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

Caroline H. Shiboski,[1] Stephen C. Shiboski,[1] Raphaèle Seror,[2] Lindsey A. Criswell,[1]
Marc Labetoulle,[2] Thomas M. Lietman,[1] Astrid Rasmussen,[3] Hal Scofield,[4] Claudio Vitali,[5]
Simon J. Bowman,[6] Xavier Mariette,[2] and the International
Sjögren's Syndrome Criteria Working Group

This criteria set has been approved by the American College of Rheumatology (ACR) Board of Directors and the European League Against Rheumatism (EULAR) Executive Committee. This signifies that the criteria set has been quantitatively validated using patient data, and it has undergone validation based on an independent data set. All ACR/EULAR-approved criteria sets are expected to undergo intermittent updates.

The ACR is an independent, professional, medical and scientific society that does not guarantee, warrant, or endorse any commercial product or service.

[1]Caroline H. Shiboski, DDS, MPH, PhD, Stephen C. Shiboski, PhD, Lindsey A. Criswell, MD, MPH, DSc, Thomas M. Lietman, MD: University of California, San Francisco; [2]Raphaèle Seror, MD, PhD, Marc Labetoulle, MD, Xavier Mariette, MD, PhD: Université Paris-Sud, AP-HP, Hôpitaux Universitaires Paris-Sud, INSERM U1184, Paris, France; [3]Astrid Rasmussen, MD, PhD: Oklahoma Medical Research Foundation, Oklahoma City; [4]Hal Scofield, MD: Oklahoma Medical Research Foundation, University of Oklahoma Health Sciences Center, and Department of Veterans Affairs Medical Center, Oklahoma City; [5]Claudio Vitali, MD: Istituto Villa San Giuseppe, Como, Italy, and Casa di Cura di Lecco, Lecco, Italy; [6]Simon J. Bowman, PhD, FRCP: University Hospitals Birmingham, NHS Foundation Trust, Birmingham, UK.

Objective. **To develop and validate an international set of classification criteria for primary Sjögren's syndrome (SS) using guidelines from the American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR). These criteria were developed for use in individuals with signs and/or 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 primary SS 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 of ≥ 1 foci/4 mm^2, each scoring 3; an abnormal ocular staining score of ≥ 5 (or van Bijsterveld score of ≥ 4), a Schirmer's test result of ≤ 5 mm/5 minutes, and an unstimulated salivary flow rate of ≤ 0.1 ml/minute, each scoring 1. Individuals with signs and/or symptoms suggestive of SS who have a total score of ≥ 4 for the above items meet the criteria for primary SS. Sensitivity and specificity against clinician-expert–derived case/non-case status in the final validation cohort were high, i.e., 96% (95% confidence interval [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 primary SS, which performed well in validation analyses and are well-suited as criteria for enrollment in 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 managed by rheumatologists, in collaboration with ophthalmologists and oral medicine/pathology specialists.

None of the 11 classification/diagnostic criteria sets for SS published between 1965 and 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 using the National Institutes of Health–funded Sjögren's International Collaborative Clinical Alliance (SICCA) registry were published after being provisionally approved by the ACR (12). These criteria were designed for classifying individuals for enrollment in 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 include similar items, the AECG criteria allow substitutions of some alternative items and the use of symptoms of dry eyes and mouth in classifying patients. The provisional ACR criteria are based solely on objective tests, and with symptoms considered 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 end points that allow measurement of the effect of new treatments, are needed for the development of new therapies. Disease activity indices for SS end points, i.e., the EULAR SS Patient Reported Index and EULAR SS Disease Activity Index (ESSDAI), have recently been developed and validated by the EULAR Sjögren's Task Force (14–17). The need for international consensus on classification criteria has recently been recognized by the SS scientific community (18). This international criteria set should be established using considerations and approaches 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 Sjögren's Syndrome Criteria Working Group. The objective was to develop classification criteria for primary SS that combined features of the ACR and AECG criteria, using methods consistent with those recommended by the ACR and EULAR. We describe herein 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 focused 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
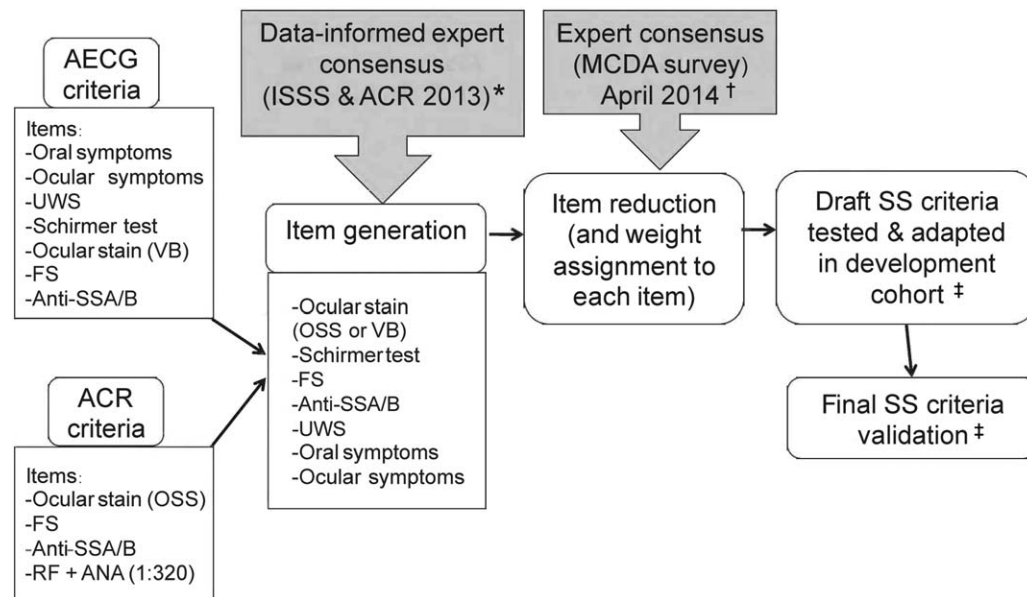
Figure 1. Overview of the methodology used for the definitive set of Sjögren's syndrome (SS) classification criteria, based on both data and expert clinical judgment. Item generation was derived from both the 2002 American-European Consensus Group (AECG) criteria and the 2012 American College of Rheumatology (ACR) criteria. UWS = unstimulated whole saliva flow rate; VB = van Bijsterveld; FS = focus score (computed from labial salivary gland biopsy in the presence of focal lymphocytic sialadenitis); OSS = Ocular Staining Score; RF = rheumatoid factor; ANA = antinuclear antibody. * = International SS Criteria Working Group meetings held during the 2013 International Symposium on Sjögren's Syndrome (ISSS) in Kyoto, Japan and the 2013 ACR Annual Meeting in San Diego, California. † = The multi-criteria decision analysis (MCDA) survey was performed using 1000Minds software. ‡ = Disease case and non-case status in both the development and the validation cohorts was derived from expert clinical judgment based on clinical vignettes.

rheumatoid arthritis (21,22) and the 2013 ACR/EULAR criteria for systemic sclerosis (23,24). The approach is outlined schematically in Figure 1 and described below.

1. We generated a preliminary list of candidate items based on the AECG and ACR criteria and guided by analyses of existing data sets (item generation). This list was finalized in 2 meetings of the International SS Criteria Working Group, held concurrently with the 2013 International Symposium on SS and the 2013 ACR Annual Meeting.

2. We used multi-criteria decision analysis (MCDA) (25) 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 primary SS disease status, as determined by clinician-expert assessment of 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 individuals whose SS case versus non–SS case status was difficult to determine (see below).

International Sjögren's Syndrome Criteria Working Group. The working group (see Appendix A) comprised 55 clinician-experts including 36 rheumatologists, 10 oral medicine/pathology specialists, and 9 ophthalmologists, as well as 2 patient advocates (from the US and Europe). The methodology team consisted of a statistician (SCS) and 2 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 US and Argentina), with the remainder from Japan.

Item generation. Extensive statistical analyses were performed within the SICCA data set, with input from the working group to better understand the similarities and differences between the AECG and ACR criteria sets. Concomitantly, statistical analyses comparing the ACR and the AECG criteria were performed within the OMRF cohort, 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 degree of concordance 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 (Schirmer's test, unstimulated whole saliva [UWS] flow rate, sialography, salivary scintigraphy), and others were included in the ACR but not in the AECG criteria (antinuclear antibody [ANA] titer and rheumatoid factor [RF] status). Also, ocular dryness was measured using the van Bijsterveld score (VBS) (26) in the AECG criteria and the Ocular Staining Score (OSS) (27) 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 the 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 the AECG and the ACR criteria, except ANA titer and RF status, would be initial candidate items. The decision to exclude ANA and RF was based on analyses showing that an extremely small number of individuals who met the ACR criteria were negative for anti-SSA/SSB (anti-Ro/La) but positive for ANA (titer ≥1:320) and RF (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 (25,28). This approach, based on pairwise ranking of alternatives (each defined using selected criteria items), has been described previously (29). The resulting item weights were normalized as percentages and used to define an additive score (see below) reflecting the likelihood of assigning disease case status.

Development and validation patient cohorts. Three prospective cohorts of individuals with signs and/or symptoms suggestive of SS have been recruited over the past 10 years by teams of investigators who are now members of the International SS Criteria Working Group. These cohorts include 1) the SICCA cohort, comprising 3,514 patients (including 1,578 individuals who meet the ACR classification criteria for primary SS) recruited from Argentina, China, Denmark, India, Japan, the UK, and the US (co–principal investigators CHS and LAC), 2) the Paris-Sud cohort, which consists of 1,011 patients (including 440 individuals who meet the AECG criteria for primary SS) recruited in Paris (principal investigator XM), and 3) the OMRF cohort, which includes 837 participants (including 279 individuals who meet the AECG criteria for primary SS) evaluated at either the Sjögren's Research Clinic at OMRF or the Sjögren's Clinic at the University of Minnesota (principal investigator K. Sivils, PhD [OMRF]).

These cohorts share several key characteristics that make them appropriate for criteria development: inclusion criteria required that participants have signs and/or symptoms suggestive of SS, warranting a comprehensive evaluation by a multidisciplinary team of SS clinicians. In addition to symptom-related data, objective tests with respect to oral, ocular, and systemic/serologic end points had been performed using similar procedures, as described below.

Oral tests. Labial salivary gland (LSG) biopsy was performed to identify focal lymphocytic sialadenitis and obtain a focus score (30). UWS flow rates were measured using standard methods (31,32).

Ocular tests. The OSS was obtained using lissamine green and fluorescein. Other ocular tests included Schirmer's test and measurement of tear break-up time. Ocular staining was assessed with the VBS in the Paris-Sud cohort, the OSS in the SICCA cohort, and both methods in the OMRF cohort. The Paris-Sud cohort investigators also used fluorescein and collected data on the individual OSS components, so the OSS 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 (i.e., scores of 1–3), the VBS was equal to the OSS, but VBS grades of 4, 5, and 6 were equivalent to OSS grades of 5,

6, and 7, respectively. For assessment of the clinical vignettes, ocular staining was expressed as the OSS, ranging from 0 to ≥7. A group of 4 ophthalmologists from France, the US, and the UK, including 3 of the authors, 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 a 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).

Serologic assays. Serologic studies included testing for anti-SSA/SSB (anti-Ro/La), ANA, RF, IgG, and complements C3 and C4.

Cohort PIs were each asked to provide a data set that consisted of a random sample of 400 individuals, with equal numbers of primary SS cases and non-cases (using their own diagnostic definition), and case status not revealed in the data set. The combined data sets thus comprised 1,200 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 (33). Vignettes described each individual with respect to age, sex, reported symptoms, clinical signs, test results including ANA titer, RF, IgG, C3, C4, anti-SSA/Ro, and anti-SSB/La status, OSS for each eye, Schirmer's test result for each eye, whether the LSG biopsy revealed focal lymphocytic sialadenitis, and focus score (see Supplementary Figure 1, on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.39859/abstract). Ocular symptoms were defined according to the AECG definition, as a positive response to at least 1 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 1 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 1,196 vignettes that were randomly distributed into 26 surveys, each containing 46 individual vignettes. Research Electronic Data Capture (REDCap) (34) was used to administer each survey to 2 clinician-experts, under blinded conditions. Twenty-six pairs of clinician-experts participated in the first survey exercise, and each pair completed 1 survey. They were instructed to review each vignette and asked if they thought the patient described had primary SS. Possible responses were "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 9 clinician-experts contributed to the second round of surveys). Concordance was then defined as 2 concordant answers of the 3, with a vignette defined as a primary SS case if there were 2 "yes" answers and as a non-SS control if there

were 2 "no" answers. Vignettes that received 3 discordant answers (yes/no/not sure) were considered "difficult-to-classify cases" and were combined into a third survey sent to 8 clinician-experts, all of whom were members of the steering committee. These difficult-to-classify cases were defined as SS cases if the majority of clinician-experts ($\geq$5 of 8) responded "yes" to a vignette, and as non-SS controls if the majority responded "no."

Randomization of vignettes across development and validation cohorts. Each of the 1,196 vignettes was assigned a unique identification number, and the vignettes were randomly divided into two groups of 598, with one to be used as development cohort and the other for validation purposes. Clinician-experts who completed the surveys were blinded with regard 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 (35) 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 on 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. Cutoff values for case designation for candidate criteria were computed using receiver operating characteristic (ROC) methods applied to clinician-expert–defined outcomes in the development data set. For each candidate item, 2 cutoff values were identified using a generalized Youden index (36). For the first cutoff value, sensitivity and specificity were weighted as equally important; for the second, specificity was weighted as twice as important as sensitivity.

We held a final meeting of the International SS Criteria 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.

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

Results

Distribution of responses and item weights in the MCDA survey. Fifty-two clinician-experts participated in the MCDA survey. Table 1 shows the item weights for each of the 7 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 that used in histograms. The results indicate that an LSG biopsy showing focal lymphocytic sialadenitis with a focus score of $\geq$1 and anti-SSA/SSB (anti-Ro/La) positivity received the highest average weights, followed by OSS, UWS, Schirmer's test result, oral symptoms, and ocular symptoms, respectively. Weight distributions for ocular/oral symptoms, Schirmer's test result/UWS, and focus score/anti-SSA/SSB (anti-Ro/La) were remarkably similar in both mode and variability.

Case status assessment in the development and validation cohorts. The first round of surveys yielded 819 concordant and 377 discordant responses (see

Table 1. Estimated weights for 3 alternate criterion scores, based on the development vignette data

Item	MCDA*	Logistic†	Modified†
Labial salivary gland with focal lymphocytic sialadenitis and focus score of $\geq$1 foci/4 mm^2	0.22	3	3
Anti-SSA/SSB (anti-Ro/La) positive	0.21	3‡	3‡
OSS $\geq$5	0.15	1	1
Schirmer's test $\leq$5 mm/5 minutes	0.12	1	1
UWS $\leq$0.1 ml/minute	0.12	0.5	1
Oral symptoms	0.09	–	–
Ocular symptoms	0.09	–	–
Total	1	8.5	9

* The multi-criteria decision analysis (MCDA) weights were based on the pairwise ranking of alternatives.
† The logistic and modified weights resulted from the clinician-expert rating of the development vignettes randomly selected from among the 3-cohort data set. The modified version of the logistic score assigned equal weights to the Ocular Staining Score (OSS), Schirmer's test, and unstimulated whole saliva flow rate (UWS) items.
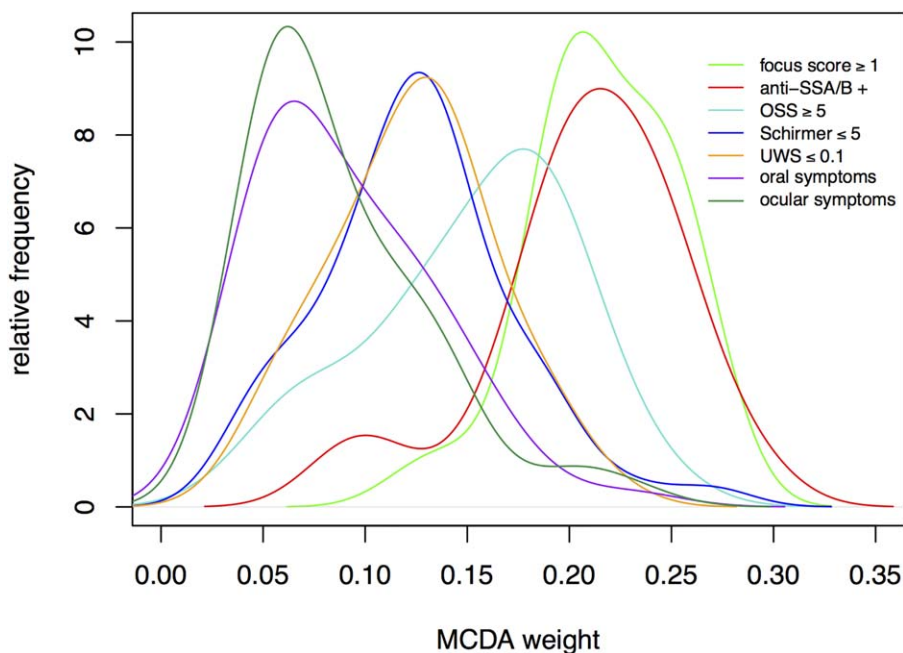‡ Based on anti-SSA/Ro only.

Figure 2. Distributions of clinician-expert–assigned weights for 7 items included in the multi-criteria decision analysis (MCDA) survey. Curves are smoothed kernel probability density estimates, and the vertical scale can be interpreted similarly to relative frequency histograms. OSS = Ocular Staining Score; UWS = unstimulated whole saliva flow rate.

Supplementary Figure 2, on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.39859/abstract). The concordant responses provided 415 primary SS cases and 377 non-SS controls. The 377 vignettes with discordant responses were included in a second round of 9 surveys assigned to 9 clinician-experts, providing a third response to each discordant vignette. This yielded an additional 151 primary SS cases and 125 non-SS controls (with 2 of the 3 responses being concordant). When reconciling identification numbers among the vignettes initially randomly assigned to be used in either cohort, the first 2 rounds of surveys yielded 288 primary SS cases and 248 non-SS controls in the development cohort, and 278 primary SS cases and 254 non-SS controls in the validation cohort.

The 72 vignettes in the second round of the survey that received 3 discordant responses were included in a third round of surveys administered to the 8 members of the steering committee who were also clinician-experts. These provided a pool of 49 difficult-to-classify cases that received a majority of concordant responses ($\geq$5 of 8) after the third round of survey: 35 primary SS cases and 14 non-SS controls.

Criteria development. Random forest variable importance rankings based on the clinician-expert classifications of the development data set 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 corresponded well with results from the MCDA survey and clearly indicated the relatively greater importance of objective measures such as the LSG focus score and antibody results in expert classification decisions. Oral and ocular symptoms did not affect classification performance, reflecting the observation that >94% of individuals had at least 1 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 in which items representing significant predictors of case status were retained (37). 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 data set vignettes (Figure 3B). Furthermore, oral and/or ocular symptoms had been part of the inclusion criteria for participation in the 3 patient cohorts; therefore, a
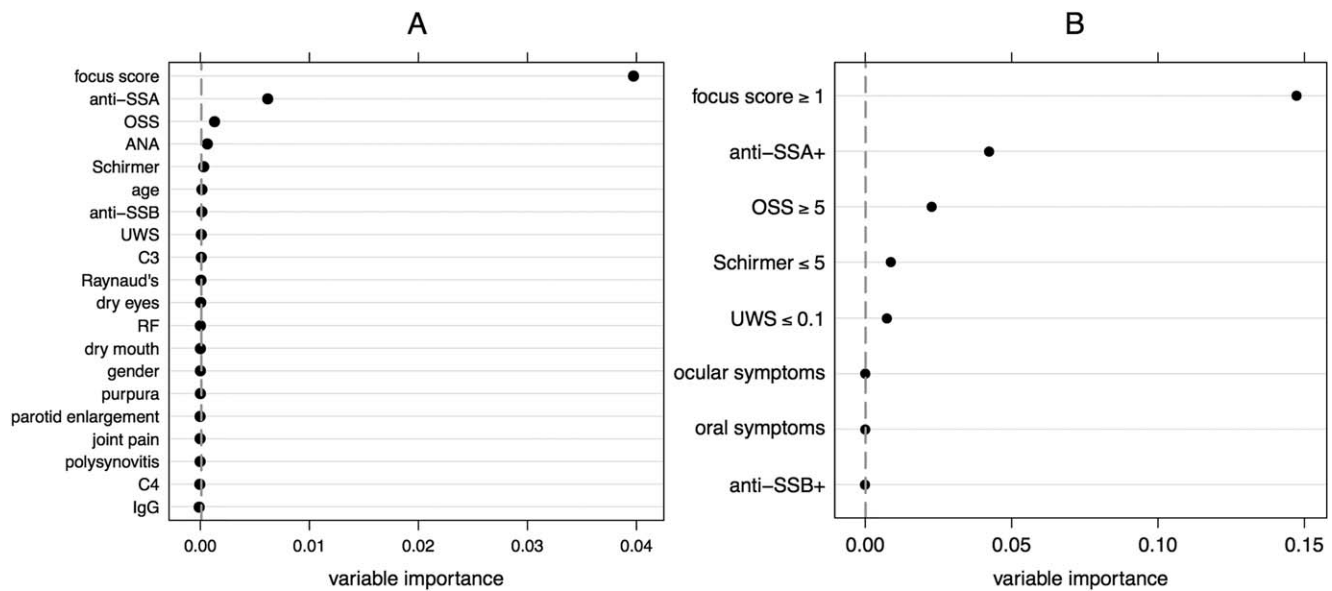
Figure 3. Importance of variables for random forest classification of clinician-expert case/non-case designations in development data vignettes. Analyses based on all vignette variables (**A**) and restricted to binary indicators consistent with the multi-criteria decision analysis survey items (**B**) were performed. OSS = Ocular Staining Score; ANA = antinuclear antibody; UWS = unstimulated whole saliva flow rate; RF = rheumatoid factor.

group decision was made that oral and/or ocular symptoms or suspicion of SS based on 1 of the domains of the 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 demonstrating that the presence of anti-SSB/La without anti-SSA/Ro antibodies had no significant association with SS phenotypic features, relative to seronegative participants (38).

ROC analysis of the MCDA score yielded an AUC value of 0.96 and 2 alternate cutoffs for case classification (Table 2). ROC analysis of the logistic score yielded an AUC value of 0.98 and 2 alternate cutoffs for case classification. We also considered a modified

Table 2. Cutoff values, sensitivity, specificity, kappa statistic, AUC values, and agreement with existing AECG and ACR criteria sets, for 3 candidate criterion scores*

Candidate criterion score, cutoff†	Specificity (95% CI)	Sensitivity (95% CI)	κ	AUC	Agreement with AECG criteria (κ)	Agreement with ACR criteria (κ)
MCDA‡				0.96		
0.46	83 (78–88)	95 (92–97)	0.79		0.90	0.78
0.58	98 (95–99)	78 (73–83)	0.75		0.70	0.74
Logistic§				0.98		
3.5	89 (84–93)	96 (93–98)	0.86		0.91	0.82
4	94 (90–96)	91 (87–94)	0.76		0.70	0.75
Modified§				0.98		
4	89 (85–93)	96 (93–98)	0.86		0.91	0.82
5	98 (95–99)	80 (74–84)	0.76		0.70	0.75

* AUC = area under the curve; AECG = American-European Consensus Group; ACR = American College of Rheumatology; 95% CI = 95% confidence interval.
† Score values greater than or equal to the cutoff value define a case. Cutoffs 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).
‡ The multi-criteria decision analysis (MCDA) weights were based on the pairwise ranking of alternatives.
§ The logistic and modified weights resulted from the clinician-expert rating of the development vignettes randomly selected from among the 3-cohort data set. The modified version of the logistic score assigned equal weights to the Ocular Staining Score, Schirmer's test, and unstimulated whole saliva flow rate items.

Table 3. American College of Rheumatology/European League Against Rheumatism classification criteria for primary Sjögren's syndrome: The classification of primary Sjögren's syndrome applies to any individual who meets the inclusion criteria,* does not have any of the conditions listed as exclusion criteria,† and has a score of ≥4 when the weights from the 5 criteria items below are summed.

Item	Weight/score
Labial salivary gland with focal lymphocytic sialadenitis and focus score of ≥1 foci/4 mm²‡	3
Anti-SSA/Ro positive	3
Ocular Staining Score ≥5 (or van Bijsterveld score ≥4) in at least 1 eye§¶	1
Schirmer's test ≤5 mm/5 minutes in at least 1 eye§	1
Unstimulated whole saliva flow rate ≤0.1 ml/minute§#	1

* These inclusion criteria are applicable to any patient with at least 1 symptom of ocular or oral dryness, defined as a positive response to at least 1 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 in whom there is suspicion of Sjögren's syndrome (SS) from the European League Against Rheumatism SS Disease Activity Index questionnaire (at least 1 domain with a positive item).
† Exclusion criteria include prior diagnosis of any of the following conditions, which would exclude diagnosis of SS and participation in SS studies or therapeutic trials because of overlapping clinical features or interference with criteria tests: 1) history of head and neck radiation treatment, 2) active hepatitis C infection (with confirmation by polymerase chain reaction, 3) AIDS, 4) sarcoidosis, 5) amyloidosis, 6) graft-versus-host disease, 7) IgG4-related disease.
‡ The histopathologic examination should be performed by a pathologist with expertise in the diagnosis of focal lymphocytic sialadenitis and focus score count, using the protocol described by Daniels et al (30).
§ Patients who are normally taking anticholinergic drugs should be evaluated for objective signs of salivary hypofunction and ocular dryness after a sufficient interval without these medications in order for these components to be a valid measure of oral and ocular dryness.
¶ Ocular Staining Score described by Whitcher et al (27); van Bijsterveld score described by van Bijsterveld (26).
# Unstimulated whole saliva flow rate measurement described by Navazesh and Kumar (32).

version of the logistic score that assigned equal weights to the OSS, Schirmer's test result, and UWS items, reflecting clinician-expert opinions that UWS should be weighted similarly to the Schirmer's test result 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 3 alternate 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 cutoff selected to weight sensitivity and specificity equally.

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

Validation of candidate criteria. 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 apparent major differences in the 2 data sets (see Supplementary Table 1, on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.39859/abstract). Initial validation of the selected criteria was based on estimated sensitivity and specificity using the clinician-expert responses in the full validation data set. 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-to-classify cases and non-cases, for which sensitivity was 83% (95% CI 66–93%) and specificity was 100% (95% CI 77–100%).

Discussion

We present herein an international set of classification criteria for primary SS, 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 1 symptom of ocular or oral dryness (based on AECG questions) (11) or suspicion of SS due to systemic features derived from the ESSDAI measure (16) with at least 1 positive domain item. The criteria do not apply to individuals with a prior diagnosis of a condition (from a prespecified list) that would exclude participation in primary SS therapeutic trials because of overlapping clinical features or interference with criteria tests. The new classification criteria are based on 5 objective tests/items. Individuals are classified as having primary SS if they have a total score of ≥4, derived from the sum of the weights assigned to each positive test/item (with focal lymphocytic sialadenitis with focus score ≥1 and anti-SSA/Ro positivity having the highest weights [3 each] and OSS ≥5 [or VBS ≥4] in at least 1 eye, Schirmer's test result ≤5 mm/5 minutes in at least 1 eye, and UWS flow rate ≤0.1 ml/minute having a weight of 1 each). We found that the criteria perform very well when validated using vignettes describing patients with primary SS status defined by expert opinion. The criteria retained high sensitivity and specificity in a subset of 49 vignettes for which case/non-case distinction was difficult.

The form of the proposed criteria improves 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 focal lymphocytic sialadenitis on LSG biopsy, and thus would have been classified as non-SS using the new criteria. However, 12 of them would have been classified as having primary SS based on both the AECG and the 2012 ACR criteria, and this would very likely have been a misclassification.

Improvements from the 2012 ACR criteria include the addition of Schirmer's test and the UWS, the use of a higher threshold for the OSS (≥5), and the optional use of the VBS as an alternative to the OSS (in cases when an ophthalmologist trained in the OSS is not available). Additional modifications include removal of 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 (i.e., eligibility of individuals to be assessed for SS using the criteria) rather than serving as criteria items, the OSS being included as an alternative to the VBS, and sialography and salivary scintigraphy being omitted. Furthermore, the new criteria consider systemic signs and B cell activation biomarkers (determined using the ESSDAI) in inclusion eligibility determination, 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 has been added, hepatitis C infection requires confirmation by polymerase chain reaction, and preexisting lymphoma is allowable, since diagnosis of SS is sometimes made after a prior lymphoma occurrence.

Strengths of our approach include the following: 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); and 3) our methods have been successfully applied in the development and validation of ACR/EULAR classification criteria for rheumatoid arthritis (21,22) and systemic sclerosis (23,24). 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 serologic assays. For example, some research suggests that it may be important to distinguish between monospecific antibody assays to Ro 60 or Ro 52 (39–42), although further validation studies will be needed before they can be used for patient classification. A shared limitation, common to criteria for many rheumatic diseases, is the use of expert clinical judgment 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 primary SS, 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 primary SS, which performed well in validation and are well-suited as entry criteria for clinical trials.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. S. C. Shiboski had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. C. H. Shiboski, S. C. Shiboski, Seror, Bowman, Mariette.

Acquisition of data. C. H. Shiboski, S. C. Shiboski, Seror, Criswell, Labetoulle, Lietman, Rasmussen, Scofield, Vitali, Bowman, Mariette.

Analysis and interpretation of data. C. H. Shiboski, S. C. Shiboski, Seror, Criswell, Labetoulle, Lietman, Rasmussen.

REFERENCES

1. Bloch KJ, Buchanan WW, Wohl MJ, Bunim JJ. Sjoegren's syndrome: a clinical, pathological, and serological study of sixty-two cases. Medicine (Baltimore) 1965;44:187–231.
2. Shearn MA. Sjögren's syndrome. Philadelphia: WB Saunders; 1971.
3. Daniels TE, Silverman S Jr, Michalski JP, Greenspan JS, Sylvester RA, Talal N. The oral component of Sjögren's syndrome. Oral Surg Oral Med Oral Pathol 1975;39:875–85.
4. Ohfuji T. Sjögren's Disease Research Committee review on research reports: annual report of the Ministry of Health and Welfare. Tokyo: Japanese Ministry of Health; 1977.
5. Manthorpe R, Frost-Larsen K, Isager H, Prause JU. Sjögren's syndrome: a review with emphasis on immunological features. Allergy 1981;36:139–53.
6. Homma M, Tojo T, Akizuki M, Yamagata H. Criteria for Sjögren's syndrome in Japan. Scand J Rheumatol Suppl 1986;61:26–7.
7. Skopouli FN, Drosos AA, Papaioannou T, Moutsopoulos HM. Preliminary diagnostic criteria for Sjögren's syndrome. Scand J Rheumatol Suppl 1986;61:22–5.
8. Fox RI, Robinson CA, Curd JG, Kozin F, Howell FV. Sjögren's syndrome: proposed criteria for classification. Arthritis Rheum 1986;29:577–85.
9. Vitali C, Bombardieri S, Moutsopoulos HM, Balestrieri G, Bencivelli W, Bernstein RM, et al. Preliminary criteria for the classification of Sjögren's syndrome: results of a prospective concerted action supported by the European Community. Arthritis Rheum 1993;36:340–7.
10. Fujibayashi T, Sugai S, Miyasaka N, Hayashi Y, Tsubota K. Revised Japanese criteria for Sjögren's syndrome (1999): availability and validity. Mod Rheumatol 2004;14:425–34.
11. Vitali C, Bombardieri S, Jonsson R, Moutsopoulos HM, Alexander EL, Carsons SE, 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.
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 Sjögren's International Collaborative Clinical Alliance cohort. Arthritis Care Res (Hoboken) 2012;64:475–87.
13. Rasmussen A, Ice JA, Li H, Grundahl K, Kelly JA, Radfar L, et al. Comparison of the American-European Consensus Group Sjögren'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.
14. Seror R, Gottenberg JE, Devauchelle-Pensec V, Dubost JJ, le Guern V, Hayem G, et al. European League Against Rheumatism Sjögren's Syndrome Disease Activity Index and European League Against Rheumatism Sjögren's Syndrome Patient-Reported Index: a complete picture of primary Sjögren's syndrome patients. Arthritis Care Res (Hoboken) 2013;65:1358–64.
15. Seror R, Mariette X, Bowman S, Baron G, Gottenberg JE, Bootsma H, et al. Accurate detection of changes in disease activity in primary Sjögren's syndrome by the European League Against Rheumatism Sjögren's Syndrome Disease Activity Index. Arthritis Care Res (Hoboken) 2010;62:551–8.
16. Seror R, Ravaud P, Bowman SJ, Baron G, Tzioufas A, Theander E, et al. EULAR Sjögren's Syndrome Disease Activity Index: development of a consensus systemic disease activity index for primary Sjögren's syndrome. Ann Rheum Dis 2010;69:1103–9.
17. Seror R, Ravaud P, Mariette X, Bootsma H, Theander E, Hansen A, et al. EULAR Sjögren's Syndrome Patient Reported Index (ESSPRI): development of a consensus patient index for primary Sjögren's syndrome. Ann Rheum Dis 2011;70:968–72.
18. Bowman SJ, Fox RI. Classification criteria for Sjögren's syndrome: nothing ever stands still! Ann Rheum Dis 2014;73:1–2.
19. Classification and Response Criteria Subcommittee of the American College of Rheumatology Committee on Quality Measures. Development of classification and response criteria for rheumatic diseases. Arthritis Rheum 2006;55:348–52.
20. Dougados M, Gossec L. Classification criteria for rheumatic diseases: why and how? Arthritis Rheum 2007;57:1112–5.
21. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO III, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis 2010;69:1580–8.
22. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO III, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62:2569–81.
23. 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;65:2737–47.
24. 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. Ann Rheum Dis 2013;72:1747–55.
25. 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.
26. Van Bijsterveld OP. Diagnostic tests in the Sicca syndrome. Arch Ophthalmol 1969;82:10–4.
27. Whitcher JP, Shiboski CH, Shiboski SC, Heidenreich AM, Kitagawa K, Zhang S, et al. A simplified quantitative method for assessing keratoconjunctivitis sicca from the Sjögren's Syndrome International Registry. Am J Ophthalmol 2010;149:405–15.
28. 1000Minds. URL: https://www.1000minds.com/solutions/decision-making-software.
29. 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;62:2582–91.
30. Daniels TE, Cox D, Shiboski CH, Schiodt M, Wu A, Lanfranchi H, et al. Associations between salivary gland histopathologic diagnoses and phenotypic features of Sjögren's syndrome among 1,726 registry participants. Arthritis Rheum 2011;63:2021–30.
31. Navazesh M. Methods for collecting saliva. Ann N Y Acad Sci 1993;694:72–7.
32. Navazesh M, Kumar SK, University of Southern California School of Dentistry. Measuring salivary flow: challenges and opportunities. J Am Dent Assoc 2008;139 Suppl:35S–40S.
33. R Project. URL: http://www.R-project.org.
34. 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;42:377–81.
35. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. BMC Bioinformatics 2008;9:307.
36. 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;16:73–81.
37. Sullivan LM, Massaro JM, D'Agostino RB Sr. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. Stat Med 2004;23:1631–60.

38. 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 Sjögren's syndrome. Ann Rheum Dis 2015;74:1557–61.
39. Defendenti C, Atzeni F, Spina MF, Grosso S, Cereda A, Guercilena G, et al. Clinical and laboratory aspects of Ro/SSA-52 autoantibodies. Autoimmun Rev 2011;10:150–4.
40. Dugar M, Cox S, Limaye V, Gordon TP, Roberts-Thomson PJ. Diagnostic utility of anti-Ro52 detection in systemic autoimmunity. Postgrad Med J 2010;86:79–82.
41. 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. Autoimmun Rev 2011;10:509–13.
42. 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;46:32–9.

APPENDIX A: THE INTERNATIONAL SJÖGREN'S SYNDROME CRITERIA WORKING GROUP