ORIGINAL RESEARCH

Reliability Exercise of Ultrasound Salivary Glands in Sjögren's Disease: An International Web Training Initiative

Baptiste Quéré · Alain Saraux (iD) · Guillermo Carvajal-Alegria · Dewi Guellec · Gaël Mouterde ·

Christophe Lamotte · Daniel Hammenfors · Malin Jonsson · Sung-Eun Choi · Min Hong-Ki ·

Alja Stel · Benjamin A. Fisher · Mark Maybury · Benedikt Hofauer · Francesco Ferro ·

Vera Milic · Dana Direnzo · Valérie Devauchelle-Pensec · Sandrine Jousse-Joulin

ABSTRACT

Introduction: Major salivary gland ultrasonography (SGUS) demonstrated its good metric properties as an outcome measure for diagnosing primary Sjögren's disease (SD). The objective was to assess SGUS reliability among sonographers with different levels of experience, using web training.

Methods: Sonographers from expert centers participated in the reliability exercise. Before exercises, training was done by videoconferencing. Reliability of the two most experienced sonographers (MES) was assessed and then compared to other sonographers. Intra-reader and inter-reader reliability of SGUS items were assessed by computing Cohen's κ coefficients.

Results: All sets were read twice by all 14 sonographers within a 4-month interval. Intra-reader reliability of MES was almost perfect for homogeneity, substantial for Outcome Measures in Rheumatology (OMERACT) scoring system (OMERACTss). Among LES (less experi-

B. Quéré · A. Saraux · D. Guellec · V. Devauchelle-Pensec · S. Jousse-Joulin
Rheumatology Department, CHU Hospital La Cavale-Blanche, Centre National de Référence des Maladies Auto-Immunes Rares (CERAINO), Brest, France

B. Quéré · A. Saraux · V. Devauchelle-Pensec · S. Jousse-Joulin
LBAI, UMR1227, Univ Brest, Inserm, Brest, France

G. Carvajal-Alegria
Rheumatology Department, Hospital, CHU Trousseau Tours, Chambray-lès-Tours, France

D. Guellec
CIC INSERM 1412, Brest, France

G. Mouterde
Rheumatology Department and IDESP, CHU Montpellier, Montpellier University, Montpellier, France

C. Lamotte
Internal Medicine Department, Hospital Huriez, Lille, France

D. Hammenfors
Department of Rheumatology, Haukeland University Hospital, Bergen, Norway

D. Hammenfors
Broegelmann Research Laboratory, University of Bergen, Bergen, Norway

enced sonographers), reliability was moderate to almost perfect for homogeneity, fair to moderate for OMERACTss, and fair to almost perfect for binary OMERACTss. Inter-reader reliability between MES was almost perfect for homogeneity, substantial for diagnosis, moderate for OMERACTss, and substantial for binary OMERACTss. Compared to MES, reliabilities of LES were moderate to almost perfect for both homogeneity and diagnosis, only fair to moderate for OMERACTss, but increased in binary OMERACTss.

Conclusions: Videoconferencing training sessions in an international reliability exercise could be an excellent tool to train experienced and less-experienced sonographers. SGUS homogeneity items is useful to distinguish normal from abnormal salivary glands parenchyma independently of diagnosis. Structural damage evaluations by OMERACT scoring system is a new comprehensive score to diagnose patients with SD and could be easily used by sonographers in a binary method.

PLAIN LANGUAGE SUMMARY

The goal of this project was to evaluate the reliability of salivary gland ultrasonography in patients with Sjögren's disease using online training in an international study. Currently, salivary gland ultrasonography is routinely used only by European expert sonographers but few studies have studied intra-reader and inter-reader reliability, among less experienced international sonographers. Many salivary gland ultrasonography scoring systems are used today, but it is difficult to know how to put them into practice. Online training on an international level allows a significant number of practitioners to use the different scoring systems including the latest OMERACT (Outcome Measures in Rheumatology) score, which is simple and comprehensive. There were two phases to this project: A first step consisted in a training session by videoconferencing to all sonographers, the second step was an inter and intra-reader reliability exercises. The results of our study showed satisfactory results, especially for parenchyma homogeneity. Regarding the comprehensive OMERACT score, the results are quite disparate, notably for less experienced sonographers and could be explained by this

M. Jonsson
Department of Clinical Dentistry-Section for Oral and Maxillofacial Radiology, University of Bergen, Bergen, Norway

S.-E. Choi
Chonnam National University Bitgoeul Hospital, Gwangju, South Korea

M. Hong-Ki
Division of Rheumatology, Department of Internal Medicine, Konkuk University Medical Center, Seoul, Korea

A. Stel
UMCG, Groningen, The Netherlands

B. A. Fisher · M. Maybury
National Institute of Health Biomedical Research Centre and Department of Rheumatology, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

B. A. Fisher · M. Maybury
Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK

B. Hofauer
Department of Otorhinolaryngology/Head and Neck Surgery, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany

F. Ferro
Rheumatology Unit, Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

V. Milic
Institute of Rheumatology, Faculty of Medicine, University of Belgrade, Beograd, Serbia

D. Direnzo
Division of Rheumatology, Johns Hopkins University, 5200 Eastern Ave., MFL Bldg. Center Tower, Suite 4100, Baltimore, MD 21224, USA

A. Saraux (✉)
Service de Rhumatologie, CHRU Cavale Blanche, Boulevard Tanguy Prigent, 29200 Brest, France
e-mail: alain.saraux@chu-brest.fr

new comprehensive scoring system. However, when binary OMERACT score (minor damage versus major damage of salivary gland parenchyma (OMERACT score 0–1 vs. 2–3) was employed, reliability increased and can be very useful for novice sonographers in routine practice because it does not require scoring of all the pathological features in Sjögren's disease. This study highlights the need to train non-experts interested in this field and demonstrates the potential for beginners to quickly become experts.

Keywords: Web-based training; Reliability exercise; Salivary glands; Sjögren's syndrome; Sjögren's disease; Ultrasound

Key Summary Points

Why carry out this study?

Salivary gland ultrasonography (SGUS) is only used by European expert sonographers.

Reliability among less experienced international sonographers has not been studied.

The objective was to evaluate the reliability of SGUS in Sjögren's disease using online training in an international study.

What was learned from the study?

Web-based training sessions could be an excellent tool to train sonographers.

SGUS homogeneity items is reliable to distinguish normal or abnormal parenchyma.

OMERACT scoring system could be used by non-expert sonographers in a binary method.

INTRODUCTION

Primary Sjögren's syndrome (pSS), also recently renamed Sjögren's disease (SD) [1], is a multi-system autoimmune inflammatory disease that affects multiple systems. It is characterized by the infiltration of lymphoid cells into the exocrine glands, particularly the salivary and lacrimal glands, leading to decreased gland function and resulting in symptoms such as dry eyes and mouth. Oral dryness is one of the most common manifestations of SD. The 2016 ACR/EULAR (American College of Rheumatology/European Alliance of Associations for Rheumatology) classification criteria for diagnosis of SD is based on five items (positivity for anti-SSA (Ro) antibodies, focal lymphocytic sialadenitis with focus score (FS) $\geq$ 1 foci/mm^2 in a minor labial salivary gland biopsy, abnormal ocular staining score $\geq$ 5, Schirmer test $\leq$ 5 mm/5 min and an unstimulated salivary flow rate $\leq$ 0.1 ml/min) but do not consider ultrasound sonography (US) of the salivary gland (SG) [2]. More recent data from a large international survey have demonstrated the usefulness of major salivary gland ultrasonography (SGUS) in classifying patients. The inclusion of SGUS with a similar weighting to other minor items improves the performance of the 2016 ACR/EULAR classification criteria [3]. The diagnosis of SD may be challenging in patients with suspected Sjögren's disease, due to the absence of pathognomonic signs and many other conditions can produce similar presentations [4]. Various imaging techniques, including sialography, parotid scintigraphy, parotid magnetic resonance imaging (MRI), and most recently SGUS, have been proposed to improve the diagnostic approach to SD.

First proposed in 1988 [5], SGUS has proven to be a valuable tool for assessing damage in patients with SD, with its usefulness demonstrated since 1992 [6]. SGUS is a simple, non-invasive, non-irradiating, low-cost and effective examination that can easily be repeated over time. SGUS is performed at both parotid glands (PG) and submandibular glands (SMG). SGUS has demonstrated its good metric properties as an outcome measure for diagnosing SD, and

may be helpful in evaluating SG parenchyma and in defining the presence of SG involvement [4, 7]. Studies have reported SGUS abnormalities (grade 2 or higher) in 62.8% of patients with SD, with no significant differences according to disease duration [8]. The SGUS score assesses greyscale (B-mode) features such as echogenicity, homogeneity, location of an (hypo)echoic area, hyperechoic bands. More recently, power Doppler mode (D-mode) has been used to evaluate inflammatory activity in SD [9–11]. These two new simple scoring systems for SGUS (B-mode and Doppler-mode) have been developed by the OMERACT (Outcome Measures in Rheumatology) ultrasound subgroup in Sjögren's disease.

Since the development of SGUS for SD diagnosis, several scoring systems have been developed, but no significant differences have been found in terms of reproducibility, sensitivity or specificity [12]. Several studies analyzed the impact of adding SGUS criteria for salivary gland involvement in SD, showing an increase in sensitivity without decreasing specificity [3, 8, 13, 14]. Currently, SGUS is routinely used only by expert sonographers. Few studies have studied intra-reader and inter-reader reliability, particularly among less experienced sonographers. De Vita et al., in 2020, from an European experience, demonstrated an excellent intra-reader reliability (Light's kappa of 0.86) and a correct inter-reader reliability (Light's kappa of 0.77) in a group of SGUS expert sonographers [15, 16].

The objective of our study was to assess the reliability of SGUS interpretation, using some independent items and a comprehensive new simplified OMERACT scoring system, among sonographers from different countries and different levels, following web-based training. This first study is a prelude for the beginning of an international SGUS study to evaluate Modification Abnormalities of Salivary glands in SD According to disease duration (MASAI study).

METHODS

We conducted an international ultrasound reliability exercise in 12 expert centers for SD from nine countries. These centers routinely perform SGUS. The exercise included a total of 14 sonographers with different levels of experience in SGUS from France (Brest, Montpellier, Lille), Norway (Bergen), Germany (Munich), England (Birmingham), Serbia (Belgrade), Holland (Groningen), Italy (Pisa), South Korea (Gwangiu, Seoul) and USA (Baltimore), in two sessions of web-based training in July and September 2019. A web-based exercise was then conducted in September and December 2019. Among them, two participants were qualified as most experienced sonographers (MES) due to their participation in previous SGUS training and their experiences as a sonographer. One expert of SGUS from the center leading the project (SJJ), did not participate in the reliability study due to their role in training and ultrasound images selection.

Web-Based Training

The training sessions were conducted by an expert of SGUS (SJJ) through videoconferencing, with each session lasting for 2 h. The purpose of the training was twofold: (1) to familiarize the participants with different pathological findings in SGUS using greyscale imaging. This included description of glandular homogeneity or inhomogeneity, location of hypoechoic areas (Supplementary data Fig. S1A) and the presence or absence of hyperechoic bands (Supplementary data Fig. S1B) to explain the new OMERACT scoring system, defined by a four-grade semiquantitative scoring system, grade 0 correspond to a normal parenchyma; grade 1, minimal change, mild inhomogeneity without anechoic/hypoechoic areas or fatty gland; grade 2, moderate change, inhomogeneity with focal anechoic/hypoechoic areas and with persistent normal parotid tissue; grade 3, severe change, diffuse inhomogeneity with anechoic/hypoechoic areas occupying the entire gland surface, or fibrous gland [17]. During the exercise, SGUS findings were assessed using a standardized file (Supplementary Table S1), with the following parameters: homogeneity (yes/no), location of hypoechoic aera (0 = none, 1 = isolated (< 25%),

2 = localized (15–50%), 3 = diffuse (> 50% of the surface), hyperechoic band (defined as hyperechoic lines without acoustic shadow located in the PG or SMG parenchyma according to a grading as follows: 0 = none, 1 = in less than 25%, 2 = between 25 and 50%, 3 = more than 50% of parenchyma), comprehensive OMERACT scoring system (graded from 0 (healthy parenchyma) to 3 (diffuse hypo/anechoic areas occupying all the surface of the glands or a fibrotic aspect of parenchyma leading SG parenchyma indistinguishable from adjacent soft tissue) [17]), binary comprehensive OMERACT (0–1 versus 2–3, respectively minor damage versus major damage of parenchyma salivary gland) and ultrasound diagnosis of Sjögren's disease (yes/no). The ultrasound diagnosis appreciation of Sjögren's disease was determined based on the conviction that SGUS images corresponded to a patient with Sjögren's disease within a set of images from patients with suspected to have SD.

Web-Based Exercise

After training sessions, two reliability exercise sessions were performed. Each participant evaluated a set of high-quality ultrasound images of salivary glands (60 greyscale static images, 30 parotid and 30 submandibular glands), selected by one expert of SGUS (SJJ) from an anonymized database of patients with suspected Sjögren's disease (Fig. 1). This study was carried out using images from patients with suspected Sjögren's syndrome included in CHU of Brest. According to French regulation, all images were anonymized so IRB approval was not required for this study.

For intra-reader reliability assessment, ultrasound images were scored by each participant at two different times, with a 4-month interval between the two training exercises. Images were presented in a different order during each session.

Statistical Analysis

Statistical analysis was performed to determine intra-reader and inter-reader reliabilities.

Cohen's κ coefficients were computed using SPSS 25.0 (SPSS Inc., Chicago, IL, USA), and the interpretation of the coefficients was as follows: slight (0–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1).

First, reading by participants was used to compare inter-reader reliability. The most experienced sonographer (MES1) was considered as reference to compare inter-reader reliability with LES.

Theory

Building upon the growing prevalence and demonstrated efficacy of SGUS among expert sonographers, this study asses the reliability of SGUS in Sjögren's disease across different levels of expertise, particularly following web–based training. This SGUS reliability exercise is a prerequisite to a future international SGUS study to evaluate modification abnormalities of salivary glands in SD according to disease duration.

RESULTS

All sonographers participated in the entire session's program. All of them read all 60 images twice with an interval of 4 months between two reliability exercise sessions for intra-observer reliability.

Intra-reader Reliability

Intra-reader reliability for the two MES was perfect for homogeneity, substantial to almost perfect for an anechoic or hypoechoic area's location, hyperechoic bands, and comprehensive OMERACT. Ultrasound diagnosis appreciation of Sjögren's disease was perfect.

Intra-reader reliability of LES was fair to almost perfect for homogeneity, fair to substantial for anechoic/hypoechoic areas location, hyperechoic bands, and comprehensive OMERACT. Ultrasound diagnosis appreciation was moderate to almost perfect. All corresponding results are presented in Table 1.
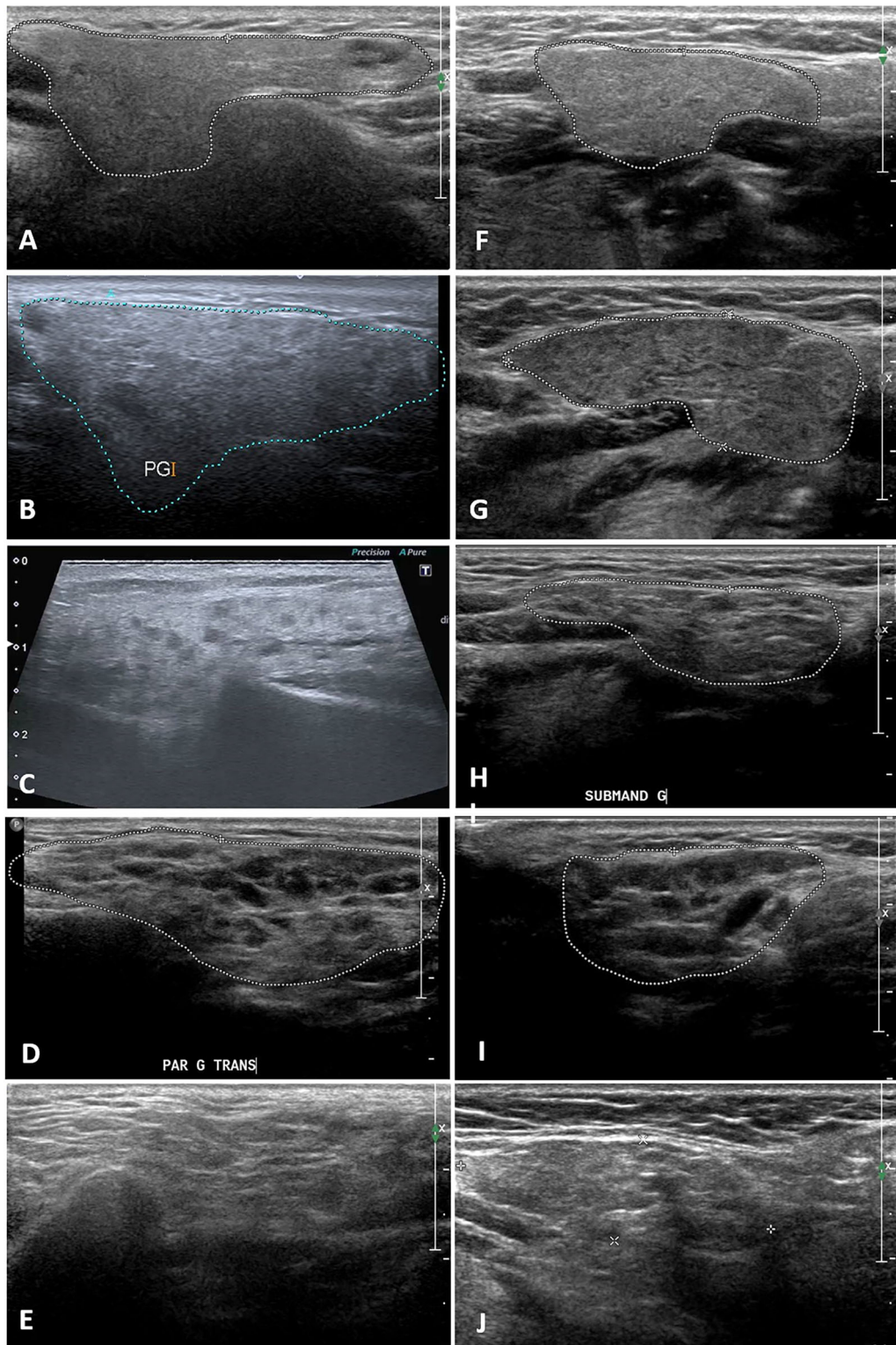
◄Fig. 1 Examples of ultrasound images from the training database in a four-grade (0–3) semiquantitative scoring system, **A–E** are parotid gland: **A** OMERACT score grade 0; **B** OMERACT score grade 1; **C** OMERACT score grade 2; **D** OMERACT score grade 3; **E** OMERACT score grade 3 fibrosis; **F–J** are submandibular gland: **F** OMERACT score grade 0; **G**: OMERACT score grade 1; **H**: OMERACT score grade 2; **I** OMERACT score grade 3; **J** OMERACT score grade 3 fibrosis. *OMERACT* Outcome Measures in Rheumatology, *PAR G TRANS* left parotid gland in transversal view, *PG* parotid gland, *SUBMAND G* left submandibular gland

Inter-reader Reliability

Inter-reader reliability between the two MES was almost perfect for homogeneity, substantial for diagnosis and moderate for OMERACT scoring system. Reliabilities of LES versus MES1 (as the index score) were moderate to almost perfect for both homogeneity and diagnosis but only fair to moderate for comprehensive OMERACT score. All data are presented in Table 2.

Detailed Analysis of Reliability Among MES

Reliability of OMERACT scoring system between MES was analyzed in detail (Table 3); 38/60 images had perfect reliability with an identical OMERACT score (eight images with comprehensive OMERACT 0; two images with comprehensive OMERACT 1; 15 images with comprehensive OMERACT 2, and 13 images with comprehensive OMERACT 3). When differences in the OMERACT grade score exist, in most cases (19/60), it consisted in a shift toward one of the surrounding categories. All results are detailed in Table 3. For the binary OMERACT scoring system (minor damage versus major damage of parenchyma salivary gland (respectively with an OMERACT score 0–1 vs. 2–3), intra-reader reliability of the most experienced sonographers was substantial to almost perfect, with a kappa value of 0.77 and with a kappa value of 0.89. Most of all, inter-reader reliability was substantial ($k = 0.65$) in binary comprehensive OMERACT system between MES.

Detailed Analysis of Reliability Among LES

Among LES, use of the binary OMERACT scoring system (i.e., minor damage versus major damage of parenchyma salivary gland in other words an OMERACT score 0–1 vs. 2–3), increased intra-reader reliability for all participants, except one. As a matter of fact, intra-reader reliability was almost perfect for three participants, substantial for three and moderate for the others (Table 4). Concerning inter-reader reliability among LES for a binary OMERACT scoring system, it was substantial for 32/66, moderate for 21/66, fair for 8/66, and almost perfect for 5/66 (no slight score). All corresponding results are described in Table 5.

DISCUSSION

In this international reliability exercise, we confirm good intra-reader and inter-reader reliability for the assessment of ultrasound grey-scale abnormalities of Salivary glands in SD using a videoconferencing training session. First of all, teaching was done in two steps, using video-conferencing in web training sessions followed by a reliability exercise on static images with the different experts. This approach was chosen to evaluate the effectiveness of videoconferencing training in an international SGUS exercise for use in future clinical trials. It is worth noting that previous studies assessing the reliability of sonographers in SGUS were solely based on expert panels and did not include sonographers with different levels of experience. In our study, we aimed to fill this gap by including sonographers with varying levels of experience. Web-based training for ultrasonography has already been described in the literature, with satisfactory results, which seems comparable to traditional classroom training [18–20].

The results of our study showed satisfactory results, especially for reliability of scores on the salivary gland parenchyma homogeneity,

Table 1 Intra-reader reliability of the Web exercise of the reading of SGUS abnormalities

Participants		Homogeneity (k score)	Hypoechoic area (k score)	Hyperechoic bands (k score)	Comprehensive OMERACT (k score)	Sjögren's Diagnosis appreciation (k score)
MES	MES 1	0.88	0.88	0.89	0.73	0.92
	MES 2	0.86	0.75	0.79	0.81	0.92
LES	LES 1	0.47	0.50	0.50	0.41	0.50
	LES 2	0.60	0.56	0.45	0.24	0.76
	LES 3	0.40	0.61	0.55	0.48	0.66
	LES 4	0.86	0.67	0.32	0.64	0.82
	LES 5	0.81	0.66	0.28	0.71	0.93
	LES 6	0.84	0.30	0.31	0.80	0.79
	LES 7	0.53	0.63	0.23	0.52	0.83
	LES 8	0.61	0.70	0.63	0.55	0.71
	LES 9	0.49	0.60	0.38	0.50	0.75
	LES 10	0.76	0.48	0.38	0.48	0.73
	LES 11	0.67	0.46	0.44	0.31	0.71
	LES 12	0.83	0.37	0.33	0.34	0.67

LES less experienced sonographers, *MES* most experienced sonographers, *OMERACT* Outcome Measures in Rheumatology

which was closely related to the SGUS diagnosis appreciation. This indicates that videoconferencing using static ultrasound grey images of salivary glands is effective for training in performing these ultrasounds. In accordance with the literature, we confirm that homogeneity is the most reliable item for both MES and LES [21]. However, it is important to note that homogeneity or heterogeneity alone is not sufficient to diagnose SD. On the contrary, the OMERACT scoring system is applicable to patients with SD, taking into account all elementary pathological elementary lesions and it should be emphasized that homogeneity is only a part of the OMERACT scoring system. Our findings suggest that the assessment of heterogeneity items, which include different elementary lesions such as hypoechoic areas and hyperechoic bands, may be relatively easier based on the observed kappa values. However, the evaluation of hypoechoic areas can sometimes be challenging, as they may not be well

defined, and the assessment of hyperechoic bands can also pose difficulties. Hyperechoic bands in normal conditions correspond to fascia between lobule in parotid glands. When numerous hyperechoic bands are present, the gland becomes heterogeneous and hyperechoic, indicating the presence of fat deposits or fibrous tissue. This glandular state (grade 3 fibrous), defined by OMERACT scoring system is difficult to assess in an independent item. In summary, while homogeneity is a reliable item in SGUS assessment, it is important to consider the complete OMERACT scoring system, as it incorporates various pathological elements. The evaluation of heterogeneity including hypoechoic areas and hyperechoic bands can be more challenging but is crucial for a comprehensive assessment of salivary gland abnormalities in SD. The lower kappa values observed for hyperechoic bands in the structural damage assessment of the OMERACT scoring system can be attributed to the challenges in defining and

Table 2 Inter-reader reliability between 14 participants. MES1 is taking as reference

	Homogeneity (*k* score)	Hypoechoic area (*k* score)	Hyperechoic bands (*k* score)	Comprehensive OMERACT (*k* score)	Diagnosis appreciation (*k* score)
Concordance between MES					
MES1/MES2	0.83	0.38	0.28	0.49	0.72
Concordance between MES1 and LES					
MES1/LES1	0.55	0.46	0.11	0.46	0.67
MES1/LES2	0.81	0.16	0.05	0.38	0.79
MES1/LES3	0.52	0.59	0.46	0.40	0.53
MES1/LES4	0.83	0.34	0.15	0.49	0.81
MES1/LES5	0.62	0.60	0.14	0.52	0.68
MES1/LES6	0.71	0.10	0.09	0.42	0.72
MES1/LES7	0.76	0.48	0.42	0.49	0.63
MES1/LES8	0.56	0.63	0.29	0.54	0.67
MES1/LES9	0.68	0.32	0.32	0.41	0.61
MES1/LES10	0.76	0.48	0.37	0.35	0.66
MES1/LES11	0.59	0.49	0.10	0.41	0.58
MES1/LES12	0.76	0.36	0.08	0.23	0.56

LES less experienced sonographers, *MES* most experienced sonographers, *OMERACT* Outcome Measures in Rheumatology

evaluating these bands consistently. This could explain the lower reliability in this aspect of SGUS evaluation. Regarding the comprehensive OMERACT scoring system, the results showed substantial intra-reader reliability and moderate inter-reader reliability for MES, while fair to moderate inter-reader reliability was observed for LES. However, when the binary OMERACT scoring system was employed, categorizing grades 0–1 as no or low parenchyma damage and grades 2–3 as major parenchyma damage, the reliability increased, with a substantial to almost perfect intra-reader reliability and substantial inter-reader reliability in MES. The good results obtained with binary OMERACT system is very useful and shows that the OMERACT scoring system could be simplified between 0

and 1 normal and 2 and 3 abnormal. However, fibrous appearance and hypoechoic areas aspect would not be described in this simplified classification. The moderate reliability observed for the 0–3 OMERACT scoring system, primarily due to a one-point difference between observers, may not significantly impact the patient's overall status in routine practice and this is corroborated by the good reliability in ultrasound diagnosis appreciation among MES but also among LES.

The OMERACT scoring system is a new simple evaluation of salivary glands for diagnosis and may also be useful in assessing change in patients with SD over time. In routine practice, this scoring system is very helpful because it does not require scoring of all the pathological

Table 3 Reliability of comprehensive OMERACT scoring system between MES 1 and 2 according to each grade

		Comprehensive OMERACT MES 2				
	Grade	0	1	2	3	Total
Comprehensive OMERACT MES1	0	8	0	0	0	8
	1	6	2	2	0	10
	2	0	1	15	9	25
	3	1	2	1	13	17
Total		15	5	18	22	60

MES most experienced sonographers, *OMERACT* Outcome Measures in Rheumatology

Table 4 Intra-reader reliability of less experienced sonographers with comprehensive OMERACT score and binary comprehensive OMERACT score

	Comprehensive OMERACT score (*k* score)	Binary comprehensive OMERACT score (0–1 vs. 2–3) (*k* score)
LES 1	0.41	0.47
LES 2	0.24	0.58
LES 3	0.48	0.60
LES 4	0.64	0.82
LES 5	0.71	0.93
LES 6	0.80	0.83
LES 7	0.52	0.73
LES 8	0.55	0.70
LES 9	0.50	0.70
LES 10	0.48	0.58
LES 11	0.31	0.40
LES 12	0.34	0.32

LES less experienced sonographers, *OMERACT* Outcome Measures in Rheumatology

features in SD described by Jousse-Joulin et al. and it can be easily applicable for novice ultra-sonographers [21]. However, results of LES are lower than MES, confirming that it is necessary to train and practice salivary glands ultrasound regularly in order to have a better accuracy. It should be noted that, due to the variation in experience level of sonographers, only kappa coefficients have been calculated and Light's kappa coefficient has not been calculated because Light's kappa would have been biased by extreme results.

Our study has some limitations. Firstly, the definition of MES based on their participation in previous SGUS training and their experience as sonographers could introduce some subjectivity and lack of consensus to determining their expertise level. Secondly, the use of static ultrasound images differs from live images acquisition, which may impact the interpretation and assessment of SGUS findings. Lastly, no color Doppler US was used because the results of the new OMERACT scoring system was not yet published [22] and it would be interesting to explore the inclusion of color Doppler in future studies to further enhance the evaluation of inflammatory activity in SD. SGUS diagnosis was a subjective appreciation of sonographers on the diagnosis of Sjögren's disease based only on SGUS images.

Finally, the reliability exercise demonstrated that SGUS can be a valuable tool for evaluating salivary gland parenchymal damage in patients with SD, particularly among MES, but can be useful for less trained sonographers. This SGUS reliability exercise is a prerequisite to a future international SGUS study to evaluate modification abnormalities of salivary glands in SD according to disease duration, using web training.

Table 5 Inter-reader reliability between all less experienced sonographers after changing OMERACT scoring system in binary items

	LES1	LES2	LES3	LES4	LES5	LES6	LES7	LES8	LES9	LES10	LES11	LES12
LES1		0.60	0.65	0.75	0.69	0.63	0.67	0.71	0.69	0.55	0.67	0.30
LES2			0.53	0.78	0.72	0.72	0.58	0.87	0.65	0.50	0.57	0.31
LES3				0.68	0.55	0.63	0.47	0.67	0.56	0.45	0.47	0.23
LES4					0.86	0.81	0.64	0.79	0.73	0.56	0.70	0.35
LES5						0.68	0.77	0.68	0.80	0.70	0.77	0.38
LES6							0.48	0.82	0.55	0.41	0.53	0.25
LES7								0.63	0.83	0.80	0.67	0.42
LES8									0.68	0.56	0.55	0.37
LES9										0.70	0.63	0.36
LES10											0.65	0.52
LES11												0.43
LES12												

LES less experienced sonographers, *OMERACT* Outcome Measures in Rheumatology

CONCLUSIONS

Videoconferencing training sessions in an international reliability exercise could be an excellent tool to train expert and non-expert sonographers. SGUS homogeneity is the most reliable item and remains a good feature to distinguish normal from abnormal SG parenchyma independently of diagnosis. However, structural damage evaluation by the new comprehensive OMERACT scoring system to diagnose SD can be easily used by non-expert sonographers in a binary method.

images were anonymized so IRB approval was not required for this study.

REFERENCES

1. Baer AN, Hammitt KM. Sjögren's disease, not syndrome. Arthritis Rheumatol. 2021;73(7):1347–8.

2. Shiboski CH, Shiboski SC, Seror R, Criswell LA, Labetoulle M, Lietman TM, et al. 2016 American College of Rheumatology/European League Against Rheumatism Classification Criteria for Primary Sjögren's Syndrome: a Consensus and Data-Driven Methodology Involving Three International Patient Cohorts. Arthritis Rheumatol. 2017;69(1):35–45.

3. Jousse-Joulin S, Gatineau F, Baldini C, Baer A, Barone F, Bootsma H, et al. Weight of salivary gland ultrasonography compared to other items of the 2016 ACR/EULAR classification criteria for Primary Sjögren's syndrome. J Intern Med. 2020;287(2):180–8.

4. Luciano N, Baldini C, Tarantini G, Ferro F, Sernissi F, Varanini V, et al. Ultrasonography of major salivary glands: a highly specific tool for distinguishing primary Sjögren's syndrome from undifferentiated connective tissue diseases. Rheumatol Oxf Engl. 2015;54(12):2198–204.

5. de Clerck LS, Corthouts R, Francx L, Brussaard C, de Schepper A, Vercruysse HA, et al. Ultrasonography and computer tomography of the salivary glands in the evaluation of Sjögren's syndrome. Comparison with parotid sialography. J Rheumatol. 1988;15(12):1777–81.

6. De Vita S, Lorenzon G, Rossi G, Sabella M, Fossaluzza V. Salivary gland echography in primary and secondary Sjögren's syndrome. Clin Exp Rheumatol. 1992;10(4):351–6.

7. Mossel E, Delli K, van Nimwegen JF, Stel AJ, Kroese FGM, Spijkervet FKL, et al. Ultrasonography of major salivary glands compared with parotid and labial gland biopsy and classification criteria in patients with clinically suspected primary Sjögren's syndrome. Ann Rheum Dis. 2017;76(11):1883–9.

8. Cornec D, Jousse-Joulin S, Pers JO, Marhadour T, Cochener B, Boisramé-Gastrin S, et al. Contribution of salivary gland ultrasonography to the diagnosis of Sjögren's syndrome: toward new diagnostic criteria? Arthritis Rheum. 2013;65(1):216–25.

9. Jousse-Joulin S, Devauchelle-Pensec V, Cornec D, Marhadour T, Bressollette L, Gestin S, et al. Brief report: ultrasonographic assessment of salivary gland response to rituximab in primary Sjögren's syndrome. Arthritis Rheumatol. 2015;67(6):1623–8.

10. Carotti M, Salaffi F, Manganelli P, Argalia G. Ultrasonography and colour Doppler sonography of salivary glands in primary Sjögren's syndrome. Clin Rheumatol. 2001;20(3):213–9.

11. Shimizu M, Okamura K, Yoshiura K, Ohyama Y, Nakamura S. Sonographic diagnosis of Sjögren syndrome: evaluation of parotid gland vascularity as a diagnostic tool. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2008;106(4):587–94.

12. Martel A, Coiffier G, Bleuzen A, Goasguen J, de Bandt M, Deligny C, et al. What is the best salivary gland ultrasonography scoring methods for the diagnosis of primary or secondary Sjögren's syndromes? Jt Bone Spine. 2019;86(2):211–7.

13. Cornec D, Jousse-Joulin S, Marhadour T, Pers JO, Boisramé-Gastrin S, Renaudineau Y, et al. Salivary gland ultrasonography improves the diagnostic performance of the 2012 American College of Rheumatology classification criteria for Sjögren's syndrome. Rheumatol Oxf Engl. 2014;53(9):1604–7.

14. Le Goff M, Cornec D, Jousse-Joulin S, Guellec D, Costa S, Marhadour T, et al. Comparison of 2002 AECG and 2016 ACR/EULAR classification criteria and added value of salivary gland ultrasonography in a patient cohort with suspected primary Sjögren's syndrome. Arthritis Res Ther. 2017;19(1):269.

15. Zabotti A, Zandonella Callegher S, Tullio A, Vukicevic A, Hocevar A, Milic V, et al. Salivary gland ultrasonography in Sjögren's syndrome: a European multicenter reliability exercise for the HarmonicSS Project. Front Med. 2020;7: 581248.

16. Finzel S, Jousse-Joulin S, Costantino F, Hánová P, Hocevar A, Iagnocco A, et al. Patient-based reliability of the Outcome Measures in Rheumatology (OMERACT) ultrasound scoring system for salivary gland assessment in patients with Sjögren's syndrome. Rheumatol Oxf Engl. 2021;60(5):2169–76.

17. Jousse-Joulin S, D'Agostino MA, Nicolas C, Naredo E, Ohrndorf S, Backhaus M, et al. Video clip assessment of a salivary gland ultrasound scoring system in Sjögren's syndrome using consensual definitions: an OMERACT ultrasound working group reliability exercise. Ann Rheum Dis juill. 2019;78(7):967–73.

18. Filippucci E, Meenagh G, Ciapetti A, Iagnocco A, Taggart A, Grassi W. E-learning in ultrasonography: a web-based approach. Ann Rheum Dis. 2007;66(7): 962–5.

19. Sekiguchi H, Bhagra A, Gajic O, Kashani KB. A general critical care ultrasonography workshop: results of a novel Web-based learning program combined with simulation-based hands-on training. J Crit Care. 2013;28(2):217.e7-12.

20. Platz E, Goldflam K, Mennicke M, Parisini E, Christ M, Hohenstein C. Comparison of web-versus classroom-based basic ultrasonographic and EFAST training in 2 European hospitals. Ann Emerg Med. 2010;56(6):660–7.

21. Jousse-Joulin S, Nowak E, Cornec D, Brown J, Carr A, Carotti M, et al. Salivary gland ultrasound abnormalities in primary Sjögren's syndrome: consensual US-SG core items definition and reliability. RMD Open. 2017;3(1): e000364.

22. Hočevar A, Bruyn GA, Terslev L, et al. Development of a new ultrasound scoring system to evaluate glandular inflammation in Sjögren's syndrome: an OMERACT reliability exercise. Rheumatology (Oxford). 2022;61(8):3341–50.