



HAL
open science

Diagnostic performances of ultrasound evaluation of major salivary glands according to the 2019 OMERACT US scoring system

François Robin, Jean-David Albert, Alain Lescoat, Amélie Martel, Aleth Perdriger, Michel Debandt, François Maillot, Guillaume Coiffier

► To cite this version:

François Robin, Jean-David Albert, Alain Lescoat, Amélie Martel, Aleth Perdriger, et al.. Diagnostic performances of ultrasound evaluation of major salivary glands according to the 2019 OMERACT US scoring system. *Arthritis Care and Research*, 2022, 74 (11), pp.1924-1932. 10.1002/acr.24631 . hal-03253476

HAL Id: hal-03253476

<https://hal.science/hal-03253476>

Submitted on 15 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DR. FRANÇOIS ROBIN (Orcid ID : 0000-0001-5134-8160)

DR. ALAIN LESCOAT (Orcid ID : 0000-0003-2081-8558)

DR. GUILLAUME COIFFIER (Orcid ID : 0000-0003-3560-128X)

Article type : Original Article

Original Article

Diagnostic performances of ultrasound evaluation of major salivary glands according to the 2019 OMERACT US scoring system.

François Robin, MD ¹, Jean-David Albert, MD ¹, Alain Lescoat, MD ^{2,3}, Amélie Martel, MD ⁴, Aleth Perdriger, MD/PhD ¹, Michel DeBandt, MD/PhD ⁵, François Maillot MD/PhD ⁴, Guillaume Coiffier, MD ^{2,6*}.

¹Rheumatology department, University Hospital Center of Rennes, France

² Internal Medicine and Clinical Immunology department, University Hospital Center of Rennes, France

³ Univ Rennes, CHU Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) – UMR_S 1085, F-35000 Rennes, France

⁴ Internal Medicine department, University Hospital Center of Tours, France

⁵ Rheumatology department, University Hospital Center of Fort-de-France, France

⁶ Rheumatology department, Hospital Center of Dinan, France

* Corresponding author: Dr Guillaume COIFFIER, MD, Rheumatology department, Hospital group of the Rance-émeraude territory, France. Address: Centre Hospitalier de Dinan, 74 boulevard Châteaubriand, BP 91056, 22101 DINAN Cedex, France. Phone : +33 2 96 85 79 23. E-mail : guillaume.coiffier@ch-dinan.fr, ORCID iD 0000-0003-3560-128X.

Manuscript: 2711 words. 22 References. 3 Tables. 3 Figures. +Supplementary Material

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/ACR.24631](https://doi.org/10.1002/ACR.24631)

Abstract: 249 words.

Abstract

Objective. To evaluate the diagnostic performance of ultrasound examination of the major salivary glands (US-SG) according to the 2019 OMERACT US scoring system for Sjogren's syndrome (SS).

Methods. This was a retrospective study based on a multicentric cohort of SS/sicca syndrome. ACR/EULAR 2016 (≥ 4 without OSS), AECG 2002 and clinician experts were considered as reference standards for SS diagnosis. An OMERACT score ≥ 2 according to 2 independent readers defined the diagnosis of SS based on US-SG assessment. Diagnostic performances and Inter-observer reproducibility of US-SG were assessed.

Results. Forty-two patients fulfilling the ACR/EULAR 2016 criteria for SS were compared to 30 controls with sicca syndrome. Twenty-five patients were diagnosed with SS according to US-SG evaluation, they were more frequently included in the SS group (52.5%) than in controls (10.0%) ($p < 0.001$). US-SG showed AUC of 0.751 (0.621-0.882) for the diagnosis of SS (ACR/EULAR 2016 classification). The inclusion of USSG in the ACR/EULAR 2016 classification improved Se (91.5% *vs* 89.4%) with limited decrease of Sp (96.0% *vs* 100%) and with an AUC of 0.975 (0.945-1.00). Similar results were observed when USSG was included in the AECG 2002 classification. Inter-observer reproducibility of the 2019 OMERACT US scoring ≥ 2 for SS diagnosis was good ($k = 0.73$ (0.64-0.81)). Histological lymphocyte infiltration of the minor salivary glands was associated with the OMERACT grading of US-SG.

Conclusion. This study confirms the good specificity of the 2019 OMERACT US classification of US-SG for the diagnosis of SS and its feasibility in daily practice.

Keywords: Ultrasound, Salivary Gland, OMERACT, Sjogren, sicca syndrome

Key messages

- Salivary gland ultrasound (US-SG) shows a good specificity for the diagnosis of Sjogren Syndrome (SS)

- The 2019 OMERACT US classification for US-SG shows a good inter-observer reproducibility
- Good correlation of the severity of gland impairment on US with histological and immunological findings.

1. Introduction

Sjogren's Syndrome (SS) is a frequent systemic autoimmune disease. Sicca syndrome is the clinical hallmark of this disorder and is often associated with more non-specific symptoms such as fatigue and musculoskeletal pain. SS is frequently associated with other organ-specific autoimmune diseases (i.e. autoimmune thyroiditis, primary biliary cirrhosis) or other connective tissue diseases (i.e. rheumatoid arthritis, systemic lupus, systemic sclerosis)[1].

SS diagnosis is based on a combination of parameters including objectivated sicca syndrome (Schirmer test, unstimulated salivary flow, Ocular Staining Score (OSS)), markers of autoimmunity (presence of anti-SSA antibodies often associated with a rheumatoid factor) and lymphocytic sialadenitis on biopsy of the accessory salivary glands [2]. Several SS classification systems have been proposed, including the most recent ACR/EULAR 2016 [3] and AECG 2002 criteria [4]. Some of the parameters included in these classification

criteria are sometimes difficult to perform in daily practice depending on the centers and their easy access to some specialist practitioners (i.e. OSS). Moreover, they are sometimes not performed despite their simplicity (i.e. unstimulated salivary flow). In addition, a recent study demonstrated that when ultrasound imaging of the major salivary glands (US-SG) was considered, OSS was not significantly associated with the diagnosis of primary SS (pSS) according to expert clinicians [6].

Recent studies have highlighted the relevance of including US-SG in existing SS classification criteria to improve their diagnostic performances [5-8]. However, different US-SG scoring systems exist with similar diagnostic performances [9] but with various impact regarding time consumption or practicability, affecting their use in daily practice, and with issue on their interobserver reproducibility [10]. OMERACT (Outcome Measures in Rheumatology) is an independent initiative of international health professionals and patient research partners interested in outcome measures and measurement methodology, particularly in the field of rheumatic diseases. Recently, OMERACT has proposed a consensus-based scoring system to harmonize and standardize the analysis of US-SG for the assessment of SS [11]. This 2019 OMERACT US scoring system includes a semi-quantitative assessment of parenchymal heterogeneity of major salivary glands with a scoring ranging from 0 (normal parenchyma) to 3 (pathological parenchyma without area of normal parenchyma). However, this work also highlights that in some cases, the salivary glands may be difficult to be distinguished from the adjacent soft tissues, raising challenges for the determination of this parenchymal score. In this specific situation, when semi-quantitative evaluation is impossible, an alternative two-item qualitative scoring system has therefore been proposed, defining a parenchymal pattern either normal (fatty gland, scored 1) or strictly pathological (fibrotic gland, scored 3). Based on this four-grade semi-quantitative scoring system or two-item qualitative scoring system, a score ≥ 2 defines a US-SG parenchyma compatible with a diagnosis of SS.

The objectives of our study were i) to evaluate the diagnostic performances of this 2019 OMERACT US scoring system within a population of patients with sicca syndrome, divided into SS according to the ACR/EULAR 2016 and/or AECG 2002 classification and controls with clinical sicca syndrome but who did not fulfilled these classification criteria, ii) to evaluate the new diagnostic performances of the ACR/EULAR 2016 and AECG 2002 for SS when 2019 OMERACT US criteria were included within these existing classifications.

2. Materials and Methods

We retrospectively reviewed the available ultrasound scans of 97 patients included in an existing French multicenter database. Twenty-five patients were excluded due to insufficient number of images or insufficient image quality for retrospective review. Image acquisition was performed by 3 different examiners (one examiner per center) with different devices (Philips, General Electrics and Esaote with a 5-18 MHz linear

probe) [9]. If US-SG images were not sufficient (no recorded image or less than 2 glands correctly evaluated) or of poor quality, patients were also excluded.

2.1. Ultrasound evaluation

Two evaluators (FR and GC, respectively 2 and 10 years of experience in the practice of US-SG) reviewed previously selected static images, blinded from patients' final diagnoses. These two evaluators received practical training in the 2019 OMERACT US scoring system during a training formation at French national referral center, with additional training on video clip as previously published [11]. A dedicated check-list was developed in accordance with the 2019 OMERACT US scoring system in order to optimize the evaluation. Each of the 4 glands was individually scored. US-SG compatible with the SS diagnosis was defined for a score of ≥ 2 for the most affected salivary gland. In case of discrepancy between the 2 examiners for final SS diagnosis based on US-SG, the images were reviewed by a third expert (JDA, 10 years of experience in the practice of US-SG) also blinded from patients' final diagnoses.

2.2. ACR/EULAR 2016 Classification

Clinical, biological and functional data were collected prospectively at the time of US-SG evaluation, by an independent investigator blinded from the results of US-SG (AM). The OSS was a frequently missing data in our population because it is difficult to obtain this specific examination from an ophthalmologist in daily care in French hospitals. The OSS score is a minor variable (1 point) in the ACR/EULAR 2016 classification. Therefore, the final diagnosis of SS was considered for a total score ≥ 4 points according to the ACR/EULAR 2016 criteria, even when OSS was missing. Patients with a total score of 2 or less according to ACR/EULAR 2016 classification, were considered as non-SS patients (controls) even when OSS was missing.

A specific attention was accorded to borderline patients, with a score of 3 according to ACR/EULAR 2016 classification, as ophthalmological assessment by the OSS could have changed the final conclusion regarding SS status.

2.3. Diagnosis of SS by expert clinicians

To evaluate the diagnostic performances of modified existing classification systems with the inclusion of USSG parameters according to 2019 OMERACT US scoring, the diagnosis of SS by a group of expert clinicians (FR, AL and GC) was considered as clinical reference standard. Medical records were reviewed retrospectively and blindly to the USSG results. The diagnosis was based on the demographic, clinical, immunological and histological data of each patient. Different levels of diagnostic probability were predefined: "certain", "probable", "unprobable", and "excluded" (Supplementary data Table S1). The diagnosis was clinically defined if at least 2 out of 3 considered experts concluded that the diagnosis of SS was "certain" or "probable" (Supplementary data Table S1).

2.4. Statistics

A *p-value* < 0.05 was considered statistically significant. The quantitative variables were expressed as mean and standard deviation in case of Gaussian distribution of the variables after verification by the Kolmogorov-

Smirnov test and their comparison was performed by the Student-t-test. In case of non-Gaussian distribution, the variables were expressed in median and inter-quartile distance and their comparison was performed by the Mann-Whitney non-parametric tests. The qualitative variables were expressed in absolute number (n) and percentage (%) and their comparison was carried out by the chi-2 test or the Fisher's exact test as appropriate.

The diagnostic performances of the 2019 OMERACT US scoring were presented through ROC curves with assessment of area under the curve (AUC) and sensitivity (Se), specificity (Sp), positive (LR+) and negative (LR-) likelihood ratio with their 95% confidence interval with 2019 OMERACT US scoring threshold ≥ 2 as the diagnostic threshold.

Interobserver reproducibility was presented by Cohen's kappa. The degree of agreement within the 2 examiners (corresponding to (true positive test + true negative test according to the 2 examiners)/total of observations) were evaluated a) for a 2019 OMERACT US score ≥ 2 as a dichotomic outcome, b) for each of the 4 final OMERACT semi-quantitative scores, and qualitative score (fatty or fibrotic gland), and c) for each scoring by saliva gland (parotid and sub-mandibular). To evaluate the level of agreement between the 2 examiners for the ordinal score of the 2019 OMERACT US as an assessment method for daily practice, we used the intra-class correlation coefficient (ICC) (two-way random). ICC can be interpreted as a low (<0.50), medium (0.50-0.75), good (0.75-0.90) or excellent (>0.90) level of agreement [12]. The statistical analyses were performed with the SPSS 20.0 software.

3. Results

3.1. Characteristics of the study population

Seventy-two patients with a clinical sicca syndrome and meeting our proofreading quality criteria of the US-SG were included. The characteristics of all patients finally included in this study are presented in Table 1. Forty-two patients met the ACR/EULAR 2016 criteria (≥ 4 without OSS) and 30 patients did not meet these criteria and were considered controls.

Patients with a diagnosis of SS had significantly lower Schirmer test and unstimulated salivary flow, and higher salivary gland histological focus score than controls. The positivity of ANAs and SSA antibodies was significantly higher in the SS patients' group than in controls (85.7% *vs* 40.0% and 66.7% *vs* 13.3% respectively) (Table 1).

Nine borderline patients (30.0%) classified in the control group had an ACR/EULAR 2016 classification score of 3 points (without OSS). As already mentioned, none of these 9 borderline patients met the AECG 2002 classification criteria for the diagnosis of SS (Figure 1) and were considered as controls. The characteristics of patients with SS *vs* controls according to the clinician expert group are available in table S2, supplementary appendix.

3.2. Distribution of the 72 patients included according to the different Sjogren's syndrome classification systems (ACR/EULAR 2016 without OSS, 2019 OMERACT US and clinician expert).

Twenty-five patients were diagnosed with SS according to US-SG evaluation (OMERACT scoring ≥ 2). They were more frequent in the SS group (52.5%) than in the control group based on the ACR/EULAR 2016 criteria (10.0%) (OR = 9.9 (2.60-37.7), $p < 0.001$). Twenty-two (88.0%) patients with abnormal US-SG (OMERACT scoring ≥ 2) met the ACR/EULAR 2016 and AECG 2002 classification criteria for the diagnosis of SS. The distribution of patients with a diagnosis of SS according to the classification used (ACR/EULAR 2016 classification without OSS, AECG 2002, clinician expert group or 2019 OMERACT US) is shown in Figure 1.

3.3. Diagnostic performance of the OMERACT US-SG according to different classification systems.

According to the 2019 OMERACT US scoring system, the diagnostic performances of US-SG for the diagnosis of SS was characterized by an AUC ROC of 0.751 (0.621-0.882), and 2019 OMERACT US score ≥ 2 had a Se of 52.4% (0.377-0.666), a Sp of 90.0% (0.744-0.965), a LR+ of 5.24 (1.72-15.9) and a LR- of 0.53 (0.35-0.84) for the diagnosis of SS, based on the ACR/EULAR 2016 classification (Figure 2). For the 9 borderline patients with a score of 3 in the ACR/EULAR 2016 classification (and who did not meet the AECG 2002 criteria), 2 patients had an 2019 OMERACT US scoring ≥ 2 . The clinical and biological characteristics of the 3 patients who did not meet the ACR/EULAR 2016 (without OSS) or AECG 2002 criteria with 2019 OMERACT US scoring ≥ 2 are available in *Supplementary Table S3*.

According to the diagnosis of SS selected by the expert clinician group, the performance of the USSG is characterized by a AUC ROC of 0.705 (0.585-0.825) with a Se of 48.9% and a Sp at 92.0% (Table 2).

3.4. Diagnostic performance of the ACR/EULAR 2016 and AECG 2002 classifications without and with inclusion of the US-SG according to the 2019 OMERACT US scoring system.

According to the diagnosis of SS selected by the expert clinician group, the performance of the ACR/EULAR 2016 classification (without OSS) is characterized by a AUC ROC of 0.979 (0.953-1.00), with for a score ≥ 4 a Se of 89.4% and a Sp at 100%. For AECG 2002 classification, AUC ROC was 0.926 (0.863-0.988), with for a score ≥ 4 a Se of 85.1% and a Sp at 100% (Figure 3 and Table 2).

After multivariate analysis including all variables of the ACR/EULAR 2016 classification (except OSS) and US-SG (2019 OMERACT US scoring system ≥ 2), only the histological criterion (focus score ≥ 1) was independently associated with the final diagnosis of SS (*Supplementary Table S4*). Therefore, we have included the US-SG parameter (2019 OMERACT US rating system ≥ 2) with a weighting of +1 point in addition to the other variables of the ACR/EULAR 2016 or AECG 2002 classification. Diagnosis of SS was retained for ACR/EULAR 2016 ≥ 4 or AECG 2002 $\geq 5/7$ including ultrasound analysis.

The inclusion of the USSG in the ACR/EULAR 2016 classification showed an increase in Se to 91.5% while Sp declined slightly to 96.0% with a AUC of 0.975 (0.945-1.00). Similar results were observed when the USSG was included in the AECG 2002 classification (Figure 3 and Table 2).

3.5. Intra-observer reproducibility of the 2019 OMERACT US scoring system

The inter-observer reproducibility of US-SG for SS diagnosis defined by an 2019 OMERACT US score ≥ 2 was good with a kappa of 0.73 (0.64-0.81) with a degree of agreement of 87.5% and a ICC of 0.769 (0.63-0.86).

Nine (12.5%) patients had a disagreement between reader 1 and reader 2 requiring further reading by a third reader (*Supplementary Table S5*). All these patients had SS in accordance with ACR/EULAR 2016 classifications (without OSS) and AECG 2002 and no 2019 OMERACT US diagnostic discordance was observed for the controls (all of them were classified as normal US-SG) (Table 3). In the end, reader 3 concluded to a US diagnosis of SS (2019 OMERACT US score ≥ 2) for 4 of the 9 patients (*Supplementary Table S6*).

When the exact 2019 OMERACT US score for each gland was considered separately, inter-observer reproducibility decreased significantly (kappa between 0.2 and 0.4) with a degree of agreement $< 50\%$, better for submandibular glands than for parotid glands. Inter-observer reproducibility decreased further when the fibrotic Gland and fatty gland patterns were considered (Table 3).

3.6. Correlation of 2019 OMERACT US ordinal score and other saliva or immunological parameters.

The level of minor salivary gland lymphocyte infiltration assessed by the focus score level was significantly associated with the stage of major salivary gland morphological abnormality according to 2019 OMERACT US (regardless of the reader) with a Kendall's Tau correlation level of 0.318 (*Supplementary Figure S1*). Anti-SSA antibodies, but not rheumatoid factor, were more frequently detected in patients with high 2019 OMERACT US score. The correlation between 2019 OMERACT US score and the severity of the objective sicca syndrome by the unstimulated salivary flow was not significant (*Supplementary Table S7*).

4. Discussion

Our study suggests that the US-SG according to the 2019 OMERACT US score is very specific for the diagnosis of SS but with disappointing sensitivity. Nonetheless, the inclusion of USSG in addition with the other variables of the ACR/EULAR 2016 and AECG 2002 classifications increased sensitivity with a lower specificity but which remained high and clinically relevant ($> 95\%$). The OMERACT score allows a consensus ultrasound diagnostic harmonization of major salivary glands applicable for daily practice for clinicians with a good reproducibility for a diagnostic threshold ≥ 2 .

In our study, when using ACR/EULAR 2016 criteria as the reference standard, 2019 OMERACT US score had a sensitivity of 52.4% and a specificity of 90.0%. Our results are partly consistent with the literature. A meta-analysis in 2018 [13] evaluating diagnostic accuracy of US for the diagnosis of SS found a pooled sensitivity of 75% and a pooled specificity of 93%. Nevertheless, these results may not be directly comparable, as the US scoring system using in these studies were different and they did not include the 2019 OMERACT US scoring system. In the DiaPSS cohort of 164 patients (80 SS and 84 controls), with clinical diagnosis used as the reference standard, 2019 OMERACT US scoring had a sensitivity of 65% and a specificity of 81% [6].

In our study, when using the diagnosis of clinical expert group as the reference standard, the addition of ultrasound parameter to the ACR/EULAR 2016 criteria allowed an improvement in sensitivity (91.5% *vs* 89.4%) but a decrease in specificity (96.0% *vs* 100%). This result is in accordance with other data from the literature: Van Nimwegen *et al.* [8] found that addition of US-SG parameter in ACR/EULAR 2016 criteria increased sensitivity (97.3 *vs* 95.9%) and decreased specificity (90.2 *vs* 92.2%) with clinical diagnosis used as the reference standard. Nonetheless, in this study, US scoring was limited to a unilateral evaluation (including parotid and submandibular glands) of hypoechogenic area. An optimal cut-off value of $\geq 1,5$ was considered as pathological. Unilateral evaluation has been already described as sufficient, without losing specificity or sensitivity [14]. More recently, Jousse-Joulin *et al.* [6] with a methodology close to the one used in our study, reported similar results (AUC was 0.96 (0.95-0.98) with for the score ≥ 4 , a Se of 95.6% (*vs* 90.2% without USSG) and a Sp of 82.6% (*vs* 84.2% without USSG). These concordant results highlights that the general use of the USSG, in combination with other commonly used tests (salivary or lacrimal function test, anti-SSA antibody, and biopsy of the labial minor salivary glands) should be considered for daily clinical practice.

The lack of sensitivity of US-SG evaluation for the diagnosis of SS does not seem to be explained by the duration of disease progression (74% of our SS patients had a sicca syndrome < 5 years), which is consistent with the results of previous studies [6,15-17]. Furthermore, US-SG appears to have a low sensitivity for change in the natural history of the disease [18]. We suggest that its disappointing sensitivity reflects the heterogeneity of the disease and that ultrasound characteristics of the major salivary glands should be considered as phenotypic markers. As already described in the literature [19], the severity of the parenchymal heterogeneity according to the 2019 US OMERACT scoring was significantly correlated with the positivity of anti-SSA antibodies, as well as with the levels of lymphocyte infiltration of the glandular parenchyma assessed by the

focus score. Similarly, another study has suggested a possible improvement of the US heterogeneity of parenchyma after Rituximab, which may indirectly confirm the important correlation between glandular involvement, immunological parameters and B lymphocyte hyperactivity in this disease [20]. As recently suggested, we can assume that beyond this issue of the diagnostic value of US-SG, the establishment of this 2019 OMERACT US salivary gland scoring could be valuable for the phenotypic classification of the patient with SS at risk of systemic complications resulting from B lymphocyte hyperactivation, such as the onset of cryoglobulinemic vasculitis or B lymphoma [21].

To be applicable and fully trustworthy, a good reproducibility for the definition of each stage of the 2019 OMERACT US classification should be warranted. Our results are consistent with literature [9,11] with a good reproductivity inter-observer (Kappa of 0.73), when focusing on SS diagnosis (OMERACT ≥ 2). Our results also suggest a poor reproducibility for some definitions. The poor reproducibility of the fatty salivary gland cannot be considered as a limitation because it is a variant of normality and it corresponds to a non-pathological 2019 OMERACT US score of 1. Whereas one of the strengths of this 2019 OMERACT US classification is to include the presence of an inhomogeneous parenchyma with numerous thick hyperechoic bands, representing a post-inflammatory fibrous transformation of the salivary gland (OMERACT score 3), our poor reproducibility for this entity could be due to a poor classification between grade 1 and grade 3. However, this poor reproducibility may be overestimated by the reading of static images, that makes interpretations more difficult than of dynamic examinations, as already described [22]. The spreading and generalization of this 2019 OMERACT US classification associated with a prior training of the practitioners and regular improvement of US equipment may improve these results in the future.

Our study has several potential limitations due to its retrospective nature with missing data (i.e. ESSDAI activity level or functional gene severity by ESSPRI). The absence of data regarding OSS, due to the French health system organization and the poor availability of ophthalmologists, did not allow us to properly evaluate the relevance of substituting OSS by US-SG for the diagnosis of SS, because US-SG made the diagnosis of SS according to the AECG criteria for only 2 of the 9 borderline patients with a 3-point ACR/EULAR 2016 score. However, OSS was not significantly associated with an expert pSS diagnosis when US-SG was considered and OSS diagnostic value in daily practice is challenging by the EULAR US-pSS Task Force (UTOPIA study) [6].

In conclusion, our study confirms the good specificity of the 2019 OMERACT US classification of the US-SG and its feasibility in daily practice for the diagnosis of SS. In the future, we hope that the systematic inclusion of US-SG in prospective cohorts of patients with SS could confirm the prognostic value of ultrasound assessment.

Acknowledgements

The authors gratefully acknowledge all patients and investigators involved in this study.

Funding: No specific funding was received from any funding bodies in the public, commercial or not-for-profit sectors to carry out the work described in this manuscript.

Disclosure statement: All other authors have declared no conflicts of interest.

References

1. Mariette X, Criswell LA. Primary Sjögren's Syndrome. *N Engl J Med* 2018;378:931-9.
2. Brito-Zerón P, Baldini C, Bootsma H, Bowman SJ, Jonsson R, Mariette X et al. Sjögren syndrome. *Nat Rev Dis Primer* 2016;2:16047.
3. 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;69:35-45.
4. Vitali C, Bombardieri C, Jonsson R, Moutsopoulos HM, Alexander EL, Carsons SE, et al. Classification criteria for Sjogren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *Ann Rheum Dis* 2002;61:554-8.
5. 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 Sjogren's syndrome. *Rheumatology* 2014;53:1604-7.
6. 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;287:180-8.
7. Takagi Y, Nakamura H, Sumi M, Shimizu T, Hirai Y, Horai Y, et al. Combined classification system based on ACR/EULAR and ultrasonographic scores for improving the diagnosis of Sjögren's syndrome. *PLoS One* 2018;13:e0195113.
8. van Nimwegen JF, Mossel E, Delli K, Delli K, van Ginkel MS, Stel AJ, et al. Incorporation of salivary gland ultrasonography into the ACR EULAR criteria for primary Sjögren's syndrome. *Arthritis Care Res* 2020;72:583-90.
9. 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? *Joint Bone Spine* 2019;86:211-7.
10. Jousse-Joulin S, Milic V, Jonsson MV, Plagou A, Theander E, Luciano N, et al. Is salivary gland ultrasonography a useful tool in Sjögren's syndrome? A systematic review. *Rheumatology* 2016;55:789-800.
11. Jousse-Joulin S, D'Agostino MA, Nicolas C, Naredo E4, Ohrndorf S5, 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* 2019;78:967-73.
12. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016;15:155-63.
13. Zhou M, Song S, Wu S, Duan T, Chen L, Ye J, et al. Diagnostic accuracy of salivary gland ultrasonography with different scoring systems in Sjögren's syndrome: a systematic review and meta-analysis. *Sci Rep* 2018;8:17128.

14. Mossel E, Arends S, van Nimwegen JF, Delli K, Stel AJ, Kroese FGM, et al. Scoring hypoechogenic areas in one parotid and one submandibular gland increases feasibility of ultrasound in primary Sjögren's syndrome. *Ann Rheum Dis* 2018;77:556-62.
15. Cornec D, Jousse-Joulin S, Pers J-O, Marhadour T, Cochener B, Boissramé-Gastrin S, et al. Contribution of salivary gland ultrasonography to the diagnosis of Sjögren's syndrome: Toward new diagnostic criteria? *Arthritis Rheum* 2013;65:216-25.
16. Baldini C, Luciano N, Tarantini G, Pascale R, Sernissi F, Mosca M, et al. Salivary gland ultrasonography: a highly specific tool for the early diagnosis of primary Sjögren's syndrome. *Arthritis Res Ther* 2015;17:146.
17. 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. *Rheumatology* 2015;54:2198-204.
18. Gazeau P, Cornec D, Jousse-Joulin S, Guellec D, Saraux A, Devauchelle-Pensec V. Time-course of ultrasound abnormalities of major salivary glands in suspected Sjögren's syndrome. *Joint Bone Spine* 2018;85:227-32.
19. Kim J-W, Lee H, Park S-H, Kim SK, Choe JY, Kim JK. Salivary gland ultrasonography findings are associated with clinical, histological, and serologic features of Sjögren's syndrome. *Scand J Rheumatol* 2018;47:303-10.
20. 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;67:1623-8.
21. Coiffier G, Martel A, Albert J-D, Lescoat A, Bleuzen A, Perdriger A, et al. Ultrasonographic damages of major salivary glands are associated with cryoglobulinemic vasculitis and lymphoma in primary Sjögren's syndrome: are the ultrasonographic features of the salivary glands new prognostic markers in Sjögren's syndrome? *Ann Rheum Dis* 2019 Aug 16. pii: annrheumdis-2019-216122.
22. 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:e000364.

Diagnostic performances of ultrasound evaluation of major salivary glands according to the 2019 OMERACT US scoring system.

Robin F ¹, Albert JD ¹, Lescoat A ², Martel A ³, Perdriger A ¹, DeBandt M ⁴, Maillot F ³, Coiffier G ^{2,6*}.

Table 1. Comparison of the characteristics of the 72 patients included with sicca syndrome. Patients with Sjogren's syndrome (SS) met the ACR/EULAR 2016 classification criteria (without OSS).

	SS ^a n = 42	Controls n = 30	<i>p</i>
Age (years) ^b	58.0 ± 13.8	56.9 ± 11.1	0.68
Female	39 (92.9)	29 (96.7)	0.64
Other systemic disease ^c	9 (21.4)	12 (40.0)	0.18
Sicca syndrom parameters			
Duration of sicca syndrom ≥ 5 years	11 (26.2)	12 (40.0)	0.22
Subjective ocular dryness	40 (95.2)	29 (96.7)	0.77
Schirmer test (mm/5 min) ^d	1.5 (0-5.0)	8.0 (2.5-15)	< 0.001
Pathological Schirmer test (< 5mm)	11/28 (39.3)	1/25 (4.0)	0.003
Subjective dry mouth	40 (95.2)	24 (80.0)	0.06
Unstimulated salivary flow (usSF) (ml/5min) ^b	1.18 ± 1.18	1.73 ± 1.08	0.09
Pathological usSF (< 0.1ml/min)	35/39 (89.7)	10/29 (34.5)	< 0.001
Biopsy of accessory salivary glands			
Focus score ^d	1.2 (1.0-2.0)	0.14 (0-0.72)	< 0.001
Focus pathological score (≥1/4mm ²)	37/40 (92.5)	4/29 (13.8)	< 0.001
Biological parameters			
Positive ANA (≥ 1/320)	36 (85.7)	12 (40.0)	< 0.001
Positive SSA antibodies	28 (66.7)	4 (13.3)	< 0.001
Positive Rheumatoid factors	18/35 (51.4)	11 (36.7)	0.23
Polyclonal hypergammaglobulinemia (≥ 16 g/L)	19 (45.2)	2 (6.7)	0.001
US-SG parameters			
US-SG SS (2019 OMERACT US scoring ≥ 2)	22 (52.4)	3 (10.0)	< 0.001
Global OMERACT scoring (reader1/reader2)			
- Grade 0	8 / 7	4 / 13	ns / <0.05
- Grade 1	14 / 12	22 / 13	<0.05 / ns
- Grade 2	11 / 5	3 / 1	ns / ns
- Grade 3	9 / 18	1 / 3	< 0.05/ < 0.05
Parotid Glands scoring (<i>right + left PGs</i>) (n = 144)			

- Grade 0	24 / 23	21 / 31	
- Grade 1	21 / 13	27 / 10	<i>ns</i> / <0.05
- Grade 2	19 / 28	2 / 1	<i>ns</i> / <i>ns</i>
- Grade 3	5 / 6	0 / 0	< 0.05/ < 0.05
- Fibrotic Gland	5 / 7	0 / 2	<i>ns</i> / <i>ns</i>
- Fatty Gland	10 / 11	10 / 16	<i>ns</i> / <i>ns</i>
Sub-mandibular Glands scoring (<i>right + left SMGs</i>) (<i>n</i> = 142)			<i>ns</i> / <i>ns</i>
- Grade 0	21 / 26	17 / 37	
- Grade 1	32 / 16	33 / 15	<i>ns</i> / <0.05
- Grade 2	18 / 8	5 / 3	<i>ns</i> / <i>ns</i>
- Grade 3	1 / 6	0 / 2	<i>ns</i> / <i>ns</i>
- Fibrotic Gland	10 / 25	1 / 1	<i>ns</i> / <i>ns</i>
- Fatty Gland	0 / 0	4 / 4	< 0.05/ < 0.05
2002 AECG classification (≥ 4 points)	42	0	<i>ns</i> / <i>ns</i>
Clinician expert group diagnosis of SS	42	5	-
			-

^a SS diagnosis according 2016 ACR classification without OSS

^b Quantitative variable of Gaussian distribution verified by the Kolmogorov-Smirnov Test, expressed as an mean \pm standard deviation

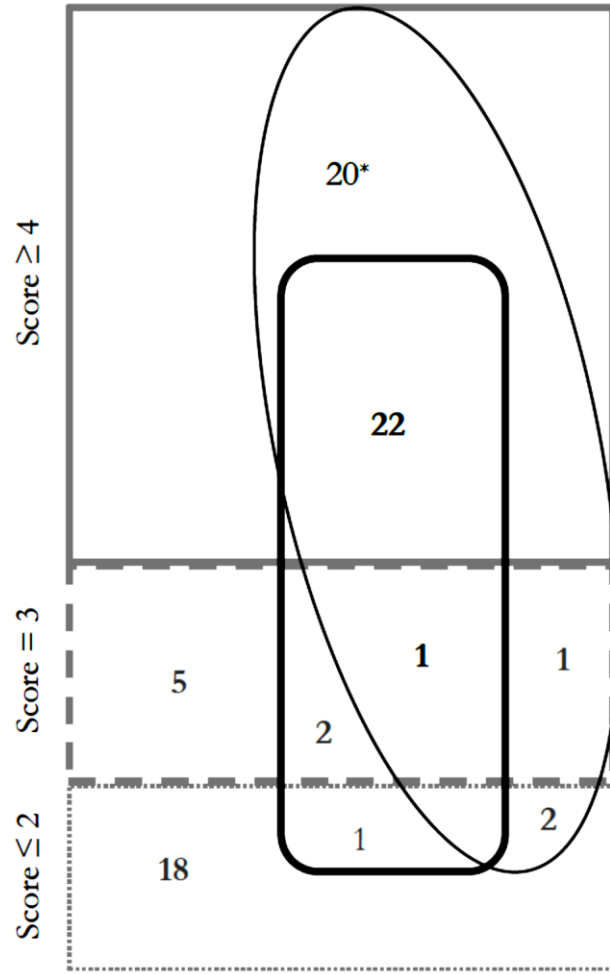
(SD)

^c SS group (4 RA, 4 UCTD, 2 LES, 1 SSc) and controls (8 RA, 3 LES, 1 UCTD)

^d Non-parametric distribution variable, expressed as median (IQR)

Figure 1. Distribution of the 72 patients included according to the different Sjogren's syndrome classification systems (ACR/EULAR 2016 without OSS, 2019 OMERACT US and clinician expert).

ACR/EULAR 2016 classification (without OSS)



The black box corresponds to the distribution of the 25 patients with 2019 OMERACT US ≥ 2

The black circle corresponds to the distribution of the 47 clinical diagnosis of SS.

* Of these patients, 18 met the 2002 AECG criteria.

Figure 2. Diagnostic performances of the 2019 OMERACT US classification of the major salivary glands for the SS using ACR/EULAR 2016 classification criteria as the reference standard.

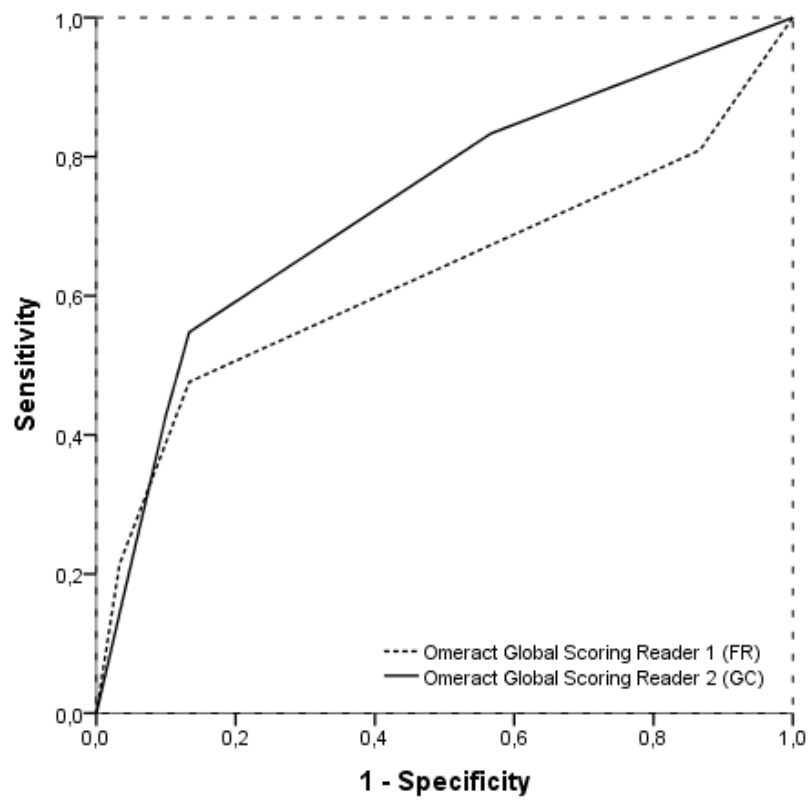


Figure 3. Comparison of the ROC curves of the diagnostic performance of the ACR/EULAR 2016 and AECG 2002 classification systems with or without inclusion of the USSG according to the 2019 OMERACT US scoring system (diagnosis of SS by the expert clinician group as a reference).

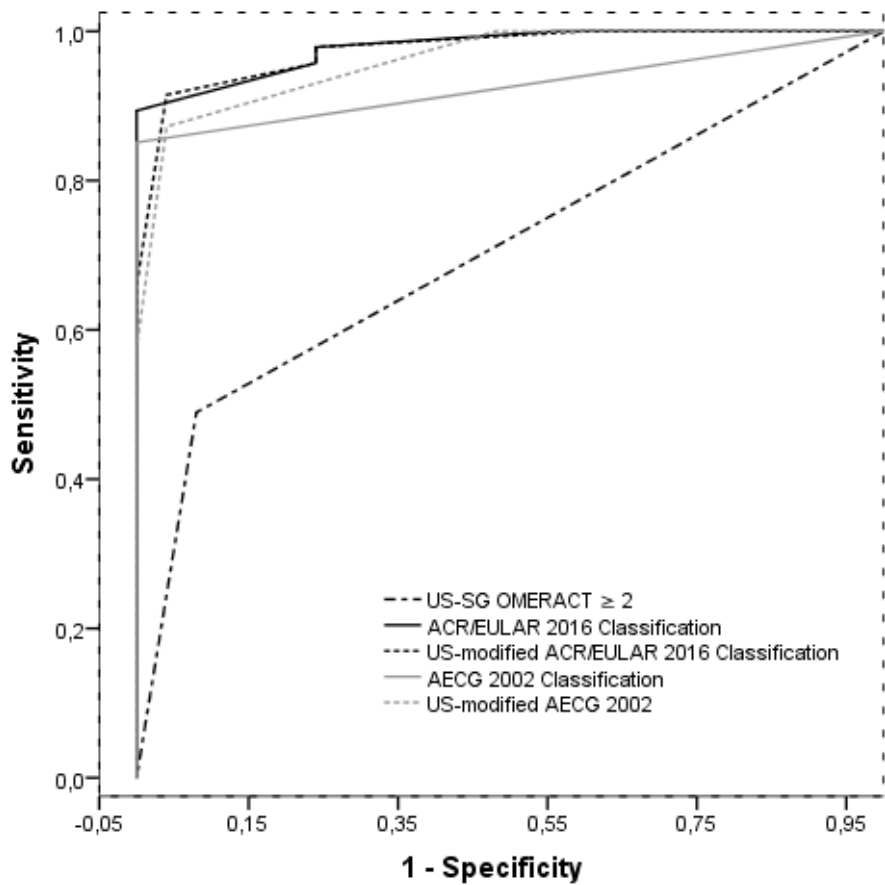


Table 2. Diagnostic performance for the diagnosis of SS (according to the clinician expert group as reference standard) of the 2019 OMERACT US scoring, the ACR/EULAR 2016 classification without OSS and AECG2002 including or not US-SG parameters.

	AUC	Cut-off	Se	Sp	LR+	LR-
2019 OMERACT US	0.705 (0.585-0.825)	≥ 2	48.9% (0.353-0.628)	92.0% (0.750-0.978)	6.1 (1.6-23.9)	0.55 (0.41-0.75)
ACR/EULAR 2016 (without OSS)	0.979 (0.953-1.00)	≥ 4	89.4% (0.774-0.954)	100% (0.867-1.0)	$+\infty$	0.11 (0.05-0.24)
		≥ 5	55.3% (0.413-0.686)	100% (0.867-1.0)	$+\infty$	0.45 (0.33-0.61)

US-modified ACR/EULAR 2016 (without OSS)	0.975 (0.945-1.00)	≥ 4	91.5% (0.801-0.967)	96.0% (0.805-0.993)	22.9 (3.4-156)	0.09 (0.04-0.23)
		≥ 5	66.0% (0.517-0.778)	100% (0.867-1.0)	+∞	0.34 (0.23-0.51)
AECG 2002	0.926 (0.863-0.988)	≥ 4	85.1% (0.723-0.926)	100% (0.867-1.0)	+∞	0.15 (0.08-0.30)
		≥ 5	38.3% (0.258-0.526)	100% (0.867-1.0)	+∞	0.62 (0.49-0.77)
US-modified AECG 2002	0.982 (0.922-0.999)	≥ 4	87.2% (0.748-0.940)	96.0% (0.805-0.993)	21.8 (3.2-149)	0.13 (0.06-0.28)
		≥ 5	57.5% (0.433-0.705)	100% (0.867-1.0)	+∞	0.43 (0.31-0.59)

Table 3. Inter-observer correlations of US variables included in the 2019 OMERACT classification

	Kappa (<i>reader1/ reader2</i>)	ICC (<i>reader1/ reader2</i>)	Agreement (<i>reader1/ reader2</i>)
US-SG Sjs (2019 OMERACT US scoring ≥ 2)	0.727 (0.643-0.811)	-	87.5 %
Global 2019 OMERACT US scoring	0.405 (0.329-0.481)	0.769 (0.63-0.86)	56.9 %
<i>Parotid Glands scoring</i>			
Right PG	0.279 (0.205-0.353)	0.392 (0.03-0.62)	44.4 %
Left PG	0.226 (0.152-0.300)	0.406 (0.05-0.63)	40.2 %
<i>Sub-mandibular Glands scoring</i>			
Right SMG	0.346 (0.275-0.417)	0.656 (0.45-0.79)	48.6 %
Left SMG	0.363 (0.297-0.429)	0.617 (0.39-0.76)	51.4 %
<i>Fatty Gland</i>			
PG (20 ^a /27 ^b)	0.167 (0.071-0.263)	0.237 (-0.22-0.52)	75.7 %
SMG (4 ^a / 4 ^b)	0.268 (0.041-0.495)	-0.039 (-0.66-0.35)	96.5 %
<i>Fibrotic Gland</i>			
PG (5 ^a / 9 ^b)	0.277 (0.107-0.447)	0.551 (0.28-0.72)	93.8 %
SMG (11 ^a /26 ^b)	0.287 (0.183-0.391)	0.600 (0.36-0.75)	83.3 %

^areader 1 (FR), ^breader 2 (GC), ICC: Intraclass Correlation Coefficient (two-way random).