JAMDA 14 (2013) 170–178

JAMDA

journal homepage: www.jamda.com

Review

Validity and Reliability of Tools to Measure Muscle Mass, Strength, and Physical Performance in Community-Dwelling Older People: A Systematic Review

Donja M. Mijnarends MSc [a,*], Judith M.M. Meijers PhD, RN [a], Ruud J.G. Halfens PhD [a], Sovianne ter Borg MSc [b], Yvette C. Luiking PhD [b], Sjors Verlaan MSc [b], Daniela Schoberer MSc [c], Alfonso J. Cruz Jentoft MD, PhD [d], Luc J.C. van Loon Prof, PhD [e], Jos M.G.A. Schols Prof, PhD [a,f]

[a] School CAPHRI, Department of Health Services Research, Maastricht University, Maastricht, The Netherlands
[b] Nutricia Advanced Medical Nutrition, Danone Research, Centre for Specialised Nutrition, Wageningen, The Netherlands
[c] Institute of Nursing Science, Medical University Graz, Graz, Austria
[d] Servicio de Geriatría, Hospital Universitario Ramón y Cajal, Madrid, Spain
[e] School NUTRIM, Department of Human Movement Sciences, Maastricht University, Maastricht, The Netherlands
[f] School CAPHRI, Department of General Practice, Maastricht University, Maastricht, The Netherlands

Keywords:
Muscle mass
muscle function
community-dwelling
measurement properties

A B S T R A C T

Background: This study critically appraises the measurement properties of tools to measure muscle mass, strength, and physical performance in community-dwelling older people. This can support the selection of a valid and reliable set of tools that is feasible for future screening and identification of sarcopenia.
Methods: The databases PubMed, Cumulative Index to Nursing and Allied Health Literature (CINAHL), and Cochrane were systematically searched (January 11, 2012). Studies were included if they investigated the measurement properties or feasibility, or both, of tools to measure muscle mass, strength, and physical performance in community-dwelling older people aged ≥60 years. The consensus-based standards for the selection of health status measurement instruments (COSMIN) checklist was used for quality appraisal of the studies.
Results: Sixty-two publications were deemed eligible, including tools for muscle mass (n = 16), muscle strength (n = 15), and physical performance (n = 31). Magnetic resonance imaging, computed tomography, and a 4-compartment model were used as gold standards for muscle mass assessment. Other frequently used measures of muscle mass are dual-energy x-ray and the bioelectrical impedance (BIA); however, reliability data of the BIA are lacking. Handheld dynamometry and gait speed or a short physical performance battery provide a valid and reliable measurement of muscle strength and physical performance, respectively.
Conclusions: It can be concluded that several tools are available for valid and reliable measurements of muscle mass, strength, and performance in clinical settings. For a home-setting BIA, handheld dynamometry and gait speed or a short physical performance battery are the most valid, reliable, and feasible. The combination of selected instruments and its use for the screening and identification of sarcopenia in community-dwelling older people need further evaluation.

The term *sarcopenia* was first introduced by Rosenberg[1] in 1989 and literally means poverty (or deficiency) of flesh. The relevance of sarcopenia as a geriatric syndrome is indicated by the statement that "no decline with age is more dramatic or potentially more functionally significant than the decline in lean body mass."[1] Over the last 6 years, several initiatives have been undertaken to find consensus on a proper definition of sarcopenia.[2] Diagnosing sarcopenia by measuring only muscle mass appeared to be insufficient. Therefore, in 2009, 2 consensus definitions were proposed, adding loss of muscle function (International Working Group on Sarcopenia) or muscle strength and physical performance (European Working Group on Sarcopenia in Older People) to its definition.[2] In 2010, another working group formulated sarcopenia as a reduced muscle mass with limited mobility.[3] Depending on the definition used, prevalence rate estimates of sarcopenia in community-dwelling older people >60 years old can vary between 3% and 52%.[4,5]

This work was supported by Nutricia Advanced Medical Nutrition, Danone Research, Centre for Specialised Nutrition.
Danone Research provided the salary and project support for D.M.M.
* Address correspondence to Donja M. Mijnarends, MSc, Maastricht University, Department of Health Services Research, School CAPHRI, P.O. Box 616, 6200 MD, Maastricht, The Netherlands.
E-mail address: d.mijnarends@maastrichtuniversity.nl (D.M. Mijnarends).

1525-8610/$ - see front matter Published by Elsevier Inc. on behalf of the American Medical Directors Association, Inc.
http://dx.doi.org/10.1016/j.jamda.2012.10.009

With adequate screening for sarcopenia among community-dwelling older people, those with an increased risk for adverse outcomes, such as physical disability, and increased risk for falls, loss of independence, and death[6–9] may be identified at an earlier stage. After this initial screening, diagnosis could take place in a clinical setting. Early identification of sarcopenia would be of great clinical relevance because the loss of muscle mass and strength with aging can be largely reversed by proper exercise and nutritional intervention.[8] The European Working Group on Sarcopenia in Older People introduced an algorithm for the identification of older people with sarcopenia based on their definition.[10] For identification of sarcopenia in a research setting, several tools were stated to measure muscle mass, strength, and physical performance. However, those tools are not specifically focused on screening among community-dwelling older people, for whom case finding should be performed. Thus, exploring the measurement properties (validity and reliability) of tools feasible for measurements of muscle mass, strength, and performance is an important step for the future development of a set of tools to screen for or diagnose sarcopenia in a valid and reliable way among community-dwelling older people.

To the best of our knowledge, no systematic review on the measurement properties of tools to measure muscle mass, strength, and physical performance in community-dwelling older people has previously been published. The objective of this systematic literature review is to critically appraise the measurement properties of tools to measure muscle mass, strength, and physical performance. Moreover, the feasibility of such tools in community-dwelling older people will be evaluated. The identification of a set of most valid and reliable tools may support the future development of a screening tool for sarcopenia in community-dwelling older people.

Methods

Online databases PubMed, Cumulative Index to Nursing and Allied Health Literature, and Cochrane were systematically searched in title and abstract. The search was limited to publications in English and Dutch. Articles were searched up to January 11, 2012. Search terms were selected from literature and expert consultation, taking into account the three parameters of sarcopenia, that is, muscle mass, strength, and performance, as mentioned in the consensus definition of the European Working Group.[10] Backward citation tracking was performed to identify additional relevant articles.

The final selection of search terms was: (1) *construct of interest*—muscle mass, fat free mass, skeletal muscle, muscle strength, lower limb strength, upper limb strength, lower extremity strength, upper extremity strength, grip strength, hand grip strength, elbow flexion strength, ankle strength, knee strength, maximal strength, physical performance, functional performance, muscle quality, muscle function, gait speed, walking speed; (2) *target population*—elderly, older adults, older people, older persons, sarcopeni*, community-dwelling, assisted living; (3) *type of measurement instrument*—tool*, instrument*, technique*, measure*, assess*, evaluat*, test; and (4) *measurement properties*—reliab*, valid*, feasib*, consistenc*, accura*, agreement, precision, psychometric propert*. Asterisks indicate search for words with alternative endings, e.g. reliable, reliability etcetera.

Study Eligibility Criteria

The following inclusion criteria were used for the selection of relevant studies: The study had to evaluate the validity, reliability, and/or feasibility of a tool to measure muscle mass, strength, physical performance, or sarcopenia; focus on community-dwelling older people or people in assisted living facilities at age $\geq$60 years; and provide a description of the method used to measure muscle mass, strength, physical performance, or sarcopenia.

Studies were excluded if they studied a specific patient population (eg, patients with Parkinson disease) or if they measured only activities of daily living (eg, Late Life Function and Disability Instrument), because those scales are focused on *functional* activities rather than on *physical* performance.

Study Appraisal and Synthesis Methods

The search hits were inserted in EndNote X2 and duplicates were removed. All titles and abstracts were independently screened by two authors (D.M. and S.t.B.) and scored as “relevant” or “not relevant” based on the inclusion and exclusion criteria mentioned earlier. The reviewers discussed their opinions to reach consensus if they disagreed about the inclusion of a study. A third reviewer (J.M.M. or Y.L.) was asked to participate in the final decision if disagreement persisted. Subsequently, full texts were assessed for inclusion by one reviewer (D.M.), according to the eligibility criteria mentioned earlier. After that, the methodologic quality of the studies was assessed by the consensus-based standards for the selection of health status measurement instruments (COSMIN) checklist.[11] The COSMIN checklist evaluates the methodologic quality of studies on measurement properties among others, content validity (evidence that the content of a test corresponds to the content of the construct it was designed to cover), construct validity (the degree to which the scores of a tool are consistent with hypotheses or are related to other variables and other tools measuring the same construct), and concurrent validity (evidence that scores from a tool correspond with the gold standard or concurrent external tools conceptually related to the measured construct). Criteria encompass, for example, handling of missing items, sample size, and appropriateness of statistical methods. A methodologic quality score (poor, fair, good, or excellent) per box was obtained by taking the lowest rating of any item in a box (“worse score counts”). One reviewer (D.M.) assessed the quality of all articles, and a second reviewer (D.S.) randomly assessed one third of the articles to validate the outcomes of the first reviewer. Studies with a poor quality score were excluded for this review; no weighting was applied to the studies rating fair, good, or excellent quality. The final selection of articles was checked by an expert in the field of sarcopenia (A.J.C.) who verified that relevant articles were included.

A tool is scored “+” when having a high reliability [intraclass correlation coefficient or weighted Kappa $\geq$ 0.70 or Pearson correlation (r) $\geq$ 0.80; high construct validity when correlation between constructs $\geq$0.50, or high concurrent validity when Pearson/Spearman correlation or area under the curve $\geq$ 0.70].[12]

Results

An overview of the process of study selection and reasons for exclusion is shown in Figure 1. After title, abstract, and full-text screening, 135 studies were found eligible and assessed for quality. Of these 135 studies, 49 were appraised for quality by a second reviewer (D.S.); disagreement between the reviewers existed over four, because of lack of clarity of appropriate statistical methods (n = 2), choice of measurement property (n = 1), or interpretation of study results (n = 1). In a consensus meeting, the two reviewers discussed their opinions, after which agreement was reached.

A final selection of 62 studies was included in this review, classified as having fair (n = 61) or good (n = 1) quality. An overview of the characteristics of the individual studies is presented in Appendixes A, B, and C. Table 1 provides an overview of the assessed measurement properties of the included studies.

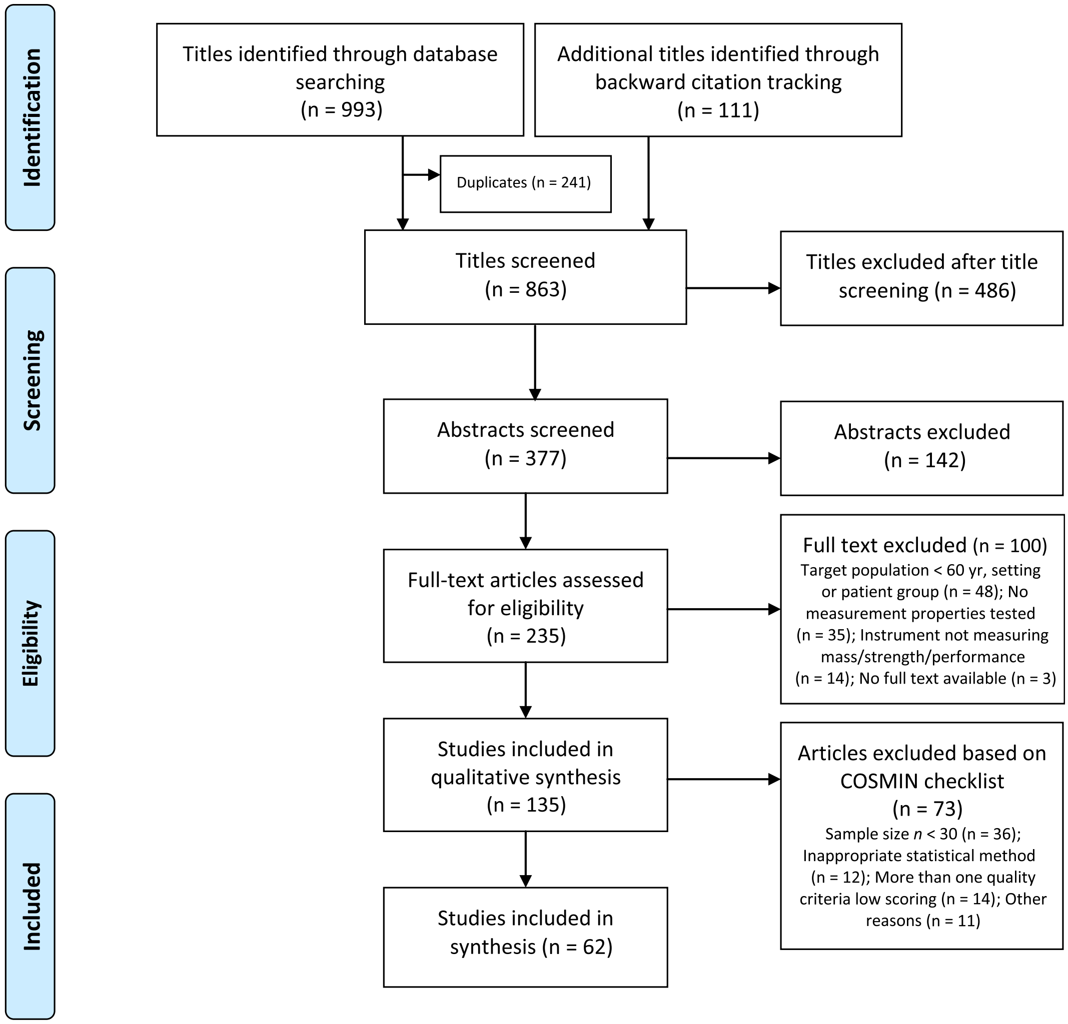

Fig. 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart showing selection procedure of articles.

The tools are described in the following sections according to the parameter: muscle mass (n = 16), strength (n = 15), and performance (n = 31).

Validity, Reliability, and Feasibility

The validity and reliability of 10 different tools to assess muscle mass were reported (Table 2). The included studies evaluated mainly the concurrent validity, only one study assessed responsiveness,[13] and no studies evaluated the reliability of the tools. As listed in Table 2, magnetic resonance imaging (MRI), computed tomography (CT), and a 4-compartment (4-C) model were used as gold standards for assessment of muscle mass. The only study describing responsiveness showed that ultrasonography was able to detect changes in muscle mass before and after training. Dual-energy X-ray (DXA) was found to be highly correlated with MRI, CT, and the 4-C model.

Bioelectrical impedance (BIA) was found to have high concurrent validity; however, significant differences in estimation of mean fat-free mass between BIA and DXA were found.[14] Furthermore, it was stated that its validity is questionable due to significant differences in

Table 1
Measurement Properties Assessed in the Included Studies (by the COSMIN Checklist)

Measurement Property	Muscle Mass			Muscle Strength			Physical Performance		
	P	F	G	P*	F	G	P*	F	G
Box A Internal consistency[30,35]	–	–	–	–	–	–	2	–	–
Box B Reliability[18,21–23,25–31,34,35,38,40,41,44–47,49–53,56,58,68–71]	–	–	–	–	14	–	–	17	–
Box C Measurement error[25,26,68]	–	–	–	1	1	–	–	1	–
Box D Content validity[58]	–	–	–	–	–	–	–	1	–
Box F Hypothesis testing[18–20,23,29,30,32–36,38,39,41–45,47,48,51,52,54,55,58,68,71,72]	–	–	–	–	8	–	3	16	1
Box H Criterion validity[13–17,20,21,24,37,41,51,53,57,69,73–83]	–	16	–	–	5	–	–	4	–
Box I Responsiveness[13,50]	–	1	–	–	–	–	–	1	–

COSMIN, consensus-based standards for the selection of health status measurement instruments; F, fair; G, good; P, poor.
No studies were scored with excellent (E).
*For some studies, more than one box was assessed; in case one box was assessed “poor” quality but the other with “fair” or “good,” the study was included in the final study selection, only taking into account data from the fair/good box.

Table 2
Measurement Properties of Muscle Mass Tools in Community-Dwelling Older Persons

Instrument	Reliability	Validity*			Portable and Executable in a Home Setting?
		Outcome	Concurrent	Comparator Instrument	
BIA					
Single frequency[14,16,73,75–77,80]		$r > 0.79$, $R^2 = 0.70$	+	TBW, 4-C model, DXA	Yes
Multifrequency[73,74,78]		ICC > 0.95, LOA 12 kg	+	DXA whole body	Yes
		ICC > 0.69	±	DXA segmental	
BOD POD[74]		LOA −11.0 to 2.4†	?	DXA	No
Calf circumference[15]		$r = 0.63$	−	DXA	Yes
CT[13,17,83]		$r > 0.83$, $R^2 = 0.96$	+	Used as gold standard vs DXA and ultrasonography	No
DXA[14–17,78–81]		$r > 0.91$	+	MRI, CT, 4-C model	No
Equation for LBW[79]		LOA 0.65 − 11.65 kg†	+	DXA	Yes
MRI[17]		$r > 0.91$	+	Used as gold standard vs DXA	No
Skin-fold thickness[16]		$R^2 = 0.62$	−	DXA	Yes
Ultrasonography[13]		$r > 0.83$	+	CT	Yes
4-C model[76,83]		$R^2 = 0.98$, $r = 0.95$	+	Used as gold standard vs DXA and BIA	No

+, high concurrent validity [Pearson/Spearman correlation or area under the curve (AUC) ≥ 0.70 or responsiveness ≥ 0.50]; (−), low validity (Pearson/Spearman correlation or AUC < 0.70); 4-C model, 4-compartment model; BIA, bioelectrical impedance; BOD POD, measure of air displacement plethysmography; CT, computed tomography; DXA, dual-energy x-ray; ICC, intraclass correlation coefficient; LBW, lean body weight; LOA, limits of agreement; MRI, magnetic resonance imaging; TBW, total body water.
*Only concurrent validity (evidence that scores from a tool correspond with the gold standard) was assessed in the included studies.
†LOA could not be interpreted because no information was provided on the minimally important change.

the estimation of muscle mass by BIA compared with DXA, and reliability data are lacking. Calf circumference[15] and skin-fold thickness[16] both showed low correlations with DXA. Feasibility criteria discussed were exposure to radiation and costs.[17]

Table 3 provides an overview of the tools to measure muscle strength and their validity and reliability. In the included studies, reliability, construct, and concurrent validity were assessed. The handheld dynamometer (HHD), by which measurements of hand grip, ankle, elbow, hip, and knee strength can be made, is valid and reliable.[18–26] It showed both high interrater and intrarater reliability, and concurrent and construct validity were shown by comparison of several types of HHDs with an isokinetic dynamometer,[24] a vigorimeter,[20,23] and sit-to-stand testing.[19] Other tools to assess muscle strength like the leg press,[27] plate spring gauge,[28] and pull down[29] showed good reliability. However, no validity data were found for these specific tools. Feasibility criteria mentioned were rate of injuries, simplicity, time of the measurement, safety, and costs.[29]

Table 4 lists the validity and reliability of tools that can be used to measure physical performance. Most studies evaluated the intrarater reliability, construct, and/or concurrent validity. Tools to assess physical performance comprised questionnaires,[30,31] several performance-based tools,[31–57] and a tool using video animation (the mobility assessment tool).[58] Some tools measure single performance items, such as gait speed or standing balance, whereas other tools include multiple items. The latter was applied in, for example, the frequently used short physical performance battery (SPPB), which includes standing balance, gait speed, and chair rises (sit-to-stand).[31,37,40,50] The mobility assessment tool is a tool that uses video clips of several types of performance, which subjects have to score as being able to do or not. Reliability and validity for gait speed measurements was confirmed in 9 studies,[31,32,37,38,40,51,53,54,56] and it was found to have high construct validity, shown by correlations with SPPB and stair climb, and predictive validity for disability.[37,42,53–55] Muscle soreness, safety, ease of administration, acceptability to patients, portability, time span, and ability to perform the test were mentioned with regard to feasibility.[35,41,49,58]

Discussion

Many tools are described that measure muscle mass, strength, and physical performance. MRI, CT, and a 4-C model were used as gold standards to measure muscle mass. Also, DXA, even though it is not the gold standard, was often used as reference method, because it is a cheaper and quicker option than the other gold standards for muscle mass. However, when comparing an instrument with a reference instrument that is not a gold standard, it is unknown to which degree the correlation between instruments is influenced by measurement errors of the reference instrument. A remarkable finding was the lack of studies examining the reliability of tools to measure muscle mass in older people. Reeves et al,[59] excluded from this review because of a small sample size, looked at the reliability of ultrasonography and its validity compared with MRI, and found good reliability and validity for ultrasonography. This adds to the evidence for high concurrent validity and responsiveness of ultrasound measurements found in this review.[13]

The leg press and HHD used on both upper and lower extremities are valid and reliable tools to measure muscle strength. The HHD is frequently used; however, Roberts et al[60] concluded in their review that protocols to measure grip strength by HHD differ, which makes comparison between studies difficult. Stark et al[61] reviewed the reliability and validity of HHD in young and older people, and also found that the various studies revealed a lack of homogeneity in methodology for the application of HHD, which underlines the need for using a standard protocol. They concluded that HHD cannot fully replace isokinetic measurements, but considering the costs of isokinetic devices and the impracticality, HHD is a good alternative. However, using hand-grip strength as a predictor of overall strength seems unjustified in the healthy older adult.[62] It can be argued that lower extremity strength might be even more relevant than upper extremity strength, because lower extremity strength is important for functional activities.[62]

Many instruments have been applied to measure (aspects of) physical performance. Gait speed is a useful tool to assess physical performance given its high reliability and concurrent validity. Participants with SPPB scores ≤10 at baseline had significantly higher odds of mobility disability at 3-year follow-up.[63] Cooper et al's[64] review concluded that walking speed, chair rises, and standing balance (components of the SPPB) were all associated with mortality. Those studies add to the clinical importance of the frequently used physical performance tools, namely, gait speed and the SPPB.

Feasibility

For quick *screening* of muscle mass, strength, and physical performance among community-dwelling older people, it would be

Table 3
Measurement Properties of Muscle Strength Tools in Community-Dwelling Older Persons

Instrument	Type of Strength	Outcome	Reliability			Validity			Portable and Executable in a Home Setting?
			Intrarater	Interrater	Measurement Error	Construct	Concurrent	Comparator instrument	
Chest press[27]	Upper limb	ICC > 0.94	+					Not applicable	No
Dumbbell[69]	Elbow flexion	$r = 0.62$				+		Elastic band	Yes
Elastic bands[69]	Elbow flexion	ICC = 0.89, $r = 0.46/0.62$	+			+/−		Dumbbell test, isokinetic assessment	Yes
Handheld dynamometer[18–26]	Grip, pinch, ankle, elbow, hip, knee, trunk flexion and extension	ICC > 0.78, $r > 0.72$	+	+	SEM 2.4	+	+	Knee extension vs STS 10 sec; hand grip vs vigorimeter, several HHD devices, and isokinetic measurements	Yes
		Construct: $r < 0.37$				−		All types of strength compared with 6-MW, BIA, grip, elbow, POMA, TUG	
Isokinetic dynamometer[71]	Ankle, knee, elbow flexion, extension	ICC 0.34–0.85	+/−			+/−			No
		$r = 0.53$						Ankle strength vs chair rise and gait speed	
		$r = 0.47$						Used as reference method	
Leg press[27,41]	Lower limb	ICC > 0.94, $r = 0.78$ men; $r = 0.71$ women	+			+		Leg press vs chair stand Used for comparison	No
Manual muscle testing[19]	Knee extension	$r > 0.64$				+		Knee extension vs STS	Yes
Vigorimeter[20]	Hand grip	ICC > 0.91, $r = 0.89–0.90$ Construct: hypothesis not confirmed	+			−	+	Jamar handheld dynamometer	Yes
Plate with spring gauge[28]	Ankle	ICC = 0.88	+					Not applicable	No
Pull down[29]	Arm, shoulder	$r = 0.97$, LOA 0.43–6.9 kg	+					Not applicable	No

+, high reliability (ICC, weighted Kappa ≥ 0.70 or Pearson correlation ≥ 0.80), high construct validity (correlation between constructs ≥ 0.50) or high concurrent validity [Pearson/Spearman correlation or area under the curve (AUC) ≥ 0.70]; (−), low reliability (ICC, weighted Kappa < 0.70 or correlation < 0.80) or low validity (Pearson/Spearman correlation or AUC < 0.70); BIA, bioelectrical impedance; HHD, handheld dynamometry; ICC, intraclass correlation coefficient; 6-MW, 6-min walk test; POMA, performance-oriented mobility assessment; SEM, standard error of measurement; STS, sit-to-stand; TUG, timed up and go.

Table 4
Measurement Properties of Physical Performance Tools in Community-Dwelling Older Persons

Instrument	Outcome	Reliability: Intrarater	Reliability: Interrater	Reliability: Measurement Error	Validity: Content Validity	Validity: Construct Validity	Validity: Concurrent Validity	Validity: Comparator Instrument	Portable and Executable in a Home Setting?
Continuous scaled physical functional performance[30]	Intra/inter-rater $r > 0.85$	+	+			+		Biceps, knee, max oxygen consumption	Yes
	Construct $r = 0.19 - 0.68$					−		Hip and shoulder strength	
Figure-8 walk[39]	$r = 0.50/0.57$					+		Gait speed	No
	$r = 0.11–0.35$					−		Step width, length, number of steps, GARS and PPT	
Fullerton Functional Fitness Test battery[47,*]	ICC 0.94–0.98	+							No
Functional reach[52,*]	ICC = 0.92; $r = 0.58/0.60/–0.24$	+				+		ADL scale, frailty scale	No
	$r = 0.24$					−		CIRS	
GAITRite mat (4.6-m mat with sensor)[46,*]	ICC = 0.91	+							No
Gait speed (2 m to 1 km)[31,36,37,40,42,43,49,53–55,*]	Reliability $r = 0.90$/ICC = 0.94 Construct AUC > 0.70 Concurrent $r = 0.74–0.93$	+	+			+	+	SPPB, discriminating level of mobility limitation, predictive validity for ADL disability and 4 m course compared with 400-m course	Yes (short distance only)
	$r = 0.05–0.39$					−		Compared with grip strength, chair stands, tandem stand	
Gait speed (6 min)[32,37,38,51,56,*]	ICC 0.88–0.94 $r > 0.71$ Construct r > 0.61	+				+	+	Stair climb time, habitual gait and maximal gait speed, chair stand, and aerobic capacity Treadmill, predictive validity for disability	No
	$r = –0.07$ and 0.10					−		BMI, general health perceptions	
Mobility assessment tool-SF[58,*]	ICC = 0.93, r =0.59–0.96	+			+	+		Parts of MAT vs total score, SPPB, and 400-m walk	Yes
Modification scale: chair rise, stair ascent, kneel, supine rise[44]	ICC = 0.92/0.98	+	+						No
Physical capacity evaluation: walking speed, grip, etc.[35,*]	Reliability: $r > 0.94$ Construct: $r = 0.74$	+				+		Health assessment questionnaire	No
Physical performance test (4-item)[57]	$r = 0.92$						+	Mini PPT: 9 item	No
Physical performance test (7-item)[33]	$r = 0.70–0.77$					+		Lower extremity muscle force, lower extremity ROM	No
	$r = 0.43–0.69$					−		Upper extremity ROM, upper extremity muscle force	
Self-reported physical function (13 items)[31]	ICC = 0.63–0.92	+/−				+		10-ft walk, chair stand	Yes
	Kappa 0.38–0.95					−		Lifting, sitting for 1 h	
SPPB[31,37,40,50,54,*]	ICC 0.88–0.92 $r = 0.74$ AUC = 0.75	+				+	+	400-m walk, mobility disability	Yes
Sit to stand 5 times[31,34,40,42,43,49]	ICC = 0.71 $r > 0.82$	+	−			+		Discriminating level of mobility limitation Timed walk, grip strength	Yes
	$r = 0.47$					−		Self-report	
Sit to stand 10 times[48]	$r = –0.02$ to 0.11					−		Peak torque, endurance, knee extension	Yes
Sit to stand 30 sec[41,45]	Reliability: ICC 0.84–0.92, $r = 0.93$	+				+		Leg press, isokinetic leg strength, 5 chair stands	Yes
	Construct: $r = 0.71–0.83$ and $r =$ 0.21–0.52					−		Lower limb strength (knee and hip)	

(*continued on next page*)

Table 4 (*continued*)

Instrument	Outcome	Reliability			Validity				Portable and Executable in a Home Setting?
		Intrarater	Interrater	Measurement Error	Content Validity	Construct Validity	Concurrent Validity	Comparator Instrument	
Stair climb[43]	Only feasibility								No
Standing balance[40,42,43,*]	Reliability: weighted kappa 0.29; Construct: AUC 0.62–0.67	−				−		Discriminating extent of mobility limitation	Yes
Tandem-stand[36,42]	Kappa $<$ 0.40					−		Single leg stand, gait speed, chair stands, grip strength; not able to discriminate level of mobility limitation	Yes
Timed up and go[42,52,56,*]	ICC 0.56–0.97 $r > 0.70$	+/−				+		Discriminating level of mobility limitation, ADL, and frailty scale	Yes
	$r = 0.38$					−		CIRS	
Trunk flexibility[43]	Only feasibility								No
UEPB (Hand test, hand signature, functional reach) and LEPB[55,72]	AUC = 0.73–0.85 $r = 0.57$					+		Discriminating low physical function, mobility limitation, ADL disability; UEPB with LEPB	No
	$r = 0.19$–0.44					−		Self-report with UEPB	
3-D accelerometer[68]	ICC $<$ 0.70 single walk, mean of two walk ICC $>$ 0.70	−/+	−/+			+		TUG	No
	r = 0.40-0.53					−		Basic ADL, grip strength	

+, high reliability (ICC, weighted Kappa ≥ 0.70 or Pearson correlation ≥ 0.80), high construct validity (correlation between constructs ≥ 0.50) or high concurrent validity (Pearson/Spearman correlation or AUC ≥ 0.70; (−), low reliability (ICC, weighted Kappa < 0.70 or correlation < 0.80) or low validity (Pearson/Spearman correlation or AUC < 0.70); ADL, activities of daily living; AUC, area under the curve; BMI, body mass index; CIRS, cumulative illness rating scale; GARS, modified gait abnormality rating scale; ICC, intraclass correlation coefficient; LEPB, lower extremity performance battery; MAT, mobility assessment tool; PPT, physical performance test; ROM, range of motion; SF, Short Form; SPPB, short physical performance battery; TUG, timed-up-and-go; UEPB, upper extremity performance battery.

*Some studies only mention "test-retest" but do not clarify interrater or intrarater.

beneficial if tools are feasible to apply in a general practitioner practice or in a home setting. With regard to muscle mass, many tools are available in clinical practice, but no well-validated and reliable tools are available for measurements of muscle mass in a home setting. BIA and the use of anthropometrics (such as calf circumference and skin-fold thickness measurements) were all found to be feasible for a home setting because the required equipment is portable. From those, BIA showed better evidence for validity, yet its validity is highly dependent on age, sex, and cultural influences,[14] because, for example, edema, diuretics, and prosthesis hamper BIA measurements. Furthermore, it is likely that the use of different reference populations and cut points for muscle mass have large effects on the outcome.[8,65,66] In a review on field and laboratory techniques to assess muscle mass, it is stated that 3-C model and 4-C methods may be required and are usually recommended in older people, but BIA is put forward as the best option for field measurements.[67] Ultrasound is a promising alternative to the BIA; however, for ultrasound to become a feasible and reliable alternative for BIA, work is warranted.

Critical Appraisal of Methodology

With regard to the methodology of this review, some aspects should be addressed. Most studies scored "fair" because they did not describe how missing items were handled. Studies were excluded when they had a sample size of less than 30, which may have narrowed our results. In addition, a correlation of 0.69 is classified as low validity, whereas a correlation of 0.71 is classified as high, despite the marginal difference. For muscle strength and performance, gold standards are not available, which hampers assessment of proper concurrent validity. It should be taken into account that for some tools, only one study on validity and reliability is available.

Conclusions and Implications of Key Findings

For a valid and reliable screening or diagnosis of sarcopenia, first one has to agree on the combination of the parameters by which sarcopenia is measured. In this article, the European Working Group on Sarcopenia in Older People criteria were chosen, including muscle mass, muscle strength, and physical performance. Gold standards used for the assessment of muscle mass were MRI, CT, and a 4-C model. A valid and reliable tool for muscle strength is the HHD; the SPPB and gait speed have good measurement properties with regard to the assessment of physical performance.

To measure muscle mass, strength and physical performance in a general practitioner practice or home-setting, BIA, HHD and gait speed over a short distance or the SPPB can be used, since those measure are transportable and executable in those specific settings. However, because the validity of BIA is not optimal, it is debatable to measure only muscle strength and physical performance for a first screening, and when scores on these parameters are below normal, *further assessment* of muscle mass by, for example, DXA, as a more valid alternative for the measurement of muscle mass, could be used. The use of a combination of tools to measure muscle mass, strength, and physical performance for the screening and diagnosis of sarcopenia in community-dwelling older people, as well as predictive value, needs further evaluation.

Acknowledgments

The authors thank Dr. C.B. Terwee from the VU University Medical Center for her support with the correct interpretation and use of the COSMIN checklist. Furthermore, they thank M.J.H. Tilly for assisting with gathering the numerous references.

References

1. Rosenberg IH. Summary comments. Am J Clin Nutr 1989;50:1231–1233.
2. Cederholm TE, Bauer JM, Boirie Y, et al. Toward a definition of sarcopenia. Clin Geriatr Med 2011;27:341–353.
3. Morley JE, Abbatecola AM, Argiles JM, et al. Sarcopenia with limited mobility: An international consensus. J Am Med Dir Assoc 2011;12:403–409.
4. Fielding RA, Vellas B, Evans WJ, et al. Sarcopenia: An undiagnosed condition in older adults. Current consensus definition: Prevalence, etiology, and consequences. International Working Group on Sarcopenia. J Am Med Dir Assoc 2011;12:249–256.
5. Landi F, Liperoti R, Russo A, et al. Sarcopenia as a risk factor for falls in elderly individuals: Results from the ilSIRENTE study. Clin Nutr 2012;31:652–658.
6. Cruz-Jentoft AJ, Landi F, Topinková E, et al. Understanding sarcopenia as a geriatric syndrome. Curr Opin Clin Nutr Metab Care 2010;13:1–7.
7. Narici MV, Maffulli N. Sarcopenia: Characteristics, mechanisms and functional significance. Br Med Bull 2010;95:139–159.
8. Rolland Y, Czerwinski S, Abellan Van Kan G, et al. Sarcopenia: Its assessment, etiology, pathogenesis, consequences and future perspectives. J Nutr Health Aging 2008;12:433–450.
9. Roubenoff R, Hughes VA. Sarcopenia: Current concepts. J Gerontol A Biol Sci Med Sci 2000;55:M716–M724.
10. Cruz-Jentoft AJ, Baeyens JP, Bauer JM, et al. Sarcopenia: European consensus on definition and diagnosis: Report of the European working group on sarcopenia in older people. Age Ageing 2010;39:412–423.
11. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international delphi study. Qual Life Res 2010;19:539–549.
12. Terwee CB, Bot SD, De Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007; 60:34–42.
13. Sipilä S, Suominen H. Quantitative ultrasonography of muscle: Detection of adaptations to training in elderly women. Arch Phys Med Rehabil 1996;77: 1173–1178.
14. Genton L, Karsegard VL, Kyle UG, et al. Comparison of four bioelectrical impedance analysis formulas in healthy elderly subjects. Gerontology 2001;47: 315–323.
15. Rolland Y, Lauwers-Cances V, Cournot M, et al. Sarcopenia, calf circumference, and physical function of elderly women: A cross-sectional study. J Am Geriatr Soc 2003;51:1120–1124.
16. Haapala I, Hirvonen A, Niskanen L, et al. Anthropometry, bioelectrical impedance and dual-energy x-ray absorptiometry in the assessment of body composition in elderly Finnish women. Clin Physiol Funct Imaging 2002;22: 383–391.
17. Chen Z, Wang Z, Lohman T, et al. Dual-energy x-ray absorptiometry is a valid tool for assessing skeletal muscle mass in older women. J Nutr 2007;137: 2775–2780.
18. Abizanda P, Navarro JL, García-Tomás MI, et al. Validity and usefulness of hand-held dynamometry for measuring muscle strength in community-dwelling older persons. Arch Gerontol Geriatr 2011;54:21–27.
19. Bohannon RW. Alternatives for measuring knee extension strength of the elderly at home. Clin Rehabil 1998;12:434–440.
20. Desrosiers J, Hebert R, Bravo G, et al. Comparison of the Jamar dynamometer and the Martin vigorimeter for grip strength measurements in a healthy elderly population. Scand J Rehabil Med 1995;27:137–143.
21. Guerra RS, Amaral TF. Comparison of hand dynamometers in elderly people. J Nutr Health Aging 2009;13:907–912.
22. Ottenbacher KJ, Branch LG, Ray L, et al. The reliability of upper- and lower-extremity strength testing in a community survey of older adults. Arch Phys Med Rehabil 2002;83:1423–1427.
23. Payette H, Hanusaik N, Boutier V, et al. Muscle strength and functional mobility in relation to lean body mass in free-living frail elderly women. Eur J Clin Nutr 1998;52:45–53.
24. Reed RL, Den Hartog R, Yochum K, et al. A comparison of hand-held isometric strength measurement with isokinetic muscle strength measurement in the elderly. J Am Geriatr Soc 1993;41:53–56.
25. Wang CY, Olson SL, Protas EJ. Test-retest strength reliability: Hand-held dynamometry in community-dwelling elderly fallers. Arch Phys Med Rehabil 2002;83:811–815.
26. Wang CY, Chen LY. Grip strength in older adults: Test-retest reliability and cutoff for subjective weakness of using the hands in heavy tasks. Arch Phys Med Rehabil 2010;91:1747–1751.
27. LeBrasseur NK, Bhasin S, Miciek R, et al. Tests of muscle strength and physical function: Reliability and discrimination of performance in younger and older men and older men with mobility limitations. J Am Geriatr Soc 2008;56: 2118–2123.
28. Menz HB, Tiedemann A, Kwan MM, et al. Reliability of clinical tests of foot and ankle characteristics in older people. J Am Podiatr Med Assoc 2003;93: 380–387.
29. Rydwik E, Karlsson C, Frändin K, et al. Muscle strength testing with one repetition maximum in the arm/shoulder for people aged 75 + - test-retest reliability. Clin Rehabil 2007;21:258–265.
30. Cress ME, Buchner DM, Questad KA, et al. Continuous-scale physical functional performance in healthy older adults: A validation study. Arch Phys Med Rehabil 1996;77:1243–1250.
31. 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;53:M295–M300.
32. Bean JF, Kiely DK, Leveille SG, et al. The 6-minute walk test in mobility-limited elders. J Gerontol A Biol Sci Med Sci 2002;57:M751–M756.
33. Beissner KL, Collins JE, Holmes H. Muscle force and range of motion as predictors of function in older adults. Phys Ther 2000;80:556–563.
34. Bohannon RW, Shove ME, Barreca SR, et al. Five-repetition sit-to-stand test performance by community-dwelling adults: A preliminary investigation of times, determinants, and relationship with self-reported physical performance. Isokinet Exerc Sci 2007;15:77–81.
35. Daltroy LH, Phillips CB, Eaton HM, et al. Objectively measuring physical ability in elderly persons: The physical capacity evaluation. Am J Public Health 1995; 85:558–560.
36. Guralnik JM, Seeman TE, Tinetti ME, et al. Validation and use of performance measures of functioning in a non-disabled older population: Macarthur studies of successful aging. Aging Clin Exp Res 1994;6:410–419.
37. Guralnik JM, Ferrucci L, Pieper CF, et al. Lower extremity function and subsequent disability: Consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. J Gerontol A Biol Sci Med Sci 2000;55:M221–M231.
38. Harada ND, Chiu V, Stewart AL. Mobility-related function in older adults: Assessment with a 6-minute walk test. Arch Phys Med Rehabil 1999;80: 837–841.
39. Hess RJ, Brach JS, Piva SR, et al. Walking skill can be assessed in older adults: Validity of the figure-of-8 walk test. Phys Ther 2010;90:89–99.
40. Hoeymans N, Wouters ER, Feskens EJ, et al. Reproducibility of performance-based and self-reported measures of functional status. J Gerontol A Biol Sci Med Sci 1997;52:M363–M368.
41. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. Res Q Exerc Sport 1999;70: 113–119.
42. Kim MJ, Seino S, Kim MK, et al. Validation of lower extremity performance tests for determining the mobility limitation levels in community-dwelling older women. Aging Clin Exp Res 2009;21:437–444.
43. Malmberg JJ, Miilunpalo SI, Vuori IM, et al. A health-related fitness and functional performance test battery for middle-aged and older adults: Feasibility and health-related content validity. Arch Phys Med Rehabil 2002; 83:666–677.
44. Manini TM, Cook SB, VanArnam T, et al. Evaluating task modification as an objective measure of functional limitation: Repeatability and comparability. J Gerontol A Biol Sci Med Sci 2006;61:718–725.
45. McCarthy EK, Horvat MA, Holtsberg PA, et al. Repeated chair stands as a measure of lower limb strength in sexagenarian women. J Gerontol A Biol Sci Med Sci 2004;59:1207–1212.
46. Menz HB, Latt MD, Tiedemann A, et al. Reliability of the GAITRite walkway system for the quantification of temporo-spatial parameters of gait in young and older people. Gait Posture 2004;20:20–25.
47. Miotto JM, Chodzko-Zajko WJ, Reich JL, et al. Reliability and validity of the Fullerton functional fitness test: An independent replication study. J Aging Phys Activity 1999;7:339–353.
48. Netz Y, Ayalon M, Dunsky A, et al. 'The multiple-sit-to-stand' field test for older adults: What does it measure? Gerontology 2004;50:121–126.
49. Ostchega Y, Harris TB, Hirsch R, et al. Reliability and prevalence of physical performance examination assessing mobility and balance in older persons in the us: Data from the third national health and nutrition examination survey. J Am Geriatr Soc 2000;48:1136–1141.
50. Ostir GV, Volpato S, Fried LP, et al. Reliability and sensitivity to change assessed for a summary measure of lower body function: Results from the women's health and aging study. J Clin Epidemiol 2002;55:916–921.
51. Rikli RE, Jones CJ. The reliability and validity of a 6-minute walk test as a measure of physical endurance in older adults. J Aging Phys Activity 1998;6: 363–375.
52. Rockwood K, Awalt E, Carver D, et al. Feasibility and measurement properties of the functional reach and the timed up and go tests in the canadian study of health and aging. J Gerontol A Biol Sci Med Sci 2000;55: M70–M73.
53. Rolland YM, Cesari M, Miller ME, et al. Reliability of the 400-m usual-pace walk test as an assessment of mobility limitation in older adults. J Am Geriatr Soc 2004;52:972–976.
54. Sayers SP, Guralnik JM, Newman AB, et al. Concordance and discordance between two measures of lower extremity function: 400 meter self-paced walk and sppb. Aging Clin Exp Res 2006;18:100–106.
55. Seino S, Kim MJ, Yabushita N, et al. Is a composite score of physical performance measures more useful than usual gait speed alone in assessing functional status? Arch Gerontol Geriatr 2012;55:392–398.
56. Steffen TM, Hacker TA, Mollinger L. Age- and gender-related test performance in community-dwelling elderly people: Six-minute walk test, Berg balance scale, timed up & go test, and gait speeds. Phys Ther 2002;82: 128–137.

57. Wilkins CH, Roe CM, Morris JC. A brief clinical tool to assess physical function: The mini-physical performance test. Arch Gerontol Geriatr 2010; 50:96–100.
58. Rejeski WJ, Ip EH, Marsh AP, et al. Development and validation of a video-animated tool for assessing mobility. J Gerontol A Biol Sci Med Sci 2010;65: 664–671.
59. Reeves ND, Maganaris CN, Narici MV. Ultrasonographic assessment of human skeletal muscle size. Eur J Appl Physiol 2004;91:116–118.
60. Roberts HC, Denison HJ, Martin HJ, et al. A review of the measurement of grip strength in clinical and epidemiological studies: Towards a standardised approach. Age Ageing 2011;40:423–429.
61. Stark T, Walker B, Phillips JK, et al. Hand-held dynamometry correlation with the gold standard isokinetic dynamometry: A systematic review. PM R 2011;3: 472–479.
62. Samuel D, Rowe P. An investigation of the association between grip strength and hip and knee joint moments in older adults. Arch Gerontol Geriatr 2012; 54:357–360.
63. Vasunilashorn S, Coppin AK, Patel KV, et al. Use of the short physical performance battery score to predict loss of ability to walk 400 meters: Analysis from the inchianti study. J Gerontol A Biol Sci Med Sci 2009;64:223–229.
64. Cooper R, Kuh D, Hardy R. Objectively measured physical capability levels and mortality: Systematic review and meta-analysis. BMJ 2010;341:c4467.
65. Bijlsma AY, Meskers CGM, Ling CHY, et al. Defining sarcopenia: The impact of different diagnostic criteria on the prevalence of sarcopenia in a large middle aged cohort. Age (Dordr) 2012 Feb 8. [Epub ahead of print].
66. Di Monaco M, Vallero F, Di Monaco R, et al. Prevalence of sarcopenia and its association with osteoporosis in 313 older women following a hip fracture. Arch Gerontol Geriatr 2011;52:71–74.
67. Norgan NG. Laboratory and field measurements of body composition. Public Health Nutr 2005;8:1108–1122.
68. Bautmans I, Jansen B, Van Keymolen B, et al. Reliability and clinical correlates of 3d-accelerometry based gait analysis outcomes according to age and fall-risk. Gait Posture 2011;33:366–372.
69. Manor B, Topp R, Page P. Validity and reliability of measurements of elbow flexion strength obtained from older adults using elastic bands. J Geriatr Phys Ther 2006;29:16–19.
70. Spink MJ, Fotoohabadi MR, Menz HB. Foot and ankle strength assessment using hand-held dynamometry: Reliability and age-related differences. Gerontology 2010;56:525–532.
71. Suzuki T, Bean JF, Fielding RA. Muscle power of the ankle flexors predicts functional performance in community-dwelling older women. J Am Geriatr Soc 2001;49:1161–1167.
72. Hazuda HP, Dhanda R, Owen SV, et al. Development and validation of a performance-based measure of upper extremity functional limitation. Aging Clin Exp Res 2005;17:394–401.
73. Aleman-Mateo H, Rush E, Esparza-Romero J, et al. Prediction of fat-free mass by bioelectrical impedance analysis in older adults from developing countries: A cross-validation study using the deuterium dilution method. J Nutr Health Aging 2010;14:418–426.
74. Bertoli S, Battezzati A, Testolin G, et al. Evaluation of air-displacement plethysmography and bioelectrical impedance analysis vs dual-energy x-ray absorptiometry for the assessment of fat-free mass in elderly subjects. Eur J Clin Nutr 2008;62:1282–1286.
75. Deurenberg P, Van der Kooij K, Evers P, et al. Assessment of body composition by bioelectrical impedance in a population aged greater than 60 y. Am J Clin Nutr 1990;51:3–6.
76. Dey DK, Bosaeus I, Lissner L, et al. Body composition estimated by bioelectrical impedance in the swedish elderly: Development of population-based prediction equation and reference values of fat-free mass and body fat for 70- and 75-y olds. Eur J Clin Nutr 2003;57:909–916.
77. Dey DK, Bosaeus I. Comparison of bioelectrical impedance prediction equations for fat-free mass in a population-based sample of 75 y olds: The NORA study. Nutrition 2003;19:858–864.
78. Ling CH, De Craen AJ, Slagboom PE, et al. Accuracy of direct segmental multi-frequency bioimpedance analysis in the assessment of total body and segmental body composition in middle-aged adult population. Clin Nutr 2011; 30:610–615.
79. Mitchell SJ, Kirkpatrick CM, Le Couteur DG, et al. Estimation of lean body weight in older community-dwelling men. Br J Clin Pharmacol 2010;69: 118–127.
80. Rech CR, Cordeiro BA, Petroski EL, et al. Validation of bioelectrical impedance for the prediction of fat-free mass in brazilian elderly subjects. Brazil Arch Endocrinol Metab 2008;52:1163–1171.
81. Roubenoff R, Baumgartner RN, Harris TB, et al. Application of bioelectrical impedance analysis to elderly populations. J Gerontol A Biol Sci Med Sci 1997; 52:M129–M136.
82. Valencia ME, Aleman-Mateo H, Salazar G, et al. Body composition by hydrometry (deuterium oxide dilution) and bioelectrical impedance in subjects aged >60 y from rural regions of Cuba, Chile and Mexico. Int J Obes Metab Dis 2003; 27:848–855.
83. Visser M, Fuerst T, Lang T, et al. Validity of fan-beam dual-energy x-ray absorptiometry for measuring fat-free mass and leg muscle mass: Health, aging, and body composition study: Dual-energy x-ray absorptiometry and body composition working group. J Appl Physiol 1999;87:1513–1520.