SYSTEMATIC REVIEW

Measures of attributes of locomotor capacity in older people: a systematic literature review following the COSMIN methodology

Germain Honvo1,2, Shaun Sabico3, Nicola Veronese3,4, Olivier Bruyère1,2, René Rizzoli1,5, Jotheeswaran Amuthavalli Thiagarajan6, Christopher Mikton7, Theresa Diaz8, Cyrus Cooper1,9, Jean-Yves Reginster1,2,3

1World Health Organization (WHO) Collaborating Center for Epidemiology of Musculoskeletal Health and Ageing, University of Liège, Liège, Belgium
2Division of Public Health, Epidemiology and Health Economics, University of Liège, Liège, Belgium
3Chair for Biomarkers of Chronic Diseases, Biochemistry Department, College of Science, King Saud University, Riyadh, Kingdom of Saudi Arabia
4Geriatric Unit, Department of Internal Medicine and Geriatrics, University of Palermo, Palermo, Italy
5Division of Bone Diseases, Geneva University Hospitals and Faculty of Medicine, Geneva, Switzerland
6Ageing and Health Unit, Department of Maternal, Newborn, Child, Adolescent Health and Ageing, World Health Organization (WHO), Geneva, Switzerland
7Demographic Change and Healthy Aging Unit, Social Determinants of Health, World Health Organization, Geneva, Switzerland
8Epidemiology, Monitoring and Evaluation Unit, Maternal, Newborn, Child, Adolescent Health and Ageing, World Health Organization, Geneva, Switzerland
9MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton, UK

Address correspondence to: Germain Honvo, University of Liège, Division of Public Health, Epidemiology and Health Economics, CHU - Sart Tilman, Quartier Hôpital, Avenue Hippocrate 13, Bât. B23, 4000 LIÈGE, Belgium, Email: germain.honvo@uliege.be

Abstract

Background: Locomotor capacity (LC) is an important domain of intrinsic capacity and key determinant of functional ability and well-being in older age. The United Nations Decade of Healthy Ageing (2021–2030) calls for strengthening data and research on healthy ageing, including the measurement of older persons’ LC. To advance the measurement and monitoring of LC, there is pressing need to identify valid and reliable measures.

Objective: To identify all the available tools that were validated for measurement of LC or of its specific attributes in older people and to assess the methodological quality of the studies and measurement properties of the tools.

Design: Systematic review.

Setting: Anywhere (Community-dwelling; long-term care facility; etc.)

Subjects: Older people.

Methods: We used highly sensitive search strategies to search the following databases: Medline, Embase, Scopus, CINAHL and PsycINFO. The study was conducted following the CONsensus-based Standards for the selection of health Measurement Instruments (COSMIN) methodology for systematic review of outcome measurement instruments.

Results: A total of 125 studies were included, which assessed tools for balance (n = 84), muscle power (n = 12), muscle strength (n = 32, including four studies about tools for balance and muscle power) and endurance (n = 1). No studies on tools for muscle function, joint function, or locomotor capacity overall, were retrieved. We identified 69 clinician-report or objective assessment tools for balance, 30 for muscle strength, 12 for muscle power and 1 endurance assessment tool. The GRADE assessment of quality of evidence showed that only a few tools have high quality evidence for both sufficient validity and reliability: The Balance Evaluation Systems Test (BESTest), the Mini-Balance Evaluation Systems Test (Mini-BESTest), the Berg Balance Scale (BBS) and the Timed Up and Go (TUG) test.
Conclusions: A few tools with high quality evidence for sufficient validity and reliability are currently available for balance assessment in older people that may be recommended for use in clinical and research settings. Further validation studies are required for muscle strength, muscle power and endurance assessment tools.

Keywords: locomotor capacity, balance, endurance, muscle strength, muscle power, muscle function, joint function, screening or assessment tools, measurement properties, older people, systematic review

Key Points
• We identified 69 tools for balance, 30 for muscle strength, 12 for muscle power and 1 endurance assessment tool.
• Only a few tools with high quality evidence for sufficient validity and reliability are available to assess balance.
• Further validation studies are required for muscle strength, muscle power and endurance assessment tools.
• Several issues to be addressed by the WHO Locomotor Capacity Working Group were identified.

Introduction

Healthy ageing is defined by the World Health Organization (WHO) as ‘the process of developing and maintaining the functional ability that enables wellbeing in older age’ [1, 2]. Functional ability comprises the health-related attributes that enable people to be and to do what they have reason to value. It is made up of the intrinsic capacity of the individual, relevant environmental characteristics and the interactions between the individual and these characteristics. Intrinsic capacity is the composite of all the physical and mental capacities of an individual, including visual and hearing capacities, cognitive and psychological capacities, vitality and locomotor capacity [3].

Locomotor capacity is an important domain of intrinsic capacity and key determinant for functional ability and wellbeing in older age. The WHO expert working group on locomotor capacity, consisting of a fifty clinicians and scientists in fields of musculoskeletal health and ageing, from all regions of the world, proposed a working definition of locomotor capacity as ‘a state (static or dynamic over time) of the musculoskeletal system that encompasses endurance, balance, muscle strength, muscle function, muscle power and a joint function of the body’ [4]. As a next step, this systematic review was conducted to identify valid, reliable and responsive measures of locomotor capacity and of its attributes.

The United Nations Decade of Healthy Ageing (2021–2030), endorsed by the World Health Assembly and the United Nations (UN) General Assembly, recognises the importance of strengthening data for measurement, monitoring and evaluation, as a facilitator of progress assessment against goals in the prioritised four action areas [5]. These key action areas include: (i) changing how we think, feel and act towards age and ageing (i.e. combatting ageism); (ii) ensuring that communities foster the abilities of older people (i.e. developing age-friendly environments); (iii) delivering person-centred integrated care and primary health services that are responsive to older people and (iv) providing access to long-term care for older people who need it [6]. A systematic assessment of best available measures of locomotor capacity is therefore essential to develop recommendations for use in population surveys and routine health information, as well as for individual assessments of patients by clinicians.

Evidence suggests that systematic literature reviews can help identify the available outcome measures in specific fields, thus providing a comprehensive overview of their measurement properties as well as supporting evidence-based recommendations for use in research and clinical practice [7]. A recent systematic review was published that has identified commonly used tests of balance and strength and evaluated their measurement properties in young seniors (aged 60–70 years) [8]. However, to the best of our knowledge, no systematic review has comprehensively assessed measurement tools for all attributes of locomotor capacity considering all stages of older age.

Research question, objectives and purpose

The research question for this systematic literature review is: What are the available and validated tools to measure the specific attributes of locomotor capacity, or locomotor capacity overall, in older people?

The objectives of the study were to comprehensively review the available outcome measurement instruments that were validated for specific attributes of locomotor capacity or for locomotor capacity overall in older people and to assess the methodological quality of the studies and measurement properties of the tools. The findings of this systematic review will support WHO in developing evidence-based recommendations for use of these tools in population surveys and data collection in health care facilities.

Methods

Guidelines and protocol registration

This systematic review was conducted following the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) methodology [7]. Recommendations in the Cochrane handbook for systematic literature reviews were also followed for screening and selection of studies [9]. The current report follows the
Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [10].

The protocol of this study was registered in the International Prospective Register of Systematic Reviews (PROSPERO: Registration number, CRD42022318959).

The Covidence online software (https://www.covidence.org/) was used to manage the entire study selection process, from title/abstract screening to full-text selection.

**Information sources and search strategies**

To conduct this systematic literature review, several bibliographic databases were comprehensively searched (from inception to April 18, 2022) using detailed and highly sensitive search strategies tailored to the syntax of each database. These databases include: Medline (via Ovid), Embase, Scopus, Cumulative Index to Nursing and Allied Health Literature (CINAHL) and PsycINFO (via Ovid). The search for individual studies in the databases was subsequently supplemented by manual search of Google and of references of relevant systematic reviews that were identified, along with references of included studies.

To guide the identification of adequate keywords to build the search strategies, the research question was framed into the ‘Participants, Intervention, Comparison, Outcome’ (PICO) format, following the framework proposed in the WHO handbook for guideline development, section on systematic review question formulation [11]. The PICO format question is as follows: What are the available and validated screening or assessment tools (I) for measuring endurance, balance, muscle strength, muscle function, muscle power and joint function or locomotor capacity overall (O) in older people (aged 60 years and older) (P)?

The terms of this PICO format question (i.e. P, I and O) were then adequately combined (with Boolean operators) to build the search strategies, using free vocabulary words and controlled terms tailored to databases. To search for records relating to screening or assessment tools (I), we used the Ovid search filter for patient-reported outcome measurement (PROM) that was developed by the Oxford ‘PROM Group Construct & Instrument Type Filters’ [12], which we adapted to fit best with our review question and to limit background noise while being sensitive. We also used additional search strings for ‘measurement tool’ (I) developed by our review team. In the end, the PubMed exclusion filter developed by Terwee et al. [13] was adapted for Ovid and used to remove irrelevant records, such as case reports and animal studies, from the search results. The exclusion filter was used exactly as indicated by Terwee et al. [13]. The search strategies developed for all databases are provided as Supplementary material to this paper (Appendix 1).

**Eligibility criteria**

**Inclusion criteria:**

Individual studies on screening or assessment tools for both objective or self-reported assessment of specific attributes of locomotor capacity (i.e. endurance, balance, muscle strength, muscle function, muscle power, joint function), or of locomotor capacity overall, in older people (aged 60 years and older), were included in this systematic literature review.

The specific selection criteria regarding the study population in articles were as follows: a) studies include older people aged 60 years of age or older or b) studies with a mean age of sample above 65 years or c) studies with at least 50% of the sample (defined as majority [14]) with older people aged 60 years or older or d) studies separately report results on participants aged 60 years or older. Original studies on development and validation of tools, aiming at evaluating one or more measurements properties, as well as studies reporting their translation, cross-cultural adaptation and validation in other languages or settings, or in older people were included. Finally, studies examining the measurement properties of more than one measurement instrument for the same attribute or for several distinct attributes of locomotor capacity were also included.

**Exclusion criteria**

Validation studies in populations with specific medical conditions (e.g. Parkinson’s disease, stroke, etc.), even if in older people, were excluded from this systematic literature review. Likewise, studies that did not report data on measurement properties of tools were excluded, as were studies in which a measurement tool was used in a validation study of another instrument (i.e. the instrument to be considered is the one that is being validated). Articles in which a measurement tool is used only for outcome measurement in an experimental study were also considered ineligible for inclusion in this review, as were review papers (systematic or not) and editorials. Finally, abstracts reporting studies on measurement tools without full-text reports, and articles in languages other than English were excluded.

**Study selection**

We followed recommendations in the Cochrane handbook for Systematic reviews to select studies based on title/abstract first, then on full manuscripts [9]. The title/abstract selection was independently done by three review authors (GH, SS, NV), and the full text selection by two members of the review team (GH, SS), with consensus meetings to discuss any disagreements. A third member of the review team (NV) was involved for final decision on full text selection, when necessary.

**Data collection and data items**

All the data were extracted by one reviewer (GH), then the extractions were independently checked by a second review author (SS) for identification and correction of inaccuracies. Items collected from the retrieved full-text articles were information for identification of the manuscript, data on the characteristics of the study population, as well as data on characteristics of the tools and on their measurement...
Assessment of measurement properties of tools

The measurement properties of the included tools were assessed by applying the updated COSMIN criteria for good measurement properties [7]. For each included studies and tools, each measurement property was rated as either sufficient (+), insufficient (−), or indeterminate (?). Measurement properties for all the included studies were assessed by the lead author (GH) and cross-checked by another member of the review team (SS).

Regarding hypotheses testing for construct validity and responsiveness, we pre-formulated hypotheses to evaluate the results of the included studies, so that all results are compared against the same set of hypotheses, as recommended by the COSMIN methodology [7]. For convergent validity, the following hypotheses were formulated: 1) Correlations (Pearson’s, Spearman’s correlations, or Intra-class correlation coefficients [ICC]) or Kappa coefficient for concordance with instruments measuring similar constructs should be >0.50; 2) Correlations with instruments measuring related but not similar constructs (e.g. a balance assessment tool validated against a gait speed test) should be between 0.30 and 0.50; 3) Correlations with instruments measuring dissimilar constructs should be <0.30. For discriminative (knowledge) validity, scores of instruments should be significantly different between relevant subgroups (e.g. patients with history of falls versus patients without history of falls, for balance assessment tools), whatever the statistical method used for comparison. In the end, for responsiveness, area under the curve (AUC) with an external measure of change used as the gold standard should be ≥0.70, as per COSMIN methodology [7].

For ICC or other correlation values, when range of values (e.g. ICC = 0.52–0.89) or multiple values for the same measurement property (e.g. ICC = 0.50 for inter-rater, and 0.88 for test–retest reliability) were available from a single study and tool, the best value was considered for measurement property rating.

Data synthesis and GRADE assessment of findings

Data extracted from the retrieved articles were summarised in tables presenting the main characteristics of the included studies and tools, as well as information on the measurement properties of the tools. Qualitative summaries of results of measurement properties were presented, based on data from all the included studies on each specific tool, according to the COSMIN guideline [7, 15]. Overall measurement property ratings were performed for each tool, considering the summary results, as recommended by the COSMIN guideline [7, 15].

For tools with at least two validation studies included, we assessed the quality of evidence on measurement properties using the modified GRADE (Grading of Recommendations, Assessment, Development and Evaluations) approach, as described by the COSMIN guideline [7, 15].

Results

Literature search result

From 31,146 records retrieved from databases search, 117 individual studies were included, after exclusions. Eight (8) additional studies were found from manual search of Google and of references of studies included from databases search, bearing the total number of included studies to 125. An overview of the flow of studies selection with reasons for full texts exclusions is presented in Figure 1.

Most of the included studies (n = 84) were on balance assessment tools [19–102], and several studies evaluated multiple balance tools (two to four tools). Twelve (12) studies assessed muscle power tools [103–114]. Muscle strength assessment tools were evaluated in 32 studies [73, 108, 112, 113, 115–142], among which three studies [135, 138, 141]
evaluated multiple tools for muscle strength. Two (2) of the studies on muscle strength also assessed another instrument for muscle power [108, 113], while one study assessed an instrument for both muscle strength and power [112], and one other study evaluated the measurement properties of a same tool to assess both muscle strength and balance [73]. Ultimately, only one study was retrieved as a validation study of a tool to assess endurance [143] in older people. The literature search returned no studies validating tools to assess muscle function, joint function or locomotor capacity overall, in older people.

**Characteristics of included studies**

The main characteristics of the included studies and populations are summarised, and separately reported for balance, muscle strength, muscle power and endurance assessment tools in Appendix 2.

Appendix 2a describes the characteristics of studies on balance tools. The mean age of participants in most of these studies was ≥70 years. Only about a quarter of studies (n = 22) had a sample size of 100 or more. Most studies included more women (> 50% of the sample) than men, while five studies included women only [26, 36, 55, 57, 95] and 11 did not report the percentage of female included. In 80 out of the 84 studies on balance assessment tools, participants were recruited in the community and/or long-term care facilities (i.e. nursing homes, residential care facilities; homes for the elderly; etc.), with the dominant setting being the community. Three studies recruited patients from other settings such as rehabilitation centers and day unit for elderly [38, 47, 53], and the setting was not reported in one study [43]. Figure 2.a shows the geographical distribution of studies, with number of studies by countries: Of the 84 studies on balance assessment tools, 35 studies were conducted in North America (USA and Canada).

The characteristics of the 32 studies on muscle strength assessment tools are presented in Appendix 2b. In 75% of these studies, the mean age of patients was >70 years. As for balance tools studies, women were more represented in the studies on muscle strength tools than men: The percentage of female was >50% in 21 studies, while five studies included 100% women [108, 131, 133, 136, 138] and one study included men only [113]; three studies included ≤50% of women in their sample, and this information was not reported in two studies. Participants were mainly recruited from the community, but also from long-term care facilities. Most of the studies originated from USA (seven studies) and Canada (four studies) (Figure 2.b).

With regard to studies on muscle power assessment tools (Appendix 2c), four included female only [106, 108–110], and one male only [113]; in almost all of the other studies, the percentage of female was >50%. The mean age was >70 years in nearly all the studies on muscle power assessment tools, which included patient recruited mainly from the community (9 of 12 studies). The three other studies included patients from long-term care facilities (two studies) and a geriatric clinic. More than 80% of these studies were conducted in USA and European countries (Figure 2.c).
The only one study retrieved on endurance assessment tools (Appendix 2d) was conducted in the USA and included 77 participants (73.1 ± 7.2 years) recruited from the community, who were mainly female (62.3%).

In summary, in terms of geographic distribution of validation studies, considering all attributes of locomotor capacity, it is worth noting that none of these studies were conducted, neither in African countries, nor in countries such as China, France, Russia and most studies come from the USA.

**Included tools and measurement properties**

The characteristics of included tools, including details about attributes measured, mode of administration, number of items (where applicable) and scoring, are described in Appendix 3. The identified tools are PerFOMs (Objective assessment) or ClinROMs (Clinician report); no patient-reported outcomes measures (PROMs) were included. Here, we summarise the measurement properties of these instruments by attributes of locomotor capacity. Tables 1–4, reports summary results of measurement properties for all included tools, with overall quality ratings against the COSMIN criteria for good measurement properties [7]. The detailed data on measurement properties for all included tools (from individual studies), with quality ratings (by study and overall) are shown in Appendix 4(a-d).

**Balance**

A total of 69 tools were identified from the 84 included papers validating balance assessment instruments. Table 1 presents the summary results of measurement properties by specific tools, with overall ratings of measurement properties (for detailed data, see Appendix 4.a). Reliability, measurement error, criterion validity and convergent and discriminative validity (construct validity) were the most frequently reported measurement properties. A very few studies evaluated content validity and structural validity, but as these measurement properties were found to be marginally reported in the included studies, they were not assessed in this systematic review. No studies reported cross-cultural validity. Fifteen (15) of the identified balance tools were validated in at least two studies, including eight tools which were assessed in at least three studies. These eight tools, ranked by numbers of validation studies are:

- **The Berg Balance Scale (BBS)**, with evidence for sufficient reliability (ICC ≥ 0.77) and construct validity (convergent and discriminative validity), using summary results from 11 studies.
- **The Balance Evaluation Systems Test (BESTest)**, with evidence for sufficient reliability (ICC ≥ 0.77), criterion validity ($r \geq 0.70$) and construct validity (convergent and discriminative validity), from six studies.
**Table 1.** Measurement properties of balance assessment tools: Summary results with overall quality ratings: Summary results of measurement properties of tools with quality ratings

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reference</th>
<th>Reliability</th>
<th>Measurement error</th>
<th>Internal consistency</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Spanish version of the BESTest (Spanish BESTest)</td>
<td>Dominguez-Olivan, 2020 [25]</td>
<td>ICC = 0.97 (+)</td>
<td>MIC not defined (?) Criteria not met (?)</td>
<td>N/R</td>
<td>$r &lt; 0.70$ (−) $r &gt; 0.50$ (+)</td>
</tr>
<tr>
<td>The Spanish version of the Mini-BESTest (Spanish Mini-BESTest)</td>
<td>Dominguez-Olivan, 2020 [25]</td>
<td>ICC = 0.79 (+)</td>
<td>MIC not defined (?) Criteria not met (?)</td>
<td>N/R</td>
<td>$r = 0.18$ (−) Hypothesis confirmed (+)</td>
</tr>
<tr>
<td>The modified Clinical test of Sensory Interaction in Balance (mCTSIB) of the Balance Platform Biodex Balance System (BBS)</td>
<td>Antoniadou, 2020 [26]</td>
<td>ICC = 0.628 (−)</td>
<td>N/R N/R</td>
<td>N/R</td>
<td>$r &gt; 0.50$ (+)</td>
</tr>
<tr>
<td>The Brazilian version of the Berg balance scale (Brazilian BBS)</td>
<td>Miyamoto, 2004 [35]</td>
<td>ICC ≥ 0.99 (+)</td>
<td>N/R N/R</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The Lateral Reach (LR) Test</td>
<td>Brauer, 1999 [36]</td>
<td>ICC = 0.999 (+)</td>
<td>N/R N/R</td>
<td>N/R</td>
<td>$r &lt; 0.70$ (−) N/R</td>
</tr>
<tr>
<td>The Six-Spot Step Test</td>
<td>Brincks, 2021 [37]</td>
<td>ICC = 0.94–0.96 (+)</td>
<td>MIC not defined (?) N/R</td>
<td>N/R</td>
<td>$r &gt; 0.50$ (+)</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reference</th>
<th>Reliability</th>
<th>Measurement error</th>
<th>Internal consistency</th>
<th>Validity</th>
<th>Hypothesis testing for construct validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gait Initiation Assessment</td>
<td>Chang, 1999 [42]</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>Hypothesis confirmed (+)</td>
</tr>
<tr>
<td>The modified Wii Fit balance board</td>
<td>Chang, 2013 [43]</td>
<td>ICC ≥ 0.93 (+)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>Hypotheses not confirmed (−)</td>
</tr>
<tr>
<td>The Stepping Threshold Test (STT)</td>
<td>Adams, 2021 [44]</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>Hypothesis not confirmed (−)</td>
</tr>
<tr>
<td>The Unstable board (DYJOC BOARD, SAKAI Medical Co., Ltd.)</td>
<td>Akizuki, 2018 [45]</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The limits of stability (LOS) test</td>
<td>Clark, 1997 [46]</td>
<td>ICC not reported (?)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>Hypothesis not confirmed (−)</td>
</tr>
<tr>
<td>The Four Square Step Test (FSST)</td>
<td>Cleary, 2017 [48]</td>
<td>ICC ≥ 0.98 (+)</td>
<td>N/R</td>
<td>N/R</td>
<td>Inconsistent (+)</td>
<td>(±), for convergent validity (+), for discriminative validity</td>
</tr>
<tr>
<td>The mediolateral balance assessment (MELBA) tool</td>
<td>Cofre Lizama, 2015 [50]</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>r &lt; 0.70 (−)</td>
<td>N/R</td>
</tr>
<tr>
<td>The Spring Scale Test (SST)</td>
<td>DePasquale, 2009 [51]</td>
<td>ICC &gt; 0.94 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>Hypotheses confirmed (+)</td>
</tr>
<tr>
<td>The Microsoft Xbox One Kinect (Kinect v2)</td>
<td>Eltoukhy, 2018 [52]</td>
<td>ICC &gt; 0.75 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>r not reported (?)</td>
<td>Hypothesis confirmed (+)</td>
</tr>
<tr>
<td>The TURN 180 test</td>
<td>Fitzpatrick, 2005 [53]</td>
<td>ICC &gt; 0.828 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>r not reported (?)</td>
<td>Hypothesis confirmed (+)</td>
</tr>
<tr>
<td>The Lower Quarter-Y-Balance Test (LQ-YBT)</td>
<td>Ranji, 2020 [54]</td>
<td>ICC ≥ 0.98 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>Hypothesis not confirmed (−)</td>
</tr>
<tr>
<td>The Narrow Path Walking Test (NPWT)</td>
<td>Gimmon, 2013 [56]</td>
<td>ICC = 0.77–0.92 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>Inconsistent (±)</td>
</tr>
<tr>
<td>One leg standing (OLS)</td>
<td>Giorgetti, 1998 [40]</td>
<td>ICC ≥ 0.75 (+)</td>
<td>N/R</td>
<td>N/R</td>
<td>Indeterminate (?)</td>
<td>(−), for convergent validity (−), for convergent validity</td>
</tr>
<tr>
<td>Tandem Gait (TG)</td>
<td>Giorgetti, 1998 [40]</td>
<td>ICC = 0.31 (−)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>(±), for convergent validity (+), for discriminative validity</td>
</tr>
<tr>
<td>The five-times-sit-to-stand test (FTSTS)</td>
<td>Goldberg, 2012 [57]</td>
<td>ICC = 0.95 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>Inconsistent (±)</td>
</tr>
<tr>
<td>The Maximum Step Length (MSL) test</td>
<td>Goldberg, 2010 [58]</td>
<td>ICC = 0.90–0.96 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>Inconsistent (±), for discriminative validity (+), for discriminative validity</td>
</tr>
<tr>
<td>The Thirty Rapid-Step test (30-RST)</td>
<td>Goldberg, 2015 [59]</td>
<td>ICC = 0.85 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>Inconsistent (±)</td>
</tr>
<tr>
<td>The Community Balance and Mobility Scale (CBM)</td>
<td>Weber, 2018 [60]</td>
<td>ICC ≥ 0.97 (+)</td>
<td>N/R</td>
<td>N/R</td>
<td>Criteria not met (?)</td>
<td>Hypothesis confirmed (+)</td>
</tr>
<tr>
<td>The <strong>German</strong>-Community Balance and Mobility Scale (German CBM)</td>
<td>Gordt, 2019 [61]</td>
<td>ICC ≥ 0.99 (+)</td>
<td>N/R</td>
<td>Criteria not met (?)</td>
<td>r &gt; 0.70 (+)</td>
<td>Hypothesis confirmed (+)</td>
</tr>
<tr>
<td>The Shortened version of the Community Balance and Mobility Scale (s-CBM)</td>
<td>Gordt, 2020 [62]</td>
<td>N/R</td>
<td>N/R</td>
<td>Criteria met (+)</td>
<td>N/R</td>
<td>Inconsistent (±)</td>
</tr>
<tr>
<td>The ‘Step-Ex’ (New Development Technologies [NDT], Stockholm, Sweden)</td>
<td>Halvarsson, 2012 [63]</td>
<td>ICC = 0.71–0.87 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>Tinetti’s POMA balance subscale</td>
<td>Harada, 1995 [30]</td>
<td>ICC ≥ 0.93 (+)</td>
<td>N/R</td>
<td>N/R</td>
<td>Indeterminate (?)</td>
<td>(±), for convergent validity (−), for convergent validity</td>
</tr>
<tr>
<td></td>
<td>Lin, 2004 [41]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
Table 1. Continued

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reference</th>
<th>Reliability</th>
<th>Measurement error</th>
<th>Internal consistency</th>
<th>Criterion validity</th>
<th>Hypothesis testing for construct validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Multi-Directional Reach Test (MDRT)</td>
<td>Hsiao, 2018 [66]</td>
<td>ICC ≥ 0.775 (+)</td>
<td>N/R</td>
<td>N/R</td>
<td>r &lt; 0.70 (-)</td>
<td>N/R</td>
</tr>
<tr>
<td>The Kinect system (Kinect for Xbox 360™, Microsoft Corp, Seattle, WA, USA)</td>
<td>Iyigun, 2018 [67]</td>
<td>ICC = 0.96 (+)</td>
<td>N/R</td>
<td>N/R</td>
<td>r ≥ 0.70 (+)</td>
<td>N/R</td>
</tr>
<tr>
<td>The Turkish version of Fullerton Advanced Balance (FAB-T) Scale</td>
<td>Klein, 2011 [68] Rose, 2006 [69]</td>
<td>ICC not reported (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>Hypotheses not confirmed (–)</td>
</tr>
<tr>
<td>The parallel walk test</td>
<td>Lark, 2009 [70]</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>Hypotheses confirmed (+)</td>
</tr>
<tr>
<td>The Balance Computerised Adaptive Testing (Balance CAT)</td>
<td>Lu, 2015 [72]</td>
<td>N/R</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>r = 0.90 (+)</td>
<td>Hypothesis confirmed (+)</td>
</tr>
<tr>
<td>The MyBalance test</td>
<td>Mansson, 2021 [73]</td>
<td>N/R</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>r not reported (?)</td>
</tr>
<tr>
<td>The Functional Gait Assessment-Brazil (FGA- Brazil)</td>
<td>Marques, 2021 [74] Kirkwood, 2021 [75] Mathias, 1986 [76]</td>
<td>ICC &gt; 0.90 (+)</td>
<td>MIC not defined (?)</td>
<td>Criteria not met (?)</td>
<td>rS = 0.80 (+)</td>
<td>Hypothesis confirmed (+)</td>
</tr>
<tr>
<td>The ‘Get-up and Go’ Test</td>
<td>Matijacic, 2010 [77]</td>
<td>ICC not reported (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>r &lt; 0.70 (-)</td>
<td>Hypothesis confirmed (+)</td>
</tr>
<tr>
<td>The apparatus for assessment of postural responses</td>
<td>Mcmanus, 2022 [78]</td>
<td>ICC ≥ 0.75 (+)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>Inconsistent (±)</td>
</tr>
<tr>
<td>A comprehensive set of inertial sensor measures of postural sway (The Balance Score (BS) &amp; The Weighted Balance Score (WBS))</td>
<td>Ng, 2021 [79]</td>
<td>ICC = 0.95 (+)</td>
<td>MIC not defined (?)</td>
<td>Criteria not met (?)</td>
<td>N/R</td>
<td>Inconsistent (±)</td>
</tr>
<tr>
<td>The Modified Version of the Community Balance and Mobility Scale (CBMS-Home)</td>
<td>Panella, 2008 [80]</td>
<td>N/R</td>
<td>N/R</td>
<td>Criteria not met (?)</td>
<td>r not reported (?)</td>
<td>Hypotheses confirmed (+)</td>
</tr>
<tr>
<td>The Pavia Instrumented Tinetti Test (PITT)</td>
<td>Pelicioni, 2022 [33]</td>
<td>ICC ≥ 0.85 (+)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>r &gt; 0.50 (+)</td>
</tr>
<tr>
<td>The Dynamic Gait Index (DGI)</td>
<td>Jønsson, 2011 [98]</td>
<td>ICC = 0.82–0.89 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reference</th>
<th>Reliability</th>
<th>Measurement error</th>
<th>Internal consistency</th>
<th>Criterion validity</th>
<th>Hypothesis testing for construct validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Functional Gait Assessment (FGA)</td>
<td>Pelicioni, 2022 [33]</td>
<td>ICC ≥ 0.80</td>
<td>N/R</td>
<td>N/R</td>
<td>r = 0.84</td>
<td>Hypotheses confirmed (+)</td>
</tr>
<tr>
<td></td>
<td>Wrisley, 2010 [99]</td>
<td>(+)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Beninato, 2016 [100]</td>
<td>(+)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The NIH Toolbox® Standing Balance Test</td>
<td>Peller, 2022 [81]</td>
<td>ICC = 0.84</td>
<td>MIC not defined</td>
<td>N/R</td>
<td>r &lt; 0.70</td>
<td>N/R</td>
</tr>
<tr>
<td>The Biodex SD (Biodex Medical Systems, Shirley NY)</td>
<td>Riemann, 2017 [82]</td>
<td>ICC = 0.74–0.86</td>
<td>MIC not defined</td>
<td>N/R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Balance Scale (by Roberts)</td>
<td>Roberts, 1987 [83]</td>
<td>N/R</td>
<td>N/R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Turkish Version of the Berg Balance Scale (BBS)</td>
<td>Sahin, 2008 [84]</td>
<td>ICC = 0.97–0.98</td>
<td>N/R</td>
<td>Criteria not met</td>
<td>r &gt; 0.50</td>
<td>(+)</td>
</tr>
<tr>
<td>The Persian version of the Berg Balance Scale (BBS)</td>
<td>Salavati, 2012 [85]</td>
<td>ICC = 0.93–0.95</td>
<td>N/R</td>
<td>Criteria not met</td>
<td>r &gt; 0.50</td>
<td>(+)</td>
</tr>
<tr>
<td>The Nintendo Wii Fit exercgame</td>
<td>Sato, 2021 [86]</td>
<td>N/R</td>
<td>N/R</td>
<td></td>
<td>r &lt; 0.70</td>
<td>N/R</td>
</tr>
<tr>
<td>The Wii Stillness (WST) Test</td>
<td>Simms, 2020 [88]</td>
<td>N/R</td>
<td>N/R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The short form of the Fullerton Advanced Balance (SF-FAB) scale</td>
<td>Sinai, 2021 [89]</td>
<td>ICC = 0.92–0.99</td>
<td>MIC not defined</td>
<td>Criteria not met</td>
<td></td>
<td>Hypothesis confirmed (+)</td>
</tr>
<tr>
<td>The ‘balance meter’</td>
<td>Stokes, 1998 [90]</td>
<td>ICC not reported</td>
<td>MIC not defined</td>
<td>N/R</td>
<td></td>
<td>Inconsistent (±)</td>
</tr>
<tr>
<td>The AMTI Accusway system for balance and postural sway measurement (Advanced Mechanical Technology Inc., Watertown, Massachusetts)</td>
<td>Swanenburg, 2008 [91]</td>
<td>ICC = 0.52–0.89</td>
<td>MIC not defined</td>
<td>N/R</td>
<td></td>
<td>N/R</td>
</tr>
<tr>
<td>A dual-task computer game-based platform (TGP)</td>
<td>Szturm, 2015 [92]</td>
<td>ICC = 0.55–0.7</td>
<td>MIC not defined</td>
<td>N/R</td>
<td></td>
<td>N/R</td>
</tr>
<tr>
<td>The Modified Bathroom Scale</td>
<td>Vermuelen, 2012 [93]</td>
<td>N/R</td>
<td>N/R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The instrumented modified Clinical Test of Sensory Interaction on Balance (i-mCTSIB) utilising the Neurocom Very Simple Rehab (VSR) Sport force plate (Natus Medical Incorporated, Pleasanton, California)</td>
<td>Watson, 2021 [94]</td>
<td>ICC = 0.898</td>
<td>MIC not defined</td>
<td>N/R</td>
<td>r not reported (?</td>
<td>Hypothesis confirmed (+)</td>
</tr>
<tr>
<td>Models for estimating decline in balance using accelerometer-based gait features</td>
<td>Simila, 2017 [95]</td>
<td>N/R</td>
<td>N/R</td>
<td></td>
<td>r not reported (?)</td>
<td></td>
</tr>
<tr>
<td>The FICSIT Balance Scales (FICSIT-3 and FICSIT-4)</td>
<td>Rossiter-Fornoff, 1995 [96]</td>
<td>ICC not reported</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>Hypothesis confirmed (+)</td>
</tr>
<tr>
<td>The Wii Balance Board™ (WBB)</td>
<td>Olvera-Chavez, 2013 [97]</td>
<td>ICC = 0.64–0.85</td>
<td>MIC not defined</td>
<td>N/R</td>
<td>r not reported (?)</td>
<td>Hypothesis confirmed (+)</td>
</tr>
<tr>
<td>The Balance Tracking System (BTrackS)</td>
<td>Levy, 2018 [101]</td>
<td>ICC = 0.83</td>
<td>MIC not defined</td>
<td>Criteria not met</td>
<td>r ≥ 0.82</td>
<td>N/R</td>
</tr>
</tbody>
</table>

(continued)
Table 1. Continued

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reference</th>
<th>Reliability</th>
<th>Measurement error</th>
<th>Internal consistency</th>
<th>Criterion validity</th>
<th>Hypothesis testing for construct validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The NeuroCom Smart Equitest Research System (Natus Medical Inc, Pleasanton, California)</td>
<td>Haaro, 2019 [102]</td>
<td>ICC ≥ 0.71 (+)</td>
<td>MIC not defined (-)</td>
<td>N/R</td>
<td>N/R</td>
<td>Hypotheses not confirmed (-)</td>
</tr>
</tbody>
</table>

$\text{Only structural validity assessed. Measurement property rating: sufficient (+), insufficient (-), inconsistent (±), indeterminate (?) ABC, The Activities-specific Balance Confidence Scale; DSE, direction-sensitive evaluation (new strategy proposed by the authors in this study); 8LBS, The 8-level balance scale; APSI, anteroposterior stability index; MLSI, mediolateral stability index; PABAK, prevalence-adjusted bias-adjusted kappa; PASE, Physical Activity Scale for the Elderly; FAB, The Fullerton Advanced Balance scale; 3MTW, The three meter tandem walk; SD, standard deviation; CTSIB, Clinical Test of Sensory Interaction on Balance; FRT, The Functional Reach Test; SLS, single leg stance; N/R, not reported.}

- **The Mini-Balance Evaluation Systems Test (Mini-BESTest)**, with evidence for sufficient reliability (ICC ≥ 0.71), criterion validity ($r_S \geq 0.83$) and convergent validity, based on results of five studies. Inconsistent results were reported for discriminative validity.
- **The Timed Up and Go (TUG) test**, validated as a balance assessment tool in five studies, with summary results showing sufficient measurement properties, for reliability (ICC ≥ 0.83) and convergent and discriminative validity.
- **The Functional reach (FR) test**, validated by four studies and showing evidence for sufficient reliability (ICC ≥ 0.73) and discriminative validity, but not for convergent validity.
- **The Four Square Step Test (FSSST)**, assessed in three studies, with sufficient reliability (ICC ≥ 0.98) and discriminative validity, but not for convergent validity.
- **The Brief-Balance Evaluation Systems Test (Brief-BESTest)**, with sufficient reliability (ICC ≥ 0.82), criterion validity ($r_S \geq 0.83$) and convergent validity, based on summary data from three studies.
- **The Functional Gait Assessment (FGA)**, with three studies, including one study that assessed only structural validity. The two other studies provided, together, evidence for sufficient reliability (ICC ≥ 0.80), criterion validity ($r = 0.84$) and construct validity.

The seven other tools with two validation studies are the following: The Functional Gait Assessment-Brazil (FGA-Brazil), The Fullerton Advanced Balance (FAB) Scale, The TURN 180 test, The One leg standing (OLS) test, The Tinetti’s Performance-Oriented Mobility Assessment (POMA) balance subscale, The Multi-Directional Reach Test (MDRT) and The Wii Balance Board™ (WBB), with various measurement properties and ratings (Table 1 and Appendix 4.a).

**Musle power**

The 12 included studies on muscle power assessment tools evaluated 12 distinct tools (1 tool per study) (Table 3 and Appendix 4.c). Evidence for sufficient criterion validity was reported for only one tool, the 30-s sit-to-stand (STS) muscle power test [103], while another tool, the sit-to-stand (STS) performance power using a linear encoder [110] showed insufficient criterion validity. The following tools showed sufficient convergent or discriminant validity, along with sufficient reliability: The sit-to-stand power test (STSp) using a portable linear transducer [105], the chair stand mean power (CSMP) test using the Fitro Dyno device [109] and a standardised heel-rise test (using trunk accelerometry) [112]. Responsiveness was not reported by studies on muscle power assessment tools.

**Muscle strength**

Our literature search identified 30 different tools for muscle strength assessment in older people. Reliability, measurement error and hypothesis testing for construct validity (i.e. convergent and discriminative validity) were the most reported measurement properties for these tools. None of the studies on tools for muscle strength reported data on responsiveness.

Only four of the tools for muscle strength assessment (Table 2 and Appendix 4.b) were validated by at least two studies, with evidence for sufficient criterion validity available for only one tool: The Handheld Dynamometry (HHHD), Lafayette Manual Muscle Tester, Model #01163 [120, 123]. All the four tools showed sufficient test–retest or inter-rater reliability, while one of them (The Nintendo Wii Balance Board) [124, 125] also showed sufficient convergent validity. Among the other tools validated by only one study, evidence for sufficient criterion validity was reported for two tools: The calf-raise senior (CRS) test [119] and The Leg Press Sled (LPS) [132], which also showed sufficient reliability. A few other tools had sufficient convergent or discriminant validity, with sufficient reliability. Ultimately, the following tools showed sufficient criterion validity for muscle strength assessment in older people: The lateral step (LS) test [138], The Tandem Gait (TG) test [138] and the Single-stance test (SS) test [138].
### Table 2. Measurement properties of muscle strength assessment tools: Summary results with overall quality ratings

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reference</th>
<th>Reliability</th>
<th>Measurement error</th>
<th>Internal consistency</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The JAMAR hand-held hydraulic dynamometer</td>
<td>Abizanda, 2012 [115] Silva, 2019 [116]</td>
<td>ICC = 0.90–0.99 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>A uni-axial load cell device</td>
<td>Alqahtani, 2019 [118]</td>
<td>ICC = 0.90–0.99 (+)</td>
<td>SDC &gt; MIC (MCID) (−)</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The calf-raise senior (CRS) test</td>
<td>Andre, 2016 [119]</td>
<td>ICC = 0.79–0.93 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The Nintendo Wii Balance Board (WBB)</td>
<td>Blomkvist, 2016 [124] Jorgensen, 2015 [125]</td>
<td>ICC = 0.96–0.97 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The Modified Sphygmomanometer Test (MST)</td>
<td>Brito, 2022 [126]</td>
<td>ICC = 0.80–0.99 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>MicroFET2 hand-held dynamometer (Hoggan Industries, Inc., West Jordan, UT, USA)</td>
<td>Buckinx, 2017 [117]</td>
<td>ICC = 0.62 – 0.87 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The isometric knee extension (IKE) test (IKE test + strain gauge)</td>
<td>Buendia-Romero, 2021 [127]</td>
<td>ICC = 0.96–0.99 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The Q Force</td>
<td>Douma, 2016 [128]</td>
<td>ICC = 0.80–0.96 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>An analog dynamometer (SENSIX®, Poitiers, France) coupled with the DELSYS System (Trigno sensor, DELSYS, INC Boston, MA)</td>
<td>Gafner, 2017 [129]</td>
<td>ICC = 0.90–0.94 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The Biodes System 3 isokinetic dynamometer (Biodex Medical Systems, Shirley, NY)</td>
<td>Hartmann, 2009 [130] Symons, 2004 [131]</td>
<td>ICC ≥ 70 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The Isokinetic dynamometer (KinCom 500H, Chattecx Corp., Hixson, TN, USA)</td>
<td>Holsgaard Larsen, 2007 [108]</td>
<td>ICC not reported (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The Leg Press Sled (LPS)</td>
<td>Hutchinson, 2006 [132]</td>
<td>ICC ≥ 0.70 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>r ≥ 0.70 (+)</td>
</tr>
<tr>
<td>The Microfet 2000 strain gauge portable dynamometer (PD)</td>
<td>Karner, 1998 [133]</td>
<td>ICC ≥ 0.70 (+)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>A load cell setup</td>
<td>Keshavarzi, 2022 [134]</td>
<td>ICC = 0.99 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The push-off test (POT)</td>
<td>Legg, 2020 [135] Legg, 2020 [135]</td>
<td>ICC = 0.92 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The functional multi-joint isokinetic dynamometer</td>
<td>Legg, 2020 [135]</td>
<td>ICC = 0.98 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The MyBalance test</td>
<td>Mansson, 2021 [73]</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The maximal isometric strength test of the trunk (measured by a precalibrated digital loading cell connected to the MuscleLab software)</td>
<td>Mesquita, 2019 [136]</td>
<td>ICC ≥ 0.70 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The one-repetition maximum (1 RM) using elastic resistance bands test</td>
<td>Nyberg, 2014 [137]</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>Correlation &gt; 0.70 (+)</td>
</tr>
</tbody>
</table>

(continued)
G. Honvo et al.

Table 2. Continued

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reference</th>
<th>Reliability</th>
<th>Measurement error</th>
<th>Internal consistency</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The lateral step (LS) test</td>
<td>Porto, 2020 [138]</td>
<td>ICC = 0.95 (+)</td>
<td>N/R</td>
<td>N/R</td>
<td>r &lt; 0.70</td>
<td>Hypothesis confirmed (+)</td>
</tr>
<tr>
<td>Tandem Gait (TG)</td>
<td>Porto, 2020 [138]</td>
<td>ICC ≥ 0.70 (+)</td>
<td>N/R</td>
<td>N/R</td>
<td>r &lt; 0.70</td>
<td>Hypothesis confirmed (+)</td>
</tr>
<tr>
<td>Single-leg stance (SS) test</td>
<td>Porto, 2020 [138]</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>Hypothesis not confirmed (−)</td>
</tr>
<tr>
<td>The one repetition maximum (1 RM) using a muscle strength training device for the arm/shoulder (Pull Down, Norway)</td>
<td>Rydwik, 2007 [139]</td>
<td>ICC not reported (?)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>Hypothesis not confirmed (−)</td>
</tr>
<tr>
<td>The five-repetition sit-to-stand (STS) test</td>
<td>Schaubert, 2005 [140]</td>
<td>ICC = 0.82 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>A standardised heel-rise test (Using trunk accelerometer)</td>
<td>Schmid, 2011 [112]</td>
<td>ICC = 0.31 and 0.79 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>r &gt; 0.50 (+)</td>
</tr>
<tr>
<td>The one-repetition maximum (1 RM) performed on the Keiser A-300 pneumatic equipment (Keiser Corp., Fresno, CA) or on selectorised weight-stack resistance exercise machines (Cybex VR2; Cybex International Inc., Medway, MA)</td>
<td>Schroeder, 2007 [113]</td>
<td>ICC not reported (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>Grip strength, measured using a Smedley-type dynamometer (T.K.K.5401, TAKEI Scientific Instruments Co., Ltd., Niigata, Japan)</td>
<td>Suzuki, 2019 [141]</td>
<td>ICC = 0.96 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>Knee extension strength, measured using a handheld dynamometer (μ-Tas F-1; Anima Inc., Tokyo, Japan)</td>
<td>Suzuki, 2019 [141]</td>
<td>ICC = 0.90 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The 30-s Chair-Stand Test</td>
<td>Jones, 1999 [142]</td>
<td>ICC ≥ 0.70 (+)</td>
<td>N/R</td>
<td>N/R</td>
<td>Hypotheses confirmed (+)</td>
<td></td>
</tr>
</tbody>
</table>

Measurement property rating: sufficient (+), insufficient (−), inconsistent (±), indeterminate (?) MDD, #minimum detectable difference; RLOA, ratio of limits of agreement; SDD, smallest detectable difference; SPPB, Short Physical Performance Battery balance; GES, Gait Efficacy Scale; F8WT, Figure of 8 Walk Test; 6MWT, Six-Minute Walk Test; MCID, minimal clinically important difference.

Endurance

The 6-Minute Walk Test is the single tool for endurance assessment, evaluated by the only one study [143]. It has sufficient test–retest reliability and sufficient convergent and discriminative validity. Neither criterion validity, nor responsiveness were reported for this tool (Table 4 and Appendix 4.d).

Methodological quality of studies (RoB) and quality of evidence (GRADE)

Appendix 5 reports the outcomes of the risk of bias assessment for each measurement property assessed in the included studies (Appendices 5a–d). For most of the tools and studies, criterion validity, convergent validity and discriminative validity were found to have ‘very good’ methodological quality. In contrary, for reliability and measurement error, the far dominant ratings were ‘adequate’ and ‘doubtful’.

Using the outcomes of the risk of bias assessment along with other criteria (inconsistency, imprecision and indirectness) of the modified GRADE approach for grading the quality of evidence for outcome measurement instruments [15], we assessed the certainty of evidence for tools with at least two validation studies. This assessment included 15 balance tools and 4 muscle strength tools (Appendix 6). Our analyses showed high quality evidence for both sufficient validity (at minimum convergent validity) and reliability for the following tools:

Table 3. Measurement properties of muscle power assessment tools: Summary results with overall quality ratings

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reference</th>
<th>Reliability</th>
<th>Measurement error</th>
<th>Internal consistency</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The 30-s sit-to-stand (STS) muscle power test</td>
<td>Alcazar, 2020 [103]</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>r ≥ 0.70 (+) N/R</td>
</tr>
<tr>
<td>The sit-to-stand (STS) muscle power test</td>
<td>Alcazar, 2018 [104]</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R r &gt; 0.50 (+)</td>
</tr>
<tr>
<td>The sit-to-stand power test (STSp), using a portable linear transducer</td>
<td>Balachandran, 2021 [105]</td>
<td>ICC = 0.96 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>Hypotheses confirmed (+)</td>
</tr>
<tr>
<td>The Vertical jump (VJ) measured by a contact mat</td>
<td>Farias, 2013 [106]</td>
<td>ICC = 0.91–0.96 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The Tendo Weightlifting Analyser (Trencin, Slovak Republic)</td>
<td>Grey, 2014 [107]</td>
<td>ICC not reported (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>r &gt; 0.50 (+)</td>
</tr>
<tr>
<td>Counter-movement jump (CMJ) test performed on a force platform (Kistler Instruments 9,281 B, Wintertur, Switzerland, 40 x 60 cm)</td>
<td>Holsgaard Larsen, 2007 [108]</td>
<td>ICC not reported (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R N/R</td>
</tr>
<tr>
<td>The chair stand mean power (CSMP) test, using the Fitro Dyne device (Fitronic S. R. O. Co, Slovakia).</td>
<td>Kato, 2015 [109]</td>
<td>ICC = 0.88–0.92 (+)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R Hypotheses confirmed (+)</td>
</tr>
<tr>
<td>The sit-to-stand (STS) performance power using a linear encoder (MuscleLab Power model MLPRO, Ergotest Technology, Langesund, Norway)</td>
<td>Lindemann, 2015 [110]</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
<td>r &lt; 0.70 (−) N/R</td>
</tr>
<tr>
<td>The Jumping Mechanography</td>
<td>Rittweger, 2004 [111]</td>
<td>ICC not reported (?)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R r &gt; 0.50 (+)</td>
</tr>
<tr>
<td>A standardised heel-rise test (Using trunk accelerometry)</td>
<td>Schmid, 2011 [112]</td>
<td>ICC = 0.78–0.80 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R r &gt; 0.50 (+)</td>
</tr>
<tr>
<td>Unilateral leg extension power (W) using the Bassey Power Rig (University of Nottingham, Nottingham, U.K.)</td>
<td>Schroeder, 2007 [113]</td>
<td>ICC not reported (?)</td>
<td>N/R</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>The Ramp Power Test</td>
<td>Signorile, 2007 [114]</td>
<td>ICC = 0.921 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>r not reported (?)</td>
</tr>
</tbody>
</table>

N/R = Not reported Measurement property rating: sufficient (+), insufficient (−), inconsistent (±), indeterminate (?) SDC, smallest detectable change; LoA, limits of agreement

Table 4. Measurement properties of endurance assessment tools: Summary results with overall quality ratings

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reference</th>
<th>Reliability</th>
<th>Measurement error</th>
<th>Internal consistency</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The 6-Minute Walk Test</td>
<td>Rikli, 1998 [143]</td>
<td>ICC ≥ 0.70 (+)</td>
<td>MIC not defined (?)</td>
<td>N/R</td>
<td>N/R Hypotheses confirmed (+)</td>
</tr>
</tbody>
</table>

ANOVA, analysis of variance. Measurement property rating: sufficient (+), insufficient (−), inconsistent (±), indeterminate (?)

- The Timed Up and Go (TUG) test: Convergent and discriminative validity and reliability.

The Brief-Balance Evaluation Systems Test (Brief-BESTest) had high quality evidence for sufficient criterion validity.
and convergent validity, but the quality of evidence was downgraded to moderate for reliability, due to small sample size (<100 subjects). None of the muscle strength assessment tools had high quality evidence for both validity and reliability.

As patients were recruited from various settings, we sought to stratify the analyses by settings (e.g. community vs long-term care facility) for tools with high quality evidence for sufficient validity and reliability to check whether the findings reported above equally apply to either setting. However, there were not sufficient data to allow such analyses for all the tools. In fact, of the six studies on the BESTest, five included patients from the community. Likewise, for the Mini-BESTest and the TUG which were each assessed in five studies, four of the studies on each of these tools included subjects from the community. Analysis by setting was therefore possible only for the BBS, which was assessed in five studies that recruited patients from the community and in six studies that included patients living in long-term care facilities. This analysis showed that the results for each setting were similar to those reported when all the included studies for the BBS were considered together (data not shown).

Discussion

This systematic review aimed to identify all the studies validating the available measurement instruments for locomotor capacity or specific attributes of locomotor capacity, as defined by the WHO Locomotor Capacity Working Group [4], and to assess the methodological quality of the studies and the measurement properties of the tools. From the 125 studies retrieved by our comprehensive literature search, we identified 69 balance assessment tools, 30 tools for muscle strength, 12 tools for muscle power and 1 endurance assessment tool, with varying numbers of validation studies for each tool. Balance assessment tools had the highest numbers of validation studies per tool, although the overwhelming majority of existing tools were validated by only one study (only 15 tools had at least two validation studies). For muscle power and endurance, only one validation study was retrieved for each tool. Of important note, our literature search did not retrieve any assessment tool or validation studies for muscle function or joint function. Although no tools were identified for assessment of locomotor capacity overall, this finding was not so surprising or disappointing, as locomotor capacity is a relatively new concept with process for consensus on conceptual and operational definitions started only recently [4].

The GRADE assessment of confidence in evidence on measurement properties for balance tools (considering tools having at least two validation studies) revealed that only very few of these tools have high quality evidence for both sufficient validity and reliability. In fact, high-quality evidence for both sufficient criterion and construct validity and for reliability was found for only two tools: The Balance Evaluation Systems Test (BESTest) and the Mini-Balance Evaluation Systems Test (Mini-BESTest) [20]. However, this evidence applies more to community-dwelling patients, as only one study on the BESTest (on six studies included) and one on the Mini-BESTest (on five studies included) recruited patients from long-term care facilities. Therefore, whether this evidence may apply as well to patients living in long-term care facilities remains to be confirmed. The Berg Balance Scale (BBS) [27] showed high-quality evidence for sufficient construct validity (convergent and discriminative validity), and for reliability, as had the TUG test [41]. However, contrary to the TUG for which this evidence applies more to community-dwelling patients (only one study of five on TUG recruited patients from a nursing home), the evidence on the measurement properties of the BBS applies equally to both community-dwelling patients and long-term care facility residents. Consistent with our findings on balance assessment tools, an expert panel acknowledged the excessive number of standing balance assessment tools and reached consensus on two balance measures, recommending that at a minimum, either the BBS or the Mini-BESTest be used for measuring standing balance in adult populations [144]. Our analyses showed that none of the tools for muscle strength assessment has high quality evidence for both sufficient validity and reliability.

Regarding endurance, the only tool identified by our systematic review is the 6-Minute Walk Test [143]. Although it has good convergent validity, discriminative validity and reliability (when applying the criteria for good measurement properties), further validation studies are needed to strengthen the evidence on the usefulness of this tool in older people. Many other tools already exist for walking endurance assessment, which have been validated or used in other age groups and populations. These include the endurance shuttle walk test [145], which was validated for the assessment of endurance capacity in patients with chronic obstructive pulmonary disease (COPD), the long distance corridor walk [146] and the 400-m Walk Test [147]. These tools may also be validated for use in healthy older people in community and long-term care facilities.

This systematic review did not identify any tool formally validated as a measure of joint function in older people. However, the goniometer, which seems to have been used in clinical practice as a ‘proxy’ for joint function assessment may be a useful tool in assessing locomotor capacity. In fact, in clinical research, this tool has rather been used to assess range of motion [148], even if there seems to be a confusion between range of motion and joint function in some publications [149, 150].

In order to come to clear conceptual and operational definitions of locomotor capacity in older people, there are some burning issues that the WHO Locomotor Capacity Working Group [4] may need to further discuss, including the usefulness of considering muscle function as an attribute of locomotor capacity, along with muscle strength and power.
Issues to be addressed by the WHO locomotor capacity working group

The findings of this systematic review reveal that the WHO Locomotor Capacity Working Group still have to clarify several aspects related to the current attributes of locomotor capacity. First, regarding balance, it is important to clarify whether static or dynamic balance are to be assessed, or both, even if most of the main tools included in this systematic review assess both aspects of balance [20]. It may also be important to clarify whether only standing balance is to be assessed in the context of locomotor capacity, or whether sitting balance [151] is also essential. Regarding this particular aspect, we assumed in this systematic review that standing balance was the type of balance to be considered in the context of assessment of locomotor capacity; therefore, studies assessing tools for sitting balance were excluded as ‘not locomotor capacity or attributes’. Second, regarding muscle strength, it may be important to clarify which specific muscle groups are to be primarily assessed, as various identified tools target various muscle groups [120, 126, 132, 134]. For example, whether handgrip strength measures should be considered in the context of locomotor capacity assessment is to be clarified, even if grip strength has been found to reflect general muscle strength [152, 153]. In fact, handgrip strength has also been identified as a measure of vitality, one of the six key domains of the WHO intrinsic capacity concept [154]; in addition, the fact that grip strength can represent global muscle strength should not eliminate the need to assess specific muscle groups, when indicated [155]. Third, when referring to endurance, it may be useful to precise that we are talking about ‘walking endurance’, and not about ‘muscle endurance’ [156]. In this systematic review, we assumed that only ‘walking endurance’ had to be considered and therefore, we did not include studies assessing the measurement properties of tools for ‘muscle endurance’. Furthermore, it is worth noting that ‘muscle endurance’ has yet been considered as one of the measures of vitality capacity in the WHO working definition of this other key domain of intrinsic capacity [157]. Forth, regarding joint function, which seems to have been assessed in practice through range of motion, one might wonder why range of motion itself would not be directly listed as an attribute of locomotor capacity, instead of joint function. Fifth, beyond the fact that our systematic literature search identified no measurement tools for muscle function, it may be important to further discuss the usefulness of considering muscle function as an attribute of locomotor capacity, knowing that muscle function has been defined as including measures of strength and power [158]. In the end (sixth), another important issue that the WHO Locomotor Capacity Working Group will need to address is to provide consensus definitions of terms used to define locomotor capacity (i.e. the attributes). These definitions may be provided in a consensus paper summarising terms commonly used to define intrinsic capacity (taking the form of a glossary), including terms used to define the other domains of intrinsic capacity.

Limitations of the study

We acknowledge some limitations of this systematic review. First, we limited our literature search to articles published in English, which may have excluded some validation studies published in other languages. However, research has reported that excluding non-English language publications from evidence-syntheses did not jeopardise the conclusions of systematic reviews [159]. Besides the issue of language restriction, our search strategies may have not captured a few validation studies from the databases searched, as in any systematic review, mainly for the attributes for which MeSH or EmTree terms are not yet available (i.e. muscle function, joint function and muscle power). However, the search strategies were detailed enough, and our literature search covered the most important and relevant databases (including Scopus that doesn’t use thesaurus terms), so that we can be quite confident that we didn’t miss any significant evidence that would alter the conclusions of this research.

Implications for future research

One important question raised by the findings of this systematic review is: Why has all this research been conducted on so many tools if, at the end, the studies bring limited evidence on the usefulness of these tools for the intended purpose? Considering this, several strong recommendations are to be formulated:

First, future validation studies should adhere to the COSMIN terminology of measurement properties [16] and to the COSMIN reporting guideline for primary studies on measurement tools [160]. Second, researchers should avoid fragmented research questions (i.e. validation studies addressing only single specific aspects of measurement properties) and consider instead thoroughly assessing all the relevant measurement properties and aspects for each single tool, with adequate sample size. Third, the findings of this systematic review underlining that future research agenda should focus on development and validation of tools to measure other attributes of locomotor capacity, for which high quality evidence for validity, reliability and responsiveness is lacking in older people. These include endurance, for which tools already exist with evidence for validity and reliability in other populations or age groups [147]. Regarding balance and specifically standing balance, we think there is no need to invest in the development of new assessment tools, given the excessive number of existing tools. Instead, researchers should focus on setting up well designed studies to provide high quality evidence on the measurement properties (i.e. complete evidence with regard to validity, reliability and responsiveness) of some of the most promising existing tools, with a particular attention to feasibility aspects (e.g. completion time, ease of administration, required equipment, etc.). Fourth, future research should also consider validation of these tools in low- and middle-income countries, particularly in African countries, and in other high-income countries where these tools are not yet validated. Fifth, research should
be initiated to provide Minimal Important Change (MIC) values for available tools in older people, as missing MIC values hampered the rating of measurement error in almost all the included studies. Sixth, as a final but not least strong recommendation for researchers, future validation studies of tools for locomotor capacity should include in a single study two subsets of sample, one including patients recruited from the community and another one formed with patients recruited from long-term care facilities. By so doing, each single study will provide at the same time, evidence on the appropriateness of the tools for patients residing in both settings.

In support to all these recommendations, we would like to remind to all researchers this important message from Doug Altman (of revered memory) in his Editorial titled ‘The scandal of poor medical research’: ‘We need less research, better research, and research done for the right reasons’ [161]. We hope that lessons learned from this systematic review and outlined here will serve future researchers in designing, conducting and reporting their research on validation of tools to assess locomotor capacity. As research needs in this setting are urgent, beyond hopes, we strongly call researchers for high quality research to provide WHO, countries and clinicians with effective tools to measure locomotor capacity, by fully complying with the COSMIN terminologies and recommendations [16, 160] and by following good research practice principles [162]. In fact, ultimately, this will contribute to the wellbeing of our older people, by helping meet the United Nations decade of healthy ageing goals [5].

Conclusion

Without strong evidence supporting the validity and reliability of measurement instruments, the choice of adequate tools to screen and monitor health status of older people may be a hazardous travel. To the best of our knowledge, this systematic review is the first that assessed the measurement properties of tools to measure all the attributes of locomotor capacity. The outcomes of this study will first support the WHO Locomotor Capacity Working Group in the process of developing both conceptual and operational definitions of locomotor capacity [4]. Ultimately, these findings will help WHO in providing evidence-based recommendations for adequate tools to be used in clinical and population settings to assess locomotor capacity; and thereby, will contribute to adequate monitoring of healthy ageing and actions taken by WHO and the United Nations in the context of the 2021–2030 Decade of Healthy Ageing initiative [5, 6]. In the absence of strong evidence for validity and reliability of tools for most of the attributes of locomotor capacity in older people, WHO may provide interim recommendations for specific tools, following paradigms for appropriately formulated discordant recommendations [163].

Supplementary Data: Supplementary data mentioned in the text are available to subscribers in Age and Ageing online.

Acknowledgements: The members of the WHO Locomotor Capacity Working Group

1. Al-Daghri Nasser, College of Science and Chair for Biomarkers of Chronic Diseases, King Saud University, Riyadh, Saudi Arabia.
2. Andrieu Sandrine, Clinical epidemiology and public health department, Gerontopôle, Toulouse University Hospital, France
3. Annweiler Cédric, Department of Geriatric Medicine, University of Angers, France
4. Aubertin-Leheudre Mylène, Exercise sciences department, Université du Québec à Montréal (UQAM), Canada
5. Bautmans Ivan, Gerontology and Frailty in Ageing research departments, Vrije Universiteit Brussel (VUB), Belgium
6. Beaudart Charlotte, Department of Health Services Research, Maastricht University, Maastricht, The Netherlands; WHO Collaborating Center for Epidemiology of Musculoskeletal Health and Ageing, University of Liège, Belgium
7. Becker Clemens, Digital Geriatric Medicine, University of Heidelberg, Germany
8. Bruyère Olivier, WHO Collaborating Center for Epidemiology of Musculoskeletal Health and Ageing, University of Liège, Belgium
9. Buckinx Fanny, WHO Collaborating Center for Epidemiology of Musculoskeletal Health and Ageing, University of Liège, Belgium
10. Campusano Claudia, Universidad de los Andes, Chile
11. Cesari Matteo, Geriatric Fellowship Program, University of Milan, Italy
12. Chandran Manju, Osteoporosis and Bone Metabolism Unit, Singapore General Hospital, Singapore
13. Cherubini Antonio, Geriatrics, Accettazione geriatrica e Centro di ricerca per l'invecchiamento, IRCCS INRCA, Ancona, Italy
14. Clark Patricia, Clinical Epidemiology Research Unit, National University of México UNAM, Mexico
15. Cooper Cyrus, MRC Lifecourse Epidemiology Unit, University of Southampton, UK
16. Cruz-Jentoft Alfonso, Geriatric Department, Hospital Universitario Ramón y Cajal, Madrid, Spain
17. Dennison Elaine, MRC Lifecourse Epidemiology Unit, University of Southampton, UK
18. Fouasson Chailloux Alban, ‘Regenerative Medicine and Skeleton’ research centre, University Hospital of Nantes, France
19. Fuggle Nick, MRC Lifecourse Epidemiology Center, University of Southampton, UK
20. Gichu Muthoni, Ministry of Health Kenya, Division of Geriatric Medicine, Kenya
21. Gielen Evelen, Unit of Gerontology and Geriatrics, Department of Public Health and Primary Care, KU Leuven, Belgium
Measures of attributes of locomotor capacity in older people

22. Guicheux Jérôme, ‘Regenerative Medicine and Skeleton’ research centre, University of Nantes, France
23. Harvey Nick, MRC Lifecourse Epidemiology Centre, University of Southampton, UK
24. Haugen Ida, Division of Rheumatology and Research, Diakonhjemmet Hospital, Norway
25. Honvo Germain, WHO Collaborating Center for Epidemiology of Musculoskeletal Health and Ageing, University of Liège, Belgium
26. Lamy Olivier, Bone Unit, Lausanne University Hospital, Switzerland
27. Landi Francesco, Geriatric Internal Medicine Unit, A. Gemelli University Hospital, Rome, Italy
28. Lane Nancy, Davis School of Medicine in Sacramento, University of California, USA
29. Lazaretti Castro Marise, Bone and Mineral Research Unit, Federal University of Sao Paulo (UNIFESP), Brazil
30. Lewiecki Mike, Bone Health TeleECHO, University of New Mexico Health Sciences Center in Albuquerque, New Mexico, USA
31. Matijevic Radmila, Rehabilitation Unit, Orthopaedic and Trauma, University of Novi Sad, Serbia
32. Messina Osvaldo Daniel, Rheumatology, C Argerich Hospital, University of Buenos Aires, Argentina
33. Mkinski Ouafa, Department of Rheumatology, Ibn Rochd University Hospital, Casablanca, Morocco
34. Mobasher Ali, Research Unit of Medical Imaging, Physics and Technology, University of Oulu, Finland
35. Njeze Ngozi, University of Nigeria Medical school, Nsukka, Nigeria
36. Pinto Daniel, Department of Physical Therapy, Marquette University, USA
37. Reginster Jean-Yves, WHO Collaborating Center for Epidemiology of Musculoskeletal Health and Ageing, University of Liège, Belgium
38. Rizzoli René, Faculty of Medicine, Geneva University Hospitals, Geneva, Switzerland
39. Rolland Yves, Gérontopôle of Toulouse, University Toulouse III Paul Sabatier, France
40. Saleh Yousef, College of Medicine, King Saud Bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia
41. Singer Andrea, Division of Women’s Primary Care, MedStar Georgetown University Hospital, USA
42. Thomas Thierry, Rheumatology Department, University Hospital of Saint-Étienne (UHSE), France
43. Van der Velde Nathalie, Amsterdam UMC, The Netherlands
44. Vellas Bruno, Gérontopôle & Department of Geriatric Internal Medicine, Toulouse University Hospital, France
45. Veronese Nicola, Geriatric Unit, Department of Medicine, University of Palermo, Italy
46. Visser Marjolein, Vrije Universiteit Amsterdam, The Netherlands
47. Zee A Han, College of Medicine, The Catholic University of Korea, Seoul, South-Korea

Declaration of Conflicts of Interest: None.

Disclaimer: The authors alone are responsible for the views expressed in this article and they do not necessarily represent the views, decisions or policies of the institutions with which they are affiliated.

Declaration of Sources of Funding: This research was funded by the European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO), through the 2022 ESCEO-Islene Araujo de Carvalho Grant, a Prize granted to Germain Honvo.

The funder had no role in the development of the study protocol, in data collection and data synthesis, or in the manuscript preparation and decision for submission.

Data Availability Statement: All the data that support the findings and conclusions of this study are available as Appendices to this manuscript.

References

38. Brooks D, Davis AM, Naglie G. Validity of 3 physical performance measures in inpatient geriatric rehabilitation.


62. Gordt K, Mikolaizak AS, Taraldsen K et al. Creating and validating a shortened version of the community balance and mobility scale for application in people who are 61 to 70 years of age. Phys Ther 2020; 100: 180–91.


73. Mansson L, Backman P, Oheberg F, Sandlund J, Selling J, Sandlund M. Evaluation of concurrent validity between a smartphone self-test prototype and clinical instruments for
Measures of attributes of locomotor capacity in older people


Silva AG, Cerqueira M, Raquel Santos A, Ferreira C, Alvelhão J, Queirós A. Inter-rater reliability, standard error of measurement and minimal detectable change of the 12-item WHODAS 2.0 and four performance tests in institutionalized ambulatory older adults. Disability & Rehabilitation 2019; 41: 366–73.


Hutchison AT, Clarke MSF. An isometric strength testing device for use with the elderly: validation compared with isokinetic measures. Physical & Occupational Therapy in Geriatrics 2006; 25: 1–12.


Keshavarzi F, Azadinia F, Talebian S, Rasouli O. Test-retest reliability of a load cell setup, Ito, and timed loaded standing tests for measuring muscle strength and endurance in older adults with and without hyperkyphosis.


