Validation of the ACR-EULAR criteria for primary Sjogren's syndrome in a Dutch prospective diagnostic cohort

van Nimwegen, Jolien F; van Ginkel, Martha S; Arends, Suzanne; Haacke, Erlin A; Vegt, Bert van der; Sillevis Smitt-Kamminga, Nicole; Spijkervet, Fred K L; Kroese, Frans G M; Stel, Alja J; Brouwer, Elisabeth

Published in:
Rheumatology

DOI:
10.1093/rheumatology/kex495

Publication date:
2018

RHEUMATOLOGY

Rheumatology 2018;57:818–825
doi:10.1093/rheumatology/kex495
Advance Access publication 12 February 2018

Original article

Validation of the ACR-EULAR criteria for primary Sjögren's syndrome in a Dutch prospective diagnostic cohort

Jolien F. van Nimwegen[1,*], Martha S. van Ginkel[1,*], Suzanne Arends[1], Erlin A. Haacke[1,2], Bert van der Vegt[2], Nicole Sillevis Smitt-Kamminga[3], Fred K. L. Spijkervet[4], Frans G. M. Kroese[1], Alja J. Stel[1], Elisabeth Brouwer[1], Arjan Vissink[4] and Hendrika Bootsma[1]

Abstract

Objectives. To validate the ACR-EULAR classification criteria for primary SS (pSS), and compare them to the American–European Consensus Group (AECG) and ACR criteria in a Dutch prospective diagnostic cohort.

Methods. Consecutive patients ($n = 129$) referred for suspicion of pSS underwent a multidisciplinary evaluation, including a labial and/or parotid gland biopsy. Patients with an incomplete work-up ($n = 8$) or associated systemic auto-immune disease ($n = 7$) were excluded. The ACR-EULAR classification was compared with expert classification, AECG and ACR classification. Additionally, the accuracy of individual ACR-EULAR items in discriminating pSS from non-pSS was evaluated. The validity of criteria sets was described separately using parotid or labial gland biopsy results for classification.

Results. Of the 114 evaluated patients, the expert panel classified 34 (30%) as pSS and 80 (70%) as non-pSS. Using labial gland biopsy results, ACR-EULAR classification showed 87% absolute agreement ($\kappa = 0.73$) with expert classification, with a sensitivity of 97% and specificity of 83%. Using the parotid gland biopsy results, the ACR-EULAR criteria performed excellently as well. Focus score, anti-SSA titre and ocular staining score showed good to excellent accuracy, whereas unstimulated whole saliva and Schirmer's test had poor accuracy. The ACR-EULAR and AECG criteria had equal validity. Compared with ACR classification, ACR-EULAR classification showed higher sensitivity but lower specificity.

Conclusion. The ACR-EULAR criteria showed good agreement with expert classification, but some patients may be misclassified as pSS. Unstimulated whole saliva and Schirmer's test showed poor discriminative value. The ACR-EULAR criteria performed equally to the AECG criteria, and had higher sensitivity but lower specificity than the ACR criteria.

Key words: Sjögren's syndrome, ACR-EULAR criteria, AECG criteria, ACR criteria, classification, validation, expert consensus

CLINICAL SCIENCE

Rheumatology key messages

- Independent of the type of biopsy, the ACR-EULAR criteria for primary SS show excellent diagnostic accuracy.
- Schirmer's test and unstimulated whole saliva show limited validity to discriminate between primary SS and non-pSS.
- The ACR-EULAR criteria allow for international consensus regarding the classification of primary SS.

[1]Departments of Rheumatology and Clinical Immunology, [2]Pathology and Medical Biology, [3]Ophtalmology and [4]Oral and Maxillofacial Surgery, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

Submitted 23 June 2017; revised version accepted 17 November 2017

*Jolien F. van Nimwegen and Martha S. van Ginkel equally contributed to this study.

Correspondence to: Jolien F. van Nimwegen, Dep. Rheumatology and Clinical Immunology, University Medical Centre Groningen, PO box 30.001, 9700 RB Groningen, The Netherlands.
E-mail: J.f.van.nimwegen@umcg.nl

Introduction

Primary SS (pSS) is a systemic autoimmune disease, characterized by lymphocytic infiltration of the exocrine glands, resulting in dryness symptoms [1]. Patients present with a spectrum of signs and symptoms, evolving over time, making clinical diagnosis and classification challenging.

Currently, multiple criteria sets are in use for classification of pSS (supplementary Table S1, available at *Rheumatology* online). Most researchers and clinicians utilize the 2002 American–European Consensus Group (AECG) criteria, which include items evaluating the presence of sicca symptoms of the eye and mouth, functional impairment of the exocrine glands, presence of anti-SSA/SSB antibodies and a focus score of $\geqslant 1$ in the salivary gland biopsy [2]. However, questions were raised about the inclusion of sicca symptoms in the AECG criteria. Therefore, in 2012, Shiboski *et al.* proposed the ACR criteria for pSS. The ACR criteria include only objective tests and were designed to be used as entry criteria for clinical trials, in order to ease comparison of results between trials [3]. The ACR criteria require the presence of two out of the following three items: focus score of $\geqslant 1$, positive serology and ocular staining score (OSS) $\geqslant 3$. Positive serology was defined as the presence of anti-SSA/SSB antibodies or RF and ANA. Agreement between the AECG and ACR criteria was 78 and 81% in two prospective diagnostic cohorts [4, 5].

Although widely used, the AECG and ACR criteria sets have not been endorsed by both the ACR and the EULAR. To be able to compare different study populations in trials and cohorts, international consensus regarding the classification of pSS is crucial. Therefore, the International Sjögren's Syndrome Criteria Working Group developed the 2016 ACR-EULAR criteria for pSS using methodology endorsed by both the ACR and EULAR [6, 7].

The ACR-EULAR criteria combine features of the AECG and ACR criteria (supplementary Table S1, available at *Rheumatology* online). Instead of including sicca symptoms as an item, the ACR-EULAR criteria added the presence of sicca symptoms or a EULAR Sjögren's syndrome disease activity index (ESSDAI) of $\geqslant 1$ as an entry criterion. In the ACR-EULAR criteria, positive serology is solely based on the presence of anti-SSA antibodies, while anti-SSB, ANA and RF positivity were not adopted. The OSS score was added to the ACR-EULAR criteria with a cut-off of $\geqslant 5$, instead of $\geqslant 3$ as used for the ACR criteria, and the van Bijsterveld score with a cut-off of $\geqslant 4$ was allowed as an alternative. Sialography and scintigraphy were not included in the ACR-EULAR criteria and some updates were made in the exclusion criteria for classification as pSS.

Before the ACR-EULAR classification criteria can be implemented reliably, it is important to validate these criteria in external, prospective cohorts with complete data on all ACR-EULAR items. Recently, the ACR-EULAR criteria were validated in a cohort of Japanese patients [8]. However, this study had several limitations. The analysis was performed in a retrospective cohort with incomplete data. In some of the patients, unstimulated whole saliva (UWS) was replaced by tests assessing stimulated whole saliva. OSS was not available, and replaced by the van Bijsterveld score, making the comparison with the ACR criteria less reliable. Moreover, clinical diagnosis was used as the gold standard instead of expert classification based on anonymized case vignettes. Considering these limitations, and taking into account that the Japanese population may not show the same characteristics as Caucasian populations, further validation of the ACR-EULAR criteria is needed.

The primary objective of our study is therefore to validate the ACR-EULAR criteria for pSS using classification according to expert opinion as the gold standard, in a Dutch prospective diagnostic cohort in a daily clinical practice setting. In addition, the performance of the individual components of the ACR-EULAR criteria was assessed, and the ACR-EULAR criteria were compared with the AECG and ACR criteria.

Methods

Study population

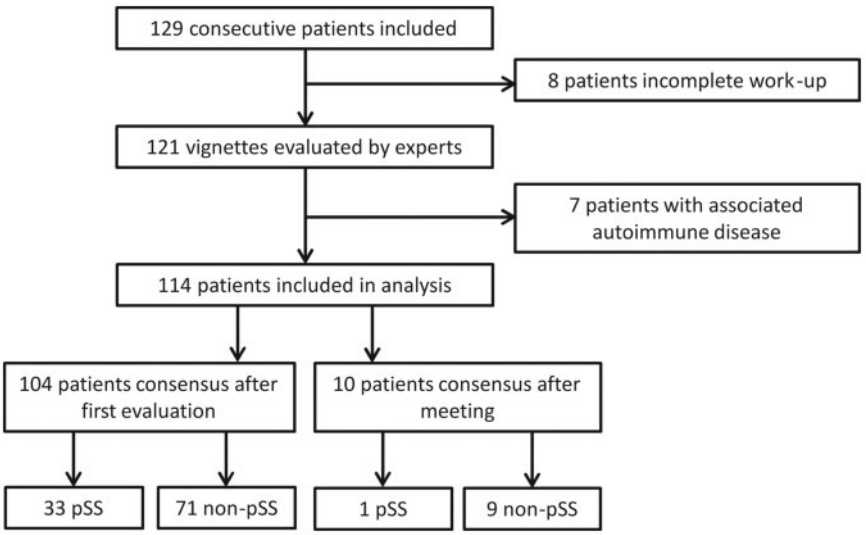
The study population consisted of consecutive patients, aged $\geqslant 18$ years, who were referred to the Sjögren Expertise Centre of the University Medical Centre Groningen, a tertiary referral centre, for suspicion of pSS between December 2013 and August 2016. Informed consent was obtained from all patients according to the Declaration of Helsinki. Patients with incomplete diagnostic work-up making it impossible to apply the AECG, ACR and ACR-EULAR criteria were excluded, as well as patients who were diagnosed with an associated systemic auto-immune disease (e.g. RA, SLE), as determined by the expert panel. The study was approved by the Medical Research Ethics Committee of the University Medical Centre Groningen (METc2013.066).

Diagnostic evaluation

Patients were evaluated by a team of clinical experts, consisting of rheumatologists, oral and maxillofacial surgeons, pathologists and one ophthalmologist, all highly experienced in diagnosing pSS. The multidisciplinary work-up included evaluation of all items of the three criteria sets [2, 3, 6, 7]. The rheumatologist performed a clinical history and physical examination, and recorded the presence of signs and symptoms of pSS and the ESSDAI score [9]. Laboratory tests included evaluation of complete blood count, ESR, CRP, ANA, anti-SSA and anti-SSB antibodies, RF, IgG, complement C3 and C4, cryoglobulinaemia and hepatitis C serology. When indicated, additional examinations such as X-rays, pulmonary function tests, thoracic high resolution CT or nailfold capillaroscopy were performed to facilitate clinical diagnosis.

Evaluation by the oral and maxillofacial surgeon included determination of sicca symptoms, a physical evaluation of the oro-facial and neck area and analysis of UWS and stimulated whole saliva. A labial and/or parotid gland biopsy was taken by the same oral and

Fig. 1 Flowchart of inclusion and expert panel evaluation



pSS: primary SS.

maxillofacial surgeon [10]. Salivary gland sialography or scintigraphy was not performed. Salivary gland biopsies were evaluated by a head and neck pathologist and trained resident for focus score (foci/4 mm^2) [11], presence of germinal centres, lymphoepithelial lesions and IgA, IgG and IgM plasma cell ratio.

Ophthalmological evaluation included determination of sicca symptoms, Schirmer's test, tear break-up time and OSS. OSS was defined using slit-lamp evaluation of lissamine green staining of the temporal and medial conjunctiva and fluorescein staining of the cornea [12].

Case ascertainment

All patients were classified as pSS or non-pSS according to the ACR-EULAR, AECG and ACR criteria [2, 3, 6, 7]. Fulfilment of the classification criteria was determined separately using the labial or parotid gland biopsy outcome for classification. Patients who did not undergo both biopsies, making it impossible to determine classification when either the labial or parotid gland biopsy results were taken into account, were excluded from that part of the analysis. Although the AECG criteria exclude patients with lymphoma, we classified patients with mucosa-associated lymphoid tissue (MALT) lymphoma who fulfilled the AECG criteria as pSS, as pSS can result in the development of MALT lymphoma [13].

The clinical diagnosis made by the treating rheumatologist was recorded. For expert classification, all cases were described in an anonymized clinical vignette, including the outcomes of all tests described above, which were reviewed by an expert panel (H.B., A.J.S., E.B.) and scored as pSS or non-primary SS (non-pSS). H.B. reviewed all vignettes, while A.J.S. and E.B. each reviewed half of the vignettes. The experts were blinded to the clinical diagnosis and classification by the other experts. In cases of discordance between the classifications by the experts, the vignette was discussed in a consensus meeting with all three experts to reach expert classification.

Statistical analysis

Statistical analyses were executed using SPSS Statistics 23 (IBM Corp., Armonk, NY, USA). Descriptive sociodemographic and disease characteristics were described as mean (S.D.), median (interquartile range) or number (%) as appropriate. The agreement between the clinical diagnosis and expert classification and between the three criteria sets was evaluated with percentage of absolute agreement and Cohen's κ coefficient. The performance of the ACR-EULAR score and individual ACR-EULAR items to predict expert classification was evaluated with the area under the ROC curve (AUC), which was interpreted as no discrimination (0–0.5), poor (0.5–0.7), fair (0.7–0.8), good (0.8–0.9) or excellent (0.9–1.0) accuracy [14]. The agreement of the three criteria sets with expert classification was evaluated with the percentage of absolute agreement, Cohen's κ coefficient, sensitivity and specificity. κ was interpreted as poor (0.0–0.2), fair (0.2–0.4), moderate (0.4–0.6), good (0.6–0.8) or excellent (0.8–1.0) agreement [15].

Results

Of the 129 consecutive patients who gave informed consent, 15 were excluded from evaluation in this study because of incomplete data or associated auto-immune diseases (Fig. 1). All remaining patients ($n = 114$) underwent a salivary gland biopsy. For most patients ($n = 100$), biopsies of both glands were obtained, but five patients underwent only a labial gland biopsy and nine patients underwent only a parotid gland biopsy.

Table 1 Comparison of ACR-EULAR, AECG and ACR classification with expert classification

Criteria	Expert classification	
	SS	Non-pSS
Criteria including labial gland biopsy		
ACR-EULAR[a]		
n	34	76
SS, *n* = 46	33	**13**
Non-pSS, *n* = 64	**1**	63
AECG[a]		
n	34	76
SS, *n* = 46	33	**13**
Non-pSS, *n* = 64	**1**	63
ACR[a]		
n	33	77
SS, *n* = 37	30	**7**
Non-pSS, *n* = 73	**3**	70
Criteria including parotid gland biopsy		
ACR-EULAR[b]		
n	34	78
SS, *n* = 37	31	**6**
Non-pSS, *n* = 75	**3**	72
AECG[b]		
n	34	78
SS, *n* = 37	31	**6**
Non-pSS, *n* = 75	**3**	72
ACR[c]		
n	34	79
SS, *n* = 32	29	**3**
Non-pSS, *n* = 81	**5**	76

Discrepant cases are bold. [a]Due to missing or inconclusive labial gland biopsies, four patients were excluded from the comparison of ACR-EULAR, AECG and ACR classification *vs* expert classification. [b]Due to missing or inconclusive parotid gland biopsies, two patients were excluded from the comparison of ACR-EULAR and AECG classification *vs* expert classification. [c]Due to missing or inconclusive parotid gland biopsies, one patient was excluded from the comparison of ACR classification *vs* expert classification. AECG: American–European Consensus Group.

Expert classification

After the first evaluation of the case vignettes, the expert panel agreed on the classification as pSS or non-pSS in 104 patients. For the remaining 10 patients, expert classification was reached during the consensus meeting. Of the 34 patients classified as pSS by the expert panel, the mean age was 52.3 (S.D. 15.3) years and 32 (94%) patients were female. Of the 80 patients classified as non-pSS, the mean age was 50.2 (S.D. 12.6) years and 69 (86%) patients were female.

The expert classification showed 89% agreement with the clinical diagnosis made by the treating rheumatologist ($\kappa = 0.77$). Eleven patients were clinically diagnosed with pSS by the treating physician, but classified as non-pSS by the experts, and one patient was clinically not diagnosed with pSS, but classified as pSS by the experts.

Comparison of criteria with expert classification

Taking the labial gland biopsies into account for classification, the ACR-EULAR score had an AUC of 0.94 (95% CI: 0.88, 1.00) to discriminate pSS from non-pSS. The ACR-EULAR criteria and AECG criteria both showed an absolute agreement of 87% ($\kappa = 0.73$) with expert classification, with 97% sensitivity and 83% specificity (Table 1). The ACR criteria showed an absolute agreement of 91% ($\kappa = 0.79$) with expert classification, with 91% sensitivity and 91% specificity.

Taking the parotid gland biopsies into account for classification, the ACR-EULAR score had an AUC of 0.97 (95% CI: 0.92, 1.00) to discriminate pSS from non-pSS. The ACR-EULAR criteria and AECG criteria both showed an absolute agreement of 92% ($\kappa = 0.82$) with expert classification, with 91% sensitivity and 92% specificity (Table 1). The ACR criteria showed an absolute agreement of 93% ($\kappa = 0.83$) with expert classification, with 85% sensitivity and 96% specificity.

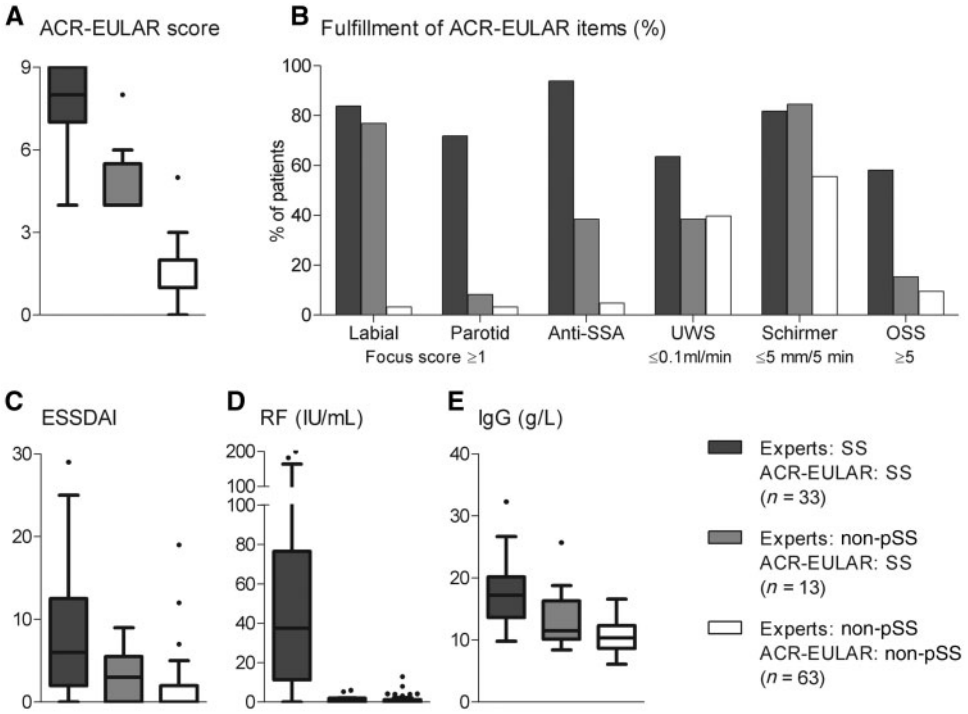
Description of patients with discrepant ACR-EULAR and expert classification

Characteristics of patients with concordance or discrepancy between the expert and ACR-EULAR classification are shown in Fig. 2 and Fig. 3, taking into account the labial or parotid gland biopsy for classification, respectively. Patients who were classified as non-pSS by the experts but pSS by the ACR-EULAR criteria showed low biological activity, and most of them had ACR-EULAR scores between 4 and 6. Interestingly, the Schirmer's test was often positive, while the OSS was mostly negative in this group of patients. Patients who were classified as pSS by the experts but non-pSS by the ACR-EULAR criteria, when taking into account the labial ($n = 1$) or parotid gland biopsy ($n = 3$) for classification, were not included in the figures. However, a detailed list of discrepant cases is provided in supplementary Table S2, available at *Rheumatology* online. Of these 17 discrepant cases, 9 were also classified differently by the two experts during the first round of evaluation. For these patients, expert classification was reached during the consensus meeting.

Performance of individual ACR-EULAR items

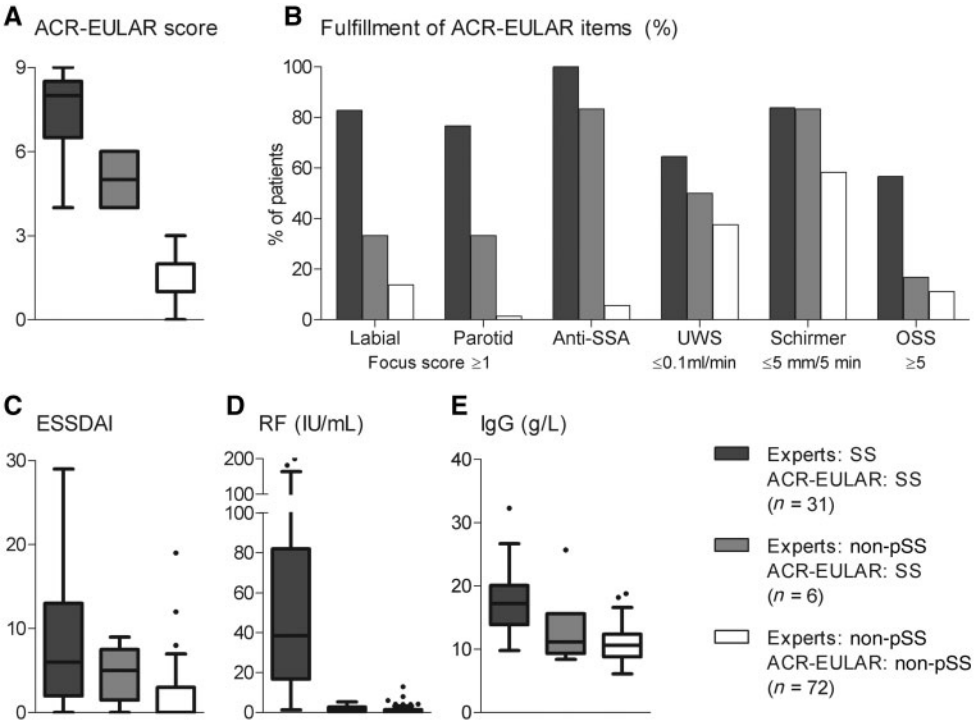
In this prospective cohort, 33 (97%) pSS patients and 78 (98%) non-pSS patients reported sicca symptoms. The ESSDAI was $\geqslant 1$ in 31 (91%) pSS patients and 40 (51%) non-pSS patients. Only one non-pSS patient did not fulfill the entry criteria of the ACR-EULAR criteria, as she had neither sicca complaints nor an ESSDAI $\geqslant 1$. None of the patients was solely SSB positive. Focus score and anti-SSA titre showed excellent accuracy and OSS showed good accuracy to discriminate pSS from non-pSS. UWS and Schirmer's test showed poor accuracy to discriminate pSS from non-pSS (Fig. 4).

Fig. 2 Characteristics of groups including the labial gland biopsy results



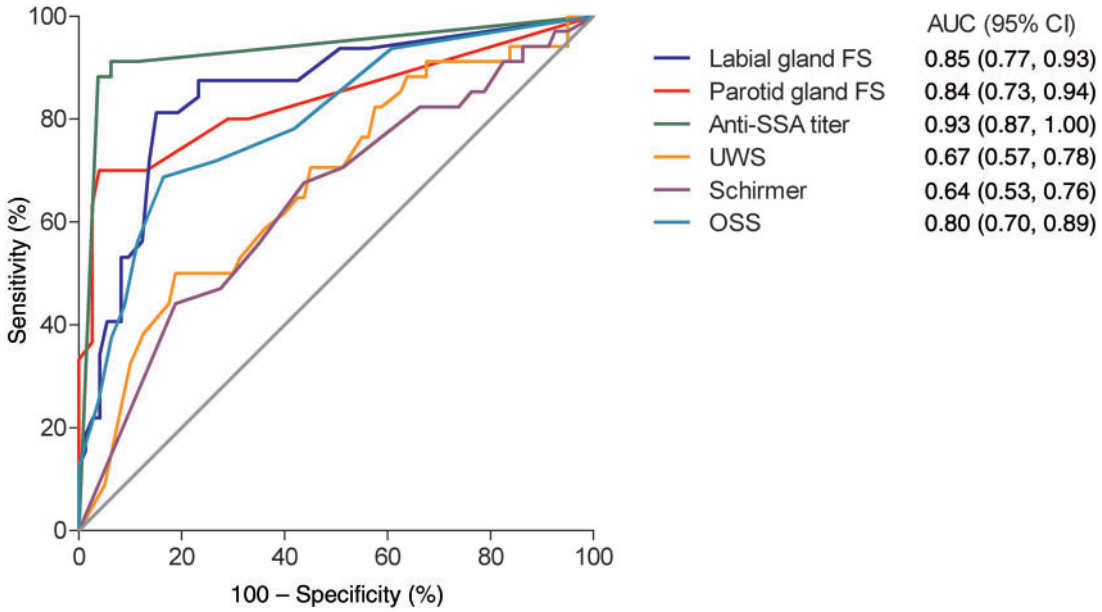
Comparison of patients who are classified as SS or non-pSS by the experts and ACR-EULAR criteria including the labial gland biopsy results. OSS: ocular staining score; UWS: unstimulated whole saliva.

Fig. 3 Characteristics of groups including the parotid gland biopsy results



Comparison of patients who are classified as SS or non-pSS by the experts and ACR-EULAR criteria including the parotid gland biopsy results. OSS: ocular staining score; UWS: unstimulated whole saliva.

FIG. 4 ROC curves of diagnostic tests included in the ACR-EULAR criteria, using expert classification as gold standard



FS: focus score (foci/4 mm^2); OSS: ocular staining score; ROC: receiver operating characteristic; UWS: unstimulated whole saliva.

TABLE 2 Comparison of ACR-EULAR with AECG and ACR classification

AECG and ACR	ACR-EULAR SS	ACR-EULAR Non-pSS
Criteria including labial gland biopsy		
AECG[a]		
n	46	64
SS, *n* = 46	45	**1**
Non-pSS, *n* = 64	**1**	63
ACR[b]		
n	45	64
SS, *n* = 37	36	**1**
Non-pSS, *n* = 72	**9**	63
Criteria including parotid gland biopsy		
AECG[c]		
n	37	75
SS, *n* = 37	36	**1**
Non-pSS, *n* = 75	**1**	74
ACR[d]		
n	37	74
SS, *n* = 32	32	0
Non-pSS, *n* = 79	**5**	74

Discrepant cases are bold. [a]Due to missing or inconclusive labial gland biopsies, four patients were excluded from the comparison of ACR-EULAR *vs* AECG classification. [b]Due to missing or inconclusive labial gland biopsies, five patients were excluded from the comparison of ACR-EULAR *vs* ACR classification. [c]Due to missing or inconclusive parotid gland biopsies, two patients were excluded from the comparison of ACR-EULAR *vs* AECG classification. [d]Due to missing or inconclusive parotid gland biopsies, three patients were excluded from the comparison of ACR-EULAR *vs* ACR classification.

Comparison of ACR-EULAR with AECG and ACR classification

Taking the labial gland biopsies into account, the ACR-EULAR criteria showed an absolute agreement of 98% ($\kappa = 0.96$) with the AECG criteria and 91% ($\kappa = 0.81$) with the ACR criteria (Table 2). Taking the parotid gland biopsies into account, the ACR-EULAR criteria showed an absolute agreement of 98% ($\kappa = 0.96$) with the AECG criteria and 95% ($\kappa = 0.90$) with the ACR criteria.

While ACR-EULAR classification was very similar to AECG classification, ACR classification was stricter, as some patients were classified as pSS by the ACR-EULAR criteria but as non-pSS by the ACR criteria. These patients had either a positive biopsy or positive serology, in combination with a positive UWS and/or Schirmer's test, but a negative OSS.

Discussion

This study evaluated the validity of the 2016 ACR-EULAR criteria for pSS, in comparison with the AECG and ACR criteria, in an external, prospective diagnostic cohort in a daily clinical practice setting. All ACR-EULAR items were evaluated, including labial and/or parotid gland biopsies in all patients. In our multidisciplinary setting, the ACR-EULAR score showed excellent accuracy with expert classification as the gold standard.

In accordance with the original validation cohort [6], the ACR-EULAR criteria showed very high sensitivity when labial gland biopsies were used. We found a specificity of 83%, which is lower than the specificity of 95% reported by Shiboski *et al.* [7]. Recently, an even lower specificity of 76.7% was found in a retrospective cohort of Japanese patients [8]. Taken together, these results

suggest that some non-pSS sicca patients may be misclassified as pSS by the ACR-EULAR criteria. This occurs mostly in patients who have an ACR-EULAR score of 4–6, based on either presence of SSA antibodies or focus score $\geqslant 1$, combined with a decreased Schirmer's test and/or UWS. In approximately half of the patients with discrepancy between the expert and ACR-EULAR classification, the experts also disagreed on the classification after the first round of evaluation of the vignettes. This illustrates that a subset of patients suspected for pSS is difficult to diagnose. Using the cut-off of $\geqslant 4$ for the ACR-EULAR score does ensure high sensitivity of the ACR-EULAR criteria, but for the clinical diagnosis, other clinical parameters have to be taken into account too, including more detailed histopathological characteristics (i.e. presence of germinal centres, lymphoepithelial lesions and plasma cell shift), the presence of comorbidities that may also partly explain the symptoms (i.e. presence of diabetes, autoimmune thyroiditis, FM) and the use of medication that may cause sicca symptoms (i.e. beta blockers, antidepressants).

In our cohort, in most patients labial and parotid gland biopsies were taken simultaneously, which gave us the unique opportunity to evaluate the performance of the ACR-EULAR criteria when including labial as well as parotid gland biopsies. We found that the ACR-EULAR criteria also have excellent accuracy when using parotid gland biopsies, with good sensitivity and specificity. Interestingly, the sensitivity of the ACR-EULAR criteria is higher when using labial gland biopsies, while the specificity is higher when using parotid gland biopsies. A detailed comparison between the labial and parotid gland biopsy from a histopathological point of view falls beyond the scope of this article and will be discussed separately (Haacke EA, manuscript in preparation).

In the analysis of the performance of individual ACR-EULAR items, the salivary gland focus score, anti-SSA and OSS showed good or excellent discriminative value. The accuracy of Schirmer's test and UWS was poor as they were positive in many non-pSS patients as well. In line with our findings, Shiboski *et al.* [3] reported limited validity of these tests in the SICCA cohort, using a latent class model. In contrast, Vitali *et al.* [16] did find acceptable validity of Schirmer's test and UWS, but the study population was different. Vitali *et al.* included selected patients, pre-defined as patients with pSS, secondary SS or controls based on clinical judgment, whereas our cohort and the SICCA cohort included consecutive patients, resulting in a population representative of daily clinical practice.

The poor performance of Schirmer's test and UWS in our cohort might be explained by non-pSS patients with exocrine gland dysfunction due to other causes, as Schirmer's test and UWS are not able to discriminate between different causes of sicca symptoms [17, 18]. The OSS shows good performance in our cohort, and we strongly recommend including evaluation of the OSS in the diagnostic work-up of SS. However, the OSS needs to be performed by a trained ophthalmologist, which is not always available. Therefore, the inclusion of Schirmer's test and UWS in the ACR-EULAR criteria has increased the feasibility of the criteria. To further improve the ACR-EULAR criteria, additional studies should evaluate whether other diagnostic tests such as salivary gland ultrasonography could complement the ACR-EULAR criteria [19].

As expected, the ACR-EULAR classification was very similar to the AECG classification, and showed equal validity in our cohort. However, the ACR-EULAR criteria have several advantages over the AECG criteria in current daily practice. For example, the sensitivity of the AECG criteria would have been lower if the three pSS patients with MALT lymphoma in our cohort had been characterized as non-pSS, according to the exclusion criteria (data not shown). Lymphoma is no longer included in the exclusion criteria of the ACR-EULAR, and other exclusion criteria have also been adjusted. Additionally, sialography and scintigraphy have been excluded from the ACR-EULAR criteria as they are no longer commonly used for the evaluation of pSS. Sialography is a painful, time-consuming procedure and is contraindicated in patients with severe salivary gland dysfunction. Scintigraphy exposes patients to radiation, has limited specificity and is not widely available [20].

Compared with the ACR criteria, the ACR-EULAR criteria show slightly lower absolute agreement with expert consensus and lower specificity. On the other hand, the ACR-EULAR criteria show higher sensitivity, similar to recent findings in Japanese patients [8]. Furthermore, the ACR-EULAR criteria are more feasible than the ACR criteria in daily clinical practice, as it is often not necessary to perform a salivary gland biopsy or ocular staining score to reach the cut-off of $\geqslant 4$ for classification as pSS. To avoid inclusion of patients who are misclassified as pSS in therapeutic trials, we do recommend performing a full diagnostic work-up [19].

An important strength of this study is the use of expert classification as gold standard. The AECG criteria are commonly used in our hospital, as shown by an agreement of 94% between the AECG criteria and the clinical diagnosis of the treating physician (data not shown). As the ACR-EULAR and AECG criteria show high agreement, the validity of the ACR-EULAR classification would be overestimated when using clinical diagnosis by the treating physician as gold standard. Our expert panel consisted of three rheumatologists with broad experience in diagnosing pSS patients. Agreement between the treating physician and the expert panel was high, but the experts were stricter than the treating physician. A possible limitation is that the expert panel consisted only of physicians working in our expertise centre. For some of the cases, one of the evaluating experts was therefore also the treating physician of the patient. We cannot exclude the possibility that despite anonymization, some cases may have been recognized by the experts, but the influence of this potential source of bias is limited as all cases were evaluated by at least two experts. We did not include sialography and scintigraphy in our diagnostic work-up, which might have influenced our results regarding the AECG classification. However, as sialography and scintigraphy are not commonly performed any more to diagnose pSS,

we believe our results are representative of how the AECG criteria are most often applied.

In conclusion, the ACR-EULAR criteria showed excellent diagnostic accuracy in our prospective cohort. The ACR-EULAR criteria also have excellent accuracy when using parotid gland biopsies, with good sensitivity and specificity. The validity of Schirmer's test and UWS, as well as addition of new items should be further evaluated. Based on our results, we strongly recommend performing OSS to evaluate ocular signs of pSS. The ACR-EULAR criteria showed validity equal to the AECG criteria, and compared with the ACR criteria, high sensitivity but lower specificity. The ACR-EULAR criteria have important advantages compared with other criteria sets, and have been endorsed by both the ACR and EULAR, allowing for international consensus regarding the classification of pSS.

Acknowledgements

The authors thank J. Bulthuis-Kuiper, J. Baldi-Ekelhoff and Dr K. Delli for their contribution to patient inclusion and data collection.

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

Disclosure statement: The authors have declared no conflicts of interest.

Supplementary data

Supplementary data are available at *Rheumatology* online.

References

1 Brito-Zerón P, Baldini C, Bootsma H *et al*. Sjögren syndrome. Nat Rev Dis Primers 2016;2:16047.

2 Vitali C, Bombardieri S, Jonsson R *et al*. Classification criteria for Sjögren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. Ann Rheum Dis 2002;61:554–8.

3 Shiboski SC, Shiboski CH, Criswell L *et al*. American College of Rheumatology classification criteria for Sjogren's syndrome: a data-driven, expert consensus approach in the Sjogren's International Collaborative Clinical Alliance cohort. Arthritis Care Res 2012;64:475–87.

4 Cornec D, Saraux A, Cochener B *et al*. Level of agreement between 2002 American-European Consensus Group and 2012 American College of Rheumatology classification criteria for Sjögren's syndrome and reasons for discrepancies. Arthritis Res Ther 2014;16:R74.

5 Rasmussen A, Ice JA, Li H *et al*. Comparison of the American-European Consensus Group Sjogren's syndrome classification criteria to newly proposed American College of Rheumatology criteria in a large, carefully characterised sicca cohort. Ann Rheum Dis 2014;73:31–8.

6 Shiboski CH, Shiboski SC, Seror R *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. Ann Rheum Dis 2017;76:9–16.

7 Shiboski CH, Shiboski SC, Seror R *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.

8 Tsuboi H, Hagiwara S, Asashima H *et al*. Comparison of performance of the 2016 ACR-EULAR classification criteria for primary Sjögren's syndrome with other sets of criteria in Japanese patients. Ann Rheum Dis 2017;76:1980–5.

9 Seror R, Ravaud P, Bowman SJ *et al*. EULAR Sjogren's syndrome disease activity index: development of a consensus systemic disease activity index for primary Sjogren's syndrome. Ann Rheum Dis 2010;69:1103–9.

10 Spijkervet FKL, Haacke E, Kroese FGM *et al*. Parotid gland biopsy, the alternative way to diagnose Sjögren Syndrome. Rheum Dis Clin North Am 2016;42:485–99.

11 Chisholm DM, Mason DK. Labial salivary gland biopsy in Sjögren's disease. J Clin Pathol 1968;21:656–60.

12 Whitcher JP, Shiboski CH, Shiboski SC *et al*. A simplified quantitative method for assessing keratoconjunctivitis sicca from the Sjogren's Syndrome International Registry. Am J Ophthalmol 2010;149:405–15.

13 Routsias JG, Goules JD, Charalampakis G *et al*. Malignant lymphoma in primary Sjogren's syndrome: an update on the pathogenesis and treatment. Semin Arthritis Rheum 2013;43:178–86.

14 Tape T. The Area under an ROC curve. http://gim.unmc.edu/dxtests/ROC3.htm (5 May 2017, date last accessed).

15 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159.

16 Vitali C, Moutsopoulos HM, Bombardieri S. The European Community Study Group on diagnostic criteria for Sjögren's syndrome. Sensitivity and specificity of tests for ocular and oral involvement in Sjögren's syndrome. Ann Rheum Dis 1994;53:637–47.

17 Delli K, Spijkervet FKL, Kroese FGM *et al*. Xerostomia. Monogr Oral Sci 2014;24:109–25.

18 Fiscella RG. Understanding dry eye disease: a managed care perspective. Am J Manag Care 2011;17(Suppl 1):S432–9.

19 Vissink A, Bootsma H. Connective tissue diseases: Refining the classification criteria for primary Sjögren syndrome. Nat Rev Rheumatol 2016;13:10–2.

20 Soto-Rojas AE, Kraus A. The oral side of Sjögren syndrome. Diagnosis and treatment. A review. Arch Med Res 2002;33:95–106.