez HHC, Ong EH, Heyzer L, Tan CN, et al. Validation of the revised multi-sensor-based electronic Short Physical Performance Battery (eSPPB) kiosk in community-dwelling older adults. Eur Geriatr Med. 2025;16(3):975–83. https://doi.org/10.1007/s41999-024-01151-2 PMID: 39969803

91. Kaewkaen K, Uttama S, Rueangsirarak W, Kaewkaen P. Test-retest reliability of the five times sit-to-stand test measured using the kinect in older adults. J Assoc Med Sci. 2019;52:138–44.

92. Heslop PA, Hurst C, Sayer AA, Witham MD. Remote collection of physical performance measures for older people: a systematic review. Age Ageing. 2023;52(1):afac327. https://doi.org/10.1093/ageing/afac327 PMID: 36721962

93. Shaw J, Abejirinde I-OO, Agarwal P, Shahid S, Martin D. Digital health and equitable access to care. PLOS Digit Health. 2024;3(9):e0000573. https://doi.org/10.1371/journal.pdig.0000573 PMID: 39321176

94. Rowland S, Brewer LC, Rosas LG. Digital health equity - a call to action for clinical and translational scientists. J Clin Transl Sci. 2024;8(1):e145. https://doi.org/10.1017/cts.2024.564 PMID: 39478785

95. Buolamwini J, Gebru T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency. PMLR; 2018. pp. 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

96. Daneshjou R, Vodrahalli K, Novoa RA, Jenkins M, Liang W, Rotemberg V, et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. Sci Adv. 2022;8(32):eabq6147. https://doi.org/10.1126/sciadv.abq6147 PMID: 35960806

97. Abràmoff MD, Tarver ME, Loyo-Berrios N, Trujillo S, Char D, Obermeyer Z, et al. Considerations for addressing bias in artificial intelligence for health equity. NPJ Digit Med. 2023;6(1):170. https://doi.org/10.1038/s41746-023-00913-9 PMID: 37700029

98. Celi LA, Cellini J, Charpignon M-L, Dee EC, Dernoncourt F, Eber R, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities-a global review. PLOS Digit Health. 2022;1(3):e0000022. https://doi.org/10.1371/journal.pdig.0000022 PMID: 36812532

99. Dankwa-Mullan I. Health equity and ethical considerations in using artificial intelligence in public health and medicine. Prev Chronic Dis. 2024;21:E64. https://doi.org/10.5888/pcd21.240245 PMID: 39173183

100. Middleton A, Fritz SL, Lusardi M. Walking speed: the functional vital sign. J Aging Phys Act. 2015;23(2):314–22. https://doi.org/10.1123/japa.2013-0236 PMID: 24812254

101. Richardson J, Letts L, Chan D, Officer A, Wojkowski S, Oliver D, et al. Monitoring physical functioning as the sixth vital sign: evaluating patient and practice engagement in chronic illness care in a primary care setting--a quasi-experimental design. BMC Fam Pract. 2012;13:29. https://doi.org/10.1186/1471-2296-13-29 PMID: 22471378

102. Ruiz-Cárdenas JD, Montemurro A, Martínez-García MDM, Rodríguez-Juan JJ. Sit-to-stand video analysis-based app for diagnosing sarcopenia and its relationship with health-related risk factors and frailty in community-dwelling older adults: diagnostic accuracy study. J Med Internet Res. 2023;25:e47873. https://doi.org/10.2196/47873 PMID: 38064268

103. Thomas J, Hall JB, Bliss R, Guess TM. Comparison of Azure Kinect and optical retroreflective motion capture for kinematic and spatiotemporal evaluation of the sit-to-stand test. Gait Posture. 2022;94:153–9. https://doi.org/10.1016/j.gaitpost.2022.03.011 PMID: 35334335

104. Tulipani LJ, Meyer B, Fox S, Solomon AJ, Mcginnis RS. The Sit-to-Stand transition as a biomarker for impairment: comparison of instrumented 30-second chair stand test and daily life transitions in multiple sclerosis. IEEE Trans Neural Syst Rehabil Eng. 2022;30:1213–22. https://doi.org/10.1109/TNSRE.2022.3169962 PMID: 35468063

105. Piano L, Geri T, Testa M. Raising and stabilization phase of the Sit-to-Stand movement better discriminate healthy elderly adults from young subjects: a pilot cross-sectional study. Arch Physiother. 2020;10:7. https://doi.org/10.1186/s40945-020-00078-8 PMID: 32318274

106. van Lummel RC, Walgaard S, Maier AB, Ainsworth E, Beek PJ, van Dieën JH. The instrumented Sit-to-Stand Test (iSTS) has greater clinical relevance than the manually recorded Sit-to-Stand Test in older adults. PLoS One. 2016;11(7):e0157968. https://doi.org/10.1371/journal.pone.0157968 PMID: 27391082