Ultrasound and elastography in the assessment of skin involvement in systemic sclerosis: A systematic literature review focusing on validation and standardization – WSF Skin Ultrasound Group

Tânia Santiago[a,b,*], Eduardo Santos[c,d], Barbara Ruaro[e], Gemma Lepri[f], Lorraine Green[g,h], Marie Wildt[i], Shinji Watanabe[j], Alain Lescoat[k,l], Roger Hesselstrand[i], Francesco Del Galdo[g,h], John D. Pauling[m,n], Annamaria Iagnocco[o], JAP da Silva[a,b]

[a] *Rheumatology Department, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal*
[b] *Coimbra Institute for Clinical and Biomedical Research (iCBR), Faculty of Medicine, University of Coimbra, Coimbra, Portugal*
[c] *Health School of the Polytechnic Institute of Viseu, Portugal*
[d] *Health Sciences Research Unit: Nursing (UICISA: E), Nursing School of Coimbra (ESEnfC), Portugal*
[e] *Department of Pulmonology, University Hospital of Cattinara, Trieste, Italy*
[f] *Rheumatology Unit, University of Florence, Florence, Italy*
[g] *Leeds Institute of Rheumatic and Musculoskeletal Medicine, Faculty of Medicine and Health, University of Leeds, Leeds, UK*
[h] *National Institute for Health Research (NIHR) Leeds Musculoskeletal Biomedical Research Unit, Leeds Teaching Hospitals NHS Trust, Leeds, UK*
[i] *Department of Rheumatology, Skåne University Hospital, Lund University, Lund, Sweden*
[j] *Department of Allergy and Rheumatology, Nippon Medical School Graduate School of Medicine, Tokyo, Japan*
[k] *Univ Rennes, CHU Rennes, Inserm, EHESP, Irset (Institut de Recherche en Santé, Environnement et Travail) - UMR_S 1085, Rennes, France,*
[l] *Department of Internal Medicine and Clinical Immunology, Rennes University Hospital, Rennes, France*
[m] *Royal National Hospital for Rheumatic Diseases (at Royal United Hospitals NHS FT), Bath, UK:*
[n] *Department of Pharmacy and Pharmacology, University of Bath, Bath, UK*
[o] *Academic Rheumatology Centre, Department of Clinical and Biological Science, University of Turin, Turin, Italy*

ARTICLE INFO

Keywords:
Skin ultrasound
Skin elastography
Systemic sclerosis

ABSTRACT

Objective: To summarize the published evidence in the literature on the role of ultrasound and elastography to assess skin involvement in systemic sclerosis (SSc).
Methods: A systematic literature review (SLR) was performed within the "Skin Ultrasound Working Group" of the World Scleroderma Foundation, according to the Cochrane Handbook. A search was conducted in Pubmed, Cochrane Library and Embase databases from 1/1/1979 to 31/5/2021, using the participants, intervention, comparator and outcomes (PICO) framework. Only full-text articles involving adults, reported in any language, assessing ultrasound to quantify skin pathology in SSc patients. Two reviewers performed the assessment of risk of bias, data extraction and synthesis, independently.
Results: Forty-six studies out of 3248 references evaluating skin ultrasound and elastography domains were included. B-mode ultrasound was used in 30 studies (65.2%), elastography in nine (19.6%), and both methods in seven (15.2%). The ultrasound outcome measure domains reported were thickness (57.8%) and echogenicity (17.2%); the elastography domain was stiffness (25%). Methods used for image acquisition and analysis were remarkably heterogeneous and frequently under-reported, precluding data synthesis across studies. The same applies to contextual factors and feasibility. Our data syntheses indicated evidence of good reliability and convergent validity for ultrasound thickness evaluation against mRSS and skin histological findings. Stiffness and echogenicity have limited evidence for validity against histological findings. Evidence for sensitivity to change, test-retest reliability, clinical trial discrimination or thresholds of meaning is limited or absent for reported ultrasound domains.
Conclusion: Ultrasound is a valid and reliable tool for skin thickness measurement in SSc but there are significant knowledge gaps regarding skin echogenicity assessment by ultrasound and skin stiffness evaluation by

Abbreviations: SSc, Systemic sclerosis; mRSS, modified Rodnan skin score; EUSTAR, European League Against Rheumatism Scleroderma Trials and Research; WSF, World Scleroderma Foundation; SLR, Systematic literature review; OMERACT, Outcome measures in rheumatologic clinical trials; OFISA, OMERACTFilterInstrument Selection Algorithm; ROI, Region of interest; ICC, Intraclass correlation coefficient.

* Corresponding author at: Serviço de Reumatologia, Centro Hospitalar e Universitário de Coimbra, Praceta Prof. Mota Pinto, 3030-075 Coimbra, Portugal.
E-mail address: tlousasantiago@hotmail.com (T. Santiago).

Introduction

Systemic sclerosis (SSc) is an autoimmune rheumatic disease characterized by vascular damage and fibrosis of the skin and/or internal organs, with high clinical burden and unsatisfactory treatment [1]. Skin fibrosis is a clinical hallmark of the disease [2], and an important marker of disease activity [3], severity [4,5] and prognosis [4,5]. The extent of skin involvement and its rate of progression are associated with internal organ involvement, functional disability, and survival [6]. This makes skin assessment a crucial issue for both clinical practice and research.

The modified Rodnan skin score (mRSS) is the current gold standard to evaluate skin involvement. It is a semi-quantitative score based on the palpation of the skin at 17 anatomical sites [2]. It is often used as primary or secondary outcome in clinical trials. The mRSS is also a major component of the composite response index in diffuse cutaneous SSc [7] and it is included in the European League Against Rheumatism Scleroderma Trials and Research (EUSTAR) disease activity score [3]. This underlines the importance of skin assessment, however the mRSS has several important limitations, including its subjectivity and intra- and inter-observer variability (around 12% and 25%, respectively) [8,9]. It is poorly sensitive to change in patients with limited cutaneous SSc and it has shown a modest performance in discriminating drug versus placebo in most clinical trials [10–12]. Novel therapies are dearly needed, but current measures of effect are too blunt to support progress. This emphasizes the need for more accurate and sensitive tools to assess skin in SSc.

The potential use of skin ultrasound and elastography to this purpose has been addressed in several studies over the last three decades [13–15]. Although these tools are considered promising, their actual value as a correlate or surrogate of skin involvement in SSc remains unclear, with limited existing integration in current clinical practice and research. The “Skin Ultrasound Working Group”, that has been recently created under the auspices of the World Scleroderma Foundation (WSF), aims to develop recommendations for the standardization of the procedures and reports on the assessment of the skin by ultrasound and elastography in SSc.

We herein present the results of a systematic literature review (SLR) performed to support the development of these recommendations by highlighting the achievements and limitations of accumulated evidence. In addition, we evaluated whether available information supports the use of skin ultrasound and elastography as an surrogate outcome measure in SSc, according to the outcome measures in rheumatologic clinical trials (OMERACT) filter 2.1 instrument selection algorithm (OFISA) [16]. Specifically, this SLR focuses on the pillars of truth, discrimination and feasibility.

Methods

Literature search

This SLR was conducted according to the Cochrane Handbook [17] and reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [18]. The SLR protocol, which was not registered but is available upon request, was established *a priori* and strictly adhered to.

In the first meeting of the Skin Ultrasound Working Group of the WSF the key research questions were framed, following the population, intervention, comparator and outcome – PICO - format. This referred to the procedures/methods used for image acquisition in skin ultrasound and elastography, as well their analysis and interpretation in patients with SSc. The population of interest consisted of adult patients (>18years) with SSc fulfilling the 1980 ACR criteria [19], 2013 ACR/EULAR [20] or classified by their cutaneous subtype as defined by LeRoy et al. [21].

Eligible studies were full research articles, of all types, conveying results of original research, including at least one defined group of patients with SSc, and reporting a structured evaluation and clear definition of the ultrasound scanning protocol. Details of the inclusion criteria are presented in supplementary material table S1.

Search strategy and study selection

The SLR was conducted by two reviewers (TS and ES). The search strategy was developed by one of the reviewers (ES), with accredited experience in this field.

Medline, Cochrane Library and Embase databases were searched without language restrictions, for papers published from 1 January 1979 until 31 May 2021. This start date was chosen because the first B-mode ultrasound image of human skin was reported in 1979 [22]. Details on complete search strategies are provided in the supplementary material table S1.

All identified citations were uploaded into an EndNote VX7 (Clarivate Analytics, PA, USA) library and the duplicates removed. The reviewers (TS and ES) screened independently all title and abstracts to identify potentially eligible studies, which were, then, reviewed in full text.

Extraction of study characteristics and results

Papers fulfilling the inclusion criteria were submitted to data extraction. Both reviewers independently retrieved data using a predefined data extraction sheet, validated by the task force and informed by the EULAR checklist recommendations for the reporting of ultrasound studies in rheumatic diseases [23]: studies main characteristics (year of publication, country, design, participants, disease characteristics), ultrasound outcome domains, equipment, blinding, scanning/acquisition procedures, ultrasound scoring system, contextual factors and feasibility. Authors of papers were contacted to request missing or additional data, where required.

Evaluation was made strictly following OFISA and Handbook [24]. The successive phases of evaluations are described below:

Evaluation of the methodological quality per measurement property, per study

The methodological quality (risk of bias) of each study was assessed by two independently reviewers (TS, ES), using the COSMIN—OMERACT Good Metrics Checklist (table S2) [24]. Quality assessment was rated using a color code: ‘Green’ if good methods were used, ‘Amber’ if there were some methodological concerns but the data were acceptable for inclusion, and ‘Red’ if there was a high risk of bias, as indicated by OMERACT.

Discrepancies between reviewers regarding study selection, data extraction and risk of bias assessment were resolved by consensus or with a third reviewer (JAPS).

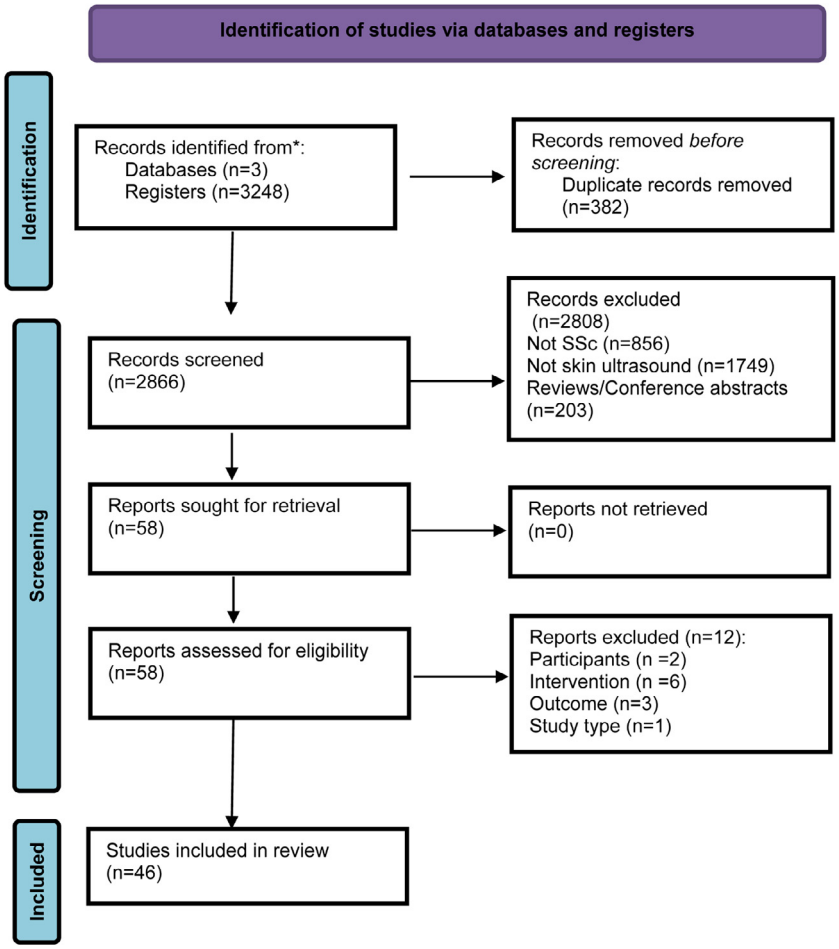

Fig. 1. PRISMA flow diagram of the SLR.

Evaluating the performance of measurement properties per study

Each study was assessed using the OMERACT provisional standards for adequate performance and assigned ratings of + (positive support for the measurement property), ± (ambivalent support, inconclusive), or - (instrument did not reach performance standards for that measurement property) (table S3) [24].

Summary ratings of individual measurement properties per domain

All studies rated amber or green were synthesized to generate an overall rating for the individual measurement properties for each domain based on the OMERACT Handbook - "Criteria for Final Rating" [24]. This rating summarises the quality and quantity of studies, the consistency of the results, and the performance of individual instruments.

A color grading was used as recommended: GREEN indicates 'Good to go', RED indicates 'Stop, do not continue', WHITE indicates 'No evidence' and AMBER indicates 'There is a concern, or caution, or weakness, but it is good enough to go forward, perhaps with a research agenda to move it to GREEN or RED'.

Results

Study selection

The literature review yielded 3248 references (Fig. 1). Of the 46 studies included, 30 (65.2%) used B-mode ultrasound, 9 (19.6%) used elastography and 7 (15.2%) used both techniques (Table 1). The ultrasound outcome domains reported were thickness in 37 (57.8%) studies, stiffness in 16 (25.4%) and echogenicity in 11 (17.2%).

Below each ultrasound domain is presented the summary of the measurement properties (SOMP) for each one – Tables 2–4. None of the studies were excluded based on overall risk of bias, but one measurement property was rated red (Table 2).

Thickness

Study characteristics

Thirty-seven studies evaluated the ultrasound-thickness of skin tissues. They were published between 1985 and 2021, most being observational and cross-sectional (Table 1). Six studies were longitudinal: 4 evaluated spontaneous change over time [25–29] and 2 changes due to interventions [30,31]. All studies were based in a single center.

In total, 1439 SSc patients (~80% female and ~60% with a limited form) were evaluated. Thirty-two studies had a control group. Cases were defined according to the 1980 ARA criteria in 20 studies (55.5%), to the 2013 ACR/EULAR classification criteria in 12 (30%) and both criteria in 2 studies (9%) [28,32]. The Le Roy criteria were used in 2 studies (5.5%) (table S2) [33,34].

Description of image acquisition and analysis

Most studies reported using linear probes with variable frequencies ranging from 10 to 50 MHz (table S3). The ultrasound equipment brand and system were mentioned in 32 (86.5%) studies. Twenty-

Table 1
. Study design, demographics and main characteristics of the studies on skin ultrasound in SSc per ultrasound domain.

Study	Country	Design	Followup (time or intervention)	Patients/ Controls n	Female%	Lim/Dif n	Probe MHz	Scoring system	Skin sites evaluated	Comparison
THICKNESS										
Serup J 1985 [46]	Denmark	Case-control	–	22/22	100	22	15 MHz	Quantitative (mm)	Finger	Ring size
Akesson A 1986 [25]	Sweden	Cohort	6,12,18mo*	40/10	75	22/18	10 MHz	Quantitative (mm)	Finger	–
Myers S 1986 [78]	USA	Case-control	–	8/11	100	8	25 MHz	Quantitative (mm)	Forearm	X-ray
Seidenari S 1996 [66]	Italy	Case-control	–	18/20	100	8/10	20 MHz	Quantitative (mm)	Forehead Cheek Hand	–
Ihn H 1995 [50]	Japan	Case-control	–	79/81	88.6	79	30 MHz	Quantitative (mm)	Chest Forearm, Hand	Skin biopsy (chest,fore-arm, hand)
Scheja A 1997 [45]	Sweden	Case-control	–	41/41	56.1	25/12/4	20 MHz	Quantitative (mm)	Forearm Hand Finger	–
Brocks 2000 [79]	Denmark	Case-control	–	20/20	100	16/4	20 MHz	Quantitative (mm)	Chest Forearm Hand Finger	mRSS
Hesselstrand R 2002 [80]	Sweden	Case-control	1–3y	11/6	72.7	5/6	20 MHz	Quantitative (mm)	Forearm	Skin biopsy (forearm)
Moore T 2003 [48]	UK	Case-control	–	39/34	79.5	26/13	22 MHz	Quantitative (mm)	17 mRSS	mRSS
Akesson A 2004 [27]	Sweden	Case-control	2–4y**	16/16	87.5	8/8	20 MHz	Quantitative (mm)	Chest Forearm Hand Finger	mRSS
Hashikabe M 2004 [30]	Japan	Cohort	Pre and after photoche-motherapy (16.3d)	13/10	92.3	6/7	20 MHz	Quantitative (mm)	Forearm Hand Finger	mRSS
Kissin 2006 [81]	USA	Case-control	–	30/12	83.3	10/20	10 MHz	Quantitative (mm)	Upperars Forearms Hands Fingers	mRSS Durometry
Hesselstrand R 2007 [54]	Sweden	Cohort	–	97/-	–	68/29	20 MHz	Quantitative (mm)	Chest Forearm Hand Finger Leg	–
Hesselstrand R 2008 [59]	Sweden	Cohort	–	106/-	81.1	76/30	20 MHz	Quantitative (mm)	Chest Forearm Hand Finger Leg	mRSS
Kaloudi 2010 [34]	Italy	Case	–	70/20	88.6	61/9	6–18 MHz	Quantitative (mm)	Fingers	mRSS HAQ
Kuhn A 2011 [31]	Germany	Open-label, non-comparative	Pre and post bosentan, 24w	10/-	58.9	4/10	20 MHz	Quantitative (mm)	Forearms Hands Fingers Leg	–
Geso L 2011 [35]	Italy	Cohort	–	22/-	57.1	14/8	6–18 MHz	Quantitative (mm)	Finger	–
Hassan I 2012 [63]	India	Case-control	–	15/15	98	NR	20 MHz	Quantitative (mm)	Forehead Chest Forearm Hand	–

(*continued on next page*)

Table 1 (*Continued*)

Study	Country	Design	Followup (time or intervention)	Patients/ Controls n	Female%	Lim/Dif n	Probe MHz	Scoring system	Skin sites evaluated	Comparison
									Finger Leg	
Sedky M 2013 [42]	Egypt	Case-control	–	40/40	95	26/14/5	5–12 MHz	Quantitative (mm)	Chest Forearm Hand Finger Leg	mRSS Severity score
Sulli A 2013 [82]	Italy	Case-control	–	57/37	89.5	43(14	18 MHz	Quantitative (mm)	Finger	mRSS NVC FBP
Hesselstrand R 2015 [28]	Sweden	Cohort	1y follow-up	75/-	83	42/33	20 MHz	Quantitative (mm)	Chest Forearm Hand Finger Leg	mRSS Serum-COMP HAMIS
Hou Y 2015 [67]	China	Case-control	–	15/15	66.6	0/15	6–18 MHz	Quantitative (mm)	17 mRSS	mRSS
Liu H 2017 [32]	China	Case-control	–	28/15	78.6	0/28	6–18 MHz	Quantitative (mm)	17 Rodnan sites	mRSS
Sousa-Neves J 2017 [37]	Portugal	Case-control	–	48/45	NR	42/6	15 MHz	Quantitative (mm)	Fingers	mRSS HAMIS
Sulli A 2017 [49]	Italy	Case-control	–	50/50	89.5	50/0	18 MHz	Quantitative (mm)	17 Rodnan sites	–
Li H 2017 [38]	China	Case-control	–	31/31	87	27/4	18 MHz	Quantitative (mm)	Chest, Forearm Hand, Finger Leg	mRSS EUSTAR-DAI
Ruaro B 2018 [53]	Italy	Case-control	–	62/62	90.3	45/17	18 MHz	Quantitative (mm)	Zygoma Hands Fingers	mRSS Skin BP
Yang Y 2018 [39]	China	Case-control	–	37/37	86.5	14/23	4–15 MHz	Quantitative (mm)	Chest Abdomen Forearms Fingers	mRSS
Ruaro B 2019 [83]	Italy	Case-control	–	8/5	87.5	8/0	18 and 22MHz	Quantitative (mm)	17 Rodnan sites	mRSS
Ruaro B 2019 [52]	Italy	Case-control	–	63/63	–	40/23	18 MHz	Quantitative (mm)	17 Rodnan sites	mRSS NVC PST
Ruaro B 2019 [58]	Italy	Case-control	.	48/48	83.3	48/0	18 MHz	Quantitative (mm)	17 Rodnan sites	–
Chen C 2020 [40]	China	Case-control	.	44/22	68.2	22/22	15 MHz	Quantitative (mm)	Hands Fingers	mRSS Skin biopsy (13SSc)
Chen Y 2020 [55]	China	Case-control	.	31/19	67.8	NR	12 MHz	Quantitative (mm)	Hand Forearm	mRSS SAM
Flower V 2020 [47]	UK	Case-control	.	53/15	88.7	45/8	18 MHz	Quantitative (mm)	Abdomen Forearm Hand Finger	mRSS Skin biopsy (10SSc;10controls)
Naredo E 2020 [41]	Spain	Case-control	.	21/6	–	6/5	50 MHz	Quantitative (mm)	Forearm Hand, Finger	mRSS Texture feature analysis
Daoudi K 2020 [43]	NL	Case-control	–	12	80/42	17/5/9	40 MHz	Quantitative (mm)	Finger	Photoacoustics (oxigenation saturation)
Vanhaecke A 2021 [57]	Belgium	Case-control	–	59/44	83	55/4	18 MHz	Quantitative (mm)	17 mRSS	–
ECHOGENICITY										
Seidenari S 1996 [66]	Italy	Case-control	–	18/20	100	8/10	20 MHz	Quantitative (scale 0–255)	Forehead Cheek Hand	–

(*continued on next page*)

Table 1 (*Continued*)

Study	Country	Design	Followup (time or intervention)	Patients/ Controls n	Female%	Lim/Dif n	Probe MHz	Scoring system	Skin sites evaluated	Comparison
Scheja A 1997 [45]	Sweden	Case-control	–	41/41	56.1	25/12/4	20 MHz	Quantitative (scale 0–255)	Forearm Hand Finger	–
Hesselstrand R 2002 [26]	Sweden	Case-control	1–3y	11/6	72.7	5/6	20 MHz	Quantitative (scale 0–255)	Forearm	Skin biopsy (forearm)
Akesson A 2004 [27]	Sweden	Case-control	2–4y**	16/16	87.5	8/8	20 MHz	Quantitative (scale 0–255)	Chest Forearm Hand Finger	mRSS
Hashikabe M 2004 [30]	Japan	Cohort	Pre and after photoche-motherapy (16.3d)	13/10	92.3	6/7	20 MHz	Quantitative (scale 0–100)	Forearm Hand Finger	mRSS
Hesselstrand R 2007 [54]	Sweden	Cohort	–	97/-	–	68/29	20 MHz	Quantitative (scale 0–255)	Chest Forearm Hand Finger Leg	–
Hesselstrand R 2008 [59]	Sweeden	Cohort	–	106/-	81.1	76/30	20 MHz	Quantitative (scale 0–255)	Chest Forearm Hand Finger Leg	mRSS
Hassan I 2012 [63]	India	Case-control	–	15/15	98	NR	20 MHz	–	Forehead Chest Forearm Hand Finger Leg	–
Hesselstrand R 2015 [28]	Sweden	Cohort	1y follow-up	75/-	83	42/33	20 MHz	Quantitative (scale 0–255)	Chest Forearm Hand Finger Leg	mRSS Serum-COMP HAMIS
Li H 2017 [38]	China	Case-control	–	31/31	87	27/4	6–18 MHz	Semi-quantitative (iso, hypo,hyperechogenic)	Chest, Forearm Hand, Finger Leg	mRSS EUSTAR-DAI
Flower V 2020 [47]	UK	Case-control	.	53/15	88.7	45/8	18 MHz	Quantitative (scale 0–255)	Abdomen Forearm Hand Finger	mRSS Skin biopsy (10SSc;10controls)
STIFFNESS										
Iagnocco A 2010 [71]	Italy	Case-control	–	18/15	100	8/10	18 mhz	color scale	Forearm Fingers	–
Geso L 2011 [35]	Italy	Case	–	22/-	57.1	14/8	6–18 mhz	color scale	Finger	mRSS RCS
Cannao P 2014 [70]	Italy	Case-control	–	6/6	100	NR	6–13 mhz	color scale	Peri-oral region	–
Hou Y 2015 [67]	China	Case-control	–	15/15	66.6	0/15	4–9 MHz	Quantitative (SWV in m/s)	17 mRSS	mRSS
Grembiale R 2016 [44]	Italy	Cohort	–	20/-	80	10/10	10–18 MHz	Quantitative (global% of hardness)	Fingers	NVC
Santiago T 2016 [74]	Portugal	Case-control	–	26/17	88.5	13/13	9 MHz	Quantitative (SWV in m/s)	mRSS (except face)	mRSS
Cildag S 2017 [72]	Italy	Cohort	–	40/-	70	0/40	15 MHz	Color scale	Forearm	–
Liu H 2017 [32]	China	Case-control	–	28/15	78.6	0/28	4–9 MHz		17 Rodnan sites	mRSS

(*continued on next page*)

Table 1 (*Continued*)

Study	Country	Design	Followup (time or intervention)	Patients/ Controls n	Female%	Lim/Dif n	Probe MHz	Scoring system	Skin sites evaluated	Comparison
								Quantitative (SWV in m/s)		
Zhang X 2017 [69]	USA	Case-control	–	41/30	68.3	NR	6.4 MHz	Quantitative (surface wave speed in m/s)	Forearm Upper arm	–
Yang Y 2018 [39]	China	Case-control	–	37/37	86.5	14/23	4–15 MHz	Quantitative (Elastic moduli in kPa)	Chest Abdomen Forearms Fingers	mRSS
Aryan A 2018 [73]	Iran	Case-control	–	36/36	83.3	16/	12 mhz	color scale	Upperam Forearm Finger	–
Chen C 2020 [40]	China	Case-control	.	44/22	83.3	22/22	4–15 MHz	Quantitative (total skin elasticity in kPa)	Hands Fingers	mRSS Skin biopsy (13SSc)
Chen Y 2020 [55]	China	Case-control	.	31/19	68.2	NR	12 MHz	Quantitative (Strain ratio)[1]	Hand Forearm	mRSS SAM
Flower V 2020 [47]	UK	Case-control	.	53/15	88.7	45/8	14MHz	Quantitative (mean SWE in KPa)	Abdomen Forearm Hand Finger	mRSS Skin biopsy (10SSc;10controls)
Santiago T 2020 [29]	Portugal	Cohort	4.9 (0.4)y	21/15	85.7	12/7	9 MHz	Quantitative (SWV in m/s)	mRSS (except face)	mRSS
Sobolewski P 2020 [68]	Poland	Case-control	–	40/28	85	29/11	5–18 MHz	Quantitative (Youngs modulus value in KPa)	20 RSS	RSS

COMP, Cartilage Oligomeric Matrix Protein.
EUSTAR-DAI, European Scleroderma Trial and Research (EUSTAR) group Disease Activity Index; HAMIS, Hand Mobility in Scleroderma.
PST, plicometer skin test; SWV, shear-wave velocity; SAM, spectral angle mapper of Hyperspectral imaging; y, years; mo, months; w, weeks; RCS, Raynaud Condition Score; RSS, classic RSS, 20 skin sites; NL, Netherlands.
* The controls were not followed up.;.
** The interval between measurements were 1–2 years.
[1] The ROI1-to-ROI2 strain ratio representing the degree of dermis stiffness in relation to the surrounding tendon or muscle was used for analysis.

eight (75.6%) studies mentioned that the probe was placed perpendicularly to the skin. Probe position was not described in the remaining studies (25%). Half of the studies reported using a layer of gel to minimize the compression induced by the transducer on the skin. Only one study reported the use of a standoff silicone interface [25]. One third of the studies reported that sonographers were blinded to local mRSS (table S4). Only a few studies mentioned the sonographers experience with skin ultrasound [35–41]. Eight studies (22.2%) reported the timing of the measurements within the day; and only 2 (5.5%) mentioned that ultrasound was performed at room temperature, without further specification [42,43].

In half of the studies, the quality criterion for acceptance of an ultrasound image was described as the adequate depiction of epidermis, dermis, and subcutis, with distinct and parallel interfaces between them (table S3). The remaining studies did not describe their quality criteria or did so insufficiently.

Overall, the skin sites examined corresponded to all or some of the Rodnan skin sites, or at least in the same body area. However, the exact definition of landmarks for each site scanned varied considerably. Two studies only assessed the palmar aspects of the fingers [43,44].

Thirteen studies (35.1%) measured solely the dermis, 13 (35.1%) assessed the epidermis and dermis, 3 studies evaluated epidermis, dermis and subcutaneous layer [37,45,46] and yet 1 study evaluated dermis and hypodermis [41] (table S3). Seven studies (19.4%) did not describe which skin layers were evaluated.

All studies reported thickness in millimetres, but the scoring system varied: the number of ultrasound images taken from each skin site was variable across studies, as well as the number of thickness measurements per skin site. The average result of these measurements was used for statistical analysis. Except for one study [47], it's unclear whether the image analysis was performed during the ultrasound evaluation or later.

Feasibility

For the ultrasound assessment of the 17 Rodnan skin sites, Moore et al. [48] reported that image acquisition took ~20minutes; and, Sulli et al. [49] took 20–25 min, including skin image acquisition and analysis. No additional feasibility parameters were reported. Following synthesis of the results and risk of bias, this measurement property was rated AMBER.

Convergent validity *against mRSS, skin biopsy, and other constructs*

The validity of ultrasound-measured thickness vs mRSS was studied in 22 and confirmed in 20 studies, by demonstration of a mild to moderate correlation between ultrasound-measured thickness and local site-specific mRSS (table S4).

Four studies assessed histological findings [26,40,47,50], but only 2 evaluated convergent validity of ultrasound vs histological parameters. These 2 studies used forearm skin biopsies from 13 [40] to 20 [47] participants, respectively, and found moderate to good correlations between skin thickness measured by ultrasound and histological findings. These preliminary analyses of convergent validity with tissue histology is encouraging but further work is desirable.

Correlations between ultrasound skin thickness and a variety of other measures were explored in several studies, including nailfold videocapillaroscopy [49,51,52], Laser Speckle Contrast Analysis [34,51,53], cartilage oligomeric matrix protein [54], EUSTAR disease activity index [38], hand mobility in scleroderma [28,37], health assessment questionnaire [34] and spectral angle mapper [55]. The committee decided that these comparisons make a smaller contribution to convergent validity, given the heterogeneous factors contributing to the parameters under evaluation.

Inter-, intra-rater and test-retest reliability

Eighteen studies performed reliability assessments (table S4). The median ICC varied from 0.84 (0.65 - 0.94) for inter-reader, and 0.92 (0.55 - 0.97) for intra-reader reliability [56,57]. One small study [58] compared the intra-reader reliability with probes of 18 and 22 MHz, and found them to be similar. Inter- and intra-rater reliability was rated GREEN. No studies evaluated test-retest reliability.

Discriminatory capacity across different participants groups

Thirty-four studies compared ultrasound-measured thickness in SSc patients vs controls (Table 2). Thickness was higher in SSc patients than in controls, at group level, in almost all Rodnan sites (mRRS≥1) and in all studies. Six studies reported on the ultrasound-measured thickness of clinically unaffected skin (mRSS=0) and all found it to be thicker in patients than in controls, at least in some Rodnan sites [34,47,49,50,53,59,60]. Two studies identified that dermal thickness decreased as the clinical phase progressed from oedematous to the atrophic phase [34,59]. Therefore this property was rated GREEN for thickness.

Sensitivity to change

Four studies reported change over time [25–28]. In these studies ultrasound dermal thickness decreased and patients became more similar to the control population between the first and the fourth year of follow up [26,28]. None of the studies made comparisons with mRSS nor did they use appropriate statistical methods to evaluate sensitivity to change (e.g., effect size of standardized response mean). Following synthesis of the results and risk of bias, this domain was rated AMBER for sensitivity to change.

Clinical trial discrimination and threshold of meaning

Two small studies investigated the impact of two interventions: photochemotherapy [30] and bosentan [31]. Photochemotherapy was associated with a significant decrease in ultrasound-measured dermal thickness. Bosentan treatment determined only a slight but not significant trend towards improvement of this parameter, from week 0 to 24, despite a statistically significant improvement in mRSS. Both studies had a small sample size and used very limited and questionable statistical analyses. This measurement property was rated AMBER.

A single study [60] reported threshold of meaning. An ultrasound-measured thickness cut-off value of 7.4 mm (sensitivity 77.4% and specificity 87.1%), was found as the minimum difference distinguishing normal from scleroderma skin, when using a composite measure of the sum of total skin thickness of five skin sites. This measurement property was rated AMBER.

Results regarding SOMP table for thickness are summarized in Table 2.

Echogenicity

Study characteristics

Eleven studies, published between 1996 and 2020, evaluated echogenicity (Table 1). Seven were observational and cross-sectional, and 4 investigated sensitivity to change over time [25,27,59,61] or due to an intervention [30]. In total, 476 SSc patients (~87% female and ~50% with a limited form) were evaluated. Seven studies had a control group. Cases were defined according to the 1980 ARA criteria in 8 studies, and 2013 EULAR/ACR in 2 and both in one study (table S2).

Description of image acquisition and analysis

All studies evaluated echogenicity together with thickness, with or without stiffness (table S3). The use of a linear probe, with frequencies from 18 to 20Mhz was reported in all studies apart from 3 which did not provide probe details [30,62,63]. The ultrasound brand and system were mentioned in all studies, but one [63].

Five studies (45.5%) reported that the sonographer was blinded for the mRSS (table S4). Only 1 study mentioned the sonographers' experience with ultrasound [32]. Four studies (36.4%) gave data regarding time of the day of the ultrasound assessment and none reported on temperature [27,30,54,59].

The quality criteria for image acceptance demanded a clear echo definition of the interfaces between the epidermis, dermis, and subcutis by A-mode ultrasound and of the echogenicity of the layers of interest, by B-mode (table S3). All studies assessed skin sites corresponding to Rodnan skin sites, or at least in same body area, although the exact definition of landmarks for each site varied.

Six (54.5%) studies evaluated epidermis and dermis, 2 (18.2%) studies focused on the dermis [30,47,64] and 1 (9%) study evaluated all three layers [45]. Two studies did not define which skin layers were evaluated (table S3).

Echogenicity was reported in a quantitative scale, except for one study [32], which used a semiquantitative scale. Most of the included studies (72.3%) measured this domain in a scale graded from 0 (black) to 255 (white) pixels, in a selected region of interest [26,28,59,65,66]. Different software applications were used to this purpose. A low value is taken as indicative of high water content, suggesting edema, and increased echogenicity as suggestive of fibrosis. Two other studies have applied a 0–100 scoring system [30,32].

In general, it is not clear whether image analysis was performed during ultrasound evaluation or later. Four studies reported on whether readers were blinded to the clinical information during image analysis and scoring [26,28,32,47].

Feasibility

None of the studies included reported the time needed for image acquisition or analysis.

Inter-, intra-rater and test-retest reliability

Four studies reported reliability assessments for ultrasound-echogenicity of the skin (table S4). Two studies (one [45] of them high risk of bias) reported ICCs of 0.92 to 0.98, for inter-reader reliability [27,45]. In 1 study, inter-reader reliability was not satisfactory (ICC <0.75) [32].

Two studies with low-risk of bias evaluated intra-reader reliability demonstrating ICC values ranging from 0.648 to 0.88, depending on skin sites [27,47]. Therefore, following synthesis of the results and risk of bias, inter- and intra-rater reliability was rated AMBER. None of the studies performed test-retest reliability.

Convergent validity against mRSS, skin biopsy, and other constructs

The validity of ultrasound-echogenicity against mRSS was evaluated in six studies and confirmed in four, by demonstration of a low to moderate correlation with local site-specific mRSS [32,59,28,54]. A study investigated ultrasound-echogenicity against features of skin biopsy (collagen content), with negative results [26,47]. In another study, the authors reported significant correlation between changes in extracellular matrix production in vitro and concomitant changes in ultrasound echogenicity over time [25]. Construct validity against skin mRSS and biopsy was rated GREEN and AMBER, respectively.

Discriminatory capacity across different participants groups

The comparison of skin ultrasound echogenicity in patients vs controls was the object of 7 studies (Table 3). In all of them the groups differed significantly in almost all skin sites assessed

Table 2
. Summary of measurement properties for ultrasound studies evaluating thickness.

Study ID	Feasibility	Truth				Discrimination				
		Convergent validity against mRSS	Convergent validity against biopsy	Inter-rater reliability	Intra-rater reliability	Test-retest reliability	Discrimination capacity across different participants groups	Sensitivity to change	Clinical trial discrimination	Threshold of meaning
Serup J 1985							+•[11]			
Akesson A 1986							+[11]			
Myers S 1986				+[1]			+•[11]	+[3]		
Seidenari S 1996							+•[11]			
Ihn H 1995			+/-	+/-[5]			+•¥[11]			
Scheja A 1997		+		+[1]			+•[11]			
Brocks 2000		+/-[6]					+•[11]			
Hesselstrand R 2002			+/-				+•[11]	+[7]		
Moore T 2003	+[2]			+	+		+•[11]			
Akesson A 2004		+/-[9]		+	+		+•[11]	+[7]		
Hashikabe M 2004							+[11]		+[8]	
Kissin 2006										
Hesselstrand R 2007							+			
Hesselstrand R 2008		+					+¥			
Kaloudi 2010		-		+	+		+			
Kunhn A 2010									+[8]	
Geso L 2011		-		+/-[10]	+		+[11]			
Hassan I 2012							+•			
Sedky M 2013		+					+•			

Sulli A 2013		+			+		+•			
Hesselstrand R 2015		+					+	+		
Hou Y 2015		+					+•			
Liu H 2017		+					+•[11]			
Sousa-Neves J 2017		+					+•[11]			
Sulli A 2017	+[2]	+			+		+•¥			
Li H 2017		+					+•¥			+/-[12]
Ruaro B 2018		+			+		+•[11]			
Yang Y 2018		+					+•[11]			
Ruaro B 2019		+			+		+•			
Ruaro B 2019		+			+		+•			
Ruaro B 2019		+			+		+•¥			
Chen C 2020		+	+	+	+		+•			
Chen Y 2020		+			+		+•[11]			
Flower V 2020		+	+		+		+•¥			
Naredo E 2020				+/-			+•			
Daoudi K 2020							+•[11]			
Vanhaecke A 2021				+	+					
Total available studies for each property	2	22	2	10	14		34	4	2	1
Total studies available for synthesis	2	22	2	9	14		34	4	2	1
Overall rating (RAGW) (put on master checklist)										
Overall rating for instrument across properties	Provisional endorsement: needs additional feasibility, test-retest reliability, sensitivity to change, clinical trial discrimination and threshold of meaning									

OMERACT Summary of evidence for measurement properties of thickness. Color designates quality of evidence: green=good methods used, use this evidence; amber=some cautions but we will use this evidence. In the rating row, color designates overall evidence-based instrument rating for the core instrument set: green= at least 2 pieces of evidence with good methods and consistent findings of adequate or better performance; amber=in between green and red; red= inadequate performance in at least 1 study that used good methods, white=no evidence.

Arithmetic signs designate the performance of the instrument according to that study (for each measurement property studied): "+" adequate or better performance, "+/−" equivocal performance, "-" less than adequate/poor performance.

[1]No data whether design of the study hold all other factors constant except for the source of variability being examined and the statistic method used.

[2]No data regarding cost of the ultrasound equipment or software, required training, time to image acquisition or analysis.

[3]No clear data about construct for change clear (either as a situation of change or an actual indicator of change); and, were the statistical methods appropriate for the testing situations.

[4]No data stated for the time interval between testing, appropriateness for this and were if there were a proportion of people expected to change in one or both groups? (Improvement or deterioration) and were hypotheses formulated regarding the anticipated mean differences in change scores between subgroups a priori and were the statistical methods adequate for the hypotheses tested (relative efficiencies, pooled treatment effect sizes, standardized mean differences).

[5]No data on whether measurements were conducted independently or if the design of the study held all other factors constant except for the source of variability being examined or were the test conditions similar for the measurements? (e.g., type of administration, environment, instructions) were the correct statistic used?

[6]No data regarding blinding of image acquisition and analysis; correlation between RSS and ultrasound-thickness rh0=0.36, $p<0.001$.

[7]No data regarding comparison with change of mRSS and no statistical methods appropriate for the testing situation.

[8]No date regarding time interval between testing stated and appropriate; no statistical methods adequate for the hypotheses tested.

[9]No data regarding blinding of image acquisition and analysis.

[10]No data whether measurements were conducted independently or if the design of the study held all other factors constant except for the source of variability being examined or were the test conditions similar for the measurements and inter-reader correlation rho=0.59.

[11]No data regarding patient selection, e.g., consecutive, or random sample patients enrolled, case-control avoided or avoid inappropriate exclusions.

[12]No data whether the anchor clearly related to the target domain of interest (i.e. good correlation between anchor and instrument) and analysis wasnt done separately for improvement and deterioration.

•Case-control studies.

¥Studies reporting ultrasound-thickness of clinically affected skin in patients thicker than in controls.

[61,66,67]. One recent study found dermal echogenicity to be increased in clinically unaffected skin sites (mRSS=0) of SSc patients compared to controls [47]. One study including patients in early disease stages found an inverse correlation between skin echogenicity and thickness, at the 5 sites evaluated [59]. This measurement property was rated as GREEN.

Sensitivity to change

Four studies evaluated sensitivity to change over time and all identified significant increases of echogenicity in five skin areas over one year of follow-up [26–28,30]. None of the four studies made direct comparisons with mRSS and statistical analysis of sensitivity to change was limited and questionable. This measurement property was rated AMBER.

Discrimination and threshold of meaning

In a single small study, including 13 SSc patients, photochemotherapy was associated with a significant increase of ultrasound echogenicity after treatment [30]. This measurement property was rated AMBER.

No threshold of meaning has been reported.

Results regarding SOMP table for echogenicity are summarized in Table 3.

Table 3
Summary of measurement properties for ultrasound studies evaluating echogenicity.

Study ID	Feasibility	Truth				Discrimination				
		Convergent validity against mRSS	Convergent validity against biopsy	Inter-rater reliability	Intra-rater reliability	Test-retest reliability	Discriminatory capacity across different participants group	Sensitivity to change	Clinical trial discrimination	Threshold of meaning
Seidenari S 1996							+[1]			
Scheja A 1997				+[2]			+[1]			
Hesselstrand R 2002			+/-[3]					+/-[4]		
Akesson A 2004		-[5]		+	+			+/-[4]		
Hashikabe M 2004								+[4]	+[5]	
Hesselstrand R 2007		+					+			
Hesselstrand R 2008		+					+			
Hassan I 2012							+[1]			
Hesselstrand R 2015		+						+		
Liu H 2017		+		+/-[6]			+			
Flower V 2020		-	-		+/-[8]		+[7]			
Total available studies for each property		6	2	3	2		7	4	1	
Total studies available for synthesis		6	2	3	2		7	4	1	
Rating (RAGW) (put on master checklist)										
Overall rating for instrument across properties	Provisional endorsement: needs additional construct validity against biopsy, test-retest reliability, sensitivity to change, clinical trial discrimination and threshold of meaning									

OMERACT Summary of evidence for measurement properties of echogenicity. Color designates quality of evidence: green=good methods used, use this evidence; amber=some cautions but we will use this evidence. In the rating row, color designates overall evidence-based instrument rating for the core instrument set: green= at least 2 pieces of evidence with good methods and consistent findings of adequate or better performance; amber=in between green and red; red= inadequate performance in at least 1 study that used good methods, white=no evidence.
Arithmetic signs designate the performance of the instrument according to that study (for each measurement property studied): "+" adequate or better performance, "+/−" equivocal performance, "-" less than adequate/poor performance.
[1]No data regarding patient selection, e.g., consecutive or random sample patients enrolled, case-control avoided or avoid inappropriate exclusions.
[2]No data whether design of the study held all other factors constant except for the source of variability being examined and the statistic method used.
[3]No data whether the design and statistical methods adequate for the hypotheses to be tested.
[4]No data whether the criterion for change is considered an adequate gold standard OR is the construct for change clear or were the statistical methods appropriate for the testing situations.
[5]No data stating time interval between testing stated and appropriateness; no statistical methods adequate for the hypotheses tested.
[6]ICC of two readers skin echogenicity classification was 0.608, $p<0.001$.
[7]Dermal echogenicity increased in clinically unaffected skin site in SSc compared with controls, at the finger, hand, forearm and abdomen.
[8]ICC for echogenicity 0.648−0.865.

Stiffness

Study characteristics

Sixteen studies, published between 2010 and 2020, evaluated stiffness, and all, but one [29], were observational and cross-sectional (Table 1). In total, 478 SSc patients (∼92% female and ∼50% with a limited form) were evaluated. Twelve studies had a control group. Cases were defined according to the 1980 ARA criteria in 8 studies, the 2013 ACR/EULAR classification criteria in 7, and both criteria in one study (4.4%).

Description of image acquisition and analysis

Seven studies assessed both stiffness and thickness [32,35,39,40,47,55,67]. Ten studies employed shear-wave elastography and six used compression elastography (table S3). Most studies used linear probes, with frequency ranging from 6.4 to 18 MHz. In three studies, a standoff gel pad was used to provide an acoustic interface to prevent local artefacts [68−70]. The remaining mentioned the use of a generous layer of gel to maintain a minimal compression of the probe to the skin.

Nine (56.3%) of the studies reported that sonographers were blinded to local-mRSS when performing image acquisition, and 7 (43.4%) stated that sonographers had previous experience with musculoskeletal ultrasound (table S4). Two studies gave data regarding the setting conditions, in particular room temperature and/or time of the day [39,68].

All, but one study [70], performed elastography in skin Rodnan sites, but the number of sites evaluated varied from 2 (forearms and fingers) [71] to all 17 Rodnan sites [32,67]. The landmarks for each skin assessment site varied. Only one study evaluated the peri-oral region [70]. Six (37.5%) studies evaluated epidermis and dermis and 2 [35,55] evaluated only the dermis. Eight studies (50%) did not specify the skin layers examined. Studies used different dedicated softwares and machine models, although most of them provided a color-graded elastogram superimposed on the B-mode image.

Elastography was described in a quantitative or in a semi-quantitative color scale [35,68,71−73] (table S3). Quantitative values were reported in different units across studies, i.e., meters per second, m/s; elastic modulus in Kilopascal, kPa; or global percentage of hardness.

There was considerable underreporting regarding the methodology used in the scoring system. Few studies detailed the size and shape (round or square) of the region of interest (ROI). The number of images scanned per site, the number of measurements per image and the exact position of the ROIs inside the elastogram was heterogeneous or unreported in many studies.

Feasibility

Only one study reported the time needed for image acquisition and analysis as being about 2 min per skin site [74]. Therefore, this property was rated AMBER.

Table 4
Summary of measurement properties for ultrasound studies evaluating stiffness.

Study ID	Feasibility	Truth				Discrimination				
		Convergent validity against mRSS	Convergent validity against biopsy	Inter-reliability	Intra-reliability	Test-retest reliability	Discriminatory capacity across different participants groups	Sensitivity to change	Clinical trial discrimination	Threshold of meaning
Iagnocco A 2010				+/-[1]	+/-[1]		+			
Geso L 2011		-		+	+					
Cannao P 2014				+[1]						
Hou Y 2015		+		+			+			
Grembiale R 2016							+			
Santiago T 2016	+/-[2]	+			+		+¥			
Cildag S 2017										
Liu H 2017							+[3]			
Zhang X 2017							+[3]			
Yang Y 2018		+					+[3]			
Aryan A 2018							+[3]			
Chen C 2020		-	-[4]	+			+¥			
Chen Y 2020		-					-[3]			
Flower V 2020		+	+		+		+			
Santiago T 2020		+			+		+	+[4]		
Sobolewski,P 2020		+		+[1]			+[3]¥			
Total available studies for each property	1	9	2	6	5		13	1		
Total studies available for synthesis	1	9	2	6	5		13	1		
Rating (RAGW) (put on master checklist)										
Overall rating for instrument across properties		**Provisional endorsement: needs additional feasibility, construct validity against biopsy, test retest reliability, Sensitivity to change, clinical trial discrimination and threshold of meaning**								

[1]No data whether design of the study held all other factors constant except for the source of variability being examined and the statistic method used.
[2]No data regarding cost of the ultrasound equipment or software, required training, time for image acquisition or analysis.
[3]No data regarding patient selection, e.g., consecutive or random sample patients enrolled, case-control avoided or avoidance of inappropriate exclusions.
[4]No data whether the design and statistical methods adequate for the hypotheses to be tested.
¥Studies reporting ultrasound-stiffness of clinically affected skin in patients stiffer than in controls.
OMERACT Summary of evidence for measurement properties of stiffness. Color designates quality of evidence: GREEN=good methods used, use this evidence; amber=some cautions but we will use this evidence. In the rating row, color designates overall evidence-based instrument rating for the core instrument set: GREEN = at least 2 pieces of evidence with good methods and consistent findings of adequate or better performance; amber=in between GREEN and red; red= inadequate performance in at least 1 study that used good methods, white=no evidence.
Arithmetic signs designate the performance of the instrument according to that study (for each measurement property studied): "+" adequate or better performance, "+/−" equivocal performance, "-" less than adequate/poor performance.

Inter-, intra-rater reliability and test-retest reliability

Nine studies reported reliability assessments (table S4). The ICCs varied from 0.72 to 0.97 for intra-reader, and 0.7 to 0.987 for inter-reader reliability [29,35,40,47,67,68,70,71,74]. Therefore, stiffness was rated GREEN for this measurement property. No studies reported test-retest reliability.

Convergent validity against mRSS, biopsy and other constructs

Ultrasound-stiffness validity against mRSS was assessed in 9 studies and confirmed in 6 (Table 4). Stiffness showed a moderate to strong positive correlation with local site-specific mRSS. Construct validity against skin histological findings, showed conflicting results in 2 studies [40,47]. In 1 of the studies no clear correlation could be established between histological skin thickness and ultrasound skin stiffness [40]. One study found that both ultrasound stiffness and local mRSS correlated strongly with histological dermal collagen content [47]. Construct validity against mRSS and biopsy was rated GREEN and AMBER, respectively.

One study evaluated the correlations between ultrasound stiffness and nailfold videocapillaroscopy patterns in 20 SSc patients [54]. The late pattern (n = 12) was independently associated with increased stiffness.

Discriminatory capacity across different participants groups

The comparison of stiffness between patients and controls was the object of 13 studies. Ultrasound stiffness was found to be higher in SSc patients than in controls in almost all Rodnan sites, in 12 studies. In 1 study there was no differences in elastography strain ratio at the forearms and fingers, between 31 SSc patients and 19 controls [55]. Patients with SSc had stiffer skin compared to controls even in areas of clinically unaffected skin i.e., local Rodnan score of zero [40,68,74]. This measurement property was rated GREEN.

Sensitivity to change

A single longitudinal study showed that skin stiffness decreased significantly and differently in almost all Rodnan skin sites, in 21 SSc patients as well as in 15 controls, over 5 years of follow-up [29]. This demonstrated a higher sensitivity to change over time than mRSS [29]. The authors did not perform statistical subgroup analysis nor

did they report standardized response mean. On this basis, this property was rated AMBER.

Discrimination and threshold of meaning

No studies have reported clinical trial discrimination or threshold of meaning. This was rated WHITE.

Results regarding SOMP table for echogenicity are summarized in Table 4.

Discussion

This SLR summarizes the measurement properties of three skin ultrasound and elastography candidate outcome domains - thickness, echogenicity, and stiffness - and informs the direction of future work necessary to endorse them as outcome domains for the assessment of skin pathology in SSc. None of these domains fully satisfies the OMERACT OFISA criteria, although the unmet needs vary between them.

The synthesis of available evidence underlying the ratings, comprising quality and quantity of studies, consistency of results, and performance of the ultrasound outcome domain across measurement properties, favours *thickness* as the outcome with more robust support. Studies demonstrated promising data concerning its construct validity against mRSS, positive results when direct compared with histologic skin thickness (in 2 studies [40,47]) and good to excellent inter- and intra-reliability. *Stiffness* demonstrated convergent validity against mRSS, good reliability (with ICCs often >0.75, for inter- an intra-rater reliability) as well as cross-sectional construct validity and some evidence (a single study [47]) for sensitivity to change against histological findings. Convergent validity of echogenicity against skin biopsy showed inconclusive data; and inter- and intra-rater reliability needs further studies to draw definite evidence. The main knowledge gaps shared by the three candidates that need to be addressed moving forward are related to feasibility and discrimination, i.e., test-retest reliability, sensitivity to change, discrimination and thresholds of meaning.

Feasibility is an area of weakness shared by all 3 ultrasound and elastography domains. Ultrasound is well accepted by patients and ultrasound machines are widely available to rheumatologists in European countries [75,76]. In addition, skin ultrasound may require minimal training to experienced sonographers. However, the time needed for image acquisition and analysis, the dedicated ultrasound software that are needed as well as their costs and technical demands were underreported in the published studies. The potential advantages and limitations of ultrasound examination in a reduced number of body areas has not been assessed.

Evidence for *convergent validity against* mRSS for the 3 ultrasound domains was derived, mainly, from low-risk bias studies with consistent findings. However, some authors argue that ultrasound outcome domains may actually reflect properties of the skin different from those assessed by the currently validated mRSS, rendering their direct comparisons questionable [74]. It is important to note that almost all studies included patients with established disease, with mean disease duration of $\geq$ 4 years. The evidence based on convergent validity and responsiveness involving subgroups of patients with different cutaneous forms (limited vs diffuse), different clinical (oedematous vs fibrotic vs atrophic) and time phases (pre vs early vs established) is scarce precludes separate analysis at this point [28,34,59]. It is important that future research addresses the performance of ultrasound measurements in the different disease subsets.

Almost three quarters of the studies used a case-control design and underreported data on enrolment, which may result in overoptimistic estimates of diagnostic accuracy. Furthermore, it is important to note that if ultrasound is to be used to aid early diagnosis, normality reference data for the relevant anatomical sites is required.

Two candidate ultrasound outcome domains (thickness and stiffness) have been reported to be reliable, with often ICC >0.75 for inter-reader and >0.85 for intra-reader. However, about one third of the published studies used mixed populations (combining SSc patients and controls) to provide reliability results (table S6).

Another area of weakness of all candidate outcomes is *discrimination*, in particular the absence of evidence for test-retest reliability and sensitivity to change. The threshold of meaning is also of crucial importance, if these ultrasound domains are to serve as outcome measures in future clinical trials: in order to have an acceptable measurement error any given measurement must provide smallest detectable change that is smaller than the minimally important difference [24].

Contextual factors have been relatively overlooked in the published studies and their importance has been recently highlighted by EULAR recommendations [23]. The potential effects of either ambient conditions (room temperature, time of the day) or patient conditions (age, gender, smoking, skin temperature, menopausal status..) on skin ultrasound outcome domains, need to be clarified before they can be ignored.

In summary, the measurement properties of the 3 candidate outcome domains to assess skin involvement in SSc show promising quality, despite a relevant number of knowledge gaps. However, underlying this evidence synthesis, and questioning its accuracy, is the remarkable heterogeneity and lack of information in a variety of technical aspects that may have decisive impact on the conclusions. These include i), probes' frequency, ii) number and precise definition of skin sites assessed, iii) skin layers evaluated, iv) scoring system (i.e., number of images scanned per site, number of measurements per image and position of ROIs within the image), and v) blinding during image acquisition and analysis. All these aspects hinder direct comparisons of studies and undermine their external validity, thus requiring further and harmonized investigation. In particular, we emphasize that the evaluation of the methodological quality (risk of bias) of the studies did not include the skin layers evaluated. However, we recognize that this is a crucial aspect that needs standardization. The use of higher frequency probes (at least 18 MHz) will allow the separate evaluation of the skin layers and this should be reported, to address differences in the involvement of dermis and hypodermis layers [41]. Finally, the recent development of a SSc-specific patient-reported outcome instrument for assessing skin involvement will enable future investigators to examine the convergent validity between ultrasound assessment of skin in SSc and how patients 'feel' and 'function' [77]. Also, a crucial future step to define sensitivity to change and clinical validity of skin ultrasound would be its inclusion as a secondary or exploratory endpoint in the context of randomised controlled trials.

Standardization of procedures and reporting seems, therefore, a crucial step to further develop and consolidate the contribution of ultrasound skin evaluation in SSc, together with research focused on the knowledge gaps identified herein. This SLR therefore provides a substrate for future recommendations and evidence-based research.

Funding

This research received a specific grant from World Scleroderma Foundation, Project title "Skin Ultrasound Working Group", 2021.

Author contributions

Study concept and design: TS, ES and JAPS; Data collection and interpretation: TS and ES; TS prepared the first version of the manuscript.

All authors revised the manuscript critically for intellectual content and gave final approval of the version to be published.

Ethical approval information

No applicable

Data sharing statement

Data sharing not applicable as no datasets generated and/or analysed for this study. All data relevant to the study are included in the article or uploaded as supplementary information.

Declaration of Competing Interest

TS: No conflict of interest
ES: No conflict of interest
BR: No conflict of interest
GL: No conflict of interest
LG: No conflict of interest
MW: No conflict of interest
SW: No conflict of interest
AL: No conflict of interest
RH: No conflict of interest
FDG: reports grants and/or personal fees from AstraZeneca, Abbvie, Boehringer Ingelheim, Capella, Chemomab, Jannsen, Kymab, Mitsubishi-Tanabe.
JDP: reports grants and/or personal fees from Boehringer Ingelheim, Janssen, Permeatus and Sojournix Pharma outside of the submitted work.
AI: No conflicts of interest
JAP da Silva: No conflicts of interest

Acknowledgements

World Scleroderma Foundation for funding this project.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.semarthrit.2022.151954.

References

[1] Allanore Y, Simms R, Distler O, et al. Systemic sclerosis. Nat Rev Dis Primers 2015;04(1):15002. doi: 10.1038/nrdp.2015.2.
[2] Khanna D, Furst DE, Clements PJ, et al. Standardization of the modified Rodnan skin score for use in clinical trials of systemic sclerosis. J Scleroderma Relat Disord 2017;2(1):11–8. doi: 10.5301/jsrd.5000231.
[3] Valentini G, Iudici M, Walker UA, et al. The European Scleroderma Trials and Research group (EUSTAR) task force for the development of revised activity criteria for systemic sclerosis: derivation and validation of a preliminarily revised EUSTAR activity index. Ann Rheum Dis 2017;76(1):270–6. doi: 10.1136/annrheumdis-2016-209768.
[4] Domsic RT, Rodriguez-Reyna T, Lucas M, Fertig N, Medsger TA. Skin thickness progression rate: a predictor of mortality and early internal organ involvement in diffuse scleroderma. Ann Rheum Dis 2011;70(1):104–9. doi: 10.1136/ard.2009.127621.
[5] Steen VD, Medsger TA. The value of the Health Assessment Questionnaire and special patient-generated scales to demonstrate change in systemic sclerosis patients over time. Arthritis Rheum. 1997;40(11):1984–91. doi: 10.1002/art.1780401110.
[6] Clements PJ, Hurwitz EL, Wong WK, et al. Skin thickness score as a predictor and correlate of outcome in systemic sclerosis: high-dose versus low-dose penicillamine trial. Arthritis Rheum. 2000;43(11):2445–54. doi: 10.1002/1529-0131(200011)43:11<2445::AID-ANR11>3.0.CO;2-Q.
[7] D. Khanna VB, E.H. Giannini, J.R. Seibold, P.A. Merkel, M.D. Mayes, M. Baron, P.J. Clements, V. Steen, S. Assassi, E. Schiopu, K. Phillips, R.W. Simms, Y. Allanore, C.P. Denton, O. Distler, S.R. Johnson, M. Matucci-Cerinic, J.E. Pope, S.M. Proudman, J. Siegel, W.K. Wong, A.U. Wells, D.E. Furst. The American College of Rheumatology provisional composite response index for clinical trials in early diffuse cutaneous systemic sclerosis. 2016;
[8] Czirjak L, Foeldvari I, Muller-Ladner U. Skin involvement in systemic sclerosis. Rheumatology (Oxford) 2008;47(Suppl 5):v44–5. doi: 10.1093/rheumatology/ken309.
[9] Amjadi S, Maranian P, Furst DE, et al. Course of the modified Rodnan skin thickness score in systemic sclerosis clinical trials: analysis of three large multicenter, double-blind, randomized controlled trials. Arthritis Rheum 2009;60(8):2490–8. doi: 10.1002/art.24681.
[10] Khanna D, Lin CJF, Furst DE, et al. Tocilizumab in systemic sclerosis: a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Respir Med 2020;8(10):963–74 10. doi: 10.1016/S2213-2600(20)30318-0.
[11] Khanna D, Denton CP, Jahreis A, et al. Safety and efficacy of subcutaneous tocilizumab in adults with systemic sclerosis (faSScinate): a phase 2, randomised, controlled trial. Lancet 2016;387(10038):2630–40. doi: 10.1016/S0140-6736(16)00232-4.
[12] Elhai M, Boubaya M, Distler O, et al. Outcomes of patients with systemic sclerosis treated with rituximab in contemporary practice: a prospective cohort study. Ann Rheum Dis 2019;78(7):979–87 07. doi: 10.1136/annrheumdis-2018-214816.
[13] Ch'ng SS, Roddy J, Keen HI. A systematic review of ultrasonography as an outcome measure of skin involvement in systemic sclerosis. Int J Rheum Dis 2013;16(3):264–72. doi: 10.1111/1756-185x.12106.
[14] Santiago T, Santiago M, Ruaro B, Salvador MJ, Cutolo M, da Silva JAP. Ultrasonography for the assessment of skin in systemic sclerosis: a systematic review. Arthritis Care Res (Hoboken) 2019;71(4):563–74 04. doi: 10.1002/acr.23597.
[15] Michael Hughes CB, Giovanna C, Andrea Delle S, Luna G, Marwin G, Gemma L, et al. The role of ultrasound in systemic sclerosis: on the cutting edge to foster clinical and research advancement. J Scleroderma Relat Disord 2020.
[16] Beaton DE, Maxwell LJ, Shea BJ, et al. Instrument Selection Using the OMERACT Filter 2.1: the OMERACT Methodology. J Rheumatol 2019;46(8):1028–35 08. doi: 10.3899/jrheum.181218.
[17] Higgins JPT TJ, Chandler J, et al. Cochrane handbook for systematic reviews of interventions version 6.0 (updated july 2019): cochrane. 2019.
[18] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. 03. doi: 10.1136/bmj.n71.
[19] Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Arthritis Rheum 1980;23(5):581–90. doi: 10.1002/art.1780230510.
[20] van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. Ann Rheum Dis 2013;72(11):1747–55. doi: 10.1136/annrheumdis-2013-204424.
[21] LeRoy EC, Black C, Fleischmajer R, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988;15(2):202–5.
[22] Alexander H, Miller DL. Determining skin thickness with pulsed ultra sound. J Invest Dermatol 1979;72(1):17–9. doi: 10.1111/1523-1747.ep12530104.
[23] Costantino F, Carmona L, Boers M, et al. EULAR recommendations for the reporting of ultrasound studies in rheumatic and musculoskeletal diseases (RMDs). Ann Rheum Dis 2021. doi: 10.1136/annrheumdis-2020-219816.
[24] Chapter 5: Instrument selection for core outcome measurement sets - workbook. Omeract handbook 2019; 1st of March Available from: https://omeracthandbook. org/workbooks-%26-resources.
[25] Akesson A, Forsberg L, Hederstrom E, Wollheim F. Ultrasound examination of skin thickness in patients with progressive systemic sclerosis (scleroderma). Acta Radiol Diagn (Stockh) 1986;27(1):91–4.
[26] Hesselstrand R, Westergren-Thorsson G, Scheja A, Wildt M, Akesson A. The association between changes in skin echogenicity and the fibroblast production of biglycan and versican in systemic sclerosis. Clin Exp Rheumatol 2002;20(3):301–8.
[27] Akesson A, Hesselstrand R, Scheja A, Wildt M. Longitudinal development of skin involvement and reliability of high frequency ultrasound in systemic sclerosis. Ann Rheum Dis 2004;63(7):791–6. doi: 10.1136/ard.2003.012146.
[28] Hesselstrand R, Carlestam J, Wildt M, Sandqvist G, Andreasson K. High frequency ultrasound of skin involvement in systemic sclerosis - a follow-up study. Arthritis Res Ther 2015;17:329. doi: 10.1186/s13075-015-0853-5.
[29] Santiago T, Santiago M, Coutinho M, Salvador MJ, Da Silva JAP. How much of skin improvement over time in systemic sclerosis is due to normal ageing? A prospective study with shear-wave elastography. Arthritis Res Therapy 2020;22(1). doi: 10.1186/s13075-020-02150-x.
[30] Hashikabe M, Ohtsuka T, Yamazaki S. Quantitative echographic analysis of photochemotherapy on systemic sclerosis skin. Arch Dermatol Res 2005;296(11):522–7. doi: 10.1007/s00403-005-0552-1.
[31] Kuhn A, Haust M, Ruland V, et al. Effect of bosentan on skin fibrosis in patients with systemic sclerosis: a prospective, open-label, non-comparative trial. Rheumatology 2010;49(7):1336–45. doi: 10.1093/rheumatology/keq077.
[32] Liu H, Hou Y, Zhu QL, et al. A preliminary study of skin ultrasound in diffuse cutaneous systemic sclerosis: does skin echogenicity matter? PLoS ONE 2017;12(3):e0174481. doi: 10.1371/journal.pone.0174481.
[33] Sulli A, Ruaro B, Alessandri E, et al. Correlations between nailfold microangiopathy severity, finger dermal thickness and fingertip blood perfusion in systemic sclerosis patients. Ann Rheum Dis 2014;73(1):247–51. doi: 10.1136/annrheumdis-2012-202572.
[34] Kaloudi O, Bandinelli F, Filippucci E, et al. High frequency ultrasound measurement of digital dermal thickness in systemic sclerosis. Ann Rheum Dis 2010;69(6):1140–3. doi: 10.1136/ard.2009.114843.
[35] Di Geso L, Filippucci E, Girolimetti R, et al. Reliability of ultrasound measurements of dermal thickness at digits in systemic sclerosis: role of elastosonography. Clin Exp Rheumatol 2011;29(6):926–32.

[36] Liu H, Hou Y, Zhu QL, et al. A preliminary study of skin ultrasound in diffuse cutaneous systemic sclerosis: does skin echogenicity matter? PLoS ONE 2017;12(3). doi: 10.1371/journal.pone.0174481.

[37] Sousa-Neves J, Cerqueira M, Santos-Faria D, Afonso C, Teixeira F. Ultrasound assessment of skin thickness performed on fingers: a tool for estimating overall severity of skin disease in systemic sclerosis patients? Acta Reumatol Port 2017.

[38] Li H, Furst DE, Jin H, et al. High-frequency ultrasound of the skin in systemic sclerosis: an exploratory study to examine correlation with disease activity and to define the minimally detectable difference. Arthritis Res Therapy 2018;20(1). doi: 10.1186/s13075-018-1686-9.

[39] Yang Y, Yan F, Wang L, et al. Quantification of skin stiffness in patients with systemic sclerosis using real-time shear wave elastography: a preliminary study. Clin Exp Rheumatol 2018;36(Suppl 113(4)):118–25.

[40] Chen C, Cheng Y, Zhu X, et al. Ultrasound assessment of skin thickness and stiffness: the correlation with histology and clinical score in systemic sclerosis. Arthritis Res and Therapy 2020;22(1). doi: 10.1186/s13075-020-02285-x.

[41] Naredo E, Pascau J, Damjanov N, et al. Performance of ultra-high-frequency ultrasound in the evaluation of skin involvement in systemic sclerosis: a preliminary report. Rheumatology (Oxford) 2020;59(7):1671–8 07. doi: 10.1093/rheumatology/kez439.

[42] Sedky MM, Fawzy SM, El Baki NA, El Eishi NH, El Bohy Ael M. Systemic sclerosis: an ultrasonographic study of skin and subcutaneous tissue in relation to clinical findings. Skin Res Technol 2013;19(1):e78–84. doi: 10.1111/j.1600-0846.2012.00612.x.

[43] Daoudi K, Kersten BE, van den Ende CHM, van den Hoogen FHJ, Vonk MC, de Korte CL. Photoacoustic and high-frequency ultrasound imaging of systemic sclerosis patients. Arthritis Res Therapy 2021;23(1). doi: 10.1186/s13075-020-02400-y.

[44] Grembiale RD, Bruno C, Tripolino C, et al. Correlation between elastosonography and nailfold microvascular alterations in systemic sclerosis patients #x202F; Clin Hemorheol Microcirc 2016;62(1):71–8. doi: 10.3233/CH-151948.

[45] Scheja A, Akesson A. Comparison of high frequency (20MHz) ultrasound and palpation for the assessment of skin involvement in systemic sclerosis (scleroderma). Clin Exp Rheumatol 1997;15(3):283–8.

[46] Serup J. Ring size measurement of the digits in females suffering from generalized scleroderma (acrosclerosis). A simple method to quantify skin and soft tissue affection. Dermatologica 1985;171(1):41–4. doi: 10.1159/000249385.

[47] Flower VA, Barratt SL, Hart DJ, et al. High frequency ultrasound assessment of systemic sclerosis skin involvement: intra-observer repeatability and relationship with clinician assessment and dermal collagen content. J Rheumatol 2021;48(6):867–76.

[48] Moore TL, Lunt M, McManus B, Anderson ME, Herrick AL. Seventeen-point dermal ultrasound scoring system—a reliable measure of skin thickness in patients with systemic sclerosis. Rheumatology (Oxford) 2003;42(12):1559–63. doi: 10.1093/rheumatology/keg435.

[49] Sulli A, Ruaro B, Smith V, et al. Subclinical dermal involvement is detectable by high frequency ultrasound even in patients with limited cutaneous systemic sclerosis. Arthritis Res Therapy 2017;19(1). doi: 10.1186/s13075-017-1270-8.

[50] Ihn H, Shimozuma M, Fujimoto M, et al. Ultrasound measurement of skin thickness in systemic sclerosis. Br J Rheumatol 1995;34(6):535–8.

[51] Sulli A, Ruaro B, Pizzorni C, et al. Dermal thickness, peripheral blood perfusion and nailfold microangiopathy in systemic sclerosis patients. Conference Abstract. Ann Rheum Dis 2013;71. doi: 10.1136/annrheumdis-2012-eular.783.

[52] Ruaro B, Pizzorni C, Paolino S, et al. Correlations between nailfold microvascular damage and skin involvement in systemic sclerosis patients. Microvasc Res 2019;125. doi: 10.1016/j.mvr.2019.04.004.

[53] Ruaro B, Sulli A, Pizzorni C, et al. Correlations between blood perfusion and dermal thickness in different skin areas of systemic sclerosis patients. Microvasc Res 2018;115:28–33. doi: 10.1016/j.mvr.2017.08.004.

[54] Hesselstrand R, Kassner A, Heinegard D, Saxne T. COMP: a candidate molecule in the pathogenesis of systemic sclerosis with a potential as a disease marker. Ann Rheum Dis 2008;67(9):1242–8. doi: 10.1136/ard.2007.082099.

[55] Chen YM, Lai KL, Chen HH, et al. Hyperspectral imaging for skin assessment in systemic sclerosis: a pilot study. Revmatologiia (Bulgaria) 2020;59(11):3201–10. doi: 10.1093/rheumatology/keaa067.

[56] Moore TL, Lunt M, McManus B, Anderson ME, Herrick AL. Seventeen-point dermal ultrasound scoring system - a reliable measure of skin thickness in patients with systemic sclerosis. Rheumatology 2003;42(12):1559–63. doi: 10.1093/rheumatology/keg435.

[57] Vanhaecke A, Cutolo M, Heeman L, et al. High frequency ultrasonography: reliable tool to measure skin fibrosis in SSc? A systematic literature review and additional pilot study. Rheumatology (Oxford) 2021. doi: 10.1093/rheumatology/keab462.

[58] Ruaro B, Sulli A, Smith V, et al. The impact of transducer frequency in ultrasound evaluation of subclinical skin involvement in limited cutaneous systemic sclerosis patients. Clin Exp Rheumatol 2019;37(Suppl 119(4):147–8.

[59] Hesselstrand R, Scheja A, Wildt M, Akesson A. High-frequency ultrasound of skin involvement in systemic sclerosis reflects oedema, extension and severity in early disease. Rheumatology (Oxford) 2008;47(1):84–7. doi: 10.1093/rheumatology/kem307.

[60] Li H, Furst DE, Sun C, et al. Skin ultrasound in SSc: an exploratory study to examine correlation with disease activity and to define the minimally detectable difference. Conference Abstract. Int J Rheum Dis 2018;21:135. doi: 10.1111/1756-185X.13361.

[61] Hesselstrand R, Carlestam J, Wildt M, Sandqvist G, Andréasson K. High frequency ultrasound of skin involvement in systemic sclerosis - a follow-up study. Arthritis Res Therapy 2015;17(1). doi: 10.1186/s13075-015-0853-5.

[62] Seidenari S, Belletti B, Conti A. A quantitative description of echographic images of sclerotic skin in patients with systemic sclerosis, as assessed by computerized image analysis on 20 MHz B-scan recordings. Acta Derm Venereol 1996;76(5):361–4.

[63] Hassan I, Sajad P, Jahangir M. High frequency skin ultrasonography in systemic sclerosis. Iran J Dermatol 2012;15(60):47–9.

[64] Hashikabe M, Ohtsuka T, Yamazaki S. Quantitative echographic analysis of photochemotherapy on systemic sclerosis skin. Arch Dermatol Res 2005;296(11):522–7. doi: 10.1007/s00403-005-0552-1.

[65] Åkesson A, Hesselstrand R, Scheja A, Wildt M. Longitudinal development of skin involvement and reliability of high frequency ultrasound in systemic sclerosis. Ann Rheum Dis 2004;63(7):791–6. doi: 10.1136/ard.2003.012146.

[66] Seidenari S, Belletti B, Conti A. A quantitative description of echographic images of sclerotic skin in patients with systemic sclerosis, as assessed by computerized image analysis on 20 MHz B-scan recordings. Acta Derm Venereol 1996;76(5):361–4.

[67] Hou Y, Zhu QL, Liu H, et al. A preliminary study of acoustic radiation force impulse quantification for the assessment of skin in diffuse cutaneous systemic sclerosis. J Rheumatol 2015;42(3):449–55. doi: 10.3899/jrheum.140873.

[68] Sobolewski P, Maślińska M, Zakrzewski J, Paluch Ł, Szymańska E, Walecka I. Applicability of shear wave elastography for the evaluation of skin strain in systemic sclerosis. Article. Rheumatol Int 2020;40(5):737–45. doi: 10.1007/s00296-020-04539-y.

[69] Zhang X, Zhou B, Kalra S, Bartholmai B, Greenleaf J, Osborn T. An ultrasound surface wave technique for assessing skin and lung diseases. Ultrasound Med Biol 2018;44(2):321–31. doi: 10.1016/j.ultrasmedbio.2017.10.010.

[70] Cannaò PM, Vinci V, Caviggioli F, et al. Technical feasibility of real-time elastography to assess the peri-oral region in patients affected by systemic sclerosis. J Ultrasound 2014;17(4):265–9. doi: 10.1007/s40477-014-0119-0.

[71] Iagnocco A, Kaloudi O, Perella C, et al. Ultrasound elastography assessment of skin involvement in systemic sclerosis: lights and shadows. J Rheumatol 2010;37(8):1688–91. doi: 10.3899/jrheum.090974.

[72] Çildağ S, Çildağ MB. The relationship between the degree of skin fibrosis by sonoelastography and the degree of pulmonary involvement in scleroderma. Turk J Med Sci 2017;47(5):1555–9 11. doi: 10.3906/sag-1702-44.

[73] Aryan A, Alaeen H, Dadgostar M, Rostamian A, Ghajarzadeh M. Sonoelastography for skin evaluation in sclerodermic patients. Int J Prev Med 2019;10(1):1–4. doi: 10.4103/ijpvm.IJPVM_43_18.

[74] Santiago T, Alcacer-Pitarch B, Salvador MJ, Del Galdo F, Redmond AC, da Silva JA. A preliminary study using virtual touch imaging and quantification for the assessment of skin stiffness in systemic sclerosis. Clin Exp Rheumatol 2016;34(Suppl 100(5)):137–41.

[75] Mandl P, Ciechomska A, Terslev L, et al. Implementation and role of modern musculoskeletal imaging in rheumatological practice in member countries of EULAR. RMD Open 2019;5(2):e000950. doi: 10.1136/rmdopen-2019-000950.

[76] Neto A, Figueiredo G, Saraiva F, Falcão S. Musculoskeletal ultrasound among rheumatologists in Portugal: state of practice and training. Acta Reumatol Port 2020;45(2):152–4 Apr-Jun 2020.

[77] Man A, Correa JK, Ziemek J, Simms RW, Felson DT, Lafyatis R. Development and validation of a patient-reported outcome instrument for skin involvement in patients with systemic sclerosis. Ann Rheum Dis 2017;76(8):1374–80. doi: 10.1136/annrheumdis-2016-210534.

[78] Myers SL, Cohen JS, Sheets PW, Bies JR. B-mode ultrasound evaluation of skin thickness in progressive systemic sclerosis. J Rheumatol 1986;13(3):577–80.

[79] Brocks K, Stender I, Karlsmark T, Ullman S, Halberg P, Jacobsen S. Ultrasonic measurement of skin thickness in patients with systemic sclerosis. Acta Derm Venereol 2000;80(1):59–60.

[80] Hesselstrand R, Westergren-Thorsson G, Scheja A, Wildt M, Åkesson A. The association between changes in skin echogenicity and the fibroblast production of biglycan and versican in systemic sclerosis. Clin. Exp. Rheumatol. 2002;20(3):301–8.

[81] Kissin EY, Schiller AM, Gelbard RB, et al. Durometry for the assessment of skin disease in systemic sclerosis. Arthritis Care Res 2006;55(4):603–9. doi: 10.1002/art.22093.

[82] Sulli A, Ruaro B, Alessandri E, et al. Correlations between nailfold microangiopathy severity, finger dermal thickness and fingertip blood perfusion in systemic sclerosis patients. Ann Rheum Dis 2014;73(1):247–51. doi: 10.1136/annrheumdis-2012-202572.

[83] Ruaro B, Soldano S, Smith V, et al. Correlation between circulating fibrocytes and dermal thickness in limited cutaneous systemic sclerosis patients: a pilot study. Rheumatol Int 2019;39(8):1369–76. doi: 10.1007/s00296-019-04315-7.