Reliability & validity of ultrasound elastography for evaluating muscle stiffness among neurological populations: a systematic review & meta-analysis

Authors: [1]Tiev Miller, MSc, [2]Michael Ying, PhD, [1]Charlotte Sau Lan Tsang, PT, [1]Meizhen Huang, PhD, & [1]Marco Y. C. Pang, PhD

[1]Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, 11 Yuk Choi Rd, Hong Kong, S.A.R.

[2]Department of Health Technology and Informatics, The Hong Kong Polytechnic University, 11 Yuk Choi Rd, Hong Kong, S.A.R.

Address all correspondence to Professor Marco Pang: marco.pang@polyu.edu.hk

PROSPERO systematic review registration #: CRD42017076571

Abstract

Background: **Ultrasound elastography is an emerging diagnostic technology used to investigate biomechanical properties of the musculoskeletal system.** ***Purpose:*** To systematically review the psychometric properties of the ultrasound elastography techniques

in evaluating muscle stiffness among neurological populations. ***Data Sources:*** A systematic search of MEDLINE, EMBASE, CINAHL and Cochrane Library databases was performed in accordance with PRISMA guidelines. ***Data Selection:*** Using Covidence software, reviewers independently screened citations for inclusion. **Peer-reviewed studies which evaluated in vivo muscle stiffness among neurological populations and reported relevant psychometric properties were considered for inclusion.** ***Data Extraction:*** Twenty-one articles were included for final review. Data relevant to measurement technique, site and neurological condition were extracted. The Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) checklist was used to rate methodological quality of included studies. **Level of evidence for specific measurement outcomes was determined using a best-evidence synthesis approach.** ***Data Synthesis:*** Reliability varied across populations, ultrasound systems and assessment conditions (i.e. joint/ body positions, active/ passive muscle, probe orientation) with most studies indicating moderate to good reliability (Intraclass Correlation Coefficient (ICC)=0.5-0.9, n=13). Meta-analysis results showed good overall correlation across studies (r=0.78, 95% Confidence Interval (CI)=0.64–0.86, p≤0.00) with no between-group difference based on population (Q=0.00, df=1, p=0.97). Convergent validity was demonstrated by strong correlations between stiffness values and measures of spasticity (n=5), functional motor recovery or impairment (n=5) and grey scale or color histogram pixel intensities (n=3). Discriminant or known-groups validity was also

established for multiple studies indicating either significant between-group differences in stiffness values (n=12) or within-group differences between more- and less-affected limbs (n=6). Responsiveness was observed in all intervention studies reporting post-treatment stiffness changes (n=6). ***Conclusions:*** **Overall, ultrasound elastography techniques show moderate reliability in evaluating in-vivo muscle stiffness, good convergent validity with relevant clinical assessments, and good divergent validity in discriminating tissue changes within and between groups.** ***Impact Statement:*** Ultrasound elastography will have clinical utility in assessing muscle stiffness, monitoring its temporal changes, and measuring the response to intervention in populations with neurological conditions.

Abstract Word Count = 329

Manuscript Word Count: 5,580

Introduction

Altered muscle tissue mechanics and morphology are resultant corollaries among populations with neurological conditions and neuromuscular dysfunction [1,2]. Macro- and microstructural changes have been observed across various strata of muscle tissue in populations with common neurological conditions such as stroke and cerebral palsy (CP) [3-6]. Although the mechanism remains unclear, additional alterations in the biomechanical properties of muscle tissue may be a secondary sequela either associated with or resulting from underlying neurological etiologies (i.e., phasic hyperreflexia, hypertonia) [7,8]**. Intramuscular collagen formation [9], extracellular matrix organization [10] and titin isoform diversity within fibers [11] are other factors thought to contribute to tissue alterations.** The assessment of biomechanical properties, such as passive muscle stiffness, may have overt clinical value in determining tissue morphology and response to treatment or rehabilitation [12]. Subjective evaluation of altered muscle properties in neurological populations using qualitative clinical assessments such as manual palpation and muscle testing or modified Ashworth (MAS) and Tardieu scales are indirect and suboptimal [13,14]. Other quantitative assessments of these properties involving dynamometry **and B-mode ultrasound** can be procedurally complex, ineffective in isolating specific tissue regions and may not be feasible across clinical settings [15,16].

Alternatively, elastography provides a direct, non-invasive stiffness quantification of individual muscle structures in real-time [17]. **As the use of diagnostic musculoskeletal ultrasound is becoming more prevalent in physical medicine and rehabilitation training programs [18] and in clinical physical therapy [19], ultrasound units with tissue imaging capabilities such as elastography may prove advantageous in monitoring transitory or progressive muscle changes associated with neurological conditions [20,21]. Elastography has also demonstrated utility in evaluating response to invasive [7,22-26] and non-invasive [27] clinical intervention strategies for reducing muscle stiffness and spasticity in neurological populations.**

Ultrasound-based elastography methods developed as an outgrowth of tissue palpation and motion tracking techniques for identifying tissue masses harder and more resistant to displacement than surrounding reference tissues. Since the development of sonoelasticity imaging and static elastography in the early 1990's for quantifying the distribution of the elastic modulus in soft tissues [28,29], ultrasound elastography (UE) methods have continued to evolve [30]. **Previous reviews have highlighted the utility of this technology for investigating the biomechanical properties of the musculoskeletal system [21,31-34].** There are several different UE techniques capable of providing either quantitative or qualitative measures of these properties [17], each differing in frequency, method of excitation and

interface [35]. There is currently no consensus with regard to which method may be optimal for assessing muscle stiffness among neurological populations in vivo.

As the development and modification of elastography for measuring specific tissue types (e.g., thyroid, breast, muscle) remains on-going [36], a general knowledge of its underlying mechanisms, technical limitations, measurement attributes and an appraisal of its potential clinical application is warranted. Briefly, the estimation of stiffness using elastography requires measuring tissue displacements in response to the application of a stress generated via mechanical, acoustic radiation or internal endogenous forces [35]. Although UE systems demonstrate great potential in clinical utility, their underlying accuracy is based upon non-biological material testing under absolute conditions [36,37]. Attempts have been made to investigate the limitation of material linearity when using UE to evaluate muscle properties among healthy and neurological populations [38-40]. However, differences in tissue type, geometry and activation remain confounding influences [41,42]. There are also operator-dependent sources of error as well as measurement range capabilities to be considered [43-45].

In the field of musculoskeletal elastography, a comprehensive review which systematically investigates the reliability and validity reported throughout the available literature is currently

lacking. **The primary objective of this review was to evaluate the evidence regarding the reliability and validity of UE for measuring muscle stiffness in neurological populations. Secondary objectives were to synthesize the information regarding measurement protocols, operators and equipment used for measuring muscle stiffness, and to assess study quality and level of evidence regarding stiffness measures. In healthy populations, UE has shown good reliability [46] and correlation with other measures assessing similar [47] or related physiological constructs [48,49] and measures of physical function [50] and disability [51]. Therefore, it was hypothesized that overall measurement reliability and validity of UE for measuring muscle stiffness in neurological populations would be good**.

Methods

Data sources & searches

A search was developed using the following concept domains as syntactic framework: (1) musculoskeletal stiffness, (2) ultrasound elastography and (3) validity and reliability. MEDLINE, EMBASE, CINAHL and Cochrane Library databases were searched. Paired and individual keywords, medical subject headings (MeSH), Embase subject headings (Emtree), field codes, boolean and proximity operators used in the search strategy syntax specific to each database are included in the supplementary appendices (Supplementary Appendix A). Results for each of the concept domains were combined to produce the final search results. In

further refining the search, an additional filter was applied limiting the results to references with human subject populations published in English language journals between January 1990 and January 2020. References of articles selected for inclusion were reviewed to identify other relevant publications for inclusion. Reference lists from review articles discussing musculoskeletal elastography were also searched. A forward search was performed before the final synthesis and analysis to include studies published after the initial search and data extraction. This review was conducted in accordance with PRISMA guidelines [52] and prospectively registered with PROSPERO (registration #: CRD42017076571).

Study selection

Using Covidence online data extraction and screening software (Cochrane software Csr. Melbourne, Victoria, Australia: Veritas Health Innovation, p. Available at www.covidence.org), two reviewers (TM, MH) independently screened titles and abstracts to determine propriety for inclusion. Hand-searched publications from reference lists were then entered into the full-text screening. Two reviewers (TM, SLT) then extracted relevant data independently. All missing or omitted data were requested from authors of included studies. Data were included under the contingency authors replied within a timeframe of 10 working days. Conflicts arising between reviewers were resolved through discussion or by consulting a third reviewer (MP, MY) to reach consensus.

Selection of studies was based on the following criteria. *Inclusion criteria:* 1. peer reviewed articles published in English between 1990-2020, 2. human subjects studied in vivo, 3. studies investigated UE measurement properties for muscle, tendon, and/or fascia stiffness, 4. data collection took place in research settings or across all stages of the continuum of care, 5. subject populations with a neurological condition (i.e., stroke, spinal cord injury (SCI), CP, etc.), 6. muscle stiffness measured with UE was a primary or secondary diagnostic objective, 7. measurement validity and/or reliability of UE was also a primary or secondary outcome, 8. studies reported reliability, convergent, discriminant, known-groups, or criterion validity as determined with the use of a comparator (i.e., previously validated diagnostic methods for evaluating tissue stiffness such as magnetic resonance elastography, electronic palpation imaging, biopsies, and histopathological samples, gelatine-based phantoms, and/or tissue equivalent phantoms) 9. study design was either a case-controlled diagnostic, prospective or retrospective cohort, cross-sectional or longitudinal, pre-post intervention or randomized-controlled trial. *Exclusion criteria*: 1. published conference proceedings (i.e., presentations, posters, symposium, etc.) 2. book reports or chapters, 3. theses or dissertations, 4. unavailable in full text, 5. incorrect timeframe (i.e., before 1990), 6. involved animals or cadaver samples without comparative in vivo measures, 7. studies focused solely on the assessment of bone,

cartilage, entheses, ligaments or joint capsules, 8. UE measured stiffness was not a primary or secondary outcome and 9. measurement properties were not a primary or secondary outcome.

Data extraction & quality assessment

The following items were extracted from included articles: 1. author information (i.e., names, title, year, location), 2. study design, 3. measurement reliability and validity, 4. diagnostic setting (i.e., inpatient, outpatient, laboratory), 5. ultrasound operator (i.e., technician, clinician, researcher), 6. ultrasound system and probe model, 6. **probe alignment in relation to muscle fiber orientation (i.e., parallel, perpendicular or oblique)**, 7. muscle site, 8. body position and joint angle during testing, 9. subject demographics and characteristics, 10. measurement units (i.e., shear modulus (SM) in kilopascals (kPa) and/or shear wave velocity (SWV) in meters per second (m/s), strain ratio (SR), sonoelastographic index/score) and 11. details of image/data acquisition and processing (i.e., region of interest (ROI), number of trials, contact interface, software).

Rating of methodological quality for included studies was performed independently by two reviewers (TM, SLT) using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Risk of Bias checklist (July 2018 version). **Although originally intended for health-related patient-reported outcomes [53], the checklist also**

facilitates the election of quality scores for studies in which measurement properties are based on operator or clinician-assessed outcomes of function [54] and disability [55] in neurological populations. Of the items encompassing the checklist, Boxes 6, 7, 8, 9a, 9b, 10a-10d were used to assess reliability, measurement error, criterion validity, convergent validity, discriminant or known-groups validity and responsiveness for each study, respectively. Other items contained within the checklist were omitted for the purposes of this review. The checklist uses a 4-point rating system in assigning ratings for each item. Ratings of 4, 3, 2 or 1 were deemed very good, adequate, doubtful or inadequate, respectively. In determining overall quality for each category, the lowest rating was used (i.e., worst score counts principle) [56,57]. A third reviewer (MP, MY) was consulted regarding any discrepancies. **The level of evidence for measurement property results was determined using a best-evidence synthesis approach [58] described in a previous systematic review of reliability, validity, and responsiveness for physical capacity tasks that assessed functioning in patients with low back pain [59]. Criteria for strong, moderate, limited, unknown, or conflicting levels of evidence are also described in detail [59]. Briefly, a positive, indeterminate or negative rating was first assigned according to an established criterion for rating measurement properties reported by Prinsen et al [60]. Levels of evidence for measurement property ratings were then assigned according to measurement property rating consistency, combined sample size, and methodological**

quality for articles with similar or comparable outcome measures (i.e., quantitative, semi-quantitative estimates of muscle stiffness).

Data synthesis & analysis

A qualitative synthesis was conducted by tabulating data according to sample population, measurement site and UE technique. Comprehensive Meta-Analysis software (CMA version 3.0, Biostat Inc., Englewood, New Jersey, USA) was used for quantitative analyses. **Subgroup analyses consisted of ≥3 homologous studies which assessed measurement reliability.** Correlation coefficients (ICC) were transformed to Fisher's Z scale for analysis [61]. The software accommodates the combination of multiple outcomes (i.e., measurement sites, probe orientations or operators) which were pooled to calculate a single metric for analysis. Proportion of variance between studies was interpreted using Higgins' I^2 statistic and 95% prediction intervals (PI) were calculated to express absolute estimates of heterogeneity for each subgroup analysis [62]. A random effects model was chosen with the assumption ICCs would vary between studies. Prior to subgroup analyses a univariate meta-regression was performed to determine the effect of independent factors related to study design (i.e., ICC model and form) on overall correlation estimates (i.e., dependent factor) [63].

Results

Screening

A flow diagram summarizing the screening process and results is provided in Figure 1. A total of 21 articles involving 326 **individuals with neurological conditions** and 177 control subjects met the criteria for final inclusion [7,8,20,22,23,25,27,64-77]. Excluded articles were either case studies [20,24] or did not report measurement validity or reliability [38].

Study characteristics

A summary of study characteristics and outcomes is provided in Table 1 and a list of all abbreviated terms for tables and figures is provided in the supplementary appendices (Supplementary Appendix B). Study designs varied in type and complexity. Most were observational studies of either cross-sectional or longitudinal design [8,23,65-78] with two measuring reliability as a primary outcome [75,77]. Six studies described the use of blinding procedures [22,25,64,68,76,77]. Measures were conducted by a radiologist, physician or physiatrist, others described as examiner, experimenter or investigator, or were not explicitly stated. Neurological conditions investigated were stroke, CP, Duchenne muscular dystrophy (DMD) and Parkinson's disease (PD). The medial gastrocnemius (MG), biceps brachii (BB) and tibialis anterior (TA) muscles were the most commonly assessed sites. Measures were collected during either passive or active muscle conditions with and without being

concomitantly monitored by electromyography (EMG). One study used a constant-current stimulator to elicit contractions [71].

Ultrasound system

A summary of system specifications, settings, software, reported units and value ranges is provided in Table 2. UE systems varied across studies. All studies reported using linear array probes with frequencies ranging from 4-15MHz. Elastography methods involved either a dynamic time-course with an acoustic radiation force application, quasi-static time-course with a mechanical force application or dynamic with mechanical force. Commonly used system settings were either standard musculoskeletal presets or shear wave elastography (SWE) mode. **Units were reported as quantitative (SM, SWV) or semi-quantitative estimates of muscle stiffness (SR or elastographic index/scores with either grey scale or color histogram pixel intensity values).**

Acquisition procedures

A summary of probe placement, fixation, applied compression, contact interface, processing software and other image and data acquisition methods are found in Table 2. For most studies, probe placement during image capture was performed in parallel alignment with fascicle orientation. Two studies investigated parallel and perpendicular alignments [75,77].

All studies used linear probes suitable for superficial structures. A transmission or stand-off gel couplant was used as the contact interface in several studies. The number of trials or single images captured for each muscle site ranged from 2-15. Selected ROIs ranged from 4.8mm in circular diameter to 30mm^2 in size and varied in placement depth, number and shape.

Reliability

A summary of reported ICCs and additional reliability information is provided in Table 3. According to the 95% CI of ICC estimations, reported values <0.5, between 0.5-0.75, between 0.75-0.9 and >0.90, were indicative of either poor, moderate, good, or excellent reliability, respectively. When considering all ICCs reported for studies, most demonstrated moderate to good reliability (ICC=0.5-0.9, n=13). **However, two studies investigating reliability in patients with CP reported large variance in confidence intervals (95% CI = 0.33-0.84) [23,64]. Another study among patients with stroke demonstrated the substantial variance in the range of reported ICCs (ICC = 0.00-0.87) [75]. Of all the included studies, this was also the only study to report estimates of measurement error. Ranges for measurement error varied substantially based on differences in probe placement and muscles sites examined (SEM = 0.61-24.81) [75]. One other study also assessed measurement reliability using different probe placements among patients with stroke,**

reporting considerably less variance in the range of ICCs (ICC = 0.55-0.85) [77] in comparison to the former [75]. Comparative reliability of active versus passive muscle conditions could not be determined from the study which investigated these conditions concomitantly [8]. **A graphical summary of subgroup analyses is provided in Figure 2**. As the results of the meta-regression showed no significant influence of ICC model and form on overall correlation (Q=1.85, df=2, p=0.40), population-based subgroup analyses were conducted. The overall correlation across subgroups was good (n=8, r=0.78, 95% CI=0.64–0.86, p≤0.00) with no significant difference between groups based on population (Q=0.00, df=1, p=0.97). For studies involving **people with CP** (n=4, r=0.78, 95% CI=0.58-0.89, p≤0.00) and **with stroke** (n=4, r=0.77, 95% CI=0.56-0.89, p≤0.00), the correlation was good. **However, estimates of absolute heterogeneity indicated a wide dispersion in reliability across studies involving patients with CP (95% PI=0.02-0.97). Proportion of variance was mostly attributable to sampling error rather than true correlation (31.5%) (I^2=31.5, p=0.22). There was larger observed dispersion in estimates of absolute heterogeneity across studies involving patients with stroke suggesting greater variance in measures (95% PI=-0.52-0.99). Proportion of variance in true correlation was also larger (53.5%) with less attributed to error (46.5%) (I^2=53.5, p=0.09).**

Convergent validity

A summary of convergent validity and study comparators is provided in Table 3. Several studies reported correlations between muscle stiffness and standardized assessments of spasticity and functional motor recovery or impairment. In two studies involving subjects with PD, Unified Parkinson's Disease Rating Scale (UPDRS) scores were positively correlated with SM values ($r=0.65$, $p\leq0.00$) [66] and negatively correlated with SR ($r=-0.78$) [68]. For **individuals with stroke**, Fugl-Meyer assessment (FMA) scores were correlated with side-to-side (i.e., paretic and non-paretic) difference in SWV values ($r^2=0.33$, $p=0.02$) [74] and values for paretic sides alone ($r=-0.58$) [69]. Paretic side SWV values were positively correlated with MAS ($r=0.66$) and TS scores ($r=0.54$) and negatively correlated with Stroke Rehabilitation Assessment of Movement (STREAM) scores ($r=-0.57$) [8,77]. For **individuals with CP**, several studies reported significant correlations not only between stiffness values and functional scores [23-25,64] but also gray scale or color histogram pixel intensities [22,25] . Correlation between stiffness and echo intensity (i.e., grey scale value) was also observed among **people with stroke** ($r^2=0.70$, $p\leq0.00$) [74].

Divergent validity

A summary of discriminant/known-groups validity is found in Table 3. In comparing people with CP to controls, stiffness was significantly greater for the CP group in several studies ($p\leq0.001$) [22,64,65,70]. SWV values were also significantly higher in more-affected

limbs (p≤0.024) [73]. However, when stratified according to motor function (i.e., GMFCS Levels I and II), there were no significant differences in SWV between groups [73]. In individuals with PD, stiffness was significantly greater compared to controls (p≤0.05) [66,68], with no difference between markedly and mildly symptomatic limbs (p≤0.05) [66]. Stiffness was significantly greater for people with DMD compared to controls across almost all muscle sites and conditions (p≤0.005) [71,72,76]. For individuals with stroke, findings varied by condition (i.e., spasticity, joint angle, muscle site, activation) with significantly greater stiffness in paretic versus non-paretic limbs (p≤0.001) [8,74] and controls during passive muscle states (p≤0.001) [8]. Differences in stiffness were also joint angle specific [67,69,77,78].

Responsiveness

A summary of responsiveness is found in Table 3. A total of five studies examined pre- to post-intervention changes in muscle stiffness [7,22,23,25,27]. Most examined changes following botulinum toxin injections in people with CP [7,22,23,25]. These studies reported significant reductions in stiffness values or scores/indices following treatment (p≤0.05). One study examined the effect of a robot-assisted stretching and joint mobility program in individuals with stroke and showed that SR values for the paretic Achilles tendon increased significantly from pre- to post-training (p=.045) [27]. Additionally, two

longitudinal studies examining the effect of disease progression in people with DMD showed SM values significantly increased between 0 (pre) and 12 months (post) ($p<0.001$) [71,72].

Quality assessment ***& level of evidence***

A summary of the quality assessment, measurement property result ratings and level of evidence synthesis is provided in Table 4. Methodological quality ratings for reliability (Box 6) were adequate for most studies, with one study rated as doubtful due to unclear description of testing conditions and time intervals between assessments [66]. Measurement error (Box 7) assessed for one study was adequate [75]. Criterion validity (Box 8) and responsiveness (Box 10a) using a criterion approach were not assessed due to a lack of concurrent comparators. Convergent validity (Box 9a) and discriminant/ known-groups validity (Box 9b) were very good or adequate for most studies. Two studies were rated as doubtful for convergent validity due to suboptimal analyses and inadequate reporting of outcomes [64,76]. Three studies were rated doubtful or inadequate for discriminant/ known-groups validity for inadequate reporting of relevant subgroup characteristics and suboptimal analyses [67,76,78]. Responsiveness using a construct approach for outcome (Box 10b), between subgroups (Box 10c) or pre-to-post intervention comparisons (Box 10d) was very good for all studies.

Several studies stated a priori hypotheses [65,66,68,69,71,73,76,77]. When not explicitly stated, the authors expectations or assumptions were compared to reported outcomes (i.e., correlations, between-group differences, pre-to-post intervention values) in determining measurement property result ratings for validity and responsiveness. Results of the best-evidence synthesis suggest there is a moderate level of evidence for negative ratings of interrater reliability (ICC < 0.70) for stiffness estimates using quantitative UE methods. For intrarater reliability using quantitative methods, the level of evidence was moderate for positive ratings (ICC > 0.70). For semi-quantitative methods, the level of evidence for positive ratings was unknown for interrater reliability and limited for intrarater reliability due low total sample size (< 50). The level of evidence for an indeterminate rating of measurement error in one study using a quantitative method was unknown. For convergent validity (hypothesis testing) of quantitative methods, the level of evidence for positive ratings (i.e., mostly in accordance with hypotheses) was strong. The level of evidence for positive ratings was moderate for semi-quantitative methods due to sample size (< 100). For discriminant/ known-groups validity of quantitative methods, the level of evidence for positive ratings was strong. Limited level of evidence for positive ratings of semi-quantitative methods was also due to lower total sample size (< 50). The level of evidence for positive ratings of responsiveness of studies involving

quantitative methods was limited. For semi-quantitative methods, the level of evidence for positive ratings was moderate due to higher sample size (> 25).

Discussion

Reliability

Studies reporting ICCs indicated mostly moderate to good measurement reliability overall. The methodological quality of most of these studies was also determined to be adequate. **However, the results of the level of evidence synthesis suggest that evidence for intrarater reliability was stronger than interrater reliability for both quantitative and semi-quantitative methods. The evidence also ranged from moderate for quantitative estimates of stiffness, to unknown for semi-quantitative estimates. Furthermore, estimates of measurement error were not reported in most studies that assessed measurement reliability. This is an important metric not only for reliability, but for interpreting clinically meaningful changes in health-related outcomes [79].** Subgroup analyses demonstrated pooled estimates of reliability were good overall with roughly equivocal correlations for CP and stroke subgroups. However, of the studies investigating reliability in **people with stroke** as a primary outcome [75,77], there was a large range in reported coefficients (ICC range=0.00-0.87). **While the dispersion in estimates of heterogeneity were large for the CP subgroup (95% PI=0.02-0.97), the stroke subgroup**

was comparatively wider (95% PI=-0.52-0.99), suggesting the need for greater precision in measurement protocols. The variability in estimates may be attributable to differences in selected muscle sites, muscle activity, operator experience, probe alignment, ultrasound system and acquisition procedures used and subject age and gender.

In adult populations without neurological conditions, stiffness measures have been shown to vary by depth, activity and joint angle. Using a curvilinear probe in deep penetration mode, Blain et al demonstrated greater reliability for the superficial erector spinae than deeper multifidus muscles [80]. Alfuraih et al also reported better reliability for more superficial than comparatively deeper muscles [81]. Generally, the greater the depth of a given anatomical structure, the greater the attenuation effect on acoustic pulse transmission and wave tracking [35], which may affect reliability [81]. In isotropic tissues such as the thyroid, signal strength may diminish at depths between 4-6cm. In anisotropic tissues such as muscle, signal diminishment may occur at lesser depths [82]. **It is unknown to what degree signal attenuation and probe type and measurement depth affect reliability in neurological populations.** Stiffness has also been shown to have a linear relationship with joint torque in passive and active muscle [42]. The use of EMG may be particularly necessary in order to monitor muscle activation status when assessing patients with spasticity (i.e., stroke). **Of the two studies investigating reliability in individuals with stroke as a primary outcome [75,77],**

only one involved the use of EMG during measures [77]. The other where stiffness was measured at multiple muscle sites and probe orientations demonstrated considerable variance in estimates of reliability and measurement error [75]. Although reliability was not a primary outcome, the study by Eby et al investigating muscle stiffness and torque response to passive elbow extension after stroke also incorporated the use of EMG and reported a comparatively smaller range of ICCs (ICC = 0.75-0.99) [78].

The influence of age on the reliability of stiffness measures is inconclusive. Only one study examined age-related differences on muscle stiffness showing that lower limb SM at different lengths was correlated with age in **individuals with DMD** ($r=0.55-0.74$, $p\leq0.05$) but not among controls ($r<0.43$) [72]. However, these findings do not suggest age has any substantive impact on reliability.

The amount of previous training or experience in musculoskeletal ultrasound and elastography also appears to influence measurement reliability. The range of reported reliability estimates was greater for studies with poorly defined operator experience (ICC range=0.00-0.94) than those with clearly defined experience (ICC range=0.65-0.92). Among these studies, operators were described as radiologists, physicians or physiatrists with 2-17 years of relevant experience.

There was also considerable variance among studies using multiple probe orientations. Mathevon et al demonstrated slightly greater variance for measures in perpendicular (ICC range=0.00-0.73) compared to parallel probe alignments (ICC range=0.27-0.87) [75]. Wu et al also showed greater intrarater and interrater reliability for parallel (ICC=0.85 and 0.76, respectively) compared to perpendicular alignments (ICC=0.71 and 0.55, respectively) [77]. As shear waves propagate longitudinally in alignment with muscle fiber direction, aligning the probe parallel to fibers may enhance measurement accuracy [83]. Perpendicular alignments, in contrast, have shown greater dispersion of shear waves [39,84].

Other aspects of measurement acquisition may also contribute to variability. ROI varied across most studies, which has been shown to influence values in regional tissue mapping studies [85]. Lack of contact interface standardization is another source of variability, as differences in the use of transmission gel have been shown to influence consistency [81]. **Discrepancies in reported values between studies (i.e., kPa, m/s, etc.) is also problematic, as these values are related but separate phenomena [86]. Tissue stiffness estimations are predicated upon static deformation models of elastic materials and described as stress (i.e., the force per unit in a given area) divided by strain (i.e., the expansion per unit of length), which is the equivalent of an elastic modulus value [36]. It is important to note that**

most UE systems operate under an assumption of material linearity wherein tissues are homogenous or structurally similar, isotropic (i.e., identical property values in all directions), and are non-viscous (i.e., identical fluid consistency) [35]. Examples which fit these assumptions in an absolute sense would be materials such as metal or glass. However, muscle tissues are heterogeneous, anisotropic and viscoelastic given the variation in their structural composition and fluid consistency [33]. For SWE systems, SWV is likely to be the most appropriate unit for interpretability across studies as this a measure of shear wave dispersion [33]. While the SM (i.e., stiffness estimation under the assumption of an absolute elasticity model) may be appropriate for isotropic tissues like the liver [87], muscle tissues are largely anisotropic due to fascicle order and orientation [88].

Intersystem differences may also contribute to variance across studies. UE systems use their own unique algorithms for capturing images and calculating stiffness values and can be generally categorized by either quasi-static or dynamic means of excitation [35,89]. Dynamic methods are complex involving a varying time-course force application in the form of vibration or an acoustic force pulse with a specific frequency (50 to 500 Hz) and often use ultrafast imaging of induced displacements or deformations [30,90]. These deformation forces are sent through tissues as either compression waves with high propagation speeds (approx. 1500 m/s) or shear waves with low propagation speeds

(approx. 1—50 m/s) [91]. Tissue displacements and wave velocities are then tracked and tissue stiffness estimations are generated based on tissue motion, frequency shifts or velocity changes [92]. While static and dynamic methods are similar in that both use an external stress and follow changes in strain, the stress applied in dynamic methods is definable and less operator-dependent, thus holding to the proportionality of Hooke's Law in the estimation of Young's modulus and providing a more quantitative measure [91,93]. Static methods provide semi-quantitative strain ratio based estimations through manual application of multi-compression cycles. However, long acquisition times and difficulties in producing artifact-free compression cycles are inherent technical challenges [84,93]. In studies exploring intersystem comparisons using dynamic and static methods in tissue mimicking phantoms or muscle in vivo, dynamic methods have demonstrated slightly greater measurement reliability [43,44,94,95]. In this review, fewer studies used systems requiring mechanical force application (i.e., tissue compression). Although these systems are associated with greater operator-dependent error, Wu et al and Mathevon et al reported the largest variance in reliability using SWE systems (ICC range=0.55-0.85 and 0.00-0.87, respectively) [75,77]. Moreover, Gao et al reported good interrater (ICC=0.84) and intrarater reliability (ICC=0.88) using a strain system [68]. Good intrarater reliability was also reported by Park et al (ICC=0.85, 0.87) and Kwon et al (ICC=0.81, 0.88) using sonoelastography systems [25,70]. **Furthermore, the operators in both**

these studies were described as physiatrists with 16 to 17 years of experience in using musculoskeletal ultrasound [25,70]. In comparison, Wu et al described the operator as a physiatrist with 2 years of experience in musculoskeletal ultrasound [77]. The operator in the Mathevon et al study was described as an investigator with no mention of additional experience or training outside of the measures collected for the study [75]. Although the risk of committing operator-related errors may be reduced with dynamic systems, it is reasonable to assume that acceptable levels of measurement reliability can be achieved with relevant training and experience in the use of other systems.

Taken together, the current evidence suggests **UE** has **moderate** reliability when used among neurological populations. Care should be taken to ensure measurement acquisition protocols are standardized. **Reliability appears to be more difficult to achieve with multiple operators than in instances with a single operator.** The use of EMG rather than visual confirmation of muscle activity status may also be an important consideration in assessing patients with characteristic spasticity or hypertonia.

Validity

As expected, convergent validity was observed in studies with strong correlations between stiffness and reduced functional motor recovery or increased impairment and spasticity.

Additionally, strong correlations between stiffness and grey scale or color histogram pixel intensities were also observed. Higher echo intensity in musculoskeletal ultrasound is generally indicative of greater organizational density of collagen rather than the presence of fluid within tissues [96]. For subjects with neurological conditions, this may represent changes in composition, fiber type distribution and intrinsic mechanical properties resulting from alterations in muscle tissue innervation [5,97]. Tissue composition alterations were also evident in one study **involving people with DMD** which showed fatty replacement and patchy edema on muscle MRI scans in addition to increased SM values [76]. However, there was no significant correlation between these parameters, perhaps due to the small cohort size**. The SWE method used was also not found to have any clear diagnostic advantage or greater sensitivity in detecting early changes to muscle in comparison to MRI [76].**

Discriminant or known-groups validity was observed in more than half of included studies suggesting that UE may be useful in monitoring muscle pathology over time. The high degree of responsiveness observed across intervention studies also suggests that UE may be useful in evaluating response to treatment. However, there is no gold standard method for assessing muscle stiffness. While concurrent methods (i.e., echo intensity, spasticity and motor function scales) showed strong correlations with stiffness, they are separate constructs. Other methods such as myotonometry and portable hardness meters may be useful comparators and have

been used to examine stiffness after stroke [98] and in patients with PD [99]. Few studies have examined the validity of using myotonometry and UE concurrently [100].

Limitations of the studies reviewed

Many studies did not clearly define operator experience which may influence measurement consistency. **Of the included studies reporting estimates of reliability, most did not provide estimates for measurement error.** Another limitation was the lack of concurrent comparators necessary for establishing criterion validity. Future research endeavors should explore the concomitant use of existing technologies (i.e., magnetic resonance elastography, myotonometry) in elucidating the validity of UE among neurological populations. **Additionally, the concomitant use of EMG may be important not only for reliability, as previous described, but also measurement validity. Although there were strong correlations between clinical assessments of spasticity (i.e., FMA, MAS) and muscle stiffness measures [8,69,74,77], these assessments are limited in their ability to distinguish between active reflex or neurogenic components of stiffness and passive non-reflex mediated components [101]. It is also unknown whether degenerative within-subject factors such as bilateral differences in motor-unit threshold and denervation of affected limbs [102] also contribute to muscle stiffness. When paired with UE, electrophysiological evaluation may be of value in this regard.**

Limitations of this systematic review

As overall estimates of reliability were similar for CP and stroke subgroups, differences in measurement protocols may have contributed substantially to the observed heterogeneity between subgroups. These methodological differences negatively influence the interpretability of the review findings. Although the number of studies **involving persons with stroke and CP** was sufficient for conducting subgroup analyses, the number of studies assessing reliability among other neurological populations (i.e., PD, DMD) were limited. To the knowledge of the authors, there is currently no research examining the use of UE in other neurological populations (i.e., SCI) not described in this review. This will require future investigation. **There was also a paucity of studies examining UE responsiveness following non-invasive treatments for individuals with neurological conditions. This may be an important, yet unaddressed, aspect of the technology which translates to routine clinical application. Studies which assess pre- to -post changes in muscle stiffness following common non-invasive therapeutic modalities are needed moving forward.**

Conclusion

Overall, UE demonstrates moderate reliability evaluating in-vivo muscle stiffness across a range of neurological populations. This method also demonstrates strong convergent

validity with relevant clinical assessments, and strong divergent validity in discriminating tissue changes within and between groups. However, further investigation regarding UE systems, image acquisition procedures and the use of concurrent assessments may be warranted to standardize measurement protocols and potentially enhance reliability and validity.

Role of funding sources

Three of the review team members (TM, SLT, MH) were supported by general research studentships provided by The Hong Kong Polytechnic University through the Department of Rehabilitation Sciences.

Conflict of interest

The authors have no conflicts of interest to declare.

References

1. Foran JR, Steinman S, Barash I, Chambers HG, Lieber RL. Structural and mechanical alterations in spastic skeletal muscle. *Dev Med Child Neurol.* 2007;47(10):713-717.
2. Lieber RL, Ward SR. Cellular Mechanisms of Tissue Fibrosis. 4. Structural and functional consequences of skeletal muscle fibrosis. *Am J Physiol Cell Physiol.* 2013;305(3):C241-C252.
3. Ito J-I, Araki A, Tanaka H, Tasaki T, Cho K, Yamazaki R. Muscle histopathology in spastic cerebral palsy. *Brain Dev.* 1996;18(4):299-303.
4. Pontén EM, Stål PS. Decreased capillarization and a shift to fast myosin heavy chain

IIx in the biceps brachii muscle from young adults with spastic paresis. *J Neurol Sci.* 2007;253(1-2):25-33.
5. Smith LR, Lee KS, Ward SR, Chambers HG, Lieber RL. Hamstring contractures in children with spastic cerebral palsy result from a stiffer extracellular matrix and increased in vivo sarcomere length. *J Physiol.* 2011;589(Pt 10):2625-2639.
6. Lieber RL, Fridén J. Spasticity causes a fundamental rearrangement of muscle-joint interaction. *Muscle Nerve.* 2002;25(2):265-270.
7. Brandenburg JE, Eby SF, Song P, Bamlet WR, Sieck GC, An KN. Quantifying Effect of Onabotulinum Toxin A on Passive Muscle Stiffness in Children with Cerebral Palsy Using Ultrasound Shear Wave Elastography. *Am J Phys Med Rehabil.* 2018;97(7):500-506.
8. Lee SSM, Jakubowski KL, Spear SC, Rymer WZ. Muscle material properties in passive and active stroke-impaired muscle. *J Biomech.* 2018.
9. Haus JM, Carrithers JA, Trappe SW, Trappe TA. Collagen, cross-linking, and advanced glycation end products in aging human skeletal muscle. *J Appl Physiol.* 2007;103(6):2068-2076.
10. Kjær M. Role of Extracellular Matrix in Adaptation of Tendon and Skeletal Muscle to Mechanical Loading. *Physiol Rev.* 2004;84(2):649-698.
11. Neagoe C, Opitz CA, Makarenko I, Linke WA. Gigantic variety: expression patterns of titin isoforms in striated muscles and consequences for myofibrillar passive stiffness. *J Muscle Res Cell Motil.* 2003;24(2):175-189.
12. Miscio G, Del Conte C, Pianca D, et al. Botulinum toxin in post-stroke patients: stiffness modifications and clinical implications. *J Neurol.* 2004;251(2):189-196.
13. Escolar DM, Henricson EK, Mayhew J, et al. Clinical evaluator reliability for quantitative and manual muscle testing measures of strength in children. *Muscle Nerve.* 2001;24(6):787-793.
14. Alhusaini AAA, Dean CM, Crosbie J, Shepherd RB, Lewis J. Evaluation of Spasticity in Children With Cerebral Palsy Using Ashworth and Tardieu Scales Compared With Laboratory Measures. *J Child Neurol.* 2010;25(10):1242-1247.
15. Seynnes OR, Bojsen-Moller J, Albracht K, et al. Ultrasound-based testing of tendon mechanical properties: a critical evaluation. *J Appl Physiol.* 2015;118(2):133-141.
16. Barber L, Barrett R, Lichtwark G. Passive muscle mechanical properties of the medial gastrocnemius in young adults with spastic cerebral palsy. *J Biomech.* 2011;44(13):2496-2500.
17. Brandenburg JE, Eby SF, Song P, et al. Ultrasound elastography: the new frontier in direct measurement of muscle stiffness. *Arch Phys Med Rehabil.* 2014;95(11):2207-2219.
18. Bockbrader MA, Thompson RD, Way DP, et al. Toward a Consensus for

Musculoskeletal Ultrasonography Education in Physical Medicine and Rehabilitation: A National Poll of Residency Directors. *Am J Phys Med Rehabil.* 2019;98(8):715-724.
19. Kooijman MK, Swinkels ICS, Koes BW, de Bakker D, Veenhof C. One in six physiotherapy practices in primary care offer musculoskeletal ultrasound - an explorative survey. *BMC Health Serv Res.* 2020;20(1):246.
20. Eby SF, Zhao H, Song P, et al. Quantifying spasticity in individual muscles using shear wave elastography. *Radiol Case Rep.* 2017;12(2):348-352.
21. Park GY, Kwon DR. Application of real-time sonoelastography in musculoskeletal diseases related to physical medicine and rehabilitation. *Am J Phys Med Rehabil.* 2011;90(11):875-886.
22. Boyaci A, Tutoglu A, Boyaci N, et al. Changes in spastic muscle stiffness after botulinum toxin A injections as part of rehabilitation therapy in patients with spastic cerebral palsy. *NeuroRehabil.* 2014;35(1):123-129.
23. Ceyhan Bilgici M, Bekci T, Ulus Y, Bilgici A, Tomak L, Selcuk MB. Quantitative assessment of muscle stiffness with acoustic radiation force impulse elastography after botulinum toxin A injection in children with cerebral palsy. *J Med Ultrason (2001).* 2018;45(1):137-141.
24. Kwon DR, Park GY, Kwon JG. The change of intrinsic stiffness in gastrocnemius after intensive rehabilitation with botulinum toxin a injection in spastic diplegic cerebral palsy. *Ann Rehabil Med.* 2012;36(3):400-403.
25. Park GY, Kwon DR. Sonoelastographic evaluation of medial gastrocnemius muscles intrinsic stiffness after rehabilitation therapy with botulinum toxin a injection in spastic cerebral palsy. *Arch Phys Med Rehabil.* 2012;93(11):2085-2089.
26. Vasilescu D, Vasilescu D, Dudea SM, Sfringeu S, Cosma D. Sonoelastography contribution in cerebral palsy spasticity treatment assessment, preliminary report: a systematic review of the literature apropos of seven patients. *Med Ultrason.* 2010;12(4):306-310.
27. Shao Y-H, Peng Z, Kong X, Wang B, Zhang H. Real-time ultrasound elastography evaluation of achilles tendon properties in patients with mild hemiplegic stroke after rehabilitation training. *J Ultrasound Med.* 2019;38(3):713-723.
28. Lerner RM, Huang SR, Parker KJ. “Sonoelasticity” images derived from ultrasound signals in mechanically vibrated tissues. *Ultrasound Med Biol.* 1990;16(3):231-239.
29. Ophir J, Céspedes I, Ponnekanti H, Yazdi Y, Li X. Elastography: A Quantitative Method for Imaging the Elasticity of Biological Tissues. *Ultrason Imaging.* 1991;13(2):111-134.
30. Bercoff J, Tanter M, Fink M. Supersonic shear imaging: a new technique for soft tissue elasticity mapping. *IEEE Trans Ultrason Ferroelectr Freq Control.* 2004;51(4):396-409.
31. Lima K, Costa Junior JFS, Pereira WCA, Oliveira LF. Assessment of the mechanical properties of the muscle-tendon unit by supersonic shear wave imaging elastography:

a review. *Ultrasonography.* 2018;37(1):3-15.
32. Pedersen M, Fredberg U, Langberg H. Sonoelastography as a Diagnostic Tool in the Assessment of Musculoskeletal Alterations: A Systematic Review. *Ultraschall in Med.* 2012;33(05):441-446.
33. Ryu J, Jeong WK. Current status of musculoskeletal application of shear wave elastography. *Ultrasonography.* 2017;36(3):185-197.
34. Wu C-H, Chen W-S, Park G-Y, Wang T-G, Lew HL. Musculoskeletal Sonoelastography: A Focused Review of its Diagnostic Applications for Evaluating Tendons and Fascia. *J Med Ultrasound.* 2012;20(2):79-86.
35. Sarvazyan A, J Hall T, W Urban M, Fatemi M, R Aglyamov S, S Garra B. An overview of elastography-an emerging branch of medical imaging. *Curr Med Imaging Rev.* 2011;7(4):255-282.
36. Sigrist RMS, Liau J, Kaffas AE, Chammas MC, Willmann JK. Ultrasound Elastography: Review of Techniques and Clinical Applications. *Theranostics.* 2017;7(5):1303-1329.
37. Barr RG. Elastography in clinical practice. *Radiol Clin North Am.* 2014;52(6):1145-1162.
38. Rasool G, Wang AB, Rymer WZ, Lee SS. Altered viscoelastic properties of stroke-affected muscles estimated using ultrasound shear waves—Preliminary data. Paper presented at: Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference.
39. Brum J, Bernal M, Gennisson JL, Tanter M. In vivo evaluation of the elastic anisotropy of the human Achilles tendon using shear wave dispersion analysis. *Phys Med Biol.* 2014;59(3):505-523.
40. Aubry S, Nueffer J-P, Tanter M, Becce F, Vidal C, Michel F. Viscoelasticity in Achilles Tendonopathy: Quantitative Assessment by Using Real-time Shear-Wave Elastography. *Radiology.* 2015;274(3):821-829.
41. Ooi CC, Malliaras P, Schneider ME, Connell DA. "Soft, hard, or just right?" Applications and limitations of axial-strain sonoelastography and shear-wave elastography in the assessment of tendon injuries. *Skeletal Radiol.* 2014;43(1):1-12.
42. Koo TK, Hug F. Factors that influence muscle shear modulus during passive stretch. *J Biomech.* 2015;48(12):3539-3542.
43. Mulabecirovic A, Vesterhus M, Gilja OH, Havre RF. In Vitro Comparison of Five Different Elastography Systems for Clinical Applications, Using Strain and Shear Wave Technology. *Ultrasound Med Biol.* 2016;42(11):2572-2588.
44. Alfuraih AM, O'Connor P, Tan AL, Hensor E, Emery P, Wakefield RJ. An investigation into the variability between different shear wave elastography systems in muscle. *Med Ultrason.* 2017;19(4):392-400.
45. Helfenstein-Didier C, Andrade RJ, Brum J, et al. In vivo quantification of the shear

modulus of the human Achilles tendon during passive loading using shear wave dispersion analysis. *Phys Med Biol.* 2016;61(6):2485-2496.
46. Lacourpaille L, Hug F, Bouillard K, Hogrel JY, Nordez A. Supersonic shear imaging provides a reliable measurement of resting muscle shear elastic modulus. *Physiol Meas.* 2012;33(3):N19-28.
47. Kelly JP, Koppenhaver SL, Michener LA, Proulx L, Bisagni F, Cleland JA. Characterization of tissue stiffness of the infraspinatus, erector spinae, and gastrocnemius muscle using ultrasound shear wave elastography and superficial mechanical deformation. *J Electromyogr Kinesiol.* 2018;38:73-80.
48. Andonian P, Viallon M, Le Goff C, et al. Shear-Wave Elastography Assessments of Quadriceps Stiffness Changes prior to, during and after Prolonged Exercise: A Longitudinal Study during an Extreme Mountain Ultra-Marathon. *PLoS One.* 2016;11(8):e0161855.
49. Gilbert F, Klein D, Weng AM, et al. Supraspinatus muscle elasticity measured with real time shear wave ultrasound elastography correlates with MRI spectroscopic measured amount of fatty degeneration. *BMC Musculoskelet Disord.* 2017;18(1):549.
50. Zhang ZJ, Ng GY, Lee WC, Fu SN. Changes in morphological and elastic properties of patellar tendon in athletes with unilateral patellar tendinopathy and their relationships with pain and functional disability. *PLoS One.* 2014;9(10):e108337.
51. Masaki M, Aoyama T, Murakami T, et al. Association of low back pain with muscle stiffness and muscle mass of the lumbar back muscles, and sagittal spinal alignment in young and middle-aged medical workers. *Clin Biomech (Bristol, Avon).* 2017;49:128-133.
52. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ.* 2009;339:b2700.
53. 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(4):539-549.
54. 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(4):359-369.
55. Oftedal S, Bell KL, Mitchell LE, Davies PS, Ware RS, Boyd RN. A systematic review of the clinimetric properties of habitual physical activity measures in young children with a motor disability. *Int J Pediatr.* 2012;2012:976425.
56. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res.*

2018;27(5):1171-1179.
57. 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(4):651-657.
58. Vet HCWd. *Measurement in medicine : a practical guide.* Cambridge New York: Cambridge University Press; 2011.
59. Jakobsson M, Gutke A, Mokkink LB, Smeets R, Lundberg M. Level of Evidence for Reliability, Validity, and Responsiveness of Physical Capacity Tasks Designed to Assess Functioning in Patients With Low Back Pain: A Systematic Review Using the COSMIN Standards. *Phys Ther.* 2018;99(4):457-477.
60. Prinsen CA, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. *Trials.* 2016;17(1):449.
61. Borenstein M. *Introduction to meta-analysis.* Chichester: John Wiley & Sons; 2009.
62. Borenstein M, Higgins JP, Hedges LV, Rothstein HR. Basics of meta-analysis: I(2) is not an absolute measure of heterogeneity. *Res Synth Methods.* 2017;8(1):5-18.
63. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods.* 2010;1(2):97-111.
64. Bilgici MC, Bekci T, Ulus Y, et al. Quantitative assessment of muscular stiffness in children with cerebral palsy using acoustic radiation force impulse (ARFI) ultrasound elastography. *J Med Ultrason (2001).* 2018;45(2):295-300.
65. Brandenburg JE, Eby SF, Song P, et al. Quantifying passive muscle stiffness in children with and without cerebral palsy using ultrasound shear wave elastography. *Dev Med Child Neurol.* 2016;58(12):1288-1294.
66. Du LJ, He W, Cheng LG, Li S, Pan YS, Gao J. Ultrasound shear wave elastography in assessment of muscle stiffness in patients with Parkinson's disease: a primary observation. *Clin Imaging.* 2016;40(6):1075-1080.
67. Gao J, He W, Du LJ, et al. Quantitative Ultrasound Imaging to Assess the Biceps Brachii Muscle in Chronic Post-Stroke Spasticity: Preliminary Observation. *Ultrasound Med Biol.* 2018.
68. Gao J, He W, Du L-J, et al. Ultrasound strain elastography in assessment of resting biceps brachii muscle stiffness in patients with Parkinson's disease: a primary observation. *Clin Imaging.* 2016;40(3):440-444.
69. Jakubowski KL, Terman A, Santana RVC, Lee SSM. Passive material properties of stroke-impaired plantarflexor and dorsiflexor muscles. *Clin Biomech (Bristol, Avon).* 2017;49:48-55.
70. Kwon DR, Park GY, Lee SU, Chung I. Spastic Cerebral Palsy in Children: Dynamic Sonoelastographic Findings of Medial Gastrocnemius. *Radiology.* 2012;263(3):794-

801.
71. Lacourpaille L, Gross R, Hug F, et al. Effects of Duchenne muscular dystrophy on muscle stiffness and response to electrically-induced muscle contraction: A 12-month follow-up. *Neuromuscul Disord.* 2017;27(3):214-220.
72. Lacourpaille L, Hug F, Guevel A, et al. Non-invasive assessment of muscle stiffness in patients with Duchenne muscular dystrophy. *Muscle Nerve.* 2015;51(2):284-286.
73. Lee SS, Gaebler-Spira D, Zhang LQ, Rymer WZ, Steele KM. Use of shear wave ultrasound elastography to quantify muscle properties in cerebral palsy. *Clin Biomech (Bristol, Avon).* 2016;31:20-28.
74. Lee SS, Spear S, Rymer WZ. Quantifying changes in material properties of stroke-impaired muscle. *Clin Biomech (Bristol, Avon).* 2015;30(3):269-275.
75. Mathevon L, Michel F, Aubry S, et al. Two-dimensional and shear wave elastography ultrasound: A reliable method to analyse spastic muscles? *Muscle Nerve.* 2018;57(2):222-228.
76. Pichiecchio A, Alessandrino F, Bortolotto C, et al. Muscle ultrasound elastography and MRI in preschool children with Duchenne muscular dystrophy. *Neuromuscul Disord.* 2018;28(6):476-483.
77. Wu CH, Ho YC, Hsiao MY, Chen WS, Wang TG. Evaluation of Post-Stroke Spastic Muscle Stiffness Using Shear Wave Ultrasound Elastography. *Ultrasound Med Biol.* 2017;43(6):1105-1111.
78. Eby S, Zhao H, Song P, et al. Quantitative Evaluation of Passive Muscle Stiffness in Chronic Stroke. *Am J Phys Med Rehabil.* 2016;95(12):899-910.
79. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol.* 2003;56(5):395-407.
80. Blain M, Bedretdinova D, Bellin MF, et al. Influence of thoracolumbar fascia stretching on lumbar back muscle stiffness: A supersonic shear wave elastography approach. *Clin Anat.* 2019;32(1):73-80.
81. Alfuraih AM, O'Connor P, Hensor E, Tan AL, Emery P, Wakefield RJ. The effect of unit, depth, and probe load on the reliability of muscle shear wave elastography: Variables affecting reliability of SWE. *J Clin Ultrasound.* 2018;46(2):108-115.
82. Goertz RS, Amann K, Heide R, Bernatik T, Neurath MF, Strobel D. An abdominal and thyroid status with Acoustic Radiation Force Impulse Elastometry – A feasibility study. *Eur J Radiol.* 2011;80(3):e226-e230.
83. Gennisson J-L, Deffieux T, Macé E, Montaldo G, Fink M, Tanter M. Viscoelastic and Anisotropic Mechanical Properties of in vivo Muscle Tissue Assessed by Supersonic Shear Imaging. *Ultrasound Med Biol.* 2010;36(5):789-801.
84. Deffieux T, Montaldo G, Tanter M, Fink M. Shear wave spectroscopy for in vivo quantification of human soft tissues visco-elasticity. *IEEE Trans Med Imaging.*

2009;28(3):313-322.
85. Le Sant G, Nordez A, Andrade R, Hug F, Freitas S, Gross R. Stiffness mapping of lower leg muscles during passive dorsiflexion. *J Anat.* 2017;230(5):639-650.
86. Urban MW, Nenadic IZ, Chen S, Greenleaf JF. Discrepancies in reporting tissue material properties. *J Ultrasound Med.* 2013;32(5):886-888.
87. Ferraioli G, Filice C, Castera L, et al. WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 3: liver. *Ultrasound Med Biol.* 2015;41(5):1161-1179.
88. Hoyt K, Kneezel T, Castaneda B, Parker KJ. Quantitative sonoelastography for the in vivo assessment of skeletal muscle viscoelasticity. *Phys Med Biol.* 2008;53(15):4063-4080.
89. Prado-Costa R, Rebelo J, Monteiro-Barroso J, Preto AS. Ultrasound elastography: compression elastography and shear-wave elastography in the assessment of tendon injury. *Insights Imaging.* 2018;9(5):791-814.
90. Cosgrove D, Piscaglia F, Bamber J, et al. EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. Part 2: Clinical applications. *Ultraschall Med (Stuttgart, Germany : 1980).* 2013;34(3):238-253.
91. Gennisson JL, Deffieux T, Fink M, Tanter M. Ultrasound elastography: principles and techniques. *Diagn Interv Imaging.* 2013;94(5):487-495.
92. Bamber J, Cosgrove D, Dietrich CF, et al. EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. Part 1: Basic principles and technology. *Ultraschall Med.* 2013;34(2):169-184.
93. Drakonaki EE, Allen GM, Wilson DJ. Ultrasound elastography for musculoskeletal applications. *Br J Radiol.* 2012;85(1019):1435-1445.
94. Ariji Y, Nakayama M, Nishiyama W, Nozawa M, Ariji E. Shear-wave sonoelastography for assessing masseter muscle hardness in comparison with strain sonoelastography: study with phantoms and healthy volunteers. *Dentomaxillofac Radiol.* 2016;45(2):20150251.
95. Ewertsen C, Carlsen JF, Christiansen IR, Jensen JA, Nielsen MB. Evaluation of healthy muscle tissue by strain and shear wave elastography - Dependency on depth and ROI position in relation to underlying bone. *Ultrasonics.* 2016;71:127-133.
96. Whittaker JL, Stokes M. Ultrasound imaging and muscle function. *J Orthop Sports Phys Ther.* 2011;41(8):572-580.
97. Sions JM, Tyrell CM, Knarr BA, Jancosko A, Binder-Macleod SA. Age- and stroke-related skeletal muscle changes: a review for the geriatric clinician. *J Geriatr Phys Ther.* 2012;35(3):155-161.
98. Chuang LL, Wu CY, Lin KC. Reliability, validity, and responsiveness of myotonometric measurement of muscle tone, elasticity, and stiffness in patients with stroke. *Arch*

Phys Med Rehabil. 2012;93(3):532-540.
99. Marusiak J, Jaskólska A, Budrewicz S, Koszewicz M, Jaskólski A. Increased muscle belly and tendon stiffness in patients with Parkinson's disease, as measured by myotonometry. *Mov Disord.* 2011;26(11):2119-2122.
100. Feng YN, Li YP, Liu CL, Zhang ZJ. Assessing the elastic properties of skeletal muscle and tendon using shearwave ultrasound elastography and MyotonPRO. *Sci Rep.* 2018;8(1):17064.
101. Lorentzen J, Grey MJ, Crone C, Mazevet D, Biering-Sorensen F, Nielsen JB. Distinguishing active from passive components of ankle plantar flexor stiffness in stroke, spinal cord injury and multiple sclerosis. *Clin Neurophysiol.* 2010;121(11):1939-1951.
102. Kallenberg LA, Hermens HJ. Motor unit properties of biceps brachii in chronic stroke patients assessed with high-density surface EMG. *Muscle Nerve.* 2009;39(2):177-185.

Figure legends

Figure 1. Flow diagram

Flow diagram illustrating article screening and selection in accordance with PRISMA guidelines [52]. A total of 21 articles were included in the final review.

Figure 2. Forest plot

Graphical summary of subgroup analyses for studies involving patients with CP or stroke. Correlation coefficients were transformed to Fisher's Z scale for analysis. Overall correlation across subgroups was r=0.78 (95% CI=0.64–0.86, p≤0.00).

Table 1. Study characteristics

Study	Sample Characteristics		Measurement			
	Neurological Group	Control Group	Muscle Site(s)	Passive/ Active	Body Position/ Joint Angle	Assessor/ Blinding
Bilgici 2018(a) - Cross sectional - (Turkey)	17 patients w/ CP (32 legs) 2 = hemiplegic MAS 1 = 3 (1.72 (0.35)), MAS 2 = 8 (2.79 (0.61)), MAS 3 = 17 (3.31 (0.50)), MAS 4 = 4 (4.37 (0.19)) m/f = 9/8 age: 9.25 (2.68)y, range: 6–14y weight: 24.53 (7.77)kg, range: 17–45kg	25 controls (50 legs) children of similar age w/o systemic/ neurological disease visiting hospital for other reasons m/f = 10/15 age: 10.40 (2.76)y, range: 7-15y weight: 36.48 (13.42)kg, range: 20-63kg	MG	Passive, no EMG	Prone position w/ feet in neutral position (0°) over the edge of the examination table	**Assessor:** Radiologist w/ 4ys of UE experience **Blinding:** Blinded to patient MAS scores
Bilgici 2018(b) – Pre-post intervention - (Turkey)	12 patients w/ CP (24 legs) m/f = 6/6 age: 8.58 (2.48)y, range: 6–14y weight: 21.83 (5.13)kg	N/A	MG	Passive, no EMG	Prone position w/ feet in neutral position (0°) over the edge of the examination table	**Assessor:** Radiologist w/ 4ys of UE experience **Blinding:** Blinded to patient MAS scores
Boyaci 2014 – Pre-post intervention - (Turkey)	16 CP children (25 legs) 9 = diplegic, 7 = hemiplegic m/f = 11/5 age: 48.87 (16.47)mo weight: 15.87 ± 3.55kg	17 children w/o neurological history (34 legs) age: 45.52 (20.13)mo weight: 17.29 (5.92)kg	MG, LG, SOL	Passive, no EMG	Prone position w/ feet over the edge of the examination couch	**Assessor:** Radiologist w/ 6ys of ultrasound experience **Blinding:** Not explicitly stated
Brandenburg 2018 - Prospective longitudinal cohort (intervention) - (USA)	9 children w/ spastic CP bilateral = 5, unilateral = 4 GMFCS: level I = 4, level II = 3, level III = 2 m/f = 5/4 median age: 60 (35-92)mo, age range: 25-105mo median BMI: 16.7 (15.9-17.9), BMI range: 15.3–18.6	N/A	LG (most affected leg)	Passive w/ sEMG (U-Control; Thought Technology Ltd.) set at lowest scale range (X1, upper threshold 3.0)	Prone position w/ feet over the edge of the examination table	**Assessor:** Not stated, assumed to be the researcher **Blinding:** No explicitly stated
Brandenburg 2016 – Cross sectional - (USA)	13 children w/ CP m/f = 7/6 median age: 5y 1m, IQ range: 4y 4m-7y 8m	13 typically developing children m/f = 7/6 median age: 5y 3m, IQ range: 4y 4m - 9y 4m	LG (most affected leg)	Passive w/ sEMG (U-Control; Thought Technology Ltd.) set at lowest scale range (X1, upper threshold 3.0)	Prone position w/ feet over the edge of the examination table	**Assessor:** Not stated, assumed to be the researcher **Blinding:** Not explicitly stated
Du 2016 – Cross sectional - (China)	46 patients w/ PD (British Brain Bank clinical criteria) m/f = 27/19 age: 47.9 (2.8)y	31 healthy controls m/f = 18/13 age: 46.7 (3.2)y	BB	Passive, no EMG	Supine position w/ limbs kept in full relaxation	**Assessor:** Not stated, assumed to be 2 researchers **Blinding:** Not explicitly stated
Eby 2016 – Cross sectional - (USA)	9 subjects w/ chronic stroke m/f = 7/2 age: 58.3y, range: 41-79y	4 healthy controls m/f = 2/2 age: 56y, range: 42-70y	BB (long head)	Passive w/ sEMG (MA-300, Motion Lab Systems) and a dynamometer	80° and 150° elbow flexion/extension and at 3 preselected joint angles ranging btw 85° and 150°	**Assessor:** Not stated, assumed to be the same rater/ researcher **Blinding:** Not explicitly stated
Gao 2018 - Cross sectional - (USA)	8 patients w/ stroke m/f = 5/3 age: 59y, range: 34-72y	8 healthy controls m/f = 4/4 age: 49y, range: 40-56y	BB	Passive, w/ the use of VTIQ quality map to monitor patient or probe motion during image capture, no EMG	Supine position w/ arm relaxed and forearm supinated	**Assessor:** A single observer (J.G.) acquired VTIQ SWV images **Blinding:** Not explicitly stated
Gao 2016 - Cross sectional - (USA)	14 patients w/ PD disease duration: 78 (13)mo, range: 6-134mo muscle rigidity scores: high (UPDRS III–IV) = 3, low (UPDRS I–II) = 11 m/f = 8/6 age: 61 (10)y, range = 41-78y	10 healthy controls m/f = 5/5 age: 60 ± 11y, range: 54-82y	BB	Passive, probe and subject arms were held stable by the researchers to minimize movement during compression cycles, no EMG	Supine position w/ arm relaxed, elbow extended and the forearm supinated (forearm elevated 15° from the bed)	**Assessor:** Physician w/ experience in MSK ultrasound. For reliability measures, 2 observers assessed 10 healthy controls **Blinding:** Blinded to UPDRS motor score and disease duration, aware of symptomatic patients w/ PD
Jakubowski 2017 - Cross sectional -	14 subjects w/ chronic stroke stroke duration: 10.6 (7.3)y, range: 4.3-	N/A	MG, TA	Passive, w/ EMG (Bagnoli, Delsys Inc.)	Seated w/ foot secured to a dynamometer (Biodex Medical systems Inc.) and knee in extension, bilateral ankle moved passively by the	**Assessor:** Described as an experimenter **Blinding:** Not explicitly stated

(USA)	29.3y FMA: 19.1 (6.1), range: 8-28 m/f = 6/8 age: 60.1 (5.9)y, range: 46–68y height: 1.7 (0.1)m, range: 1.5-1.8m body mass: 77.6 (12.5)kg, range = 58.0–96.4kg				experimenter in 6 positions: neutral 90°, 15° PF, max DF, max PF, and 2 intermediate angles where the torque on the paretic side was btw max DF and neutral or max PF and neutral, respectively	
Kwon 2012 - Cross sectional - (South Korea)	15 children w/ spastic CP (27 legs) diplegia = 12, hemiplegia = 3 m/f = 10/5 age: 58.7 (20.6)mo	13 children w/o neurological or MSK disability (26 legs) m/f = 4/9 age: 46.9 (20.2)mo	MG, SOL	Passive, no EMG	Prone w/ feet hanging from the edge of an examination plinth	**Assessor:** Physiatrist (G.Y.P.) w/ 16ys of MSK ultrasound and 3ys of UE experience **Blinding:** Not explicitly stated
Lacourpaille 2017 – Longitudinal - (France)	10 patients w/ DMD (genetically confirmed) age: 13.6 (6.3)y, range: 7-23y	9 age matched healthy controls	MG, TA, VL, BB, TB, and ADM	Passive for stiffness measures and active for EMD measures, a constant-current stimulator (Digitimer DS7A, Digitimer) was used to elicit contractions	Seated. MG: knee flexed at 90° (shortened) or fully extended (stretched) w/ ankle in neutral, TA: knee extended fully w/ ankle in neutral or 20° PF, VL: knee fully extended or flexed at 90°, BB: elbow flexed at 90° or overextended w/ hand in neutral, TB: arm extended or abducted and flexed at 90°, ADM: hand in pronation w/ 5th finger in maximal abduction or in alignment w/ 5th metacarpal	**Assessor:** Not stated, assumed to be the researcher. **Blinding:** Not explicitly stated
Lacourpaille 2015 - Cross sectional - (France)	14 patients w/ DMD age: 13.3 (5.9)y, range 5–22y	13 age-matched healthy controls age: 12.8 (5.5)y, range 6-24y	MG, TA, VL, BB, TB, and ADM	Passive, no EMG	Lying on a plinth. MG: knee flexed at 90° (shortened) or fully extended (stretched) w/ ankle in neutral, TA: knee extended fully w/ ankle in neutral or 20° PF, VL: knee fully extended or flexed at 90°, BB: elbow flexed at 90° or overextended w/ hand in neutral, TB: arm extended or abducted and flexed at 90°, ADM: hand in pronation w/ 5th finger in maximal abduction or in alignment w/ 5th metacarpal	**Assessor:** Described as an examiner, assumed to be the researcher **Blinding:** Not explicitly stated
Lee 2018 - Cross sectional - (USA)	14 subjects w/ chronic stroke stroke duration: 10.2 (8.4)y, range: 2.1-27.3y FMA: 19.6 (15.0), range: 4-48 MAS range: 0-3 TS: 1–3 muscle quality, 62°-145° catch angle for 3 speeds m/f = 5/9 age: 58.9 (7.4)y height: 1.68 (0.10)m body mass: 85.5 (18.2)kg	8 age- and sex-matched controls w/o neurological or muscular disorders m/f = 4/4 age: 57.4 (7.4)y height: 1.68 (0.11)m body mass: 75.0 (12.0)kg	BB	Active and Passive w/ sEMG (Bagnoli Delsys, Inc.)	Seated in a dynamometer (Biodex Medical Systems Inc.), upper arm resting on a plastic support, forearm secured in a fiberglass cast w/ wrist and forearm in neutral position and placed in a ring-mount interface mounted on the table, shoulder positioned w/ humerus abducted 45° and elbow at 90° of flexion	**Assessor:** Not clearly stated, assumed to be the researcher. **Blinding:** Not explicitly stated
Lee 2016 - Cross sectional - (USA)	8 subjects w/ spastic hemiplegic CP GMFCS: level I = 3, level II = 5 m/f = 5/3 age: 9.4 (3.7)y height: 1.31 (0.17)m body mass: 33.3 (12.8)kg	N/A	MG and TA	Passive w/ EMG	Seated in an IntelliStretch rotary actuator (IntelliStretch Rehabilitation Robot, Rehabtek LLC) to monitor ankle angle and torque continuously, knee placed in max extension w/ foot strapped to the device, MG and TA measures taken in 5 ankle positions (neutral, maximum DF, maximum PF, and 2 intermediary angles)	**Assessor:** Not clearly stated, assumed to be the researcher. **Blinding:** Not explicitly stated
Lee 2015 - Cross sectional - (USA)	16 subjects w/ chronic stroke stroke duration: 11.6 (11.4)y, range: 1.9–42.2y FMA: 19 (15), range: 4-48 MAS range: 0-3 TS: 1-3 muscle quality, 62°-145° catch angle for 3 speeds m/f = 6/10 age: 60.7 (8.0)y height: 1.71 (0.15)m body mass: 85.5 (18.2)kg	N/A	BB	Passive w/ sEMG	Seated in a dynamometer (Biodex Medical Systems Inc.), upper arm resting on a plastic support, forearm secured in a fiberglass cast w/ wrist and forearm in neutral position and placed in a ring-mount interface mounted on the table, shoulder positioned w/ humerus abducted 45° and elbow at 90° of flexion	**Assessor:** Not clearly stated, assumed to be the researcher. **Blinding:** Not explicitly stated
Mathevon 2018 – Observational/ Reliability - (France)	14 subjects w/ stroke stroke duration: 39mo, range: 6-255mo paretic side left/right: 6/8 m/f = 10/4 age: 56.9 (10.8)y height: 170.9 (8.9)cm weight: 80.9 (13.2)kg	N/A	MG and TA	Passive and during DF, no EMG	TA at rest and max DF: Supine. MG at rest: prone w/ feet below the table and no hip rotation at maximum passive-ankle DF w/ full knee extension. Stretching was performed manually by a second investigator (L.F.A.) in all trials.	**Assessor:** Described as an experimental investigator, assessed each patient twice at an interval of 1 week at the same time of the day **Blinding:** Not explicitly stated

Park 2012 – Pre-post intervention - (South Korea)	17 children w/ CP diplegia = 12, hemiplegia = 5 m/f = 10/7 age: 57 (22)mo, range: 26-110mo weight: 12.5kg, range: 11-15kg	N/A	MG	Passive, no EMG	Prone w/ feet hanging from the edge of an examination table	**Assessor:** Physiatrist w/ 17ys of MSK ultrasound and 4ys of RTS experience **Blinding:** image analysis conducted by another physiatrist
Pichiecchio 2018 - Cross sectional - (Italy)	5 children w/ DMD (clinical and molecular diagnosis of dystrophinopathy) m/f = 5/0 median age: 48mo, range 38-59mo BMI <30	5 age-matched healthy controls m/f = 4/1 median age: 39mo, range: 39-47mo BMI <30	GM, RF, VM, VL, AM, TA, and MG	Passive, no EMG	Supine on an examination bed (or in their mother's arms) and then in prone position	**SWE Assessors:** 2 radiologists (C.B., F.C.), w/ 5y of MSK ultrasound experience **Blinding:** blinded to subject characteristics **MRI Assessors:** 2 radiologists (A.P., F.A.), one w/ considerable expertise in neuromuscular disorders and the other, a resident, w/ 4ys of MRI experience
Shao 2019 - Pre-post intervention (China)	24 patients w/ mild hemiplegic stroke and impaired plantar flexion (MAS ≥ 2) stroke duration: 10.1 (3.7)mo m/f = 14/10 age: 60.7 (8.8)y, range: 41–75y	N/A	AT	Passive, no EMG	Prone w/ feet hanging from the edge of an examination table	**Assessor:** MSK radiologist w/ > 6y of MSK ultrasound imaging experience **Blinding:** Not explicitly stated
Wu 2016 - Cross sectional - (Taiwan)	31 patients w/ acute stroke stroke duration < 3m in 29 patients stroke duration: 8.4 (7.6)wks infarct = 19, hemorrhage = 12 paretic right/left = 11/20 m/f = 21/10 age: 60.3 (13.0)y height: 164.7 (6.9)m weight: 63.4 (12.3)kg	21 healthy controls m/f = 14/7 age: 31.2 (7.9)y height: 168.5 (7.1)m weight: 168.5 (7.1)kg	BB	Passive w/ sEMG used to monitor muscle activity	Supine on an examination bed w/ shoulders and elbows in a relaxed neutral position. SWV obtained at 0° (full extension) and 90° of elbow flexion using an custom elbow stabilizer	**Assessor:** Physiatrist w/ 2ys of MSK ultrasound experience (> 1000 cases), familiarized w/ the study protocol for optimization of SWV measures by examining 15 unimpaired subjects not included in the study (preliminary test) **Blinding:** Physical appearance of patients and controls prevented blinding of the rater to paretic limbs

Table 2. System specifications and acquisition procedures

Study	System Specifications					Acquisition Method				
	System	Probe	Settings/ Software	Units Reported	Probe Alignment	ROI	Number of Trials	Transducer Pressure/ Placement	Contact Interface	Image Acquisition/ Processing
Bilgici 2018(a)	Acuson S2000 US system (Siemens Medical)	linear array (9L4)	N/A	SWV (m/s) (range of 0-9)	parallel	ROI box 0.5 cm^2 in size placed in the mid-section of the MG corresponding to the largest circumference	5 measures for each muscle w/ the mean value used for analysis	minimal compression applied w/ the probe weight	N/A	Value ‘‘x.xx’’ displayed in case of faulty measures, repeated until valid values were obtained
Bilgici 2018(b)	Acuson S2000 US system (Siemens Medical)	linear array (9L4)	VTIQ software	SWV (m/s) (range of 0-9)	parallel	ROI box 0.5 cm^2 in size placed in the mid-section of the MG corresponding to the largest circumference	5 measures for each muscle w/ the mean value used for analysis	minimal compression applied w/ the probe weight	N/A	Value ‘‘x.xx’’ displayed in case of faulty measures, repeated until valid values were obtained, VTIQ software
Boyaci 2014	Mylab Twice US system w/ sonoelastography and Doppler (Esaote)	broadband linear array (12MHz)	N/A	ELX 2/1 Index and color pattern pixel intensities	parallel	ROI of 7.5×7.5 mm^2 in size, ELX 2/1 index calculated as a ratio of the elastic properties of the MG, LG and SOL w/in the ROI	N/A	N/A	N/A	Pixel color pattern in 0 to 255 range, w/ median blue, green, and red pixel histogram intensities analyzed w/ ImageJ software (National Institutes of Health)
Brandenburg 2018	Aixplorer US scanner (SuperSonic Imagine)	linear array (SL15-4; SuperSonic Imagine)	MSK preset	SM (kPa)	parallel	circular ROI w/ mean diameter of 4.8–5.0mm w/in the elastogram	measures at 3 foot positions and repeated twice, average of 3 measures used for analysis	minimal pressure on skin by one of the examiners	N/A	Open-source imaging plug-in for the DICOM reader (OsiriX Imaging Software) used to measure modulus values w/in the ROI
Brandenburg 2016	Aixplorer US scanner (version 4.0; SuperSonic Imagine)	linear array (SL15–4; SuperSonic Imagine)	MSK preset, SWE Optimization: Standard, HD/ Frame Rate: Balanced, Zoom: 120%, Smoothing: 5, Persistence: High	SM (kPa)	parallel	circular ROI w/ mean diameter of 4.8–5.0mm w/in the elastogram	measures at 3 foot positions and repeated twice, average of 3 measures used for analysis	minimal pressure on skin by one of the examiners	N/A	Open-source imaging plug-in for the DICOM reader (OsiriX Imaging Software) used to measure modulus values w/in the ROI
Du 2016	Aixplorer US scanner (SuperSonic Imagine)	linear array (4–15 MHz)	MSK setting, Q-Box software	SM (kPa)	parallel	5mm diameter circle in the ROI center w/ homogeneous color distribution for a 3 seconds minimum	3 measures per muscle and side w/ the average 3 measures used for analysis	gently placed on skin w/o compression	N/A	Q-Box software
Eby 2016	Verasonics US scanner (Verasonics Inc.)	linear array (L7-4, Philips Healthcare)	CUSE and SWE settings	SM (kPa) w/ color bar range 0-16 (m/s)	assumed to be parallel	approx. 160 mm^2 ROI using CUSE	3 measures at 80° and 150° elbow angles	custom holder securely fixed to the arm maintaining continuous contact pressure	transmission gel couplant applied	Shear wave motion recorded using a high frame rate technique and calculated from image data based on one-dimensional autocorrelation
Gao 2018	Acuson S3000 HELX (Siemens Medical)	linear array (9L4)	pre-locked settings w/ mechanical index 1.1, depth 4cm, scanning frequency 7 MHz, Map D/Space time (0), total gain 0dB, dynamic range 70dB, single focus, harmonic imaging kept constant, VTIQ software	SWV (m/s)	parallel	ROI of 1.5 × 1.5 mm, 1-3 cm depth from the skin w/in a color coded SWV map	10 trials w/ 5 ROIs of 1.5cm depth and 5 ROIs of 2.0cm depth from skin surface at both 90° passive elbow flexion and max passive elbow extension (up to 180°)	N/A	transmission gel couplant applied	VTIQ software
Gao 2016	Logic E9 US scanner (General Electric)	linear array (L9-3)	grayscale imaging settings for SE optimized for speckle tracking, high frame rate (N40 frames per second), single focus, turn off speckle reduction, low scanning frequency (6 MHz)	SR	parallel	Reference strain standardized to 5mm axial region in subcutaneous tissue (distance from skin to muscle)	3 compression cycles w/ 2 minute time interval btw trials	sand bag (1.5 kg) tied to probe for constant compression, 5 second compression cycles used to during tissue deformation sequence	transmission gel couplant applied	Strain of target muscle and reference tissues estimated w/ 2-D speckle tracking software (EchoInsight, Epsilon Imaging)
Jakubowski 2017	Aixplorer US system (SuperSonic Imagine)	linear array (4–15 MHz, Super Linear 15-4)	N/A	SWV (m/s)	parallel	ROI of 30mm width w/ depth set to muscle thickness	3 trials per ankle position	custom neoprene sleeve used to minimize probe movement	N/A	images manually cropped and exported for off-line processing using custom-written program in Matlab (Mathworks)

Kwon 2012	Acuson S2000 US system w/ B-mode and DS (Siemens Medical)	multi frequency linear (4–9 MHz)	N/A	SWV (m/s), SR, DS score and color pixel intensity	parallel	ROI of 5 x 5 mm^2 for SWV acquisition, ROI for pixel analysis set to cover entire muscle, excluding hyperechoic epimysium	2 trials w/ 2 representative images taken during each	compression adjusted according to quality factor display, factor ≥ 60 indicated optimal compression	N/A	color pattern (0 to 255 pixel range) of images analyzed w/ Image J software (National Institutes of Health), median blue and red pixel intensities obtained w/ color histogram
Lacourpaille 2017	Aixplorer US scanner (version 7, SuperSonic Imagine)	linear array (4–15 MHz)	SWE and research mode to acquire raw radio-frequency signals at 4 kHz, force and US data synchronized w/ transistor to transistor logic pulses	SM (kPa)	parallel	ROI corresponded to the largest muscular region w/o fascia	10 trails for each muscle and position, then averaged to obtain representative values	probe placed on thickest part of muscle belly	N/A	SSI recordings exported from system software (Version 7.0, SuperSonic Imagine) as mp4 and sequenced in jpeg format w/ custom program in Matlab (version 10.0, Mathworks Inc.), color maps were converted to SM
Lacourpaille 2015	Aixplorer US scanner (SuperSonic Imagine)	linear array (4–15 MHz)	SWE mode	SM (kPa)	not stated, not inferable due to lack of images and figures depicting alignment	N/A	10 trails for each muscle and position, then averaged to obtain representative values	probe placed on thickest part of muscle belly	N/A	N/A
Lee 2018	Aixplorer US system (SuperSonic Imagine)	Not explicitly stated	N/A	SWV (range 0-16 m/s)	parallel	circular region w/ varying diameters w/in an ROI of 12 x 12mm btw superficial and deep aponeuroses	3 trials per % max voluntary contraction	custom neoprene sleeve used to minimize probe movement	N/A	images cropped and muscle area w/in the ROI analysed w/ a custom program in Matlab (Mathworks Inc.)
Lee 2016	Aixplorer US system (SuperSonic Imagine)	linear array (4–15 MHz, Super Linear 15-4)	SSI software (Q Box)	SWV (m/s)	parallel	circular region w/ varying diameters w/in an ROI of 12 x 12mm, all manually cropped areas w/in the muscle used to calculate SWV spatial averages	60 trials (2 legs, 2 muscles, 15 trials per muscle)	custom neoprene sleeve used to minimize probe movement, placed over muscle belly mid-region	N/A	image were analysed w/ custom program in Matlab (Mathworks Inc.), Q Box software used to generate SWV values w/ a quality factor > 0.8.
Lee 2015	Aixplorer US system (SuperSonic Imagine)	linear array (4–15 MHz, SuperLinear 15–4)	SSI software (Q Box)	SWV (m/s)	parallel	ROI of 12 x 12mm placed in the mid-section btw superficial and deep aponeuroses	N/A	custom neoprene sleeve used to minimize probe movement	N/A	SWV and quality factor values in the ROI extracted and analysed w/ custom program in Matlab (Mathworks)
Mathevon 2018	Aixplorer US scanner (version 6.1.1, SuperSonic Imagine)	linear array (4–15 MHZ, SuperLinear 15–4)	2D muscle measures w/ B-mode US, data processed w/ SSI software (Q-Box)	SM (kPa)	parallel (sagittal) and perpendicular (axial)	3 circular ROI (Q-boxes) positioned over the largest contiguous area possible	mean value of 5 measures over 5 seconds used for analysis	compression limited by a thick layer of gel and the support of the examin table	transmission gel couplant applied	data processed w/ system software (Q-Box)
Park 2012	Antares US system w/ B-mode and RTS (Siemens Medical)	multi frequency linear (5-13 MHz)	image color pattern analyzed w/ Image J software (National Institutes of Health)	RTS score	parallel	ROI set to cover the entire muscle, excluding hyperechoic perimysium	N/A	manually adjusted compression, perimysium appeared yellow to red on RTS, w/ standardized color scale encoding	N/A	image color pattern analyzed w/ ImageJ software (National Institutes of Health), RTS score graded semi-quantitatively as follows: 1 (purple to green: soft), 2 (green to yellow), 3 (yellow to red), and 4 (red: hard)
Pichiecchio 2018	Toshiba Aplio 500 SWE US scanner (Toshiba Medical)	multi frequency linear array (4-15 MHz)	shear wave module (Toshiba Medical)	SM (kPa)	Assumed to be parallel	Circular ROI of 5mm w/ homogeneous stiffness values w/in an area void of tendon, fascial tissue	N/A	probe placed in a fixed position	transmission gel couplant applied	N/A
Shao 2019	Acuson S2000 US system (Siemens Medical)	linear array (5–14 MHz, Siemens)	elastographic unit w/ strain quality indicator (proprietary software, Siemens)	SR and elasticity score	parallel	Depth 3x AT thickness, SR ROI 2cm above insertion w/ corresponding deep fat reference area	3 trials to calculate SR	appropriate pressure applied	10mm gel pad (SONAR-AID)	Score/ grade calculated accordingly: SR=B/A (fat-to-tendon SR), elasticity: 3=red (soft), 2=green/ yellow (medium), 1=blue (hard)
Wu 2016	Acuson S2000 US system (Siemens Medical)	linear array (7-9 MHz, 9L4, Siemens)	Virtual Touch Quantification Technology (VTQ) (Siemens)	SWV (m/s)	parallel (longitudinal) and perpendicular (transverse)	ROI of 0.5 x 0.5cm in the mid-muscle belly	5 trials per arm in each elbow posture and averaged for analysis	probe held stationary during acquisition	transmission gel couplant applied	N/A

1

Table 3. Reliability and validity summary

Study	Reliability	Validity			
		Convergent/ Concurrent	Discriminant/ Known Groups	Responsiveness	Comparator(s)
Bilgici 2018(a)	**Interrater Reliability:** ICC=.65 (agreement, 95% CI=.33–.84, p=.001)	**Convergent Validity:** SWV values of MG correlated w/ ankle spasticity MAS scores (p=.001)	**Known Groups Validity:** Mean SWV values for the MG in patients w/ CP (3.17±.81, min–max 1.291-4.540) were significantly higher than controls (1.45±.25, min–max .938–2.080, p=.001)	N/A	Controls and MAS
Bilgici 2018(b)	**Interrater Reliability:** ICC=.65 (agreement, 95% CI=.33–.84, p=.001)	**Convergent Validity:** SWV values positively correlated w/ MAS score (q=.578, p=0.003)	N/A	**Responsiveness:** Significant difference in mean SWV values pre-BTA (3.20±.14) and post-BTA (2.45±.21, p=.001)	SWV and MAS measured pre-BTA and 1 mo post
Boyaci 2014	N/A	**Convergent Validity:** ELX 2/1 index was correlated w/ median red pixel intensity in the CP group (r=.516, p=.008). Mean MAS for the ankle decreased, from 3.4 to 2.6 (p<.05). Mean MAS scores decreased significantly from pre- (3.44±.58) to post-trial (2.60±.64, p<.001), while mean GMFM scores increased significantly from pre- (54.28±19.19) to post-trial (59.03±16.49, p=.001)	**Known Groups Validity:** ELX 2/1 indices of the CP group (MG=2.27±.88, LG=1.84±.85) were significantly higher than controls (MG=1.12±.27, LG=1.17±.39, p<.05)	**Responsiveness:** ELX 2/1 indices in the GM and GL muscles in the CP group decreased significantly post-treatment (p<.05)	Controls, pixel intensity, ELX 2/1 index, MAS and GMFM scores measured pre-procedure and 1 month post
Brandenburg 2018	N/A	**Convergent Validity:** Spearman rank correlations used to explore the relationship btw SM and continuous variables (BoNT-A dose, and ankle DF passive ROM). However, no outcomes were reported	N/A	**Responsiveness:** Despite no significant change in ankle ROM or spasticity, there was a significant difference in LG SM after BoNT-A injections	MAS, max ankle DF ROM, GMFCS
Brandenburg 2016	N/A	**Convergent Validity:** No correlation btw SM values and GMFCS level, MAS grade, or history of calf muscle botulinum toxin injection. There was a significant difference in median ankle ROM btw the CP group (-12° to 20°) and controls (5° to 31°)	**Known Groups Validity:** SM at all 3 foot positions were significantly greater for CP children (20° PF = 15.0 (11.6, 17.5),10° PF = 19.1 (15.0, 23.6), 0° PF = 28.9 (24.6, 44.2)) than controls (20° PF = 7.8 (6.1, 11.0),10° PF = 9.6 (7.3, 15.6), 0° PF = 14.9 (10.9, 20.9)). CP children had greater variability in SM, indicated by larger SD for measures at all ankle positions	N/A	Healthy controls (age- and gender-matched typically developing children), MAS, maximal ankle DF ROM, GMFCS
Du 2016	(9 control subjects only) **Interrater Reliability:** ICC(3,2)=.74 (95% CI=.68–.78) **Intrarater Reliability:** ICC(3,1)=.78 (95% CI=.75–.82)	**Convergent Validity:** Positive linear correlation found btw SM values and UPDRS motion scores in patients w/ PD (r=.646, p=.000)	**Known Groups Validity:** Significant difference in SM of the BB btw PD (54.94±20.91) and controls (24.44±5.09, p<.05) **Discriminant Validity:** No significant difference btw remarkably (54.94±20.91) and mildly symptomatic arms (47.77±24.00, p<.05)	N/A	Remarkably and mildly symptomatic arms in patients w/ PD, btw patients and controls, and btw SM and UPDRS
Eby 2016	**Intrarater Reliability:** ICC(1,1) range=.76-.99 at both 80° and 150° for all subjects ICC(1,1) range=.75-.97 for 3 preselected joint angles (ICCs indicated consistent stiffness throughout testing for dominant sides of controls, but largely inconsistent stiffness for other study conditions)	**Convergent Validity:** No association was found btw stroke mechanism, location, or hemisphere and MAS or FMA scores. SM values and torque during all 40°/sec trials: A= controls, w/ minimal torque and SM responses; y=17.394x + 7.898, R^2=.103. B= S1, w/ strong torque and SM responses to passive extension; y=36.856x + 18.197, R^2=.829, and C= S6, S3, S7, w/ strong torque response and minimal SM responses; y=2.712x + 6.676, R^2=.181.	**Known Groups Validity:** Torque and passive stiffness increased minimally for controls and was most pronounced in contralateral limbs of subjects w/ stroke. Several patterns of SM and torque responses to passive elbow extension were identified, w/ a subset of several subjects displaying very strong torque response w/ minimal stiffness response	N/A	Healthy controls and velocity dependent torque
Gao 2018	**Intrarater Reliability:** ICC=.94 (agreement, p<.001) for non-spastic BB ICC=.82 (agreement, p<.01) for spastic BB **Software Usage:** VTIQ indicates SWV quality and reliability w/ homogeneous green color scale throughout image maps	**Convergent Validity:** A strong negative correlation was found btw SWV and passive ROM (R^2=−.88, p<.0001) in spastic upper limbs. The correlation btw mean SWV and other MAS and TS parameters was weak (p>.05).	**Known Groups Validity:** At 90°, there was no significant difference in SWV btw controls and spastic BB or btw non-spastic and spastic BB. However, there was a significant difference btw controls and non-spastic BB based on Bonferroni correction. At max elbow extension, the difference in SWV btw healthy and spastic BB, and btw non-spastic and spastic BB, was significant (all values p<.01), but not significant btw controls and non-spastic BB (p>.05)	N/A	Controls (btw subjects), non-paretic arm (w/in subjects), SWV, ROM, MAS, TS and echogenicity across 3 groups (healthy, non-spastic and spastic BB muscles)
Gao 2016	**Intrarater Reliability:** ICC=.88 **Interrater Reliability:** ICC=.84	**Convergent Validity:** A negative correlation was found btw SR values and UPDRS scores (r=−.78)	**Known Groups Validity:** There was a significant difference in SR btw patients w/ PD (2.65±.36) and controls (3.30±.27). There was also no significant gender or side-to-side difference in SR among controls	N/A	Controls and UPDRS for muscle rigidity
Jakubowski 2017	N/A	**Convergent Validity:** Weak correlations were found btw MG SWV and joint stiffness for non-paretic (r=.384, p=.001) and paretic sides (r=.363, p=.002). Ankle angle, joint torque, and fascicle strain were significantly correlated w/ MG and TA SWV. (SWV vs ankle angle: MG=.705, TA=−.574, p<.001), (SWV vs	**Discriminant Validity:** There were significant increases of 27.7% and 26.9% in SWV for the paretic compared to the non-paretic MG at 90° (p=.033) and 15° PF (p=.001). However, no significant difference was found btw-sides for the TA. Paretic MG and TA SWV at torque-matched position btw max PF and	N/A	Joint torque, ankle angle, fascicle strain, and estimates of clinical measures (FMA, passive/ active ROM), joint stiffness, muscle

		joint torque GM=.626, TA=−.475, p<.001), (SWV vs fascicle strain: MG=.665, TA=.397, p<.001). FMA score was correlated w/ SWV for the paretic TA at 15° PF (r=−.586, p=.035). No correlations were found for the MG at any ankle angle. As muscle thickness increased, SWV decreased for both non-paretic (r=−.565, p=.035) and paretic MG (r=−.648, p=.012) at 90°	neutral was 18.7% (p=.109) greater and 14.7% (p=.109) less than non-paretic sides. SWV of paretic MG and TA at torque-matched position btw neutral and max DF were 16.8% (p=.033) greater and 16.3% (p=.109) less than non-paretic sides. Stiffness estimates of paretic TA from torque and angle measures were significantly greater by 23.1% (p=.033) than non-paretic TA. No significant difference for MG		architecture (thickness, fascicle length, pennation angle) and stiffness on each leg and muscle independently
Kwon 2012	**Intrarater Reliability:** ICC=.812 (repeated measures, CP group DS scores) ICC=.886 (repeated measures, control group DS scores)	**Convergent Validity:** MAS score was positively correlated w/ the DS score (r=.712) and SWV (r=.710) and negatively correlated w/ SR (r=-0.766, p=.001)	**Known Groups Validity:** MG DS score was significantly higher for the CP group than controls (2.5±.5 vs 1.1±.3, p=.01). MG SWV was also significantly higher for the CP group than controls (2.5±.7 vs 1.3±.4). SR was significantly lower for the CP group than that controls (.5±.4 vs 1.3±.8, p=.01)	N/A	Controls, MAS, SR (MG to SOL muscles), and DS score interpreted as: 1 (purple/ green =soft), 2 (green/ yellow= mostly soft), 3 (yellow/ red= mostly hard), 4 (red= hard)
Lacourpaille 2017	N/A	**Convergent Validity:** For controls, evoked max torque increased at T+12mo (+11.2±7.6%, d=2.1, p<.001) but Tm (p=.382) and EMD (p=.999) did not change. In contrast, DMD children showed no change in evoked max torque (p=.222) but both EMD (+12.9±11.3%, d=2.5, p<.001) and Tm (+10.1±21.6%, d=1.27, p=.003) were significantly longer at T+12 than T0.	**Known Groups Validity:** Muscle stiffness increased at T+12mo in DMD children for the TA (+75.1±93.5%, p=.043), MG (+144.8±180.6%, p=.050) and TB (+35.5±32.2%, p=.005). Analysis of SM maps for MG (stretched position) and TA (shortened position) and group (DMD, controls) showed significant time × group interaction for TA (p=.043) and TB (p=.005) and a significant time × length × group interaction for MG SM (p=.050).	**Responsiveness:** TA (+75.1±93.5%, d=1.04, p=.009) and TB SM (+35.5±32.2%, d=.29, p<.001) were significantly higher at T+12 than T0 in DMD children, regardless of the muscle length, w/ no change for controls (all p>.369). Also a significant increase in GM SM (lengthened) at T+12 (p<.001) compared to T0 (+123.6±180.2%, d=1.05; in DMD but not controls (p>.715)	Controls, electromechanical delay (EMD) and electrically induced maximal torque (elbow flexion)
Lacourpaille 2015	N/A	**Convergent Validity:** SM of the MG at both muscle lengths correlated w/ age in patients w/ DMD (lengthened: r=.74, p=.005 shortened: r=.55, p=.050). No significant correlation was found for controls (r<.43 in all cases). SM of GM at both lengths correlated w/ age in patients w/ DMD (long: r=.74, p=.005, short: r=.55, p=.050) No significant correlation was found for healthy participants (r<.43 in all cases).	**Known Groups Validity:** SM was significantly higher in patients w/ DMD compared to controls for all muscles (main effect for group, p<.033 in all cases), except for ADM (p=.394). For lengthened muscle, SM of the TA (p=.005) and BB (p=.017) were significantly higher in patients w/ DMD than controls, but no difference btw groups for short TA (p=.991) and BB (p=.999). A significant group x muscle length interaction was found for TA (p=.026) and BB (p=.048).	**Responsiveness:** Effect of DMD on changes in SM from T0 to T+12mo was moderate to large for all muscles (Cohen's d range=.48-.99) except the ADM (d=.33)	Controls, muscle length and age
Lee 2018	No ICC reported for this study **Preliminary Reliability Assessment:** Repeatability and reliability of SWV was previously tested among 3 controls and 3 subjects w/ stroke twice on separate days. ICC=.932, CV=4.5% (controls) ICC=.821, CV=6.5% (non-paretic side) ICC=.715, CV=9.2% (paretic side)	**Convergent Validity:** As voluntary activation increased, SWV increased non-linearly, w/ an average power fit of R^2=.83±.09 for the non-paretic side, R^2=.61±.24 for the paretic side, and R^2=.24±.15 for controls. Passive SWV on the paretic side was correlated w/ MAS scores for the elbow extensors (p=.044)	**Known Groups Validity:** Mean passive SWV across all subjects were significantly different btw the non-paretic (2.34±.41) and paretic sides (3.30±1.20, p<.001) and btw paretic sides and controls (2.24±.18, p<.001). There was no significant difference in SWV btw-sides or controls in active muscles (10, 25, 50, 75, 100% max voluntary contraction). **Discriminant Validity:** Non-paretic arms showed significantly greater passive and active elbow ROM than paretic arms (non-paretic: active=180±5° to 44±11°, passive=182±4° to 37±6°), (paretic: active= 145±27° to 59±10°, passive=175±8° to 45±7°), (active: p=.006), (passive: p=.001)	N/A	Voluntary muscle activation (active muscle condition only), MAS and ROM
Lee 2016	No ICC reported for this study **Preliminary Reliability Assessment:** Repeatability and reliability of SWV was previously tested among 3 unimpaired subjects twice on separate days. ICC=.932, CV=4.5%	**Convergent Validity:** MG and TA muscle thickness and fascicle length w/ the ankle at 90°, were significantly reduced on the more-affected side. There was no significant correlation btw SWV of the MG or TA and ankle ROM as indicated by max DF angle	**Known-Groups Validity:** There was no significant difference in SWV btw CP children w/ GMFCS Levels I and II. **Discriminant Validity:** MG and TA SWV was significantly higher in the more- than less-affected limb (MG: S1, S3, S5–S8, TA: S3–S8) at 90° in 6/8 subjects. Average difference was 14% (MG) and 20% (TA) greater in the more- than less-effected limbs, (more-affected MG=5.05±.55), (less-affected MG=4.46±.57, p=.024), (more-affected TA=3.86±.79) and (less-affected TA=3.22±.40, p=.03)	N/A	SWV, muscle thickness, fascicle strain, torque, and ankle angle and clinical assessments (ankle ROM and GMFCS Levels I and II)
Lee 2015	No ICC reported for this study **Preliminary Reliability Assessment:** Repeatability and reliability was previously tested among 3 unimpaired subjects twice on separate days. ICC=.932, CV=4.5% (SWV) ICC=.882, CV=4.2% (Echo Intensity)	**Convergent Validity:** Btw-sides differences in SWV and echo intensity were strongly correlated (R^2=.703, p=.002). A significant linear relationship was observed for the btw-sides difference in SWV and stroke duration (R^2=.301, p=.03) as well as a linear relationship btw SWV and elapsed time since stroke onset. Btw-sides difference in SWV was significantly correlated w/ FMA score (R^2=.327, p=.02)	**Discriminant Validity:** Measures of SWV were 69.5 % greater in paretic (3.67±1.28) than non-paretic BB (2.23±.40, p=.001). Echo intensity was 15.5% greater in paretic (109.48±21.90) than non-paretic BB (83.03±28.96, p=.004).	N/A	Btw-sides comparison, FMA score, echo intensity

	Software Usage: Each value was accompanied by a "quality factor" indicating SWV reliability				
Mathevon 2018	**Intrarater Reliability:** **Axial (rest)**: (CV= 53.91, SEM=20.02, ICC=.00, paretic MG), (CV=42.15, SEM=9.29, ICC=.73, non-paretic MG), (CV=40.74, SEM=22.25, ICC=.64, paretic TA), (CV=37.00, SEM=24.81, ICC=.51, non-paretic TA) **Sagittal (rest)**: (CV= 9.86, SEM=0.61, ICC=.87, paretic MG), (CV=17.64, SEM=1.70, ICC=.36, non-paretic MG), (CV=15.60, SEM=2.09, ICC=.43, paretic TA), (CV=18.50, SEM=4.35, ICC=.27, non-paretic TA) **Sagittal (max passive DF):** (CV=40.58, SEM=18.21, ICC=.11, paretic MG), (CV=24.05, SEM=8.32, ICC=.30, non-paretic MG) **Other:** CV for MG thickness and pennation angle were acceptable at rest and in DF on both sides. CV for TA thickness were acceptable for both sides, but not pennation angle on either side	N/A	N/A	N/A	Muscle thickness and pennation angle
Park 2012	**Intrarater Reliability** (pre-intervention, repeated measures): ICC=.855 (RTS scores) ICC=.906 (red pixel intensity) ICC=.915 (blue pixel intensity) **Intrarater Reliability** (post-intervention, repeated measures): ICC=.878 (RTS scores) ICC=.912 (red pixel intensity) ICC=.926 (blue pixel intensity)	**Convergent Validity:** RTS score was positively correlated w/ mean red pixel intensity in the MG (r=.756) and negatively correlated w/ mean blue pixel intensity (r=-.605). MAS score was positively correlated w/ RTS score (r=.778). There was no correlation btw RTS and GMFM scores.	N/A	**Responsiveness:** Mean RTS score of the MG before and at 4 weeks after intervention decreased significantly from 3.4 to 1.5 (p=.05).	Color pixel intensity, MAS and GMFM scores
Pichiecchio 2018	N/A, no values from the statistical analyses were reported	**Convergent Validity:** DMD children showed fatty replacement and patchy edema on muscle MRI and increased stiffness on SWE. However, there was no significant correlation btw stiffness values and MRI scores. Muscle MRI T1-w images showed fatty replacement in 3/5 children in the GM, w/ thigh and leg muscles affected in 2/5 children. Hyperintensity of STIR was identified in 4/5 children	**Known Groups Validity:** Stiffness was moderately higher in DMD children compared to controls in the RF, VL, AM and GM muscles	N/A	Controls, MRI and NSAA
Shao 2019	N/A	**Convergent Validity**: positive correlation btw SR and 10MWT (r=.83, p=.009), and TUG scores (r=.87, p=.012), and moderately positive correlation btw elasticity score and 10MWT (r=.58, p=.048), and TUG scores (r=.62, p=.011)	N/A	**Responsiveness:** pre-training, SR were significantly lower for impaired (2.92±.83) than healthy AT (3.50±.64) (p=.009). Post-training, SR of impaired AT increased significantly by 9 wks (3.39±.75) compared to pre-training (2.92±.83) (p=.045)	10MWT, TUG, AT length and thickness
Wu 2016	**Intrarater Reliability:** ICC=.852 (95% CI=.565– .955, longitudinal axis) ICC=.711 (95% CI=.260–.907, transverse axis) **Interrater Reliability** ICC=.768 (95% CI=.373–.927, longitudinal axis) ICC=.552 (95% CI=.002–.846, transverse axis)	**Convergent Validity:** At 90°, paretic side SWV correlated positively w/ post stroke duration (r=.467, p=.008), MAS (r=.662, p=.001)and TS (r=.536, p=.002) and negatively w/ STREAM score (r=-.572, p=.001)	**Known Groups Validity:** SWV was significantly greater on the paretic side than the non-paretic side at both 90° (2.23±.15 vs. 1.88±.08, p=.036) and 0° (3.28±.11 vs. 2.93±.06, p=.002). For controls, SWV did not significantly differ btw-sides at 0° (p=.311) or 90° (p=.436) or btw males and females at 0° (2.94±.03 vs 2.96±.07, p=.831) or 90° (1.92±.06 vs 1.73±.06, p=.063). For the subgroup analysis conducted among 9 patients w/ no change in muscle tone (MAS score=0), SWV was significantly lower on the paretic than non-paretic sides at 90° (1.49±.06 vs 1.76±.16, p=.046), but not at 0° (2.90±.14 vs 2.98±.11, p=.529)	N/A	Controls, Subgroup analysis for differences in SWV btw spastic and non-spastic patients w/ stroke (MAS score=0), Stroke duration, MAS, MTS, and STREAM

Table 4. Quality assessment and level of evidence synthesis

UE Method	Outcome Measure(s)	Study	Sample Size (Patients/Controls)	Quality Rating (COSMIN)[53]	Results Rating[60]	Level of Evidence[58]
Reliability						
Quantitative	SWV (m/s)	Bilgici 2018(a)	17/25	Adequate	Interrater (-)	Total N = 60/55, Interrater, Moderate (-); Total N = 63/46, Intrarater, Moderate (+)
	SWV (m/s)	Bilgici 2018(b)	12/0	Adequate	Interrater (-)	
	SM (kPa)	Du 2016	0/9	Doubtful	Interrater (+), Intrarater (+)	
	SM (kPa)	Eby 2016	9/4	Adequate	Intrarater (+)	
	SWV (m/s)	Gao 2018	8/8	Adequate	Intrarater (+)	
	SWV (m/s)	Kwon 2012	15/13	Adequate	Intrarater (+)	
	SM (kPa)	Mathoven 2018	14/0	Adequate	Intrarater (-)	
	SWV (m/s)	Wu 2016	31/21	Adequate	Interrater (-), Intrarater (+)	
Semi-Quantitative	SR	Gao 2016	14/10	Adequate	Interrater (+), Intrarater (+)	Total N = 14/0, Interrater, Unknown (+); Total N = 46/23, Intrarater, Limited (+)
	SR, DS score, color pixel intensity	Kwon 2012	15/13	Adequate	Intrarater (+)	
	RTS score	Park 2012	17/0	Adequate	Intrarater (+)	
Measurement Error						
Quantitative	SM (kPa)	Mathoven 2018	14/0	Adequate	(?)	Total N = 14/0, Unknown (?)
Hypothesis Testing (Convergent Validity)[a1]						
Quantitative	SWV (m/s)	Bilgici 2018(a)	17/25	Doubtful	(+)	Total N = 197/97, Strong (+)
	SWV (m/s)	Bilgici 2018(b)	12/0	Very good	(+)	
	SM (kPa)	Brandenburg 2018	9/0	Adequate	(-)	
	SM (kPa)	Brandenburg 2016	13/13	Adequate	(-)	
	SM (kPa)	Du 2016	46/31	Adequate	(+)	
	SM (kPa)	Eby 2016	9/4	Adequate	(+)	
	SWV (m/s)	Gao 2018	8/8	Adequate	(+)	
	SWV (m/s)	Jakubowski 2017	14/0	Very good	(+)	
	SWV (m/s)	Kwon 2012	15/13	Adequate	(+)	
	SM (kPa)	Lacourpaille 2017	10/9	Very good	(+)	
	SM (kPa)	Lacourpaille 2015	14/3	Adequate	(+)	
	SWV (m/s)	Lee 2018	14/8	Adequate	(+)	
	SWV (m/s)	Lee 2016	8/0	Very good	(+)	
	SWV (m/s)	Lee 2015	16/0	Adequate	(+)	
	SM (kPa)	Pichiecchio 2018	5/5	Doubtful	(-)	
	SWV (m/s)	Wu 2016	31/21	Very good	(+)	
Semi-Quantitative	ELX 2/1 Index, color pixel intensity	Boyaci 2014	16/17	Adequate	(+)	Total N = 86/40, Moderate (+)
	SR	Gao 2016	14/10	Adequate	(+)	
	SR, DS score, color pixel intensity	Kwon 2012	15/13	Adequate	(+)	
	RTS score	Park 2012	17/0	Very good	(+)	
	SR, elasticity score	Shao 2019	24/0	Very good	(+)	
Hypothesis Testing (Discriminative/ Known-Groups Validity)[a2]						
Quantitative	SWV (m/s)	Bilgici 2018(a)	17/25	Very good	(+)	Total N = 198/123, Strong (+)
	SM (kPa)	Brandenburg 2016	13/13	Very good	(+)	
	SM (kPa)	Du 2016	46/31	Very good	(+)	
	SM (kPa)	Eby 2016	9/4	Doubtful	(+)	
	SWV (m/s)	Gao 2018	8/8	Doubtful	(+)	

	SWV (m/s)	Jakubowski 2017	14/0	Very good	(+)	
	SWV (m/s)	Kwon 2012	15/13	Very good	(+)	
	SM (kPa)	Lacourpaille 2017	10/9	Very good	(+)	
	SM (kPa)	Lacourpaille 2015	14/3	Adequate	(+)	
	SWV (m/s)	Lee 2018	14/8	Very good	(+)	
	SWV (m/s)	Lee 2016	8/0	Very good	(+)	
	SWV (m/s)	Lee 2015	16/0	Very good	(+)	
	SM (kPa)	Pichiecchio 2018	5/5	Inadequate	(-)	
	SWV (m/s)	Wu 2016	31/21	Very good	(+)	
Semi-Quantitative	ELX 2/1 Index, color pixel intensity	Boyaci 2014	16/17	Very good	(+)	Total N = 45/40, Limited (+)
	SR	Gao 2016	14/10	Very good	(+)	
	SR, DS score, color pixel intensity	Kwon 2012	15/13	Adequate	(+)	
Responsiveness						
Quantitative	SWV (m/s)	Bilgici 2018(b)	12/0	Very good[d]	(+)	Total N = 36/22, Limited (+)
	SM (kPa)	Brandenburg 2018	9/0	Adequate[d]	(-)	
	SM (kPa)	Lacourpaille 2017	10/9	Very good[c]	(+)	
	SM (kPa)	Lacourpaille 2015	14/13	Very good[c]	(+)	
Semi-Quantitative	ELX 2/1 Index, color pixel intensity	Boyaci 2014	16/17	Very good[b,d]	(+)	Total N = 57/17, Moderate (+)
	RTS score	Park 2012	17/0	Very good[b,d]	(+)	
	SR, elasticity score	Shao 2019	24/0	Very good[b,d]	(+)	

[a1]COSMIN Box 9a - Comparison with other outcome measures (Convergent validity)
[a2]COSMIN Box 9b - Comparison between subgroups (Discriminative or known-groups validity)
[b]COSMIN Box 10b - Responsiveness (Construct approach - comparisons with other outcome measures)
[c]COSMIN Box 10c - Responsiveness (Construct approach - comparisons between subgroups)
[d]COSMIN Box 10d - Responsiveness (Construct approach - pre-to-post intervention)
Results Rating: (+) = Positive, (?) = Indeterminate, (-) = Negative