

The reliability and validity of ultrasound to quantify muscles in older adults: a systematic review

Willemke Nijholt^{1,2*}, Aldo Scafoglieri³, Harriët Jager-Wittenaar^{1,4}, Johannes S.M. Hobbelen^{1,5} & Cees P. van der Schans^{1,2,6}

¹Research Group Healthy Ageing, Allied Health Care and Nursing, Hanze University of Applied Sciences, Groningen, The Netherlands; ²Department of Rehabilitation Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; ³Faculty of Medicine and Pharmacy, Vrije Universiteit Brussel, Brussels, Belgium; ⁴Department of Oral and Maxillofacial Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; ⁵Department of General Practice and Elderly Care Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; ⁶Department of Health Psychology Research, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Abstract

This review evaluates the reliability and validity of ultrasound to quantify muscles in older adults. The databases PubMed, Cochrane, and Cumulative Index to Nursing and Allied Health Literature were systematically searched for studies. In 17 studies, the reliability ($n = 13$) and validity ($n = 8$) of ultrasound to quantify muscles in community-dwelling older adults (≥ 60 years) or a clinical population were evaluated. Four out of 13 reliability studies investigated both intra-rater and inter-rater reliability. Intraclass correlation coefficient (ICC) scores for reliability ranged from -0.26 to 1.00 . The highest ICC scores were found for the vastus lateralis, rectus femoris, upper arm anterior, and the trunk (ICC = 0.72 to 1.000). All included validity studies found ICC scores ranging from 0.92 to 0.999 . Two studies describing the validity of ultrasound to predict lean body mass showed good validity as compared with dual-energy X-ray absorptiometry ($r^2 = 0.92$ to 0.96). This systematic review shows that ultrasound is a reliable and valid tool for the assessment of muscle size in older adults. More high-quality research is required to confirm these findings in both clinical and healthy populations. Furthermore, ultrasound assessment of small muscles needs further evaluation. Ultrasound to predict lean body mass is feasible; however, future research is required to validate prediction equations in older adults with varying function and health.

Keywords Ultrasonography; Muscles; Sarcopenia; Body composition; Muscular atrophy

Received: 2 January 2017; Revised: 27 February 2017; Accepted: 20 March 2017

*Correspondence to: Willemke Nijholt, Research Group Healthy Ageing, Allied Health Care and Nursing, Hanze University of Applied Sciences, School of Health Care Studies, Eyssonusplein 18, 9714 CE Groningen, The Netherlands. Phone number: +31.6.10 94 12 90, Email: w.nijholt@pl.hanze.nl

Introduction

Globally, the proportion of older people within the worldwide population is increasing. It is estimated that, in 2050, approximately 400 million people will be aged 80 years and older.¹ During ageing, body composition changes with a 1–2% loss of muscle mass per year after the age of 50.^{2–5} This loss, together with impaired physical performance, is referred to as sarcopenia.^{2,6} Sarcopenia is associated with development of functional disability, such as slow walking speed, and may lead to a lower quality of life and dependency.^{2,7–11} The prevalence of sarcopenia in healthy older adults (mean age (SD) = 74.4 (3.2) years) is estimated to be between 0% and

15%.¹² In community-dwelling older adults with mobility problems (mean age (SD) = 80.5 (7.0) years), the prevalence is higher, with estimates between 2% and 34%. Differences in cut-off values, operational criteria, and differences in assessment methods may possibly explain the large variation in prevalence rates. For instance, prevalence rates of 33% to 34% in community-dwelling older adults of sarcopenia were found when only low muscle mass or low handgrip strength was used as diagnostic criteria for sarcopenia. When applying the diagnostic criteria of the European Working Group on Sarcopenia in Older People, the prevalence of sarcopenia in community-dwelling older adults with mobility problems is approximately 25%.¹²

Muscle mass depletion is an important characteristic of sarcopenia. Traditionally, computed tomography (CT) and magnetic resonance imaging (MRI) are considered gold standards for assessing muscle mass.^{13,14} However, both methods are not feasible for the assessment of muscle mass in daily practice. CT uses ionizing radiation and therefore is not performed on a routine basis, and MRI is expensive and has limited availability. Dual-energy X-ray absorptiometry (DXA) is also a widely used technique to determine muscle mass in a research setting; however, DXA also has limited availability. Ultrasound is potentially a good alternative for CT, MRI, and DXA, as it is a non-ionizing imaging technique that provides dynamic assessment of soft tissue structures, is portable, and also highly accessible. Furthermore, ultrasound has been shown to be reliable for assessing selected foot structures, which suggests that ultrasound has the potential to accurately assess (small) muscle groups.¹⁵

Currently, it is difficult to diagnose sarcopenia in daily practice because there is a lack of valid and/or feasible tools for the assessment of muscle mass. Ultrasound might play an important role in the diagnosis of sarcopenia, because it may offer an objective measure of the amount of muscle mass. Previous reviews concluded that ultrasound is valid for measuring muscle size in a younger population compared with measurement instruments such as MRI and CT.^{16,17} Ultrasound is also a reliable measure of muscle size in healthy individuals.¹⁸ However, until now, it is unclear whether ultrasound is a reliable and valid technique to assess muscle size in older adults. Furthermore, the use of ultrasound to predict whole body muscle mass in older adults has not been previously reviewed. Therefore, this systematic review aims to evaluate the reliability and validity of ultrasound for assessing muscle size in older adults. Moreover, this study evaluates the validity of ultrasound-derived equations for the prediction of muscle mass in older adults.

Methods

We systematically searched the PubMed, Cochrane, and Cumulative Index to Nursing and Allied Health Literature databases for studies in English, German, and Dutch. Studies were searched up until 20 January 2016. Outcomes of interest were conclusions about the reliability, concurrent validity, or feasibility of ultrasound to quantify muscles. In the search strategy, a combination of terms related to sarcopenia, older adults, and ultrasonography was used: (i) sarcopenia: muscular atroph*, muscle atroph*, muscle mass*, muscle size*, muscle diameter*, muscle volume*, muscle thickness*, muscle wasting; (ii) older adults: aged, aging, older adult, elder*, older person*, older people, senior*, ageing; (iii) ultrasonography: ultrasound, ultraso* imaging, medical sonography, echography. The complete

search strategy is available from the author. In addition to the search in the databases of Cumulative Index to Nursing and Allied Health Literature, PubMed, and Cochrane, other relevant studies were selected using backward citation tracking.

Study eligibility criteria

Studies evaluating the reliability, validity, and/or feasibility of ultrasound to assess muscle mass of the limbs and abdomen in the older population (mean age ≥ 60 years, or inclusion criteria ≥ 60 years and older) were eligible for inclusion in this study. Animal studies, studies using cadaver specimens, and (systematic) reviews were excluded.

Study appraisal and synthesis methods

Refworks (ProQuest LLC 2016) was used to insert the search hits from the databases. After deleting duplicates, titles and abstracts were independently screened by two authors (W. N. and A. S.). On the basis of the inclusion and exclusion criteria, studies were independently scored as relevant or not relevant. Disagreements regarding the relevance of the studies were solved by consensus. Both assessors (W. N. and A. S.) subsequently and independently assessed the included full text studies. The methodological quality of the included studies was assessed using two checklists: one checklist for the reliability and validity studies¹⁷ and one checklist for the studies on the validity of ultrasound-derived prediction equations.¹⁹ The methodological quality of the reliability and validity studies was assessed using a checklist developed by Pretorius and Keating.¹⁷ The checklist contains 10 items focusing on the reliability and validity of ultrasound to measure muscles (Appendix S1). A higher score signified higher methodological quality.¹⁷ The methodological quality of the validity of ultrasound-derived prediction equations was assessed by the consensus-based standards for the selection of health status measurement instruments (COSMIN) checklist. The COSMIN checklist consists of nine boxes; each box entails one measurement property, for example, reliability and criterion validity. Each box consists of 5 to 18 criteria, which can be used to assess methodological quality. Eventually, a quality score was determined by taking the lowest rating of each criterion in a box. The quality score was defined to be poor, fair, good, or excellent.¹⁹

In all of the steps of the selection procedure and during assessment of methodological quality, agreement between the two independent reviewers was calculated using the Cohen's kappa coefficient.²⁰ A score of <0.40 is regarded as poor, 0.40 – 0.75 as fair to good, and a score >0.75 as an excellent agreement between both observers.²¹

Results

An overview of the process of study selection is shown in *Figure 1*. After screening by title and abstract, 50 studies were assessed for eligibility. The inter-rater agreement regarding title and abstract screening was fair to good (Cohen's kappa = 0.60 (95% confidence interval = 0.48–0.72)). From the 50 studies, 16 studies fulfilled the eligibility criteria. Inter-rater agreement of assessment of full text studies was fair to good (Cohen's kappa = 0.68 (95% confidence interval = 0.48–0.88)). The included studies were categorized as reliability studies ($n = 13$), validity studies ($n = 6$), and ultrasound-derived prediction equation studies ($n = 2$).

Methodological quality

The quality of the included reliability and validity studies was good, with quality scores ranging between 7 and 10. Overall, more than 7 out of the 10 questions scored 'yes' for all of the studies. The most consistent shortcomings were missing data on the composition of the sample and insufficient information on the scanning procedure of ultrasound (Table 1). The quality of the two ultrasound-derived prediction equations scored as good. The reference used can be considered as a reasonable criterion method for the assessment of muscle mass, both studies used good sample sizes (both studies $n = 77$), and appropriate statistical analyses were performed in the studies.

Transducer	A device that generates and receives the ultrasound waves.
Linear transducer	A transducer in which the width of the image is the same at all tissue levels. Therefore, a linear transducer has good near-field resolution and is most often used for small, superficial structures, for example, muscles.
Curved transducer	A transducer in which the width of the image increases with deeper penetration. Therefore, a curved transducer is most often used for deep scanning.
Scanning plane	The direction in which the scan is generated. The two scanning planes used in this manuscript are (i) sagittal, which refers to longitudinal orientation and (ii) transverse, which refers to the axial orientation.
Muscle dimension	The dimension in which the muscle is measured; thickness (in millimeter or centimeter), cross-sectional area (in cm^2) or volume (in cm^3).

Reliability

As listed in Table 2, 13 studies investigated the reliability of ultrasound. Of these, four studies reported data on both the intra-rater and the inter-rater reliability.^{22–25} Eight studies involved healthy older adults,^{24–31} two studies involved

stroke patients,^{18,22} and three studies involved patients with chronic conditions, such as chronic obstructive pulmonary disease, osteoarthritis, and coronary artery disease.^{23,32,33} Two studies explicitly stated that the markings on the skin were removed prior to the second scan, to prevent bias in the measurement.^{18,25} Out of the 13 studies, four studies used a curved-array transducer.^{23,25,30,34} Included studies reported different outcome measures; one study assessed muscle volume,²⁴ three studies assessed cross-sectional area,^{23,27,29} and nine studies assessed muscle thickness (MT).^{18,22,25,26,28,30–33}

Intra-rater reliability

The intra-rater reliability of ultrasound was investigated in 13 studies. The majority of the studies measured the muscle in the transverse plane.^{18,23,25–27,29–31,33} The interval between repeated measurements varied from several minutes^{18,30} to 14 days.^{23,28} Nine out of 13 studies evaluated thigh muscles.^{23,24,26–29,31–33,35} Calf muscles,^{18,22,28} abdominal muscles,^{18,30} and spinal muscles²⁵ were also evaluated. Overall, reliability estimates ranged from -0.26 to 1.00 . The highest intraclass correlation coefficient (ICC) scores were found for the vastus lateralis (ICC = 0.852 to 0.999), the rectus femoris (ICC = 0.72 to 0.997), the upper arm anterior (ICC = 0.81 to 0.99), and the trunk (0.73 to 1.00).

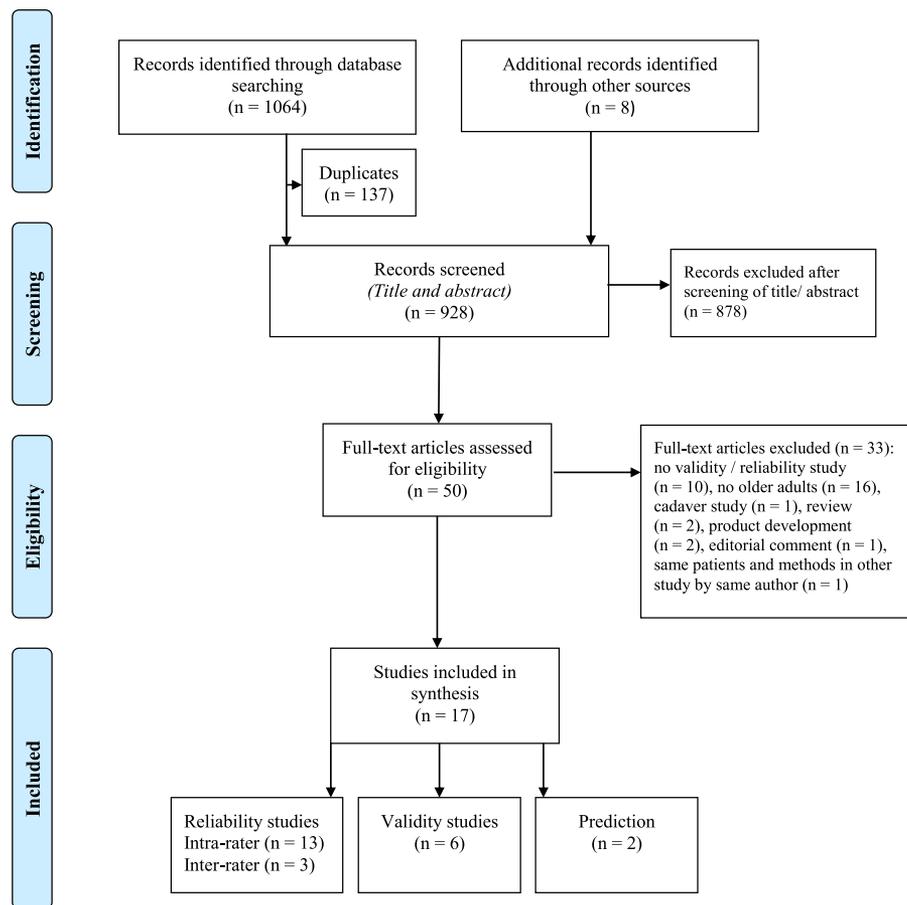
Inter-rater reliability

Four studies investigated both intra-rater and inter-rater reliability.^{22–25} One study assessed both measurement and procedural reliability. Reliability estimates for measurement reliability was higher than procedural reliability (ICC = 0.98 and ICC = 0.86, respectively).²⁵ The four studies evaluated different muscles: medial gastrocnemius,²² rectus femoris,^{23,24} and the lumbar multifidus muscle.²⁵ Two studies measured the muscle in the transverse plane.^{23,25} Reliability estimates ranged from 0.88 to 0.998.

Validity

All of the included studies evaluated concurrent validity with DXA,³⁶ MRI,^{24,29} CT,^{33,37} or ultrasound²³ used as criterion methods (Table 3). The same construct was measured with ultrasound and the reference methods, except for one study, which compared muscle size with body composition parameters.³⁶ All of the studies evaluated thigh muscles with a linear transducer. Only one study measured thigh muscle volume in the sagittal plane.²⁴ The other studies assessed muscle thickness^{33,36,37} or cross-sectional area^{23,29,37} in the transverse plane. All studies found that ultrasound is valid for the assessment of muscles, with ICC scores ranging from 0.92 to 0.999,^{23,24,29,33,36} and $r = 0.761$ to $r = 0.911$.³⁷

Figure 1 Preferred reporting items for systematic reviews and meta-analyses flowchart showing selection procedure.



Validity of ultrasound-derived prediction equations

Two studies evaluated the validity of ultrasound to predict muscle mass in older adults as compared with DXA.^{38,39} One study specifically focused on the prediction of leg muscle mass. That study was conducted with 52 healthy adults of which 22 were men (mean age 62.1 ± 8.6 years) and 30 were women (mean age 66.3 ± 5.9 years). The proposed prediction equation included the sum of four MT: thigh anterior and posterior and lower leg anterior and posterior (leg muscle mass = $0.01464 \times (MT_{\text{sum}} \times \text{length of segment}) - 2.767$). The results indicated a good validity of ultrasound for predicting leg muscle mass compared with DXA ($r^2 = 0.96$).³⁸ The second study was conducted in 77 healthy older adults (mean age = 64.8 ± 7.2 years). Two prediction equations were proposed in that study. Equation 1 (muscle mass = (sex (female = 0, male = 1) \times 7.217) + (MT_{thigh anterior} \times 1.985) + (MT_{thigh posterior} \times 2.355) + (MT_{lower leg anterior} \times 3.633) + (MT_{lower leg posterior} \times 2.670) $-$ 6.759) included MT of the thigh (anterior and posterior) and the lower leg (anterior and posterior). The results showed good validity of the ultrasound-derived prediction equation

($r^2 = 0.929$, standard error of the estimate = 2.5 kg). Equation 2 utilized the product of MT and limb length (LL) to predict muscle mass.

In this equation, the following sites were included: upper arm anterior, thigh anterior, thigh posterior, and lower leg posterior (muscle mass = (sex (female = 0, male = 1) \times 5.233) + ((MT \times LL)_{upper arm anterior} \times 0.006630) + ((MT \times LL)_{thigh anterior} \times 0.05153) + ((MT \times LL)_{thigh posterior} \times 0.05579) + ((MT \times LL)_{lower leg posterior} \times 0.07097) + 1.774). The validity of Equation 2 was good compared with DXA ($r^2 = 0.955$, standard error of the estimate = 2.0 kg).³⁹

Discussion

The main finding of this systematic review is that ultrasound is a reliable and valid tool for the assessment of muscle size in older adults. However, the validity of ultrasound-derived prediction equations for the estimation of muscle mass in older adults cannot be established, because only two studies examined the validity of ultrasound-derived prediction equations in older adults.

Table 1 Quality assessment of the included studies

Study	Blinding		Sample		Reproducibility				Study procedures				Score
	Blind assessor	Data ≥80% of cohort reported	Representative sample	Sufficient information reported	Data analyses clearly defined	Proper time frame	Instructions on muscle state	Scanning point clearly described	Minimization of contact pressure	Perpendicular position of transducer			
Agyapong <i>et al.</i> , 2014 ²⁶	x	-	l	x	x	x	x	x	-	x	x	7	
Bemben, 2002 ²⁷	x	x	x	x	x	-	x	x	x	x	x	9	
Berger <i>et al.</i> , 2015 ³⁶	x	x	l	x	x	x	x	x	x	x	x	9	
Cho <i>et al.</i> , 2014 ²²	x	x	x	x	x	x	x	x	x	x	x	10	
English <i>et al.</i> , 2011 ¹⁸	x	-	x	x	x	x	x	x	x	x	x	9	
Hammond <i>et al.</i> , 2014 ²³	x	-	x	x	x	x	x	x	x	x	x	9	
MacGillivray <i>et al.</i> , 2009 ²⁴	x	x	x	-	x	x	x	x	-	x	x	8	
Raj <i>et al.</i> , 2012 ²⁸	x	x	x	x	x	x	-	x	x	-	x	8	
Reeves <i>et al.</i> , 2004 ²⁹	x	x	x	x	x	x	x	x	x	x	x	9	
Sions <i>et al.</i> , 2014 ²⁵	x	x	x	x	x	x	x	x	x	x	x	10	
Sipila and Suominen, 1993 ³⁷	x	x	x	x	x	-	x	x	-	-	-	7	
Staepli <i>et al.</i> , 2010 ³²	x	x	x	x	x	x	x	x	x	x	x	10	
Stetts <i>et al.</i> , 2009 ³⁰	x	x	x	x	x	-	x	x	-	x	x	8	
Strasser <i>et al.</i> , 2013 ³¹	x	x	x	x	x	-	x	x	x	x	x	9	
Thomae <i>et al.</i> , 2012 ³³	x	-	x	x	x	x	x	x	x	x	x	9	

Our finding that ultrasound is reliable for the measurement of muscle size in older adults extends the conclusion of a previous review in a younger population.³⁵ We found that the reliability of ultrasound in older adults is comparable with reliability estimates found in younger adults. Furthermore, we also found that ultrasound is a reliable tool in a clinical population, for example, chronic obstructive pulmonary disease, coronary artery disease, and post-stroke and acute stroke patients, a finding in contrast to previous literature.³⁵ Reliability estimates in a clinical population appear to be equal to reliability estimates in healthy older adults. These estimates were not only similar for intra-rater and inter-rater reliability but also across different body sites. Even though we found high ICC scores for the reliability of ultrasound in clinical populations, ultrasound imaging in a clinical population may be more challenging because of increased echogenicity, that is, the reflectance of the emitted ultrasound signal, and decreased definitions of bone and muscle. Therefore, work on the feasibility of ultrasound in a clinical and ageing population is warranted.

The included studies concluded that ultrasound is a reliable tool for the assessment of muscle size. Important to acknowledge is that this conclusion is based on the assessment of large muscle groups like the *Musculus quadriceps*. Low ICC scores were found in the assessment of relatively small muscles, such as lateral forearm and lower limb muscles.¹⁸ The low ICC scores can possibly be explained by the fact that evaluating small muscles with ultrasound might be challenging because of limited spatial resolution.⁴⁰ Hence, the reliability of ultrasound for the assessment of small muscles needs further evaluation. We found that ultrasound also showed good validity for the assessment of muscle size compared with DXA, MRI, and CT. A remarkable finding of this review, however, was the lack of studies examining the validity of ultrasound-derived prediction equations for whole body muscle mass in older adults. To the best of our knowledge, one previous review was published on the validity of ultrasound for the assessment of muscles.¹⁷ That review, however, did not specifically focus on older adults and did not include studies on the validity of ultrasound for predicting lean body mass. We found that ultrasound is a valid tool to assess muscle mass in older adults in clinical practice. Only one study showed good validity of ultrasound-derived prediction equations compared with DXA. We found three other studies regarding the validity of ultrasound-derived prediction equations for muscle mass, but we excluded these from this systematic review because of the age of the study sample. Nevertheless, these studies found good validity for ultrasound-derived prediction equations for the assessment of muscle mass.^{14,41,42} This adds to the evidence for high validity of ultrasound for the prediction of muscle mass. However, for ultrasound to become a valid alternative for DXA or bioelectrical impedance analysis for diagnosing sarcopenia, cross-validation of the proposed prediction equations in older adults is warranted.

Table 2 Overview of the included reliability studies

Study	Demographics ^a	Interval in days	Transducer type	Scanning plane	Muscles	Muscle dimension	Reliability estimates ^b
Intra-rater reliability^c							
Agyapong et al., 2014 ²⁶	Community-dwelling older adults n = 32 (NR:NR) age = NR (NR)	7	Linear	Transverse	Anterior thigh muscles	Thickness	ICC = 0.88 (0.77–0.94)
Bemben, 2002 ²⁷	Postmenopausal women n = 38 (0:38) age = 58.9 (0.7)	0	Linear	Transverse	Rectus femoris Biceps brachii	CSA	SEM = 2.11 mm Rectus femoris: ICC = 0.88 (NR) SEM = 0.13 cm ² Biceps brachii: ICC = 0.99 (NR)
	Older adults n = 85 (34:51) age = 65.0 (0.4)						SEM = 0.16 cm ² Rectus femoris: ICC = 0.72 (NR) SEM = 0.12 cm ²
Inter-rater reliability^c							
Cho et al., 2014 ²²	Post-stroke patients n = 30 (15:15) age = 64.7 (5.7)	7	Linear	Sagittal	Medial gastrocnemius	Thickness	Rater 1: ICC = 0.982 (0.968–0.991) Rater 2: ICC = 0.992 (0.986–0.996)
English et al., 2011 ¹⁸	Acute stroke patients n = 29 (21:8) age = 64.0 (16.8)	0	Linear	Transverse	Anterior upper arm Posterior upper arm Lateral forearm Abdomen Anterior thigh Posterior thigh Anterior lower leg Posterior lower leg Rectus femoris	Thickness	ICCs ranging from –0.26 to 0.95 (NR) Upper LOA ranging from 2.73 to 26.01 mm. Lower LOA ranging from –2.93 to –27.69 mm. Rater 1: ICC = 0.971 (NR) LOA = –1.10 to 1.36 cm ² Rater 2: ICC = 0.942 (NR) LOA = –1.75 to 1.59 cm ² ICC = 0.997 (NR)
Hammond et al., 2014 ²³	Ambulatory COPD patients n = 17 (NR:NR) age = 66.0 (NR)	2–14	Curved	Transverse	Rectus femoris	CSA	SEM = 0.00 cm ³ Vastus lateralis:
MacGillivray et al., 2009 ²⁴	Community-dwelling older adults n = 11 (NR:NR) median age = 79	NR	Linear	Sagittal	Rectus femoris	Volume	ICC = 0.96 (0.90–0.98) for both sites 1 and 2 95% ratio LOA = 17.25% (site 1) and 10.59% (site 2) Medial gastrocnemius:
Raj et al., 2012 ²⁸	Community-dwelling older adults n = 21 (11:10) age = 68.1 (5.2)	7–14	Linear	Sagittal	Vastus lateralis Medial gastrocnemius	Thickness	

(Continues)

Table 2 (continued)

Study	Demographics ^a	Interval in days	Transducer type	Scanning plane	Muscles	Muscle dimension	Reliability estimates ^b
Intra-rater reliability^c							
Reeves <i>et al.</i> , 2004 ²⁹	Healthy adults n = 6 (3:3) age = 76.8 (3.2)	NR	Linear	Transverse	Vastus lateralis	CSA	ICC = 0.97 (0.75–0.96) 95% ratio LOA = 12.56% ICCs between 0.997 and 0.999 for scans 1 to 10 SEM = from 0.15 to 0.40 cm ² Rater 1: ICC = 0.92 (0.83–0.96)
Sions <i>et al.</i> , 2014 ²⁵	Community-dwelling older adults n = 30 (8:22) age = 71.8 (NR)	10	Curved	Transverse	Multifidus muscle	Thickness	SEM = 0.21 cm Rater 2: ICC = 0.90 (0.78–0.95) SEM = 0.22 cm ICC = 0.888 (0.778–0.945)
Staepli <i>et al.</i> , 2010 ³²	Patients with osteoarthritis: preoperative n = 10 (NR:NR) age = 59.6 (6.0) postoperative n = 20 (NR:NR) age = 61.5 (5.3) Community-dwelling older adults n = 12 (3:9) age = 72.0 (9.36)	3–10	Linear	Sagittal	Vastus lateralis	Thickness	SEM = 0.09 cm
Stetts <i>et al.</i> , 2009 ³⁰		0	Curved	Transverse	Transversus abdominus Internal oblique External oblique	Thickness	Intra-image: ICCs ranging from 0.95 to 1.00 SEM = from 0.02 to 0.08 cm Inter-image: ICCs ranging from 0.77 to 0.97 SEM = 0.01 to 0.03 cm Rectus femoris: ICC = 0.876 (NR) Vastus intermedius: ICC = 0.928 (NR) Vastus lateralis: ICC = 0.852 (NR) Vastus medialis: ICC = 0.949 (NR) ICC = 0.97 (0.92–0.99) SEM = 0.02 cm
Strasser <i>et al.</i> , 2013 ³¹	Community-dwelling older adults n = 26 (NR:NR) age = 67.8 (4.8)	1	Curved	Transverse	Rectus femoris	Thickness	
Strasser <i>et al.</i> , 2013 ³¹		1	Curved	Transverse	Vastus medialis Vastus intermedius Vastus lateralis	Thickness	
Thomaes <i>et al.</i> , 2012 ³³	Older coronary artery disease patients without cardiovascular incident in the last year n = 25 (NR)	2	Linear	Transverse	Rectus femoris	Thickness	

(Continues)

Table 2 (continued)

Study	Demographics ^a	Interval in days	Transducer type	Scanning plane	Muscles	Muscle dimension	Reliability estimates ^b
Intra-rater reliability^c							
Cho et al., 2014 ²²	age = 68.6 (4.6) Post-stroke patients n = 30 (15:15)	7	Linear	Sagittal	Medial gastrocnemius	Thickness	ICC = 0.967 (0.932–0.984)
Hammond et al., 2014 ²³	age = 64.7 (5.7) Ambulatory patients COPD n = 15 (NR:NR)	NR	Curved	Transverse	Rectus femoris	CSA	ICC = 0.998 (NR)
MacGillivray et al., 2009 ²⁴	age = NR (NR) Community-dwelling older adults n = 11 (NR:NR)	NR	Linear	Sagittal	Rectus femoris	Volume	LOA = -0.17 to 0.30 cm ² ICC = 0.982 SEM = -0.13 cm ³
Sions et al., 2014 ²⁵	Median age = 79 Community-dwelling older adults n = 30 (8:22) age = 71.8 (NR)	10	Curved	Transverse	Multifidus muscle	Thickness	Inter-examiner measurement reliability: ICC = 0.98 (0.97–0.99) SEM = 0.08 cm Within-day procedural reliability: ICC = 0.88 (0.74–0.94) SEM = 0.26 cm Between-day procedural reliability: ICC = 0.86 (0.70–0.93) SEM = 0.29 cm

COPD, chronic obstructive pulmonary disease; CSA, cross-sectional area; ICC, intraclass correlation coefficient; LOA, limits of agreement; NR, not reported; SEM, standard error of measurement.

^aStudies are arranged in type of study and in alphabetical order.

^bn = sample size of the study (Male:Female). Mean age is reported. Value in parentheses is the standard deviation.

^cFindings are reported in ICC values, except where otherwise specified. Values in parentheses are 95% confidence intervals.

^dIntra-rater reliability is defined as all types of reliability measures within observer; inter-rater reliability is defined as all types of reliability measures between observers.

Table 3 Overview of the included validity studies

Study	Demographics ^a	Reference method	Scanning plane	Muscle	Muscle dimension	Validity estimates ^b
Berger <i>et al.</i> , 2015 ³⁶	Community-dwelling older adults <i>n</i> = 51 (25:26) age (females) = 72.5 (5.8) age (males) = 74.5 (6.5)	DXA	Transverse	Rectus femoris	Thickness	Right: <i>r</i> = 0.9687 Left: <i>r</i> = 0.9667
Hammond <i>et al.</i> , 2014 ²³	Ambulatory COPD patients <i>n</i> = 15 (NR:NR) age = NR (NR)	Ultrasound linear transducer	Transverse	Rectus femoris	CSA	ICC = 0.982 (NR)
MacGillivray <i>et al.</i> , 2008 ²⁴	Community-dwelling older adults <i>n</i> = 11 (NR:NR) median age = 79	MRI	Sagittal	Rectus femoris	Volume	ICC = 0.997 (NR)
Reeves <i>et al.</i> , 2004 ²⁹	Healthy adults <i>n</i> = 6 (3:3) age = 76.8 (3.2)	MRI	Transverse	Vastus lateralis	CSA	ICCs between 0.998 and 0.999 for scans 6 to 10
Sipila and Suominen, 1993 ³⁷	Older adults <i>n</i> = 36 (0:36) Trained athletes <i>n</i> = 21 (0:21) age = 73.7 (5.6) Healthy controls <i>n</i> = 15 (0:15) age = 73.6 (2.9)	CT	Transverse	Quadriceps	Thickness, CSA	Thickness <i>r</i> = 0.761 CSA <i>r</i> = 0.911
Thomaes <i>et al.</i> , 2012 ³³	Older coronary artery disease patients without cardiovascular incident in the last year <i>n</i> = 20 (NR) age = 68.3 (7.3)	CT	Transverse	Rectus femoris	Thickness	ICC = 0.92 (0.81–0.97)

CSA, cross-sectional area; CT, computed tomography; DXA, dual-energy X-ray absorptiometry; ICC, intraclass correlation coefficient; NR, not reported; MRI, magnetic resonance imaging;

Studies are arranged in type of study alphabetical order and in alphabetical order.

^a*n* = sample size of the study (Male:Female). Mean age is reported. Value in parentheses is the standard deviation.

^bFindings are reported in ICC values, except where otherwise specified. Values in parentheses are 95% confidence intervals.

Despite high scores on methodological quality, we found that information regarding the scanning procedure was unclear in most of the reliability studies. In particular, information was lacking on the scanning position and marking of the skin. For that reason, the findings in this review might be an overestimation of the true reliability of ultrasound. Nevertheless, we included one study that investigated both measurement and procedural reliability of ultrasound. That study found high ICC scores for both measurement and procedural reliability (ICC = 0.98 and ICC = 0.86, respectively).¹⁸ Nonetheless, it is of utmost importance to investigate the reliability of the entire ultrasound scanning procedure, as this reflects the assessment procedure in clinical practice. Therefore, the reliability of the ultrasound procedure to evaluate muscle size in older adults requires further evaluation.

All of the included studies used ICC scores to assess the agreement between raters or devices, which is considered to be the preferred statistical method to assess reliability.^{43–45} However, we found that 4 out of 13 reliability studies did not provide any information on the type of ICC used in the study. It is important to report complete information on the type of ICC used in the study as this influences the results and is

needed for an appropriate interpretation of the results.⁴⁶ Furthermore, in addition to type of ICC used in the study, data on the magnitude of the error, for example, standard error of the measurement and limits of agreement, should be reported.

The findings in this systematic review should be considered in the light of some limitations. First, although we used a comprehensive tool for the assessment of methodological quality of the ultrasound-derived prediction equation studies, this instrument was originally developed to assess methodological quality of health measurement instruments. Therefore, some items were not applicable for specific studies. Nevertheless, the quality of the included studies is expected to be adequate because the items regarding methodological quality were scored as good. Second, because of the lack of information on the type of ICC used in the studies, a meta-analysis could not be conducted. Finally, given the strict inclusion criteria, studies that did not mention (an equivalent of) muscle mass were excluded from this review.

In conclusion, this systematic review shows that ultrasound is a reliable and valid tool for the assessment of muscle size in older adults. However, more high-quality

research is needed to confirm these findings in both clinical as well as healthy populations. Furthermore, more research is required to validate prediction equations in older adults with varying function and health. This further validation is required to investigate whether use of ultrasound in the screening and diagnosis of sarcopenia is feasible.

Acknowledgement

The authors certify that they comply with the ethical guidelines for publishing in the *Journal of Cachexia, Sarcopenia and Muscle*: update 2015.⁴⁷

References

1. United Nations. Department of Economic and Social Affairs. *Population Division (2015) 2015*; World Population Ageing 2015 (ST/ESA/SER.A/390).
2. Lauretani F, Russo CR, Bandinelli S, Bartali B, Cavazzini C, Di Iorio A, et al. Age-associated changes in skeletal muscles and their effect on mobility: an operational diagnosis of sarcopenia. *J Appl Physiol (1985)* 2003;**95**:1851–1860.
3. Hiona A, Leeuwenburgh C. The role of mitochondrial DNA mutations in aging and sarcopenia: implications for the mitochondrial vicious cycle theory of aging. *Exp Gerontol* 2008;**43**:24–33.
4. Marcell TJ. Sarcopenia: causes, consequences, and preventions. *J Gerontol A Biol Sci Med Sci* 2003;**58**:M911–M916.
5. Anton SD, Woods AJ, Ashizawa T, Barb D, Buford TW, Carter CS, et al. Successful aging: advancing the science of physical independence in older adults. *Ageing Res Rev* 2015;**24**:304–327.
6. Rosenberg IH. Summary comments. *Am J Clin Nutr* 1989;**50**:1231–1233.
7. Landi F, Cruz-Jentoft AJ, Liperoti R, Russo A, Giovannini S, Tosato M, et al. Sarcopenia and mortality risk in frail older persons aged 80 years and older: results from iSIRENTE study. *Age Ageing* 2013;**42**:203–209.
8. Landi F, Calvani R, Cesari M, Tosato M, Martone AM, Bernabei R, et al. Sarcopenia as the biological substrate of physical frailty. *Clin Geriatr Med* 2015;**31**:367–374.
9. Hirani V, Blyth F, Naganathan V, Le Couteur DG, Seibel MJ, Waite LM, et al. Sarcopenia is associated with incident disability, institutionalization, and mortality in community-dwelling older men: the Concord Health and Ageing in Men Project. *J Am Med Dir Assoc* 2015;**16**:607–613.
10. Rizzoli R, Reginster J, Arnal J, Bautmans I, Beaudart C, Bischoff-Ferrari H, et al. Quality of life in sarcopenia and frailty. *Calcif Tissue Int* 2013;**93**:101–120.
11. Cruz-Jentoft AJ, Baeyens JP, Bauer JM, Boirie Y, Cederholm T, Landi F, 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.
12. Reijnierse EM, Trappenburg MC, Leter MJ, Blauw GJ, Sipilä S, Sillanpää E, et al. The impact of different diagnostic criteria on the prevalence of sarcopenia in healthy elderly participants and geriatric outpatients. *Gerontology* 2015;**61**:491–496.
13. Mitsopoulos N, Baumgartner RN, Heymsfield SB, Lyons W, Gallagher D, Ross R. Cadaver validation of skeletal muscle measurement by magnetic resonance imaging and computerized tomography. *J Appl Physiol (1985)* 1998;**85**:115–122.
14. Sanada K, Kearns CF, Midorikawa T, Abe T. Prediction and validation of total and regional skeletal muscle mass by ultrasound in Japanese adults. *Eur J Appl Physiol* 2006;**96**:24–31.
15. Crofts G, Angin S, Mickle KJ, Hill S, Nester C. Reliability of ultrasound for measurement of selected foot structures. *Gait Posture* 2014;**39**:35–39.
16. Koppenhaver SL, Hebert JJ, Parent EC, Fritz JM. Rehabilitative ultrasound imaging is a valid measure of trunk muscle size and activation during most isometric submaximal contractions: a systematic review. *Australian Journal of Physiotherapy* 2009;**55**:153–169.
17. Pretorius A, Keating J. Validity of real time ultrasound for measuring skeletal muscle size. *Physical Therapy Reviews* 2008;**13**:415–426.
18. English CK, Thoires KA, Fisher L, McLennan H, Bernhardt J. Ultrasound is a reliable measure of muscle thickness in acute stroke patients, for some, but not all anatomical sites: a study of the intrarater reliability of muscle thickness measures in acute stroke patients. *Ultrasound Med Biol* 2012;**38**:368–376.
19. 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. *Qual Life Res* 2012;**21**:651–657.
20. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;**20**:37–46.
21. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*, 2nd ed. New York: John Wiley & Sons; 1981. 38–46.
22. Cho KH, Lee HJ, Lee WH. Reliability of rehabilitative ultrasound imaging for the medial gastrocnemius muscle in poststroke patients. *Clin Physiol Funct Imaging* 2014;**34**:26–31.
23. Hammond K, Mampilly J, Laghi FA, Goyal A, Collins EG, McBurney C, et al. Validity and reliability of rectus femoris ultrasound measurements: comparison of curved-array and linear-array transducers. *J Rehabil Res Dev* 2014;**51**:1155–1164.
24. MacGillivray TJ, Ross E, Simpson HA, Greig CA. 3D freehand ultrasound for in vivo determination of human skeletal muscle volume. *Ultrasound Med Biol* 2009;**35**:928–935.
25. Sions JM, Velasco TO, Teyhen DS, Hicks GE. Ultrasound imaging: intraexaminer and interexaminer reliability for multifidus muscle thickness assessment in adults aged 60 to 85 years versus younger adults. *J Orthop Sports Phys Ther* 2014;**44**:425–434.
26. Agyapong-Badu S, Warner M, Samuel D, Narici M, Cooper C, Stokes M. Anterior thigh composition measured using ultrasound imaging to quantify relative thickness of muscle and non-contractile tissue: a potential biomarker for musculoskeletal health. *Physiol Meas* 2014;**35**:2165–2176.
27. Bemben MG. Use of diagnostic ultrasound for assessing muscle size. *J Strength Cond Res* 2002;**16**:103–108.
28. Raj IS, Bird SR, Shield AJ. Reliability of ultrasonographic measurement of the architecture of the vastus lateralis and gastrocnemius medialis muscles in older adults. *Clin Physiol Funct Imaging* 2012;**32**:65–70.

Online supplementary material

Additional Supporting Information may be found online in the supporting information tab for this article.

Appendix S1. Checklist.

Conflict of interest

W.N., A.S., H.J.-W., H.H., and C.v.d.S. declare that they have no conflicts of interest.

29. Reeves ND, Maganaris CN, Narici MV. Ultrasonographic assessment of human skeletal muscle size. *Eur J Appl Physiol* 2004;**91**:116–118.
30. Stetts DM, Freund JE, Allison SC, Carpenter G. A rehabilitative ultrasound imaging investigation of lateral abdominal muscle thickness in healthy aging adults [corrected] [published erratum appears in *J GERIATR PHYS THER* 2009;**32**(3):110]. *J Geriatr Phys Ther* 2009;**32**:16–22.
31. Strasser EM, Draskovits T, Praschak M, Quittan M, Graf A. Association between ultrasound measurements of muscle thickness, pennation angle, echogenicity and skeletal muscle strength in the elderly. *Age* 2013;**35**:2377–2388.
32. Staehli S, Glatthorn JF, Casartelli N, Maffioletti NA. Test–retest reliability of quadriceps muscle function outcomes in patients with knee osteoarthritis. *J Electromyogr Kinesiol* 2010;**20**:1058–1065.
33. Thomaes T, Thomis M, Onkelinx S, Coudyzer W, Cornelissen V, Vanhees L. Reliability and validity of the ultrasound technique to measure the rectus femoris muscle diameter in older CAD-patients. *BMC Med Imaging* 2012;**12**: 7-2342-12-7.
34. Strasser EM, Draskovits T, Praschak M, Quittan M, Graf A. Association between ultrasound measurements of muscle thickness, pennation angle, echogenicity and skeletal muscle strength in the elderly. *Age* 2013;**35**:2377–2388.
35. English C, Fisher L, Thoires K. Reliability of real-time ultrasound for measuring skeletal muscle size in human limbs in vivo: a systematic review. *Clin Rehabil* 2012;**26**:934–944.
36. Berger J, Bunout D, Barrera G, de la Maza MP, Henriquez S, Leiva L, et al. Rectus femoris (RF) ultrasound for the assessment of muscle mass in older people. *Arch Gerontol Geriatr* 2015;**61**:33–38.
37. Sipilä S, Suominen H. Muscle ultrasonography and computed tomography in elderly trained and untrained women. *Muscle Nerve* 1993;**16**:294–300.
38. Takai Y, Ohta M, Akagi R, Kato E, Wakahara T, Kawakami Y, et al. Validity of ultrasound muscle thickness measurements for predicting leg skeletal muscle mass in healthy Japanese middle-aged and older individuals. *J Physiol Anthropol* 2013;**32**, 12-6805-32-12.
39. Takai Y, Ohta M, Akagi R, Kato E, Wakahara T, Kawakami Y, et al. Applicability of ultrasound muscle thickness measurements for predicting fat-free mass in elderly population. *J Nutr Health Aging* 2014;**18**:579–585.
40. Mickle KJ, Nester CJ, Crofts G, Steele JR. Reliability of ultrasound to measure morphology of the toe flexor muscles. *Journal of Foot And Ankle Research* 2013;**6**:1.
41. Abe T, Kawakami Y, Kondo M, Fukunaga T. Comparison of ultrasound-measured age-related, site-specific muscle loss between healthy Japanese and German men. *Clin Physiol Funct Imaging* 2011;**31**: 320–325.
42. Miyatani M, Kanehisa H, Kuno S, Nishijima T, Fukunaga T. Validity of ultrasonograph muscle thickness measurements for estimating muscle volume of knee extensors in humans. *Eur J Appl Physiol* 2002;**86**:203–208.
43. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998;**26**:217–238.
44. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med* 2000;**30**:1–15.
45. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil* 1998;**12**:187–199.
46. Weir JP. Quantifying test–retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005;**19**:231–240.
47. von Haehling S, Morley JE, Coats AJS, Anker SD. Ethical guidelines for publishing in the Journal of Cachexia, Sarcopenia and Muscle: update 2015. *J Cachexia Sarcopenia Muscle* 2015;**6**:315–316.