

# UCSF

## UC San Francisco Previously Published Works

### Title

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.

### Permalink

<https://escholarship.org/uc/item/3db8z9gp>

### Journal

Arthritis & rheumatology (Hoboken, N.J.), 69(1)

### ISSN

2326-5191

### Authors

Shiboski, Caroline H  
Shiboski, Stephen C  
Seror, Raphaèle  
et al.

### Publication Date

2017

### DOI

10.1002/art.39859

Peer reviewed



Published in final edited form as:

*Arthritis Rheumatol.* 2017 January ; 69(1): 35–45. doi:10.1002/art.39859.

## 2016 ACR-EULAR Classification Criteria for primary Sjögren's Syndrome: A Consensus and Data-Driven Methodology Involving Three International Patient Cohorts

Caroline H. Shiboski, DDS, MPH, PhD<sup>1</sup>, Stephen C. Shiboski, PhD<sup>2</sup>, Raphaële Seror, MD, PhD<sup>3</sup>, Lindsey A. Criswell, MD, MPH, DSc<sup>4</sup>, Marc Labetoulle, MD<sup>3</sup>, Thomas M. Lietman, MD<sup>5</sup>, Astrid Rasmussen, MD, PhD<sup>6</sup>, Hal Scofield, MD<sup>6,7</sup>, Claudio Vitali, MD<sup>8</sup>, Simon J. Bowman, PhD FRCP<sup>9</sup>, Xavier Mariette, MD, PhD<sup>3</sup>, and the International Working Group on SS Classification Criteria\*

<sup>1</sup>Department of Orofacial Sciences, School of Dentistry, University of California San Francisco, California, USA

<sup>2</sup>Department of Epidemiology and Biostatistics, School of Medicine, University of California San Francisco, California, USA

<sup>3</sup>Université Paris-Sud; Assistance Publique – Hôpitaux de Paris, Hôpitaux Universitaires Paris-Sud; INSERM U1184; Le Kremlin-Bicêtre; Paris, France

<sup>4</sup>Rosalind Russell / Ephraim P. Engleman Rheumatology Research Center, Department of Medicine, School of Medicine, University of California San Francisco, California, USA

<sup>5</sup>Francis I. Proctor Foundation for Research in Ophthalmology, University of California San Francisco, California, USA

<sup>6</sup>Oklahoma Medical Research Foundation, Oklahoma, USA

<sup>7</sup>University of Oklahoma Health Sciences Center, and Department of Veterans Affairs Medical Center, Oklahoma, USA

<sup>8</sup>Istituto Villa San Giuseppe, Como, and Casa di Cura di Lecco, Lecco, Italy

<sup>9</sup>Rheumatology Department, University Hospitals Birmingham, Edgbaston, Birmingham, UK

### Abstract

**Objective**—To develop and validate an international set of classification criteria for primary Sjögren's Syndrome (pSS) using guidelines from the American College of Rheumatology (ACR)

**Address for correspondence and reprints:** Caroline H. Shiboski, DDS, MPH, PhD, Leland A. and Gladys K. Barber Distinguished Professor in Dentistry, Chair Department of Orofacial Sciences, Box 0422, Room S612, 513 Parnassus Avenue, University of California San Francisco, San Francisco, CA 94143-0422, Phone: (415)476-5976; Fax: (415)476-4204, caroline.shiboski@ucsf.edu.

\***Europe:** C Baldini, S Bombardieri (Italy), H Bootsma (Netherlands), S Challacombe (UK), V Devauchelle (France), S De Vita (Italy), T Dörner (Germany), JE Gottenberg (France), B Fisher (UK), R Jonsson (Norway), B Kirkham (UK), AA Kruijze (Netherlands), G Larkin (UK), Fai Ng (UK), M Ramos-Casals (Spain), S Rauz (UK), TR Radstake (Netherlands), A Saraux (France), M Schiødt (Denmark), E Theander (Sweden), A Tzoufias (Greece), A Vissink (Netherlands),

**US:** E Akpek, J Atkinson, AN Baer, S Carsons, N Carteron, T Daniels, B Fox, J Greenspan, G Illei, D Nelson, A Parke, S Pillemer, B Segal, K Sivits, W St Clair, D Stone, F Vivino, A Wu

**South America:** AM Heidenreich (Argentina), H Lanfranchi (Argentina), C Vollenweider (Argentina)

**Asia:** T Sumida (Japan), H Umehara (Japan), K Kitagawa (Japan)

**Patient advocacy:** Kathy Hammitt (US); Maggy Pincemin (France)

and the European League Against Rheumatism (EULAR). These criteria target individuals with signs/symptoms suggestive of SS.

**Methods**—We assigned preliminary importance weights to a consensus list of candidate criteria items using multi-criteria decision analysis. We tested and adapted the resulting draft criteria using existing cohort data on pSS cases and non-SS controls, with case/non-case status derived from expert clinical judgment. We then validated the performance of the classification criteria in a separate cohort of patients.

**Results**—The final classification criteria are based on the weighted sum of 5 items: anti-SSA(Ro) antibody positivity and focal lymphocytic sialadenitis with a focus score  $\geq 1$  foci/mm<sup>2</sup>, each scoring 3; an abnormal ocular staining score  $\geq 5$  (or van Bijsterveld score  $\geq 4$ ), a Schirmer test  $\leq 5$  mm/5 min, and an unstimulated salivary flow rate  $\leq 0.1$  mL/min, each scoring 1. Individuals (with signs/symptoms suggestive of SS) who have a total score  $\geq 4$  for the items above, meet the criteria for pSS. Sensitivity and specificity against clinician-expert derived case/non-case status in the final validation cohort were high: 96% (95% CI: 92%, 98%), and 95% (95% CI: 92%, 97%), respectively.

**Conclusion**—Using methodology consistent with other recent ACR/EULAR-approved classification criteria, we developed a single set of data-driven consensus classification criteria for pSS, that performed well in validation, and are well-suited as entry criteria for clinical trials.

Sjögren's syndrome (SS) is a multisystem autoimmune disease characterized by hypofunction of salivary and lacrimal glands and possible systemic multi-organ manifestations. It is primarily overseen by rheumatologists, in collaboration with ophthalmologists and oral medicine/pathology specialists. None of the 11 classification/diagnostic criteria published for SS from 1965 to 2002,(1–11) had been endorsed by the American College of Rheumatology (ACR) or European League Against Rheumatism (EULAR). During the past decade, the most commonly used classification criteria have been the American European Consensus Group (AECG) criteria,(11) which have proven useful in research and clinical practice. In 2012, new classification criteria developed within the NIH-funded Sjögren's International Collaborative Clinical Alliance (SICCA) registry were published after being provisionally approved by ACR.(12) These criteria were designed for entry into clinical trials, and the target population used for their development and validation consisted of individuals with signs and symptoms suggestive of SS. Subsequent analyses to compare the ACR and AECG criteria performed in a cohort of patients at the Oklahoma Medical Research Foundation (OMRF) revealed a high level of concordance.(13) Although both criteria sets involve similar items, the AECG criteria allow substitutions for alternatives and the use of symptoms of dry eyes and mouth in classifying patients. The provisional ACR criteria are based solely on objective tests, and consider symptoms as inclusion criteria for the target population to whom the criteria should apply.

While some treatments may improve symptoms and prevent complications of SS, currently there is no cure. However, the recent development of new therapeutic options for the management of various autoimmune diseases is promising for SS patients. Well-defined entry criteria, and endpoints that allow measurement of the effect of new treatments are needed for the development of new therapies. Disease activity indices for SS endpoints have

recently been developed and validated by the EULAR Sjögren's Task Force: EULAR SS Patient Reported Index (ESSPRI) and EULAR SS Disease Activity Index (ESSDAI).(14–17) The need for international consensus on classification criteria has recently been recognized by the SS scientific community.(18) Furthermore, this international criteria set should be established using guidelines published by both ACR and EULAR in order to be approved by both organizations.(19, 20)

In 2012, investigators from the SICCA team and the EULAR Sjögren's Task Force formed the International SS Criteria Working Group. The objective was to develop classification criteria for primary SS (pSS) that combined features of the ACR and AECG criteria, using methods consistent with ACR and EULAR guidelines. We describe here the development and validation of the resulting criteria, which have been approved by the ACR and EULAR. Consistent with our goal of producing criteria to aid in recruitment for clinical trials, we focus on primary rather than secondary SS. Patients with the latter would typically not be eligible for experimental treatments for SS.

## METHODS

### Overview

Our methods rely on both data and expert clinical judgment, and mirror those used for the development and validation of the 2010 ACR-EULAR criteria for rheumatoid arthritis,(21) and the 2013 ACR-EULAR criteria for systemic sclerosis.(22) The approach is outlined below (Figure 1):

1. A preliminary list of candidate items was generated based on the AECG and ACR criteria, and guided by analyses of existing datasets (item generation). This list was finalized in two meetings of the International SS Criteria Working Group, held concurrently with the 2013 International Symposium on SS and the 2013 ACR meeting.
2. We used multi-criteria decision analysis (MCDA)(23) to reduce the number of candidate criteria items, assign preliminary weights (item reduction and weight assignment), and help define a draft criteria set.
3. We tested and adapted the draft criteria using a development cohort with pSS disease status derived from expert clinical judgment based on clinical vignettes.
4. We then tested the performance of the classification criteria in a similarly defined but separate validation cohort of patients.
5. We also tested the performance of the classification criteria in a subset of difficult cases as described below.

### International SS Criteria Working Group

The working group comprised 55 clinician-experts including 36 rheumatologists, 10 oral medicine/pathology specialists, and 9 ophthalmologists; and two patient advocates (from the USA and Europe). The methodology team consisted of a statistician (SCS) and two epidemiologists (CHS and RS). Approximately half of the clinician-experts were from

Europe (Denmark, France, Greece, Italy, the Netherlands, Norway, Spain, Sweden, and the UK), and among the other half, most were from North and South America (the USA and Argentina), with the remaining from Japan.

### Item generation

Extensive statistical analyses were performed within the SICCA dataset with input from the working group to better understand the similarities and differences between the AECG and ACR criteria sets. Concomitantly, statistical analyses were performed within the OMRF cohort comparing the ACR and the AECG criteria, and a high level of concordance was identified (91% concordance among 646 OMRF participants, including 244 who met both sets of criteria and 343 who did not meet either).(13) Considering the high level of concordance observed between the AECG and ACR criteria, and the fact that the components in both criteria sets overlap to some degree, there was general agreement on many of the key items for inclusion. However, some tests were included in the AECG but not in the ACR criteria (the Schirmer test, unstimulated whole saliva (UWS) flow rate, sialography, salivary scintigraphy), and others were included in the ACR but not in the AECG criteria (the antinuclear/ANA antibody titer and rheumatoid factor/RF). Also, the ocular dryness was measured using the van Bijsterveld score (VBS) in the AECG criteria, and the Ocular Staining Score (OSS) in the ACR criteria, although these tests both measure ocular staining, the former with lissamine green and the latter with lissamine green (for conjunctiva) and fluorescein (for cornea). The comparative analyses performed both in the SICCA and OMRF cohorts, and presented to the working group, guided the generation of a final list of candidate items. It was agreed that all items originally included in both AECG and ACR criteria, except for ANA titer and RF, would be initial candidate items. The decision to exclude ANA and RF was based on analyses that revealed that an extremely small number of cases who met the ACR criteria were anti-SSA/B(Ro/La) negative but ANA (titer 1:320) and RF positive.(13)

### Item reduction and weight assignment

Relative ranking of selected items reflecting clinician-expert opinions was based on a web-based MCDA survey administered using *1000Minds* software.(23, 24) This approach, based on pairwise ranking of alternatives, each defined using selected criteria items, has been described.(25) The resulting item weights were normalized as percentages, and used to define an additive score (described below) reflecting the likelihood of assigning disease case status.

### Development and validation patient cohorts

Three prospective cohorts of individuals with signs and symptoms suggestive of SS have been recruited over the past 10 years by teams who are now members of the International SS Criteria Working Group. These include 1) the SICCA cohort, comprised of 3514 participants (including 1578 individuals who meet the ACR classification criteria for pSS) recruited from Argentina, China, Denmark, India, Japan, the UK and the USA (co-principal investigators (PIs): C. Shiboski and L. Criswell, at the University of California San Francisco); 2) the Paris-Sud cohort that includes 1011 participants (including 440 individuals who meet the AECG criteria for pSS) recruited in Paris, France (PI: X. Mariette at Paris-Sud University,

Bicêtre hospital in Paris); and 3) the OMRF cohort, that includes 837 participants (including 279 individuals who meet the AECG criteria for pSS) evaluated at either the Sjögren's Research Clinic at OMRF or the Sjögren's Clinic in the University of Minnesota (PI: K. Sivils, OMRF).

These cohorts share several key characteristics that make them appropriate for criteria development: Inclusion criteria required that participants have signs and symptoms suggestive of SS, warranting a comprehensive work-up by a multi-disciplinary team of SS clinicians. In addition to symptom-related data, objective tests with respect to oral, ocular, and systemic/serological endpoints had been collected using similar procedures:

- Oral tests: labial salivary gland (LSG) biopsy to identify focal lymphocytic sialadenitis (FLS) and focus score (FS)(26); UWS flow rates.(27, 28)
- Ocular tests: OSS using lissamine green and fluorescein, and other ocular tests such as Schirmer test and tear break-up time. For the ocular staining test, the Paris-Sud cohort used the VBS,(29) while SICCA used the OSS,(30) and OMRF used both. The Paris-Sud cohort also used fluorescein and collected data on the individual OSS components, so it could be computed subsequently. Thus data from the Paris-Sud and OMRF cohorts could be analyzed to establish a conversion algorithm between both scores as follows: for lower scores, 1–3, the VBS was equal to the OSS, but VBS of 4, 5, or 6 were equivalent to OSS scores of 5, 6, or 7, respectively. For the clinical vignettes, the ocular staining test was expressed as the OSS ranging from 0 to 7 and above. A group of four ophthalmologists from France, the US, and the UK formed an ad-hoc working group that interpreted the analyses performed on the Paris-Sud data (ML and TML) and on the OMRF data (AR). Together, they derived the conversion algorithm between the OSS and the VBS described above. In addition, since the VBS of 4 (previously used in the AECG criteria) was equivalent to an OSS of 5, the group agreed to modify the OSS threshold to 5 in the new criteria set. This threshold has also been shown, as part of subsequent analyses of the SICCA data, to be more specific for diagnostic purposes than the previous score of 3 (data not shown).
- Serological assays: including anti-SSA/B(Ro/La), ANA titers, RF, IgG, presence of complement C3 and C4.

Cohort PIs were each asked to provide a dataset that consisted of a random sample of 400 individuals with equal numbers of pSS cases and non-cases (using their own diagnostic definition), and without revealing case status in the dataset. The combined datasets thus comprised 1200 individuals with well-characterized data on the phenotypic features of SS. Clinical vignettes describing each individual's relevant features in text form were computer-generated using a program written in R version 3.2.(31) Vignettes described each individual with respect to age, gender, reported symptoms, clinical signs, and provided test results including ANA titers, RF, IgG, C3, C4, anti-SSA(Ro), anti-SSB(La), OSS for each eye, Schirmer for each eye, whether or not the LSG biopsy revealed FLS, and a FS (supplemental Figure 1). Ocular symptoms were defined according to the AECG definition, as a positive response to at least one of the following questions: 1) Have you had daily, persistent,

troublesome dry eyes for more than 3 months? 2) Do you have a recurrent sensation of sand or gravel in the eyes? 3) Do you use tear substitutes more than 3 times a day? Oral symptoms were defined as a positive response to at least one of the following questions: 1) Have you had a daily feeling of dry mouth for more than 3 months? 2) Do you frequently drink liquids to aid in swallowing dry food?

### **Assessment of SS case/control status**

We excluded 4 vignettes selected randomly from the study population to obtain 1196 vignettes that were randomly distributed into 26 surveys, each containing 46 individual vignettes. Research Electronic Data Capture (REDCap),(32) was used to administer each survey blindly to two clinician-experts. Twenty six pairs of clinician-experts participated in the first survey exercise, and each pair completed one survey. They were instructed to review each vignette, and asked if they thought the patient described had pSS. Responses included: “yes”, “no”, and “not sure”. Concordant yes/no responses were used to assign case/non-case status; concordant “not sure” responses were interpreted as non-gradable vignettes. All vignettes with discordant answers (yes/no; yes/not sure; or no/not sure) were included in a second round of surveys that were each sent to a third clinician-expert (a total of nine clinician-experts contributed to the second round of surveys). Concordance was then defined as two concordant answers out of three with a vignette defined as a pSS case if the answers included two “yes”. Similarly, a vignette was defined as a non-SS control if the answers included two “no”. The vignettes which received three discordant answers (yes/no/not sure) were considered “difficult cases”, and were combined into a third survey sent to eight clinician-experts, members of the steering committee. These difficult cases were defined as SS cases if the majority of clinician-experts (five out of eight) responded “yes” to a vignette, and as non-SS controls if the majority (five out of eight) responded “no”.

### **Randomization of vignettes across development and validation cohorts**

Each of the 1196 vignettes was assigned a unique identification number (ID), and were randomly divided into two groups of 598, one to be used as development cohort, and the other for validation purposes. Clinician-experts who completed the surveys were blinded to the origin (development or validation set) of the clinical vignettes.

### **Testing and adaptation of the draft criteria**

We conducted exploratory analyses of the clinician-expert rankings derived from the MCDA survey to characterize distributions of item-specific weights. Results were summarized graphically, and using summary statistics. We also performed analyses linking vignette items from the development cohort with corresponding clinician-expert outcome classifications, restricted to individuals with clinician-expert-assigned case/non-case outcomes. Conditional random forest classifiers(33) were used to obtain variable importance rankings for (1) all vignette items, and (2) binary indicators corresponding to the items and used in the MCDA survey.

Based on results from exploratory analyses, we defined several candidate classification criteria focusing the items selected by clinician-experts for the MCDA survey. Criteria were defined based on scores computed as weighted sums of binary indicators of presence/



absence of items, with weights reflecting relative importance. In addition to the MCDA-derived weights, we used logistic regression models fitted to the development sample to derive alternate weights from item-specific coefficients. Cut-off values for case designation for candidate criteria were computed using receiver operating characteristic (ROC) methods applied to clinician-expert-defined outcomes in the development dataset. For each candidate, two cut-off values were identified using a generalized Youden index.<sup>(34)</sup> The first weighted sensitivity and specificity as equally important, and the second weighted specificity as twice as important as sensitivity.

We held a final meeting of the SS Working Group to present and discuss testing and adaptation of the draft criteria results. A summary report was subsequently sent to all members, including those who could not attend the meeting. A REDCap survey was administered to the entire panel of clinician-experts, seeking consensus on the final draft criteria prior to validation.

### Validation

Validation of candidate criteria was based on ROC analyses using the validation sample, restricted to individuals with clinician-expert-assigned case/non-case outcomes. We separately assessed classification performance in the subset of difficult cases. Performance was summarized using estimated sensitivity and specificity with accompanying 95% confidence intervals, and area under the curve (AUC) statistics.

## RESULTS

### MCDA survey: distribution of responses and item weights

Fifty-two clinician-experts participated in the MCDA survey. Table 1 presents the item weights for each of the seven items. Note that weights are normalized to sum to 1, yielding a proportion interpretation. Figure 2 presents the distribution of item weights across experts. The curves in the figure are smoothed kernel density estimates that have a relative frequency interpretation similar to histograms. Results indicate that an LSG biopsy result of FLS with FS = 1 and anti-SSA/B(Ro/La) positivity received the highest average weights, followed by OSS, UWS, Schirmer, oral symptoms and ocular symptoms, respectively. Weight distributions for ocular/oral symptoms, Schirmer/UWS and FS/anti-SSA/B(Ro/La) were remarkably similar in both mode and variability.

### Case status assessment in development and validation cohorts

The first round of surveys yielded 819 concordant and 377 discordant responses (supplemental Figure 2). The concordant responses provided 415 pSS cases and 377 non-SS controls. The 377 vignettes with discordant responses were included in a second round of nine surveys assigned to nine clinician-experts, providing a third response to each discordant vignette. This yielded an additional 151 pSS cases and 125 non-SS controls (with two out of three concordant responses). When reconciling ID numbers among the vignettes initially randomly assigned to be used in either cohort, the first two rounds of surveys yielded 288 pSS cases and 248 non-SS controls in the development cohort, and 278 pSS cases and 254 non-SS controls in the validation cohort.



The 72 vignettes in the second round of survey that received three discordant responses were included in a third round of surveys administered to the eight members of the steering committee who were also clinician-experts. These provided a pool of 49 difficult cases that received a majority of concordant responses (5/8) after the third round of survey: 35 pSS cases and 14 non-SS controls.

### Criteria development

Random forest variable importance rankings based on the clinician-expert classifications of the development dataset vignettes are shown in Figure 3. Results based on all vignette variables, as well as the binary indicators consistent with items included in the MCDA survey are shown. Rankings agree well with results from the MCDA survey, and clearly indicate the relatively greater importance of objective measures such as the LSG FS and antibody results in expert classification decisions. Oral and ocular symptoms did not affect classification performance, reflecting the observation that over 94% of individuals had at least one symptom.

An initial criteria score was developed as a weighted sum of the 7 items in the MCDA survey, based on the average weights reported in Table 1. We used logistic regression models to develop an alternate empirical criteria score for the development data, focusing on the items used in the MCDA survey, but including indicators for anti-SSA(Ro) and anti-SSB(La) positivity as separate variables. Scores were computed using weights based on rescaled regression coefficients from a model retaining items representing significant predictors of case status.(35) Oral and ocular symptoms, and anti-SSB(La) positivity were excluded because they did not affect classification performance based on the random forest variable importance rankings from the clinician-expert classifications of the development dataset vignettes (Figure 3B). Furthermore, oral and/or ocular symptoms had been used among inclusion criteria for participation in the three patient cohorts, and therefore a group decision was made that oral and/or ocular symptoms or suspicion of SS based on one of the domains of ESSDAI would be preliminary requirements for applying the new SS classification criteria. The decision to exclude anti-SSB(La) as an item was also based on group discussions and on a study published by Baer and colleagues(36) that demonstrated that the presence of anti-SSB(La) without anti-SSA(Ro) antibodies, had no significant association with SS phenotypic features, relative to seronegative participants.

ROC analysis of the MCDA score yielded an AUC of 0.96, and two alternate cut-offs for case classification (Table 2), ROC analysis of the logistic score yielded an AUC value of 0.98, and two alternate cut-offs for case classification. We also considered a modified version of the logistic score that assigned equal weights to the OSS, Schirmer and UWS items, reflecting clinician-expert opinions that UWS should be weighted similarly to the Schirmer test, and for greater consistency with the results of the MCDA survey (Table 1). The ROC analysis yielded similar results to the logistic score (AUC=0.98; Table 2).

Table 2 also presents kappa statistics measuring agreement between outcome classifications based on the three alternative criterion scores and classifications with the existing AECG and ACR criteria. Results indicate high levels of agreement, with the strongest values

obtained from the logistic and modified logistic scores with a cut-off selected to weigh sensitivity and specificity equally.

The REDCap survey seeking consensus on the final draft criteria, yielded 98% clinician-expert-consensus towards the modified logistic score as the basis for final draft criteria, with case status based on a score  $\geq 4$ , and agreement to move forward with validation of these criteria. Table 3 presents the final criteria definition.

### Criteria validation

We compared the validation and development data with respect to key variables, including their associations with outcome classification. Overall agreement was quite high, indicating no major differences apparent in the two datasets (supplemental table).

Initial validation of the selected criteria was based on estimated sensitivity and specificity using the clinician-expert responses in the full validation dataset. Sensitivity was 96% (95% CI: 92%, 98%), and specificity was 95% (95% CI: 92%, 97%). Validation was also performed in the subset of 49 difficult cases and non-cases, for which sensitivity was 83% (95% CI: 66%, 93%), and specificity was 100% (95% CI: 77%, 100%).

## DISCUSSION

We present an international set of classification criteria for pSS, developed and validated using approaches approved by both ACR and EULAR committees that oversee classification criteria. These criteria are applicable to any patient with at least one symptom of ocular or oral dryness (based on AECG questions),(11) or suspicion of SS due to systemic features derived from the ESSDAI(16) measure with at least one positive domain item. The criteria do not apply to anyone with a prior diagnosis of a pre-specified list of conditions that would exclude participation in pSS therapeutic trials because of overlapping clinical features or interference with criteria tests. The new classification criteria are based on five objective tests/items, and a total score  $\geq 4$ , derived from the sum of the weights assigned to each positive test/item: with FLS with FS  $\geq 1$  and positive anti-SSA(Ro) serology having the highest weights (3 each) and OSS  $\geq 5$  (or VBS  $\geq 4$ ) on at least one eye, Schirmer test  $\leq 5$  mm/5min on at least one eye, and UWS flow rate  $\leq 0.1$  mL/min, having a weight of 1 each. We found that the criteria perform very well when validated using vignettes derived from patients with pSS status defined by expert opinion. The criteria retained high sensitivity and specificity in a subset of 49 difficult cases/non-cases.

The form of the proposed criteria improve upon previous criteria, in that they are based on a weighted sum of items, with weights derived from consensus expert opinion and analyses of patient data. Also, positive serology for anti-SSB/(La) in the absence of anti-SSA(Ro) is no longer considered a criteria item. For instance, in the validation cohort, 15 individuals were anti-SSB(La)-positive in the absence of anti-SSA(Ro) and FLS in the LSG biopsy, thus would have been classified as non-SS using the new criteria. However, 12 of these would have tested positive based on both the AECG and 2012 ACR criteria, and would very likely have been misclassified. Improvements from the 2012-ACR criteria include the addition of the Schirmer test and the UWS, the use of a higher threshold for the OSS ( $\geq 5$ ) and the

optional use of the VBS as an alternative to the OSS (in cases where an ophthalmologist trained in the OSS is not available). Additional modifications include removal of the high-titer ANA and positive RF as items. Improvements from the 2002-AECG criteria include oral and ocular symptoms being considered part of eligibility determination rather than serving as items, the OSS being included as an alternative to the VBS, sialography and salivary scintigraphy being omitted. Furthermore, the new criteria consider systemic signs and B-cell activation biomarkers (using the ESSDAI) as inclusion criteria, which will allow diagnosis of systemic and earlier forms of the disease when sicca features are not already present. Compared with the AECG criteria, exclusionary conditions have also been updated. IgG4-related disease was added, HCV infection is restricted to patients with positive PCR, and pre-existing lymphoma is allowable, since diagnosis of SS is sometimes made after a prior lymphoma occurrence.

Strengths of our approach include 1) assignment of criteria item weights combined consensus methods for quantifying expert opinion with confirmatory statistical analysis of real patient vignettes classified by clinician-experts; 2) the working group was international, and represented a range of clinical specialties (65% rheumatologists, 18% oral medicine/pathology specialists, and 16% ophthalmologists); 3) our methods have been successfully applied in the development and validation of ACR/EULAR classification criteria for RA(21) and systemic sclerosis.(22) Another advantage of these methods is that they are adaptable to future modifications of the criteria that may arise with the adoption of new diagnostic tests, such as parotid ultrasonography, or improved serological assays. For example, some research suggests that it may be important to distinguish between monospecific antibody assays to Ro60 or Ro52,(37–40) although further validation studies will be required before they can be used for patient classification. A shared limitation, common to criteria for many rheumatic diseases, is the use of expert clinical judgement in the absence of an objective "gold standard" for defining the disease, and the associated effect of the resulting "circularity" on measured performance of criteria sets.

The primary application of classification criteria is recruitment into clinical trials and studies. Although our study focused on classification of pSS, the proposed criteria may be applicable to SS associated with other autoimmune diseases. However, further research is needed to confirm this.

The landscape of SS has changed in recent years, due to both the recently validated disease activity indices, and the availability of new therapeutic agents. Using methodology consistent with other recent ACR/EULAR-approved classification criteria, we developed a single set of data-driven consensus classification criteria for pSS, that performed well in validation, and are well-suited as entry criteria for clinical trials.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We would like to express our appreciation to Steve Taylor and Kathy Hammitt from the Sjögren's Syndrome Foundation for hosting three of the meetings of the International SS Criteria Working Group, Dr Frédéric Desmoulins for his important work in preparation of the Paris-Sud cohort dataset, and Mi Lam for her contribution in preparation of the SICCA dataset.

We are very grateful to Paul Hansen and Franz Ombler, the developers and owners of the *1000Minds* software (<https://www.1000minds.com>), who granted us an Academic Award, providing both access and technical support to their software.

We also express our greatest appreciation to all participants who enrolled in the three patient cohorts used for development and validation of the criteria, and to the clinician-experts members of the international working group for attending meetings, providing valuable input as part of these meetings, and responding to several rounds of surveys, including grading multiple vignettes.

### Funding and financial disclosure

The patient cohorts involved in this research were funded by 1) National Institutes of Health (National Institute for Dental and Craniofacial Research (NIDCR), National Eye Institute, and Office of Research on Women's Health) contract N01 DE32636 and NIDCR contract HHSN26S201300057C for the Sjögren's International Collaborative Clinical Alliance cohort; and 2) National Institutes of Health grant numbers AR053483, AR050782, DE018209, DE015223, AI082714, GM104938, and 1P50 AR060804, the Oklahoma Medical Research Foundation, the Phileona Foundation, and the Sjögren's Syndrome Foundation for the Oklahoma Medical Research Foundation (OMRF) cohort. *1000Minds* (<https://www.1000minds.com>) granted us an Academic Award, providing both access and technical support to their software.

The authors received no financial support or other benefits from commercial sources for the work reported on in the manuscript, or any other financial interests, which could create a potential conflict of interest or the appearance of a conflict of interest with regard to the work.

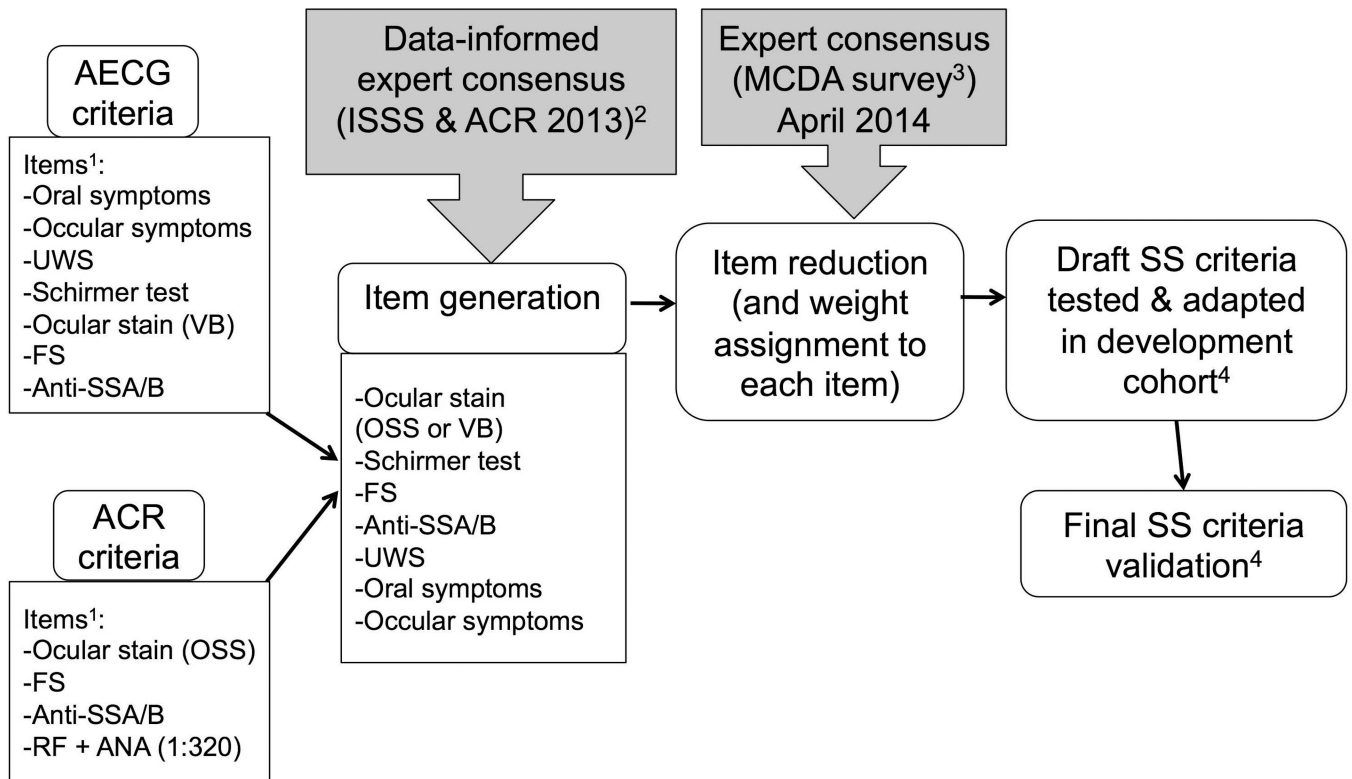
## References

1. Bloch KJ, Buchanan WW, Wohl MJ, Bunim JJ. Sjogren's Syndrome. A Clinical, Pathological, and Serological Study of Sixty-Two Cases. *Medicine (Baltimore)*. 1965; 44:187–231. [PubMed: 14315274]
2. Shearn, MA. Sjögren's syndrome Vol 2, Major Problems in Internal Medicine. Philadelphia: WB Saunders; 1971.
3. Daniels TE, Silverman S Jr, Michalski JP, Greenspan JS, Sylvester RA, Talal N. The oral component of Sjogren's syndrome. *Oral Surg Oral Med Oral Pathol*. 1975; 39(6):875–85. [PubMed: 1055974]
4. Ohfuji, T. Review on research reports. Annual report of the ministry of Health and Welfare. Sjögren's disease Research Committee; Japan: 1977.
5. Manthorpe R, Frost-Larsen K, Isager H, Prause JU. Sjogren's syndrome. A review with emphasis on immunological features. *Allergy*. 1981; 36(3):139–53. [PubMed: 7015910]
6. Homma M, Tojo T, Akizuki M, Yamagata H. Criteria for Sjogren's syndrome in Japan. *Scand J Rheumatol Suppl*. 1986; 61:26–7. [PubMed: 3473640]
7. Skopouli FN, Drosos AA, Papaioannou T, Moutsopoulos HM. Preliminary diagnostic criteria for Sjogren's syndrome. *Scand J Rheumatol Suppl*. 1986; 61:22–5. [PubMed: 3473635]
8. Fox RI, Robinson CA, Curd JG, Kozin F, Howell FV. Sjogren's syndrome. Proposed criteria for classification. *Arthritis Rheum*. 1986; 29(5):577–85. [PubMed: 3718551]
9. Vitali C, Bombardieri S, Moutsopoulos HM, Balestrieri G, Bencivelli W, Bernstein RM, et al. Preliminary criteria for the classification of Sjogren's syndrome. Results of a prospective concerted action supported by the European Community. *Arthritis Rheum*. 1993; 36(3):340–7. [PubMed: 8452579]
10. Fujibayashi T, Sugai S, Miyasaka N, Hayashi Y, Tsubota K. Revised Japanese criteria for Sjogren's syndrome (1999): availability and validity. *Modern Rheumatology*. 2004; 14:425–34. [PubMed: 24387718]

11. Vitali C, Bombardieri S, 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(6):554–8. [PubMed: 12006334]
12. Shiboski SC, Shiboski CH, Criswell LA, Baer AN, Challacombe S, Lanfranchi H, et al. American College of Rheumatology Classification Criteria for Sjögren's Syndrome: A Data-Driven, Expert Consensus Approach in the SICCA Cohort. *Arthritis Care Res (Hoboken).* 2012; 64:475–87. [PubMed: 22563590]
13. Rasmussen A, Ice JA, Li H, Grundahl K, Kelly JA, Radfar L, 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.* 2013
14. Seror R, Gottenberg JE, Devauchelle-Pensec V, Dubost JJ, Le Guern V, Hayem G, et al. European League Against Rheumatism Sjogren's Syndrome Disease Activity Index and European League Against Rheumatism Sjogren's Syndrome Patient-Reported Index: a complete picture of primary Sjogren's syndrome patients. *Arthritis Care Res (Hoboken).* 2013; 65(8):1358–64. [PubMed: 23436737]
15. Seror R, Mariette X, Bowman S, Baron G, Gottenberg JE, Bootsma H, et al. Accurate detection of changes in disease activity in primary Sjogren's syndrome by the European League Against Rheumatism Sjogren's Syndrome Disease Activity Index. *Arthritis Care Res (Hoboken).* 2010; 62(4):551–8. [PubMed: 20391511]
16. Seror R, Ravaud P, Bowman SJ, Baron G, Tzioufas A, Theander E, 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(6):1103–9. [PubMed: 19561361]
17. Seror R, Ravaud P, Mariette X, Bootsma H, Theander E, Hansen A, et al. EULAR Sjogren's Syndrome Patient Reported Index (ESSPRI): development of a consensus patient index for primary Sjogren's syndrome. *Ann Rheum Dis.* 2011; 70(6):968–72. [PubMed: 21345815]
18. Bowman SJ, Fox RI. Classification criteria for Sjogren's syndrome: nothing ever stands still! *Ann Rheum Dis.* 2014 Jan; 73(1):1–2. [PubMed: 24092418]
19. Classification and Response Criteria Subcommittee of the American College of Rheumatology Committee on Quality Measures. Development of Classification and Response Criteria for Rheumatologic Diseases. *Arth Rheum.* 2006; 55:348–52. [PubMed: 16739201]
20. Dougados M, Gossec L. Classification criteria for rheumatic diseases: why and how? *Arthritis Rheum.* 2007; 57(7):1112–5. [PubMed: 17907226]
21. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO 3rd, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Ann Rheum Dis.* 2010; 69(9):1580–8. [PubMed: 20699241]
22. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an american college of rheumatology/european league against rheumatism collaborative initiative. *Arthritis Rheum.* 2013; 65(11):2737–47. [PubMed: 24122180]
23. Hansen P, Ombler F. A new method for scoring multi-attribute value models using pairwise rankings of alternatives. *J Multi-Crit Decis Anal.* 2008; 15:87–107.
24. [cited 2015] 1000minds. Available from: <https://www.1000minds.com/solutions/decision-making-software>
25. Neogi T, Aletaha D, Silman AJ, Naden RL, Felson DT, Aggarwal R, et al. The 2010 American College of Rheumatology/European League Against Rheumatism classification criteria for rheumatoid arthritis: Phase 2 methodological report. *Arthritis Rheum.* 2010; 62(9):2582–91. [PubMed: 20872596]
26. Daniels TE, Cox D, Shiboski CH, Schiodt M, Wu A, Lanfranchi H, et al. Associations between salivary gland histopathologic diagnoses and phenotypic features of Sjogren's syndrome among 1,726 registry participants. *Arthritis Rheum.* 2011; 63(7):2021–30. [PubMed: 21480190]
27. Navazesh M. Methods for collecting saliva. *Ann N Y Acad Sci.* 1993; 694:72–7. [PubMed: 8215087]

28. Navazesh M, Kumar SK. University of Southern California School of D. Measuring salivary flow: challenges and opportunities. *J Am Dent Assoc.* 2008; 139(Suppl):35S–40S. [PubMed: 18460678]
29. van Bijsterveld OP. Diagnostic tests in the Sicca syndrome. *Archives of ophthalmology.* 1969; 82(1):10–4. [PubMed: 4183019]
30. Whitcher JP, Shiboski CH, Shiboski SC, Heidenreich AM, Kitagawa K, Zhang S, et al. A simplified quantitative method for assessing keratoconjunctivitis sicca from the Sjogren's Syndrome International Registry. *Am J Ophthalmol.* 2009; 149(3):405–15. [PubMed: 20035924]
31. Team, RC. A language and environment for statistical computing. R Foundation for Statistical Computing. 2015. [cited; Available from: <http://www.R-project.org/>]
32. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009; 42(2):377–81. [PubMed: 18929686]
33. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. *BMC Bioinformatics.* 2008; 9:307. [PubMed: 18620558]
34. Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology.* 2005; 16(1):73–81. [PubMed: 15613948]
35. Sullivan LM, Massaro JM, D'Agostino RB Sr. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med.* 2004; 23(10):1631–60. [PubMed: 15122742]
36. Baer AN, McAdams DeMarco M, Shiboski SC, Lam MY, Challacombe S, Daniels TE, et al. The SSB-positive/SSA-negative antibody profile is not associated with key phenotypic features of Sjogren's syndrome. *Ann Rheum Dis.* 2015; 74(8):1557–61. [PubMed: 25735642]
37. Defendenti C, Atzeni F, Spina MF, Grosso S, Cereda A, Guercilena G, et al. Clinical and laboratory aspects of Ro/SSA-52 autoantibodies. *Autoimmunity reviews.* 2011; 10(3):150–4. [PubMed: 20854935]
38. Dugar M, Cox S, Limaye V, Gordon TP, Roberts-Thomson PJ. Diagnostic utility of anti-Ro52 detection in systemic autoimmunity. *Postgrad Med J.* 2010; 86(1012):79–82. [PubMed: 20145055]
39. Ghillani P, Andre C, Toly C, Rouquette AM, Bengoufa D, Nicaise P, et al. Clinical significance of anti-Ro52 (TRIM21) antibodies non-associated with anti-SSA 60kDa antibodies: results of a multicentric study. *Autoimmunity reviews.* 2011; 10(9):509–13. [PubMed: 21447407]
40. Menendez A, Gomez J, Escanlar E, Caminal-Montero L, Mozo L. Clinical associations of anti-SSA/Ro60 and anti-Ro52/TRIM21 antibodies: Diagnostic utility of their separate detection. *Autoimmunity.* 2013; 46(1):32–9. [PubMed: 23039326]





**Figure 1.**

Overview of methodology used for the definitive set of Sjögren's syndrome classification criteria based on both data and expert clinical judgment. Item generation was derived from both the 2002 American-European Consensus Group (AECG) criteria and the 2012 American College of Rheumatology (ACR) criteria

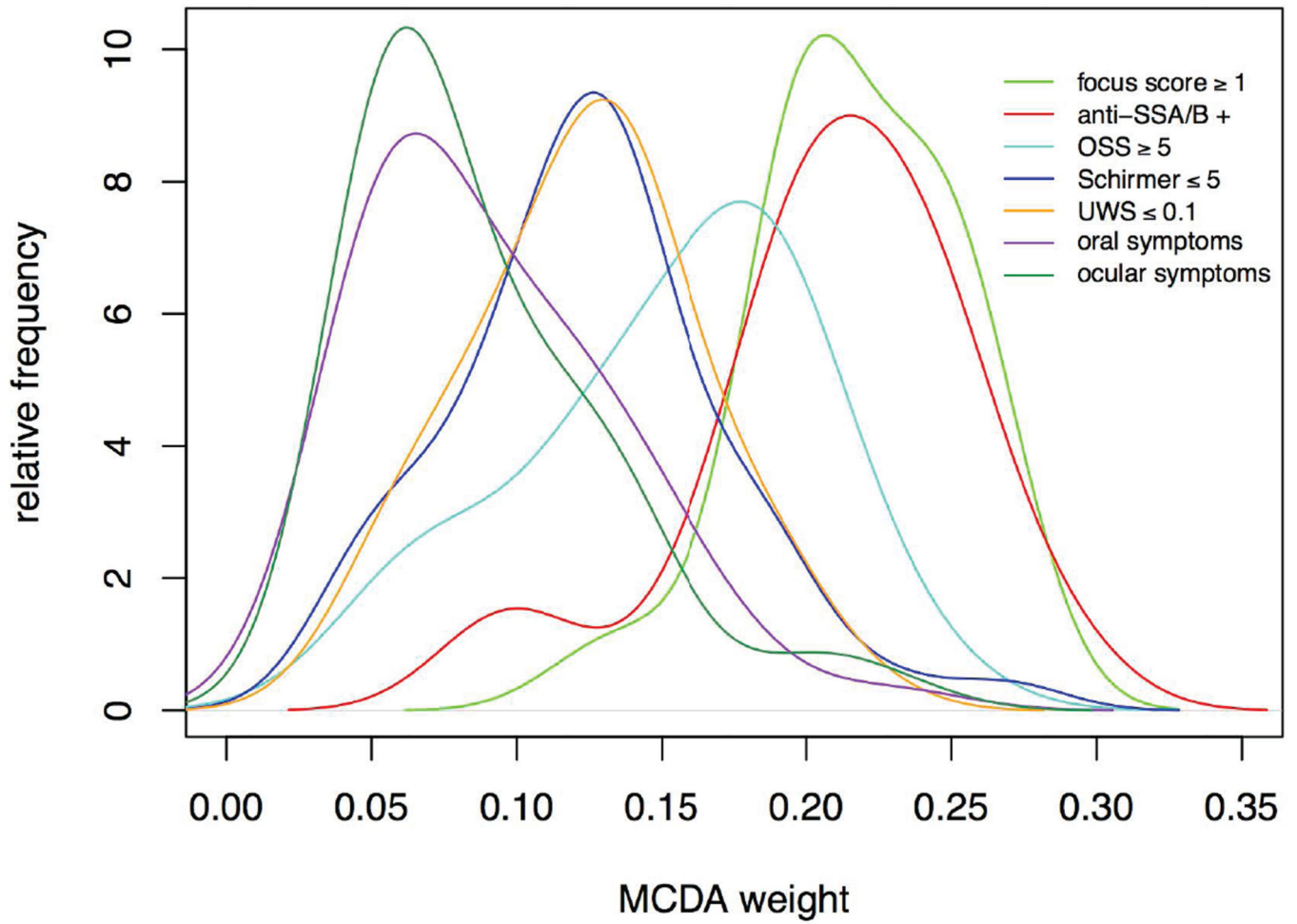
<sup>1</sup> UWS: Unstimulated whole saliva flow rate; VB: van Bijsterveld; OSS: Ocular staining score; FS: focus score computed from labial salivary gland biopsy in the presence of focal lymphocytic sialadenitis; RF: rheumatoid factor; ANA: antinuclear antibody titer

<sup>2</sup> International SS Criteria Working Group meetings held during the 2013 International Symposium on Sjögren's syndrome (ISSS) in Kyoto, Japan, and the 2013 American College of Rheumatology (ACR) annual meeting in San Diego, California

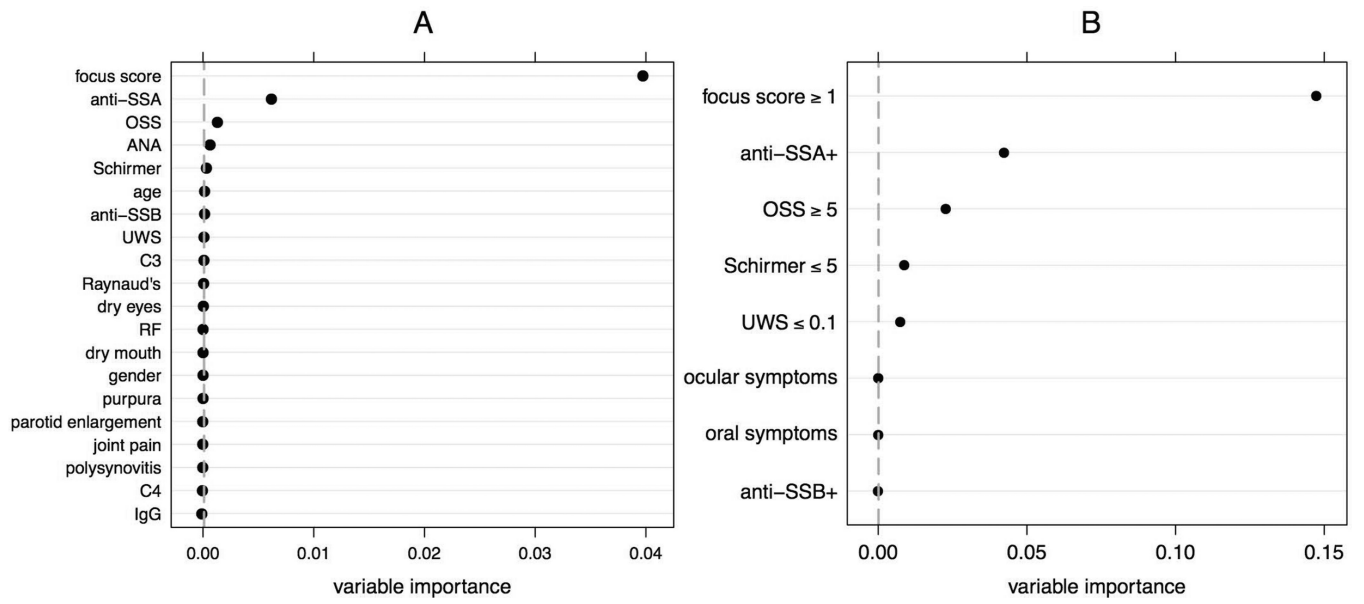
<sup>3</sup> MCDA: multi-criteria decision analysis survey performed using *1000Minds* software

<sup>4</sup> Disease case and non-case status were derived from expert clinical judgment based on clinical vignettes for both development and validation cohorts





**Figure 2.** Distributions of clinician expert assigned weights for seven items included in the MCDA survey. Curves are kernel-smoothed probability density estimates and the vertical scale can be interpreted similarly to relative frequency histograms.



**Figure 3.** variable importance for random forest classification of clinician expert case/non-case designations in development data vignettes. **A:** based on all vignette variables; **B:** restricted to binary indicators consistent with the MCDA survey items.

**Table 1**

Estimated weights for three alternate criterion scores, based on the development vignette data.

| Items <sup>1</sup>    | MCDA <sup>2</sup> | Logistic <sup>3</sup> | Modified <sup>3</sup> |
|-----------------------|-------------------|-----------------------|-----------------------|
| LSG with FLS and FS 1 | 0.22              | 3                     | 3                     |
| Anti-SSA/B(Ro/La)+    | 0.21              | 3                     | 3                     |
| OSS 5                 | 0.15              | 1                     | 1                     |
| Schirmer 5 mm/5min    | 0.12              | 1                     | 1                     |
| UWS 0.1 ml/min        | 0.12              | 0.5                   | 1                     |
| Oral Symptoms         | 0.09              | -                     | -                     |
| Ocular Symptoms       | 0.09              | -                     | -                     |
| Total                 | 1                 | 8.5                   | 9                     |

<sup>1</sup>LSG with FLS and FS 1: Labial salivary gland with focal lymphocytic sialadenitis and focus score 1 foci/4 mm<sup>2</sup>; OSS: ocular staining score; UWS: unstimulated whole saliva flow rate

<sup>2</sup>MCDA: multi-criteria decision analysis. The MCDA weights were based on the pairwise ranking of alternatives

<sup>3</sup>The logistic and modified weights resulted from the clinical expert rating of the development vignettes randomly selected among the 3 cohorts dataset. The modified version of the logistic score assigned equal weights to the OSS, Schirmer and UWS items. Logistic and Modified scores based on anti-SSA(Ro) only

Cut-off values, sensitivity, specificity, Kappa, area under the curve (AUC) values, and agreement (Kappa) with existing American-European Consensus Group (AECG) and American College of Rheumatology (ACR) criteria sets for three candidate criterion scores

**Table 2**

| Candidate criterion scores <sup>1</sup> | Cut-off <sup>2</sup> | Specificity (95% CI) | Sensitivity (95% CI) | Kappa | AUC  | Kappa AECG | Kappa ACR |
|---|----------------------|----------------------|----------------------|-------|------|------------|-----------|
| MCDA                                    | 0.46                 | 83 (78, 88)          | 95 (92, 97)          | 0.79  | 0.96 | 0.90       | 0.78      |
| Logistic                                | 0.58                 | 98 (95, 99)          | 78 (73, 83)          | 0.75  |      | 0.70       | 0.74      |
|   | 3.5                  | 89 (84, 93)          | 96 (93, 98)          | 0.86  | 0.98 | 0.91       | 0.82      |
| Modified                                | 4                    | 94 (90, 96)          | 91 (87, 94)          | 0.76  |      | 0.70       | 0.75      |
|   | 4                    | 89 (85, 93)          | 96 (93, 98)          | 0.86  | 0.98 | 0.91       | 0.82      |
|   | 5                    | 98 (95, 99)          | 80 (74, 84)          | 0.76  |      | 0.70       | 0.75      |

<sup>1</sup>MCDA: multi-criteria decision analysis. The MCDA weights were based on the pairwise ranking of alternatives. The logistic and modified weights resulted from the clinical expert rating of the development vignettes randomly selected among the 3 cohorts. The modified version of the logistic score assigned equal weights to the OSS, Schirmer and UWS items.

<sup>2</sup>Score values cut-off define a case. Cut-offs were chosen in each case to weight sensitivity and specificity equally (first row for each criterion score), or to weight specificity to be twice as important as sensitivity (second row for each criterion score).

**Table 3**  
**ACR-EULAR Classification Criteria for primary Sjögren's syndrome (pSS)**

The classification of SS applies to any individual who meets the inclusion criteria,<sup>1</sup> does not have any condition listed as exclusion criteria,<sup>2</sup> and who has a score  $\geq 4$  when summing the weights from the following items:

| Item   | Weight / Score |
|--|----------------|
| Labial salivary gland with focal lymphocytic sialadenitis and focus score $\geq 1$ . <sup>3</sup>    | 3              |
| Anti-SSA (Ro) +  | 3              |
| Ocular staining score $\geq 5$ (or van Bijsterveld score $\geq 4$ ) on at least one eye <sup>4</sup> | 1              |
| Schirmer $\leq 5$ mm/5min on at least one eye  | 1              |
| Unstimulated whole saliva flow rate $\leq 0.1$ ml/min <sup>5</sup>                                   | 1              |

<sup>1</sup>Inclusion criteria: these criteria are applicable to any patient with at least one symptom of ocular or oral dryness (defined as a positive response to at least one of the following questions: 1) Have you had daily, persistent, troublesome dry eyes for more than 3 months? 2) Do you have a recurrent sensation of sand or gravel in the eyes? 3) Do you use tear substitutes more than 3 times a day? 4) Have you had a daily feeling of dry mouth for more than 3 months? 5) Do you frequently drink liquids to aid in swallowing dry food?); or suspicion of SS from ESSDAI questionnaire (at least one domain with positive item)

<sup>2</sup>Exclusion criteria: Prior diagnosis of any of the following conditions would exclude diagnosis of SS and participation in SS studies or therapeutic trials because of overlapping clinical features or interference with criteria tests:

- History of head and neck radiation treatment
- Active Hepatitis C infection (with positive PCR)
- Acquired immunodeficiency syndrome
- Sarcoidosis
- Amyloidosis
- Graft versus host disease
- IgG4-related disease

Note: Patients who are normally taking anticholinergic drugs should be evaluated for objective signs of salivary hypofunction and ocular dryness after a sufficient interval off these medications for these components to be a valid measure of oral and ocular dryness

<sup>3</sup>The histopathologic examination should be performed by a pathologist with expertise in the diagnosis of focal lymphocytic sialadenitis, and focus score count (based on number of foci per 4 mm<sup>2</sup>) following a protocol described in Daniels et al 2011 (26)

<sup>4</sup>Ocular staining score described in Whitcher et al 2010 (30). van Bijsterveld score described in van Bijsterveld 1969 (29)

<sup>5</sup>Unstimulated whole saliva described in Navazesh & Kumar, 2008 (27)