# SYSTEMATIC REVIEW

# Measures of perceived mobility ability in community-dwelling older adults: a systematic review of psychometric properties

Marla Beauchamp[1,2,†], Qiukui Hao[1,2,†], Ayse Kuspinar[1,2], Gésine Alder[2], Keitaro Makino[2,3], Mina Nouredanesh[1,2], Yunli Zhao[4], Christopher Mikton[5], Jotheeswaran Amuthavalli Thiyagarajan[6], Theresa Diaz[7], Parminder Raina[2,4,8]

[1]School of Rehabilitation Science, McMaster University, Hamilton, ON, Canada
[2]McMaster Institute for Research on Aging, McMaster University, Hamilton, ON, Canada
[3]Department of Preventive Gerontology, National Center for Geriatrics and Gerontology, Obu, Japan
[4]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada
[5]Demographic Change and Healthy Aging Unit, Social Determinants of Health, World Health Organization, Geneva, Switzerland
[6]Ageing and Health Unit, Department of Maternal, Newborn, Child and Adolescent Health and Ageing, World Health Organization, Geneva, Switzerland
[7]Department of Maternal, Newborn, Child and Adolescent Health and Ageing, World Health Organization, Geneva, Switzerland
[8]Labarge Centre for Mobility in Aging, McMaster University, Hamilton, ON, Canada

Address correspondence to: Marla Beauchamp, School of Rehabilitation Science, McMaster University, 1400 Main St. W., IAHS, L8S 1C7 Hamilton, ON, Canada. Email: beaucm1@mcmaster.ca
[†]Co-first authors

## Abstract

**Objectives**: The objective of this systematic review was to synthesise the psychometric properties of measures of perceived mobility ability and related frameworks used to define and operationalise mobility in community-dwelling older adults.

**Methods**: We registered the review protocol with PROSPERO (CRD42022306689) and included studies that examined the psychometric properties of perceived mobility measures in community-dwelling older adults. Five databases were searched to identify potentially relevant primary studies. We qualitatively summarised psychometric property estimates and related operational frameworks. We conducted risk of bias and overall quality assessments, and meta-analyses when at least three studies were included for a particular outcome. The synthesised results were compared against the Consensus-based Standards for the Selection of Health Measurement Instruments criteria for good measurement properties.

**Results**: A total of 36 studies and 17 measures were included in the review. The Late-Life Function and Disability Index: function component (LLFDI-FC), lower extremity functional scale (LEFS), Mobility Assessment Tool (MAT)-short form (MAT-SF) or MAT-Walking, and Perceived Driving Abilities (PDA) Scale were identified with three or more eligible studies. Most measures showed sufficient test–retest reliability (moderate or high), while the PDA scale showed insufficient reliability (low). Most measures had sufficient or inconsistent convergent validity (low or moderate) or known-groups validity (low or very low), but their predictive validity and responsiveness were insufficient or inconsistent (low or very low). Few studies used a conceptual model.

**Conclusion**: The LLFDI-FC, LEFS, PDA and MAT-SF/Walking can be used in community-dwelling older adults by considering the summarised psychometric properties. No available comprehensive mobility measure was identified that covered all mobility domains.

**Keywords:** mobility, self-reported, perceived mobility, measurement properties, older people

**Key Points**

- LLFDI-FC, LEFS, PDA and MAT-SF/Walking can be used in community-dwelling older adults by considering the summarised psychometric properties.
- There is no existing comprehensive mobility measure that covers all the mobility domains described by the World Health Organization.
- Further studies are needed on the responsiveness and predictive validity of self-report mobility measures.

## Introduction

The World Health Organization (WHO) has described healthy ageing as 'the process of developing and maintaining the functional ability that enables wellbeing in older age' [1]. Functional ability comprises the health-related attributes that enable people to be and to do what they value. To assess the impact of healthy ageing actions at the national, regional and global levels, WHO proposed functional ability as an indicator of healthy ageing.

In the recently released baseline report for the Decade of Healthy Ageing 2021–2030, WHO identified the ability to be mobile as being a critical indicator of the functional ability of older people across high-, middle- and low-income countries. According to the WHO report, mobility encompasses 'movement in all its forms, whether powered by the body or a vehicle' [2]. Mobility is critical for healthy ageing with respect to both physical and psychological well-being as it is a prerequisite for participation in most physical and social activities [3] and is a consistently strong predictor of changes in quality of life over time [4–6]. It is also a key element of frailty and sarcopenia assessments and provides helpful information for tailoring interventions [7].

Most previous reviews of psychometric properties of mobility measures have focused on performance-based measures (i.e. measures of locomotor capacity for mobility) or populations with specific conditions [8–11]. One (2016) review [12] summarised the measurement properties of 34 patient-reported mobility questionnaires in various adult populations and found that only 10 of them exclusively measured mobility activities. This review also noted that most of the existing patient-reported mobility measures (18/34) at that time were not based on the WHO International Classification of Functioning, Disability and Health (ICF) conceptual framework. This highlights that although many mobility frameworks have been proposed, there remains no gold standard for defining and operationalising mobility.

In light of the lack of conceptual clarity related to mobility measurement, in this WHO *Age and Ageing* special issue on measurements of healthy ageing, we proposed a unified framework for mobility measurement that builds on existing frameworks and evidence [13], but that is also in line with current terminology in the World Report on Ageing and Health [14]. Our framework outlined three distinct facets of mobility to consider for comprehensive mobility measurement in older people: (i) perceived mobility (i.e. self-reported ability or difficulty with task performance); (ii) actual mobility (i.e. frequency, duration or counts of mobility in daily life); (iii) locomotor capacity for mobility (i.e. performance on physical tests of mobility). This review focuses on the first category of mobility measurement: perceived mobility measures.

To our knowledge, there has been no previous systematic review of the psychometric properties of perceived mobility measures in community-dwelling older adults. The objective of this review was to synthesise the psychometric properties of available measures of perceived mobility in community-dwelling older adults and to summarise related frameworks used to define and operationalise mobility in the existing literature.

## Methods

### Mobility measures

This review focuses on self-report measures of perceived mobility for older community-dwelling adults. We registered the review protocol with PROSPERO (CRD42022306689). The WHO World Report on Ageing and Health describes the ability to be mobile as 'movement in all its forms by the body (with or without an assistive device) and/or a vehicle' and proposes the following sub-domains: getting up from a chair or moving from a bed to a chair, walking for leisure, exercising, completing daily tasks, driving a car and/or using public transportation [2, 15]. We used this definition of ability to be mobile to identify eligible measures. If a measure included both mobility- and non-mobility-related items, we only included the measure if 80% or more of the items were related to mobility or if it had a sub-domain solely related to the mobility construct.

As we were interested in measures that address older people's perception of their ability to be mobile, i.e. what they say they can do [16], we included scales that asked 'Are you able to. . .', 'Do you have any difficulty. . .', 'Do you require assistance', 'How difficult is it to. . .' and so on. Measures that focused only on quantifying actual mobility levels (i.e. frequency or duration of an activity, calories expended, etc.) or solely on activities of daily living (e.g. continence, grooming) were excluded. Likewise, performance-based tests of locomotor capacity for mobility (e.g. gait speed) were also excluded and are covered in a separate review in this issue.

### Literature search and study identification

We used MEDLINE, EMBASE, CINAHL, PsycINFO and Ageline to conduct a comprehensive search to identify

potentially eligible studies that reported psychometric properties of mobility measures up to March 2022. We included studies of community-dwelling older adults with at least 50% or more of participants aged 60 or over unless a separate analysis on participants aged 60 years and over was performed. The full search strategies are in Supplementary Appendix A.

### Study selection

We included studies with any design that focused on psychometric properties of perceived mobility measures among community-dwelling older adults with or without reference to a specific framework. Eligible studies met all the following criteria: (i) focused on at least one self-reported perceived mobility measure; (ii) reported on one or more of the following psychometric properties: reliability (including test–retest reliability, inter-rater reliability, intra-rater reliability measured by correlation coefficients), construct validity (including face validity, structural validity, cross-cultural validity, hypotheses-testing), criterion validity (concurrent validity, predictive validity), internal consistency (Cronbach's alpha), measurement error (standard error of measurement (SEM)), responsiveness and interpretability measured by distribution or anchor-based methods; and (iii) included older adults (with a cut-off of 60 or over) living in the community. We excluded the following records: (i) only in abstract format, conference proceedings, editorials or commentaries; (ii) published in languages other than English; (iii) reviews.

Four reviewers independently manually screened 20% of titles and abstracts [17]. The screening orders were determined by Rayyan's compute rating feature for systematic reviews. The compute rating was calculated by Rayyan based on previous decisions during the manual screening and default machine learning algorithms. Previous work has shown that this method has high specificity (99–100%) for excluding irrelevant citations [18]. Then, using Covidence, five reviewers independently performed full-text screening of all included records. Disagreements between reviewers in the full-text screening stage were resolved by discussion or, if needed, by consultation with a third reviewer.

### Data abstraction

One reviewer extracted data using a pre-designed data extraction sheet. All the extracted data were checked by a second reviewer. The following information was extracted: first author name; publication date; country(ies) and Low-to-Middle-High-Income status according to World Bank Data; study design; characteristics of the population (e.g. mean or median age, sample size, sex, rural or urban); mobility measure details (e.g. name of the measure, definition/framework/construct, sub-domains); psychometric property details (reliability, validity, consistency, measurement error, responsiveness and interpretability) and related frameworks. We followed COSMIN's definition of responsiveness, which is 'the ability of a measure to detect change over time'. In the included studies, different

methods were used to define the minimal important difference (MID) or to test responsiveness. We used the reported results regardless of how the authors of individual studies have defined them and recognise that the method variability might result in different face values for MID and interpretations of responsiveness.

### Quality assessment of studies included

When at least three studies were included for an outcome measure, the quality of each eligible study was evaluated using the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) checklist [19]. The checklist includes 3 to 35 items for 10 psychometric properties in separate boxes and allows for rating the methodological quality as very good, adequate, doubtful and inadequate or not applicable. The methodological quality of the corresponding psychometric property was rated as the lowest rating [20]. Then, we evaluated the overall quality of evidence using the modified Grading of Recommendations, Assessment, Development and Evaluation (GRADE) [19, 21].

According to the COSMIN handbook for the systematic review of psychometric properties, we assessed the risk of bias, imprecision, inconsistency and indirectness in rating the overall quality of evidence [22]. We rated the overall quality of evidence as high to begin and adjusted the ratings based on the following existing issues. If there were any studies of doubtful or inadequate quality available, or there was only one study of adequate quality, we downgraded the evidence by one or two levels for risk of bias. For imprecision, we considered the sample size of related studies, and for inconsistency, we evaluated the proportions of studies that met pre-specified hypotheses for good measurement properties. If studies included only partial participants older than 60 years or from community settings, we downgraded the evidence for indirectness. Finally, the evidence was rated as high, moderate, low or very low. Additional details on the GRADE criteria can be found in Supplementary Appendix F.

### Synthesis of results

The major review findings were qualitatively summarised in tables, including psychometric property estimates, related operational frameworks and methods. We assessed the synthesised results against the revised COSMIN criteria [22] for good measurement properties to determine a rating of sufficient (+): at least 75% of hypotheses were met, insufficient (−): at least 75% of hypotheses were not met, inconsistent (±): mixed results (sufficient and insufficient <75%) or indeterminate (?): all results across eligible studies were indeterminant.

For test–retest reliabilities, meta-analyses were performed to pool available intraclass correlation coefficients (ICCs) after Fisher's conversion based on the point estimate and sample size to stabilise the variances and approximate the ICC to be consistent with the normal distribution. We used DerSimonian and Laird random-effects models in the R
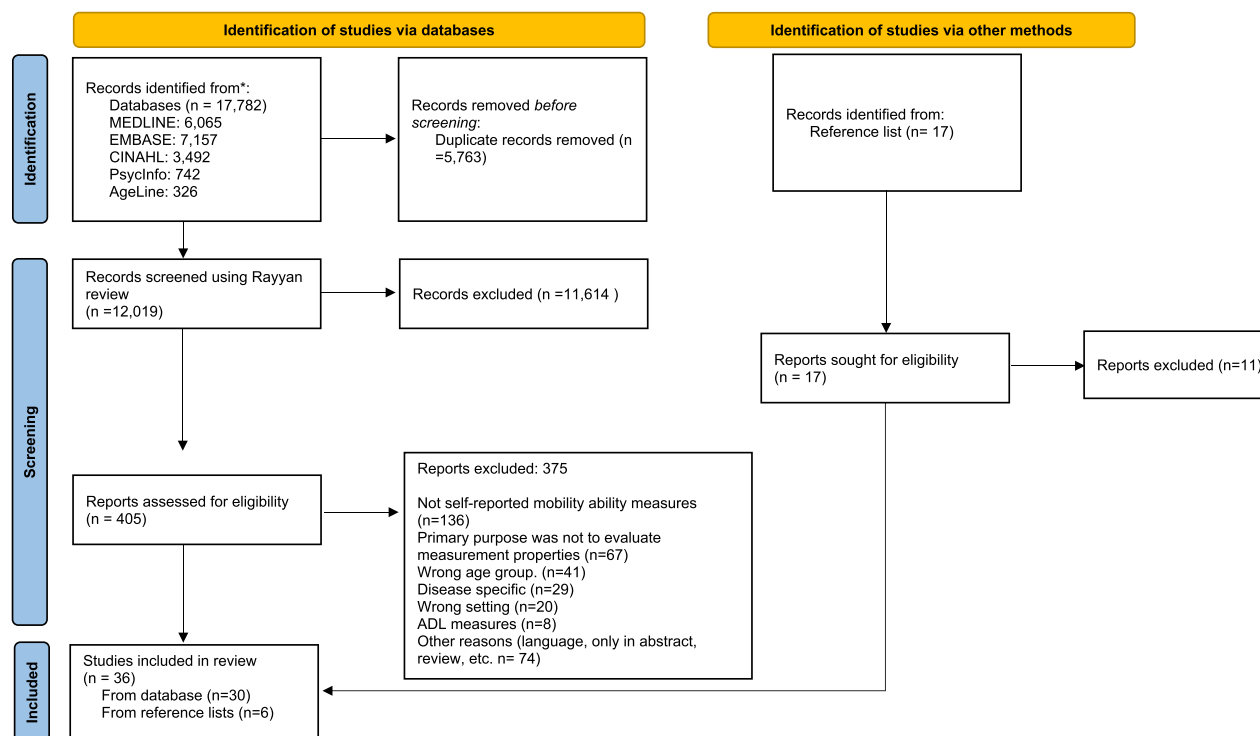
**Figure 1.** PRISMA flow diagram for selecting eligible studies. *Adapted from:* Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71.

'metacor' package to conduct the meta-analysis [23]. The evidence was limited to performing the planned subgroup analyses according to age (<75 years versus >75 years), sex (male or female) and Low-to-Middle-High-Income status.

## Results

### Study selection and characteristics

The literature search yielded 12,019 citations after removal of duplicate records. Following the screening of title and abstracts, 11,614 citations were excluded and 405 potentially relevant studies from the database search were retrieved for full-text screening. Of these, 375 were excluded for various reasons, leaving 30 eligible studies. Seventeen relevant publications were retrieved from the hand search and reviewing of reference lists, and six proved eligible. Therefore, a total of 36 studies were included in this review [24–59] covering a total of 17 measures (see Supplementary Appendix B for characteristics of included measures). Figure 1 summarises the search and screening process for eligible studies.

### Perceived mobility measures

Nine of these studies reported the psychometric properties for the Late-Life Function and Disability Index: function component (LLFDI-FC) overall scales and subscales [24, 25, 31, 35, 41, 45, 46, 53, 55]; four studies reported psychometric properties for the lower extremity functional scale (LEFS)

[47, 48, 56, 58]; five studies for the Mobility Assessment Tool (MAT)-short form (MAT-SF) [34, 39, 49, 51] or MAT-Walking [44]; and three studies for the Perceived Driving Abilities (PDA) Scale [27, 29, 42]; and 15 other studies reported on another 13 mobility scales or items [26, 28, 30, 32, 33, 36–38, 40, 43, 50, 52, 54, 57, 59] (Supplementary Appendix B). Most mobility measures did not mention related theoretical frameworks or models in their development. Frameworks or models mentioned included Nagi's disablement model (LLFDI-FC), The International Classification of Functioning, Disability and Health (ICF) framework (Patient-Reported Outcomes Measurement Information System, PROMIS and Pepper Assessment Tool for Disability, PAT-D), Activity Space Model of Disability (self-reported measures of walking, running and lifting abilities) [40], and development of Mobility Limitation framework (Preclinical and manifest mobility limitation) [43]. All these frameworks focus on limitations in physical functioning.

The appendix includes a summary of identified measures (B), hypotheses for measures with good psychometric properties (C), characteristics of the 36 eligible studies (D), the GRADE assessment details (E) and detailed results tables (F–H).

1 Late-Life Function and Disability Index: function component

Nine studies on the LLFDI-FC were identified in this review. Six of nine studies were conducted in the USA in
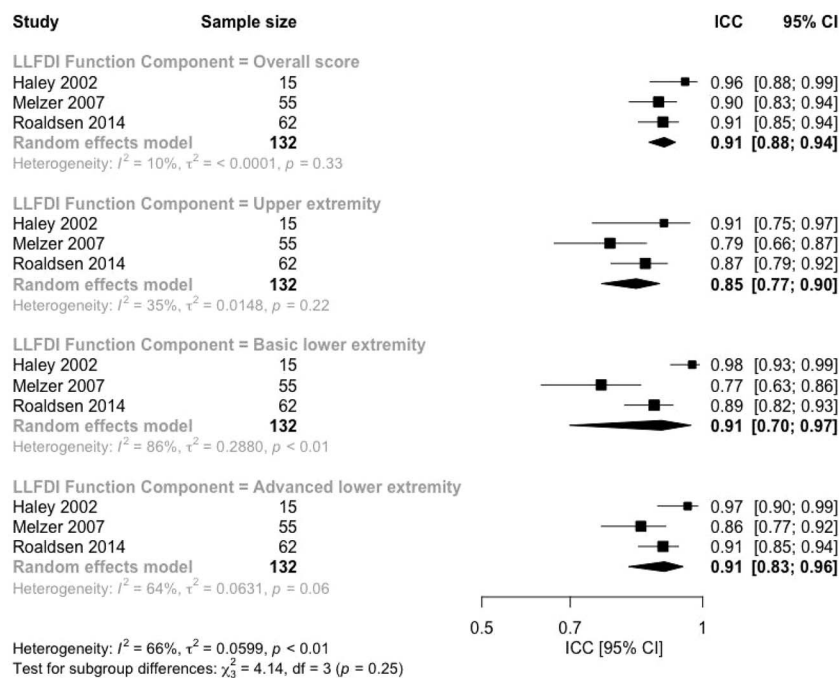
**Figure 2.** Test–retest reliability for Late-Life Function and Disability Index: function component.

English [24, 25, 35, 41, 46, 55]. Two studies were conducted in Israel (Hebrew version) [31, 45] and one study was conducted in Sweden using a Swedish version [53].

a. Internal consistency and reliability

High-quality evidence from two studies [35, 53] demonstrated that LLFDI-FC overall score and subscales showed sufficient internal consistency: Cronbach alphas of 0.96–0.97 for the overall score, 0.86–0.94 for the upper extremity subscale, 0.94–0.96 for the basic lower extremity subscale and 0.95–0.96 for the advanced lower extremity subscale.

High-quality evidence from three studies [35, 45, 53] demonstrated that LLFDI-FC overall score and subscales showed sufficient test–retest reliability: pooled ICC = 0.91; 95% confidence interval (CI): 0.88–0.94 for the overall scale, 0.85; 95% CI: 0.77–0.90 for the upper extremity subscale, 0.91; 95% CI: 0.70–0.97 for the basic lower extremity subscale, 0.91; 95% CI: 0.83–0.96 for the advanced lower extremity subscale (Figure 2, Table 1).

Moderate-quality evidence from three studies [24, 25, 53] reported inconsistent standard error of measurement (SEM) for the LLFDI-FC overall score (1.59 to 2.9), basic lower extremity (1.88 to 4.4) and advanced lower extremity (2.6 to 4.3). High-quality evidence showed insufficient SEM for the upper extremity subscale (4.2 to 5.1) (Table 3).

b. Validity

Moderate-quality evidence from four studies [31, 41, 53, 55] supported the sufficient or inconsistent cross-sectional convergent validity for LLFDI-FC overall score (Table 2). The correlations between the LLFDI-FC and other health status, function or disability scales ranged from 0.52 to 0.83 [41, 53]. Adequate correlations were observed between LLFDI-FC overall scores and performance-based measures (absolute $r = 0.55$ to 0.69) [41, 45, 55]. There was moderate-quality evidence to demonstrate that the LLFDI-FC upper extremity and advanced lower extremity scales had sufficient or inconsistent convergent validity, while the quality of evidence was low for the basic lower extremity scale (Table 2).

Low-quality evidence from two studies [35, 45] showed that the LLFDI-FC overall score and subscales had sufficient known-groups validity (functional limitation as measured by the physical functioning subscale (PF-10) of the SF-36 or cane users) (Table 2). Low-quality evidence from two studies [24, 46] showed that the LLFDI-FC had insufficient predictive validity for adverse outcomes (falls, emergency department visits, hospitalisation, low self-reported health) (Table 2).

c. Responsiveness

Low- or very low-quality evidence from two studies [24, 25] showed that the LLFDI-FC overall and advanced lower-extremity subscale had inconsistent responsiveness, while the upper extremity subscale and advanced lower-extremity subscale had insufficient responsiveness (Table 3).

d. Change scores and interpretability

Four studies reported change scores for the LLFDI-FC [24, 25, 41, 53] with a minimal detectable change (MDC) with 95% confidence (MDC$_{95}$) of 4.3 and MDC$_{90}$ of 3.7 for the overall score, and an absolute and relative (%) smallest real difference of 8.0 (14%). The minimal clinically important difference was reported as 2 for the overall score, 4 for upper extremity and advanced lower-extremity subscales,

**Table 1.** Self-reported mobility measures review: Reliability

| Reliability | # of tests/hypotheses | Summary or pooled result | Overall rating | Quality of evidence |
|---|---|---|---|---|
| LLFDI-FC: Overall score | 3 | 100% of hypotheses were confirmed; *see forest plot overall ICC = 0.91 (0.88–0.94) | Sufficient (+) | High |
| LLFDI-FC: upper extremity | 3 | 100% of hypotheses were confirmed; ICC = 0.85 (0.77–0.90) | Sufficient (+) | High |
| LLFDI—basic L/E | 3 | 100% of hypotheses were confirmed; ICC = 0.91 (0.70–0.97) | Sufficient (+) | High |
| LLFDI—advanced L/E | 3 | 100% of hypotheses were confirmed; ICC = 0.91 (0.83–0.96) | Sufficient (+) | High |
| LEFS | 3 | 100% of hypotheses were confirmed; ICC = 0.97 (0.91–0.99) | Sufficient (+) | Moderate[a] |
| PDA current | 1 | 0% of hypotheses were confirmed; ICC = 0.65 | Insufficient (±) | Low[b] |
| PDA changes | 1 | 0% of hypotheses were confirmed; ICC = 0.66 | Insufficient (±) | Low[b] |
| MAT-SF | 3 | 100% of hypotheses were confirmed; ICC = 0.91 (0.81–0.96) | Sufficient (+) | High |
| MAT-Walking | 1+ for each sub-score | 100% of hypotheses were confirmed; ICC = 0.85–0.89 | Sufficient (+) | Low[b] |

[a]Due to serious indirectness; [b]due to very serious imprecision.

and 3 for the basic lower-extremity subscale. Two studies [41, 53] noted that no floor or ceiling effects were observed for the LLFDI-FC.

There was a strong correction between the self-reported and interview-administered LLFDI-FC overall score ($r$ = 0.95) [41]. One study [31] reported that the basic lower extremity showed acceptable reliability between self-reported scores and clinician assessment scores with an ICC of 0.83 (95% CI: 0.69 to 0.91), while the upper extremity subscale showed poor reliability with an ICC of 0.3 (95% CI: −0.18 to 0.62).

### 2 Lower extremity functional scale

Four studies on the LEFS were identified in this review [47, 48, 56, 58]. Three studies were conducted in English in the USA [58], Canada [56] and Australia [48], and one study [47] was conducted in Slovenia using a Slovenian version.

#### a. Internal consistency and reliability

Very low-quality evidence from one study [47] reported that the LEFS had sufficient internal consistency with a Cronbach's alpha of 0.99 (Table 1). Moderate-quality evidence from three studies [47, 48, 58] showed that the LEFS had sufficient test–retest reliability with a pooled ICC of 0.97 (Table 1) and two studies [47, 48] reported sufficient SEM (0.97 and 3.6).

#### b. Validity

Low-quality evidence showed that the LEFS had inconsistent convergent validity (Table 2). Two studies [48, 58] reported that LEFS had moderate correlations with self-report measures of physical function ($r$ = 0.60 to 0.78) and performance-based measures ($r$ = 0.40 to 0.69). Very low-quality evidence from one study [56] supports LEFS

had sufficient known-groups validity. There were no studies on predictive validity of the LEFS.

#### c. Change scores and interpretability

One study [58] revealed that LEFS (change/SD = 1.2) is more sensitive to change after physical therapy than other function or balance measures (change/SD = 0.55 to 1.1). An MDC (not reported the confidence level) in healthy older adults of 2.69 was reported in one study [47] and an $MDC_{95}$ of 9.9 (upper 95% CI = 12.5) was reported for older adults with hip osteoarthritis [48]. One study [48] noted that floor or ceiling effects were not observed for the LEFS.

### 3 Perceived Driving Abilities Scale

Three studies included in this review assessed the PDA Scale and all were administered in Canada in the English language [27, 29, 42].

#### a. Internal consistency and reliability

High-quality evidence from two studies [27, 42] showed that the PDA had sufficient internal consistency with Cronbach's alpha of 0.92–0.94 for current abilities and 0.77–0.87 for perceived changes in driving ability (Table 1). Low quality of evidence (Table 1) from one study [27] showed insufficient test–retest reliabilities for PDA current (ICC = 0.65) and PDA change (ICC = 0.66).

#### b. Validity

Moderate-quality evidence from two studies [27, 42] showed inconsistent convergent validity for the PDA (Table 2). The PDA current scale correlated with driving frequency, comfort and avoidance scales ($r$ = 0.46 to 0.59) [27, 42], and actual driving behaviour (distance, duration and other measures) ($r$ = 0.13 to 0.39) [27]. The PDA change

**Table 2.** Self-reported mobility measures review: Validity

| Convergent validity | # of tests/hypotheses | Summary or pooled result | Overall rating | Quality of evidence |
|---|---|---|---|---|
| LLFDI-FC: overall score | 16 | 100% of hypotheses were confirmed | Sufficient (+) | Moderate Due to serious indirectness |
| LLFDI-FC: upper extremity | 8 | 75% of hypotheses were confirmed | Sufficient (+); inconsistent | Moderate Due to serious indirectness |
| LLFDI—basic L/E | 8 | 87.5% of hypotheses were confirmed | Sufficient (+); inconsistent | Low Due to serious inconsistency and serious indirectness |
| LLFDI—advanced L/E | 6 | 66.7% of hypotheses were confirmed | Inconsistent (±) | Moderate Due to serious inconsistency |
| LEFS | 13 | 69% of hypotheses were confirmed | Inconsistent (±) | Low Due to serious inconsistency and serious indirectness |
| PDA current | 6 | 50% of hypotheses were confirmed | Inconsistent (±) | Moderate Due to serious inconsistency |
| PDA changes | 6 | 50% of hypotheses were confirmed | Inconsistent (±) | Moderate Due to serious inconsistency |
| MAT-SF | 15 | 60% of hypotheses were confirmed | Inconsistent (±) | Moderate Due to serious inconsistency |
| MAT-Walking | 4 | 50% of hypotheses were confirmed | Sufficient (+) for total walking distance; inconsistent (±) for walking speed | Moderate Due to serious inconsistency |
| **Known-groups validity** | **Number of tests** | **Summary or pooled result** | **Overall rating** | **Quality of evidence** |
| LLFDI-FC: overall score | 2 | 100% of hypotheses were confirmed | Sufficient (+) | Low Due to very serious risk of bias |
| LLFDI-FC: upper extremity | 2 | 100% of hypotheses were confirmed | Sufficient (+) | Low Due to very serious risk of bias |
| LLFDI—basic L/E | 2 | 100% of hypotheses were confirmed | Sufficient (+) | Low Due to very serious risk of bias |
| LLFDI—advanced L/E | 2 | 100% of hypotheses were confirmed | Sufficient (+) | Low Due to very serious risk of bias |
| LEFS | 1 | 100% of hypotheses were confirmed | Sufficient (+) | Very low Due to imprecision and very serious risk of bias |
| MAT-SF | 3 | 100% of hypotheses were confirmed | Sufficient (+) | Low Due to very serious risk of bias |
| **Predictive validity** | **# of tests/hypotheses** | **Summary or pooled result** | **Overall rating** | **Quality of evidence** |
| LLFDI-FC: overall score | 5 | 0% of hypotheses were confirmed | Insufficient (−) | Low Due to very serious risk of bias |
| LLFDI—basic L/E | 5 | 0% of hypotheses were confirmed | Insufficient (−) | Low Due to very serious risk of bias |
| LLFDI—advanced L/E | 5 | 0% of hypotheses were confirmed | Insufficient (−) | Low Due to very serious risk of bias |
| **Structural validity** | **# of tests/hypotheses** | **Summary or pooled result** | **Overall rating** | **Quality of evidence** |
| MAT-SF | 1 | 100% of hypotheses were confirmed | Sufficient (+) | Low Due to very serious risk of bias |

scale also correlated with driving comfort scales ($r = 0.43$ to 0.49) [42], and driving frequency ($r = -0.13$) and avoidance ($r = 0.21$) scales [27].

Chen *et al*. (2021) reported that the accuracy of the PDA was 28.7%, with 52.8% of participants overestimating their ability and 18.5% of participants underestimating their ability [29]. There were no data on predictive validity.

4 Mobility Assessment Tool: MAT-SF and MAT-Walking

Five studies reporting on the MAT were included in this review [34, 39, 44, 49, 51]. Four of them were conducted in

the USA (English) [39, 44, 49, 51] and one study [34] was conducted in Brazil (Portuguese) and Colombia (Spanish).

a. Internal consistency and reliability

High-quality evidence from three cohorts in two studies [34, 49] showed that MAT-SF had sufficient test–retest reliability with a pooled ICC of 0.91, 95% CI: 0.81 to 0.96 (Table 1). Low-quality evidence from one study [44] showed that MAT-Walking sub-score had sufficient test–retest reliability with ICCs of 0.85 for both usual-pace and fast-pace walking distance and 0.89 for total walking distance.

**Table 3.** Self-reported mobility measures review: Responsiveness

| Responsiveness | Number of tests | Summary or pooled result | Overall rating | Quality of evidence |
|---|---|---|---|---|
| LLFDI-FC: overall score | 2 | 50% of hypotheses were confirmed; score changes correlated with GRC (Rho = 0.3) | Inconsistent (±) | Very low<br>Due to very serious risk of bias and serious inconsistency |
| LLFDI-FC: upper extremity | 1 | 0% of hypotheses were confirmed | Insufficient (−) | Low<br>Due to very serious risk of bias |
| LLFDI—basic L/E | 2 | 0% of hypotheses were confirmed | Insufficient (−) | Low<br>Due to very serious risk of bias |
| LLFDI—advanced L/E | 2 | 50% of hypotheses were confirmed; score changes correlated with GRC (Rho = 0.3) | Inconsistent (−) | Very low<br>Due to very serious risk of bias and serious inconsistency |
| MAT-Walking | 1 | 100% of hypotheses were confirmed (total and fast walking distance showed larger effects than usual walking distance) | Sufficient (+) | Low<br>Due to very serious risk of bias |

b. Validity

Low-quality evidence from one study [51] showed the MAT-SF had sufficient structural validity with acceptable model fit for the overall unidimensional and goodness of fit for item level (Table 2). Moderate evidence from three studies [34, 49, 51] showed the MAT-SF had sufficient but inconsistent convergent validity (Table 2). The MAT-SF correlated with the Pepper Assessment Tool for Disability subscales [49] ($r = 0.44$ to $0.60$). Low-quality evidence from two studies [49, 51] showed that the MAT-SF had sufficient known-groups validity (chronic disease or failure to complete the 400-m walk test) (Table 2).

Moderate evidence from one study [44] showed the MAT-SF had sufficient but inconsistent convergent validity: the MAT-Walking was related to speed data using objective walking speed ($r = 0.36$ and $0.45$) and other physical activity measures ($r = 0.66$ and $0.65$) [44]. No data were available for predictive validity.

c. Responsiveness

Low-quality evidence from one study [44] showed that MAT-Walking had sufficient responsiveness (treatment effect: $\eta^2 = 0.18$ to $0.38$, $P < 0.001$) after a 6-month intervention [44].

d. Interpretability

The MAT-SF did not show any floor or ceiling effects among older adults aged 70–89 years [51]. Proxy-reported mobility assessment using the MAT-SF also showed good reliability (ICC = 0.81) and the patient-reported MAT-SF showed a high correlation with proxy-reported MAT-SF ($r = 0.81$) [39].

5 Other mobility measures

The Questionnaire Rising and Sitting down (QR&S) items are predominantly loaded on a single component and have good internal consistency (Cronbach's alpha = 0.96) [54]. De Laat *et al*. [30] reported on QR&S test–retest reliability (ICC = 0.83, $n = 22$), with a relative SEM of 6.7% and the smallest detectable difference of 18.6%. The same study also demonstrated that the QR&S had good construct validity.

The PROMIS physical function may also be used to assess perceived mobility. The 10-item simulated computerised adaptive testing was better than static versions in information content and reliability in PROMIS [33], and the PROMIS $t$-score correlated with the modified physical performance test overall scores ($r = 0.59$) and component scores ($r = -0.28$ to $0.64$) [37]. Supplementary Appendix H summarises the reported psychometric properties for other self-reported mobility measures.

## Discussion

This systematic review summarised the psychometric properties of self-report measures that focus on community-dwelling older adults' perceived mobility ability. We found several measures, including the LLFDI-FC, LEFS, PDA, MAT-SF/Walking and others, that reflected the concept of mobility as described by the WHO World Report on Ageing and Health, and that have been evaluated in community-dwelling older adults. Our findings on their psychometric properties can be used by researchers and clinicians to inform the selection of the most appropriate measures for assessing mobility in community-dwelling older adults.

Previous systematic reviews on mobility measures were conducted in different populations than our current review [12] or they focused only on one specific measure (such as the LLFDI [60] or LEFS [61]), or type of assessment [62]. This review included a broad range of outcome measures based on the WHO concept of 'ability to be mobile' and summarised measures that were valid among community-dwelling older adults. Previous summaries of the psychometric properties of the LLFDI and LEFS in other reviews [60, 61] are consistent with our main findings; however, in this review, we also assessed the synthesised results using COSMIN criteria for good measurement properties and GRADE for the overall quality of the evidence. Although the LLFDI-FC, MAT

and LEFS were the most extensively evaluated measures in our review and had sufficient convergent validity, our results highlight an important gap with respect to predictive validity. The LLFDI-FC was the only measure with any evidence of its predictive validity, and reported AUCs did not reach COSMIN criteria (AUC of 0.7) (low-quality evidence). However, we acknowledge that the requirement of an AUC of 0.7 or higher for predictive ability, stipulated by the COSMIN criteria, may not be appropriate for mobility measurements and non-diagnostic work in observational studies in the community due to the complex contextual factors involved in mobility measurements. More work is necessary to determine the most appropriate criteria for evaluating the predictive ability of multifaceted mobility measures. Given the established role of mobility measurement for screening and risk stratification, there is an urgent need for longitudinal studies explicitly designed to evaluate the predictive validity of these mobility measures.

Most measures of perceived mobility in this review demonstrated sufficient internal consistency and reliability; however, the evidence was low and limited regarding their responsiveness. Responsiveness is important to consider in studies with the purpose of assessing the effectiveness of interventions in clinical trials [63]. Providing higher quality evidence on responsiveness for mobility measures will require future studies using clear hypotheses for treatment effect size and robust designs.

Some technology-dependent mobility measures emerged in this review (e.g. MAT-SF, PROMIS physical function), which had better internal consistency, reliability and convergent validity than the traditional questionnaire format, but may require more resources to administer [33]. Future well-designed studies are needed to directly compare psychometric properties between traditional and technology-based measures of perceived mobility.

It was notable that few measures of mobility in this review cited using a conceptual framework, model or operational definition. Of the frameworks mentioned, Nagi's disablement model and the ICF were the most common. However, these frameworks are not specific to mobility and do not distinguish among different aspects of mobility that can impact measurement properties. To advance the science of mobility measurement as an indicator of healthy ageing, in this issue of *Age and Ageing*, we proposed a new framework that outlines three facets of mobility measures [13]. Future studies may benefit from using our unified mobility measurement framework alongside the current WHO descriptions of relevant functional ability domains [14] to develop and test conceptually sound and comprehensive mobility measures.

This systematic review had several strengths. We did not restrict mobility to movement across different places, specific conditions or rehabilitation settings. We used a broad search that included both physical function and transportation measures. We closely followed the COSMIN methodology for systematic reviews of patient-reported outcome measures [22] that included developing the search, rating the risk of

bias, assessing the overall quality of evidence and synthesising results based on explicit criteria.

Nonetheless, our review has some important limitations. First, there is no medical subject heading in MEDLINE for mobility that is consistent with our conceptualisation. To identify all potentially relevant mobility measures, we used broad search strategies that yielded an unmanageable number of records for manual screening and necessitated the use of machine learning approaches. We also restricted our search to older adults and self-reported outcome measures and we did not search the grey literature. Therefore, we may have missed some relevant studies or outcome measures. Second, in line with COSMIN guidelines, we only included studies that examined psychometric properties as one of their main purposes, which may miss relevant information, especially for assessment of predictive validity and responsiveness. Third, we did not assess the impact of publication bias on the overall quality of evidence due to a lack of guidance on this issue, but publication bias may have substantial influences on psychometric properties. Fourth, we also did not assess the content validity of the included measures, which was out of the scope of this review, and the limited evidence did not allow us to conduct our planned subgroup analyses by age, sex and Low-to-Middle-High-Income status. Finally, we did not find any comprehensive mobility measures that covered all mobility domains outlined by the WHO. Future studies are likely needed to develop comprehensive mobility measures that reflect all the sub-domains of mobility in one measure.

Although further prospective studies are required to develop a comprehensive conceptually relevant self-perceived mobility measure with all acceptable psychometric properties, in the meantime, a simple mobility indicator may be useful to advance the WHO decade of healthy ageing and influence health-care policies, particularly in low- and middle-income countries. It is important to note that in low- and middle-income countries, where access to personal vehicles may be limited, other simple or basic mobility devices beyond cars may be more relevant. Therefore, when assessing the transportation domain of mobility measures in low- and middle-income countries, one needs to consider the availability of mobility devices, such as personal cars, public buses, canes, rollators, railings and other small mobility vehicles for shopping.

In conclusion, we summarised the psychometric properties of self-report measures that focus on perceived mobility ability in community-dwelling older adults. The LLFDI-FC, LEFS, PDA and MAT-SF/Walking can be used to assess perceived mobility in community-dwelling older adults; however, further data are needed on their responsiveness and predictive validity.

## 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.

---

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

**Acknowledgements:** We acknowledge the contributions of Neera Bhatnagar, McMaster University librarian, in assisting us with the literature search.

**Declaration of Conflicts of Interest:** M.B. holds a Tier 2 Canada Research Chair in Mobility, Ageing, and Chronic Disease. P.R. holds the Raymond and Margaret Labarge Chair in Research and Knowledge, is the Director of the McMaster Institute for Research on Ageing and the Labarge Centre for Mobility in Ageing, and holds a Tier 1 Canada Research Chair in Geroscience.

**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.

**Data Availability Statement:** This paper's underlying data are available in the article and its supplementary data files.

---

## References

1. Team WHO. Healthy Ageing and Functional Ability. https://www.who.int/news-room/questions-and-answers/item/healthy-ageing-and-functional-ability (accessed 12 February 2023).
2. Organization WH. World Report on Ageing and Health. https://apps.who.int/iris/bitstream/handle/10665/186463/9789240694811_eng.pdf?sequence=1&isAllowed=y2015 [cited 5 August 2022].
3. Rantanen T. Promoting mobility in older people. J Prev Med Public Health 2013; 46: S50–4.
4. Shafrin J, Sullivan J, Goldman DP, Gill TM. The association between observed mobility and quality of life in the near elderly. PloS One 2017; 12: e0182920. https://doi.org/10.1371/journal.pone.0182920.
5. Davis JC, Bryan S, Best JR *et al.* Mobility predicts change in older adults' health-related quality of life: evidence from a Vancouver falls prevention prospective cohort study. Health Qual Life Outcomes 2015; 13: 101. https://doi.org/10.1186/s12955-015-0299-0.
6. Freiberger E, Sieber CC, Kob R. Mobility in older community-dwelling persons: a narrative review. 2020; 11.
7. Billot M, Calvani R, Urtamo A *et al.* Preserving mobility in older adults with physical frailty and sarcopenia: opportunities, challenges, and recommendations for physical activity interventions. Clin Interv Aging 2020; 15: 1675–90.
8. Yang L, Liao LR, Lam FM, He CQ, Pang MY. Psychometric properties of dual-task balance assessments for older adults: a systematic review. Maturitas 2015; 80: 359–69.
9. Nazari G, Lu S, MacDermid JC. Psychometric properties of performance-based functional tests in patients with shoulder pathologies: a systematic review and meta-analysis. Arch Phys Med Rehabil 2020; 101: 1053–63.
10. Trautwein S, Maurus P, Barisch-Fritz B, Hadzic A, Woll A. Recommended motor assessments based on psychometric properties in individuals with dementia: a systematic review. European review of aging and physical activity : official journal of the European Group for Research into Elderly and Physical Activity 2019; 16: 20. https://doi.org/10.1186/s11556-019-0228-z.
11. Küçükdeveci AA, Elhan AH, Erdoğan BD *et al.* Use and detailed metric properties of patient-reported outcome measures for rheumatoid arthritis: a systematic review covering two decades. RMD Open 2021; 7: e001707. https://doi.org/10.1136/rmdopen-2021-001707.
12. Pinto-Carral A, Fernández-Villa T, Molina de la Torre AJ. Patient-reported mobility: a systematic review. Arch Phys Med Rehabil 2016; 97: 1182–94.
13. Beauchamp MK, Hao Q *et al.* A unified framework for the measurement of mobility in older persons. Age Ageing 2023: afad125. https://doi.org/10.1093/ageing/afad125.
14. World Health Organization. World report on ageing and health. World Health Organization. 2015. https://apps.who.int/iris/handle/10665/186463. (accessed 12 February 2023).
15. Team WHO. Decade of Healthy Ageing: Baseline Report. https://www.who.int/publications/i/item/9789240017900 2021 [cited 16 December 2021].
16. Glass TA. Conjugating the "tenses" of function: discordance among hypothetical, experimental, and enacted function in older adults. Gerontologist 1998; 38: 101–12.
17. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016; 5: 210. https://doi.org/10.1186/s13643-016-0384-4.
18. Valizadeh A, Moassefi M, Nakhostin-Ansari A *et al.* Abstract screening using the automated tool Rayyan: results of effectiveness in three diagnostic test accuracy systematic reviews. BMC Med Res Methodol 2022; 22: 160. https://doi.org/10.1186/s12874-022-01631-8.
19. Mokkink LB, de Vet HCW, Prinsen CAC *et al.* COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 2018; 27: 1171–9.
20. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 2012; 21: 651–7.
21. Guyatt G, Oxman AD, Akl EA *et al.* GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011; 64: 383–94.
22. Mokkink LB, Prinsen CA, Patrick DL, *et al*. Guideline for systematic reviews of outcome measurement instruments. https://www.cosmin.nl/tools/guideline-conducting-systematic-review-outcome-measures/. 2018. (accessed 20 March 2022).
23. Laliberté E. Metacor: Meta-Analysis of Correlation Coefficients. https://cran.r-project.org/web/packages/metacor/index.html2019 (accessed 12 January 2023).
24. Beauchamp MK, Jette AM, Ward RE *et al.* Predictive validity and responsiveness of patient-reported and performance-based measures of function in the Boston RISE study. GERONA 2015; 70: 616–22.

25. Beauchamp MK, Ward RE, Jette AM, Bean JF. Meaningful change estimates for the late-life function and disability instrument in older adults. Journals of Gerontology Series A-Biological Sciences & Medical Sciences 2019; 74: 556–9.

26. Bergland A, Jarnlo G, Laake K. Validity of an index of self-reported walking for balance and falls in elderly women. Advances in Physiotherapy 2002; 4: 65–73.

27. Blanchard RA, Myers AM. Examination of driving comfort and self-regulatory practices in older adults using in-vehicle devices to assess natural driving patterns. Accid Anal Prev 2010; 42: 1213–9.

28. Chen H, Rejeski WJ, Gill TM *et al.* A comparison of self-report indices of major mobility disability to failure on the 400-m walk test: the LIFE study. J Gerontol A Biol Sci Med Sci 2018; 73: 513–8.

29. Chen YT, Gelinas I, Mazer B *et al.* Personal and clinical factors associated with older drivers' self-awareness of driving performance. Can J Aging 2021; 40: 82–96.

30. de Laat FA, Rommers GM, Geertzen JH, Roorda LD. Construct validity and test-retest reliability of the questionnaire rising and sitting down in lower-limb amputees. Arch Phys Med 2011; 92: 1305–10.

31. Feuering R, Vered E, Kushnir T, Jette AM, Melzer I. Differences between self-reported and observed physical functioning in independent older adults. Disabil Rehabil 2014; 36: 1395–401.

32. Freedman VA, Kasper JD, Cornman JC *et al.* Validation of new measures of disability and functioning in the National Health and aging trends study. J Gerontol 2011; 66A: 1013–21.

33. Fries JF, Cella D, Rose M, Krishnan E, Bruce B. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. Journal of Rheumatology 2009; 36: 2061–6.

34. Guerra RO, Oliveira BS, Alvarado BE *et al.* Validity and applicability of a video-based animated tool to assess mobility in elderly Latin American populations. Geriatr Gerontol Int 2014; 14: 864–73.

35. Haley SM, Jette AM, Coster WJ *et al.* Late life function and disability instrument: II. Development and evaluation of the function component. J Gerontol 2002; 57: M217–22.

36. Hoeymans N, Wouters ER, Feskens EJ, van den Bos GA, Kromhout D. Reproducibility of performance-based and self-reported measures of functional status. Journals of Gerontology Series A-Biological Sciences & Medical Sciences 1997; 52: M363–8.

37. Houck J, Jacobson R, Bass M, Dasilva C, Baumhauer JF. Improving interpretation of the patient-reported outcomes measurement information system (PROMIS) physical function scale for specific tasks in community-dwelling older adults. J Geriatr Phys Ther 2020; 43: 142–52.

38. Hung M, Hon SD, Cheng C *et al.* Psychometric evaluation of the lower extremity computerized adaptive test, the modified Harris hip score, and the hip outcome score. Orthop J Sports Med 2014; 2: 2325967114562191. https://doi.org/10.1177/2325967114562191.

39. Kim S, Miller ME, Lin M *et al.* Self- vs proxy-reported mobility using the mobility assessment tool-short form in elderly preoperative patients. Eur 2018; 15: 5. https://doi.org/10.1186/s11556-018-0194-x.

40. Kopec JA, Russell L, Sayre EC, Rahman MM. Self-reported ability to walk, run, and lift objects among older Canadians. Rehabil Res Pract 2017; 2017: 1–9.

41. Lapier TK. Utility of the late life function and disability instrument as an outcome measure in patients participating in outpatient cardiac rehabilitation: a preliminary study. Physiother Can 2012; 64: 53–62.

42. MacDonald L, Myers AM, Blanchard RA. Correspondence among older drivers' perceptions, abilities, and behaviors. Topics in Geriatric Rehabilitation 2008; 24: 239–52.

43. Manty M, Heinonen A, Leinonen R *et al.* Construct and predictive validity of a self-reported measure of preclinical mobility limitation. Archives of Physical Medicine & Rehabilitation 2007; 88: 1108–13.

44. Marsh AP, Janssen JA, Ip EH *et al.* Assessing walking activity in older adults: development and validation of a novel computer-animated assessment tool. J Gerontol 2015; 70: 1555–61.

45. Melzer I, Kurz I, Sarid O, Jette AM. Relationship between self-reported function and disability and balance performance measures in the elderly. J Rehabil Res Dev 2007; 44: 685–91.

46. O'Hoski S, Bean JF, Ma J *et al.* Physical function and frailty for predicting adverse outcomes in older primary care patients. Archives of Physical Medicine & Rehabilitation 2020; 101: 592–8.

47. Paravlic AH, Pisot S, Mitic P, Pisot R. Validation of the Oxford knee score and lower extremity functional score questionnaires for use in Slovenia. Archives of Orthopaedic & Trauma Surgery 2020; 140: 1515–22.

48. Pua YH, Cowan SM, Wrigley TV, Bennell KL. The lower extremity functional scale could be an alternative to the western Ontario and McMaster universities osteoarthritis index physical function scale. J Clin Epidemiol 2009; 62: 1103–11.

49. Rejeski WJ, Ip EH, Marsh AP, Barnard RT. Development and validation of a video-animated tool for assessing mobility. J Gerontol 2010; 65: 664–71.

50. Rejeski WJ, Ip EH, Marsh AP, Miller ME, Farmer DF. Measuring disability in older adults: the international classification system of functioning, disability and health (ICF) framework. Geriatr Gerontol Int 2008; 8: 48–54.

51. Rejeski WJ, Marsh AP, Anton S *et al.* The MAT-sf: clinical relevance and validity. J Gerontol A Biol Sci Med Sci 2013; 68: 1567–74.

52. Reuben DB, Siu AL, Kimpau S. Predictive validity of self-report and performance-based measures of function and health. Journals of Gerontology 1992; 47: M106–10.

53. Roaldsen KS, Halvarsson A, Sarlija B, Franzen E, Stahle A. Self-reported function and disability in late life - cross-cultural adaptation and validation of the Swedish version of the late-life function and disability instrument. Disability & Rehabilitation 2014; 36: 813–7.

54. Roorda LD, Molenaar IW, Lankhorst GJ, Bouter LM, Measuring Mobility Study Group. Improvement of a questionnaire measuring activity limitations in rising and sitting down in patients with lower-extremity disorders living at home. Arch Phys Med Rehabil 2005; 86: 2204–10.

55. Sayers SP, Jette AM, Haley SM, Heeren TC, Guralnik JM, Fielding RA. Validation of the late-life function and disability instrument. J Am Geriatr Soc 2004; 52: 1554–9.

56. Stratford PW, Kennedy DM, Riddle DL. New study design evaluated the validity of measures to assess change after hip or knee arthroplasty. J Clin Epidemiol 2009; 62: 347–52.

57. Tager IB, Swanson A, Satariano WA. Reliability of physical performance and self-reported functional measures in an older population. J Gerontol A Biol Sci Med Sci 1998; 53A: M295–300.

58. Verheijde JL, White F, Tompkins J *et al.* Reliability, validity, and sensitivity to change of the lower extremity functional scale in individuals affected by stroke. Pm & R 2013; 5: 1019–25.

59. Wang C-Y, Ming-Hsia H, Hui-Ya C, Ren-Hau L. Self-reported mobility and instrumental activities of daily living: test-retest reliability and criterion validity. Journal of Aging & Physical Activity 2012; 20: 186–97.

60. Beauchamp MK, Schmidt CT, Pedersen MM, Bean JF, Jette AM. Psychometric properties of the late-life function and disability instrument: a systematic review. BMC Geriatr 2014; 14: 12. https://doi.org/10.1186/1471-2318-14-12.

61. Mehta SP, Fulton A, Quach C, Thistle M, Toledo C, Evans NA. Measurement properties of the lower extremity functional scale: a systematic review. J Orthop Sports Phys Ther 2016 Mar; 46: 200–16.

62. Morgan A, Bégin D, Heisz J, Tang A, Thabane L, Richardson J. Measurement properties of remotely or self-administered lower extremity mobility performance measures in adults: a systematic review. Phys Ther 2022; 102: 102. https://doi.org/10.1093/ptj/pzac078.

63. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med 1993; 118: 622–9.