Arthritis Care & Research
Vol. 64, No. 4, April 2012, pp 475–487
DOI 10.1002/acr.21591

SPECIAL ARTICLE

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

S. C. SHIBOSKI,[1] C. H. SHIBOSKI,[1] L. A. CRISWELL,[1] A. N. BAER,[2] S. CHALLACOMBE,[3] H. LANFRANCHI,[4] M. SCHIØDT,[5] H. UMEHARA,[6] F. VIVINO,[7] Y. ZHAO,[8] Y. DONG,[9] D. GREENSPAN,[1] A. M. HEIDENREICH,[4] P. HELIN,[5] B. KIRKHAM,[3] K. KITAGAWA,[6] G. LARKIN,[3] M. LI,[9] T. LIETMAN,[1] J. LINDEGAARD,[10] N. McNAMARA,[1] K. SACK,[1] P. SHIRLAW,[3] S. SUGAI,[6] C. VOLLENWEIDER,[4] J. WHITCHER,[1] A. WU,[1] S. ZHANG,[9] W. ZHANG,[11] J. S. GREENSPAN,[1] AND T. E. DANIELS,[1] FOR THE SJÖGREN'S INTERNATIONAL COLLABORATIVE CLINICAL ALLIANCE (SICCA) RESEARCH GROUPS

This criteria set has been approved by the American College of Rheumatology (ACR) Board of Directors as Provisional. This signifies that the criteria set has been quantitatively validated using patient data, but it has not undergone validation based on an external data set. All ACR-approved criteria sets are expected to undergo intermittent updates.

The American College of Rheumatology is an independent, professional, medical and scientific society which does not guarantee, warrant, or endorse any commercial product or service.

Objective. **We propose new classification criteria for Sjögren's syndrome (SS), which are needed considering the emergence of biologic agents as potential treatments and their associated comorbidity. These criteria target individuals with signs/symptoms suggestive of SS.**

Methods. **Criteria are based on expert opinion elicited using the nominal group technique and analyses of data from the Sjögren's International Collaborative Clinical Alliance. Preliminary criteria validation included comparisons with classifications based on the American–European Consensus Group (AECG) criteria, a model-based "gold standard" obtained from latent class analysis (LCA) of data from a range of diagnostic tests, and a comparison with cases and controls collected from sources external to the population used for criteria development.**

Results. **Validation results indicate high levels of sensitivity and specificity for the criteria. Case definition requires at least 2 of the following 3: 1) positive serum anti-SSA and/or anti-SSB or (positive rheumatoid factor and antinuclear antibody titer ≥1:320), 2) ocular staining score ≥3, or 3) presence of focal lymphocytic sialadenitis with a focus score ≥1 focus/4 mm^2 in labial salivary gland biopsy samples. Observed agreement with the AECG criteria is high when these are applied using all objective tests. However, AECG classification based on allowable substitutions of symptoms for objective tests results in poor agreement with the proposed and LCA-derived classifications.**

Conclusion. **These classification criteria developed from registry data collected using standardized measures are based on objective tests. Validation indicates improved classification performance relative to existing alternatives, making them more suitable for application in situations where misclassification may present health risks.**

INTRODUCTION

Sjögren's syndrome (SS) is a multisystem autoimmune disease characterized by hypofunction of the salivary and lacrimal glands. It is among the group of diseases overseen by rheumatologists; however, its diagnosis and management require 3 areas of specialty practice: rheumatology, ophthalmology, and oral medicine. The multidisciplinary

Supported by the NIH (National Institute for Dental and Craniofacial Research, National Eye Institute, and Office of Research on Women's Health; contract N01-DE32636).

[1]S. C. Shiboski, PhD, C. H. Shiboski, DDS, MPH, PhD, L. A. Criswell, MD, MPH, DSc, D. Greenspan, BDS, DSc, ScD, T. Lietman, MD, N. McNamara, OD, PhD, K. Sack, MD, J. Whitcher, MD, MPH, A. Wu, DDS, J. S. Greenspan, BDS, PhD, FRCPath, T. E. Daniels, DDS, MS: University of California, San Francisco; [2]A. N. Baer, MD: Johns Hopkins University, Baltimore, Maryland; [3]S. Challacombe, BDS, PhD,

Significance & Innovations

- New classification criteria for Sjögren's syndrome are needed to better support etiologic and genetic research and therapeutic trials for this prevalent autoimmune disease.
- Criteria used for enrollment into clinical trials need to be clear, be easy to apply, and have high specificity, considering the potentially serious adverse effects and comorbidities of biologic agents.
- We propose classification criteria for Sjögren's syndrome that are developed from registry data collected using standardized instruments and diagnostic tests, and are based entirely on objective measures.

aspect of the disease represents a challenge for the definition and validation of classification criteria because there is no single gold standard test for diagnosing SS, and it is not feasible to use a single clinician's diagnosis for the case/control definition. The closest substitute is based on expert assumptions about the characteristics of SS, specifically that it: 1) is a systemic, multiorgan autoimmune disease, 2) has a chronic or progressive course, and 3) is characterized by, but not limited to, secretory dysfunction.

While there have been 11 classification or diagnostic criteria published for SS since 1965 (1–11), none have been endorsed by the American College of Rheumatology (ACR) or the European League Against Rheumatism (EULAR). The American–European Consensus Group (AECG) criteria (11) have better specificity than their predecessor (9), as they require evidence of autoimmunity from positive anti-SSA/Ro and/or anti-SSB/La serology or focal lymphocytic sialadenitis (FLS) with a focus score (FS) $\geq$1 in a labial salivary gland (LSG) biopsy sample. However, they have been criticized for including subjective tests (symptoms), physiologic measures that lack specificity, and alternate objective tests that are not diagnostically equivalent. For example, Schirmer's test may be used instead of the rose bengal ocular stain, even though they differ in sensitivity and specificity (11). Further, the inclusion of symptoms of dry mouth and/or eyes can lead to misclassification of asymptomatic patients. In addition, physiologic measures, such as unstimulated whole salivary (UWS) flow, unanesthetized Schirmer's test, and salivary scintigraphy, are useful for assessment of salivary or tear function, but lack specificity for SS.

The need for new classification criteria is clear considering the current lack of standardization inherent to the use of multiple older criteria in the field and the emergence of biologic agents as potential treatments. Considering the potentially serious adverse effects and comorbidities of these agents, criteria used for enrollment into clinical trials will need to be clear, easy to apply, and have high specificity. They also must rely upon well-established objective tests that are clearly associated with the systemic/autoimmune, oral, and ocular characteristics of the disease, and include alternate tests only when they are diagnostically equivalent. Furthermore, it is desirable for new classification criteria for SS to be endorsed by professional rheumatology organizations across the world (such as the ACR and EULAR) to increase their credibility and maximize standardization when enrolling participants into clinical trials.

The Sjögren's International Collaborative Clinical Alliance (SICCA) is funded by the National Institutes of Health (12) to develop new classification criteria for SS, better define the SS phenotype, and collect/store clinical data and biospecimens to support future research. We propose new classification criteria for SS, following the ACR guidelines (13) to the extent possible for a condition requiring multiple clinical specialties for diagnosis. Below we describe our approach to criteria development and validation.

MATERIALS AND METHODS

We used a consensus methodology derived from the nominal group technique (14) to: 1) define the target population to whom the classification criteria should apply; 2) identify the initial list of criteria components that have face validity, would be measured as part of the SICCA project, and could be considered in the classification criteria (item generation and reduction); and 3) select preliminary classification criteria. We then engaged in a series of validation exercises. Our overall approach relied on analyses of current SICCA data and consisted of 4 phases.

FDSRCS, FRCPath, FMedSci, B. Kirkham, MD, FRCP, FRACP, G. Larkin, FRCoPath, P. Shirlaw, BDS, FDRCPS: King's College London, London, UK; [4]H. Lanfranchi, DDS, PhD, A. M. Heidenreich, MD, C. Vollenweider, MD: University of Buenos Aires, Buenos Aires, Argentina; [5]M. Schiødt, DDS, P. Helin, MD, PhD: Rigshospitalet, Copenhagen, Denmark; [6]H. Umehara, MD, PhD, K. Kitagawa, MD, PhD, S. Sugai, MD: Kanazawa Medical University, Ishikawa, Japan; [7]F. Vivino, MD: Penn Presbyterian Medical Center and University of Pennsylvania, Philadelphia; [8]Y. Zhao, MD: PUMC Hospital, Beijing, China; [9]Y. Dong, MD, M. Li, MD, S. Zhang, MD: Peking Union Medical College Hospital, Beijing, China; [10]J. Lindegaard, MD, PhD: Glostrup Hospital, Glostrup, Denmark; [11]W. Zhang, MD: Chinese Academy of Medical Science and Peking Union Medical College Hospital, Beijing, China. Collaborators of the Sjögren's International Collaborative Clinical Alliance are shown in Appendix A.

Drs. S. C. Shiboski and C. H. Shiboski contributed equally to this work.

Dr. Vivino has received consultant fees and/or honoraria (less than $10,000 each) from Daiichi-Sankyo and Parion Sciences. Dr. Wu owns stock and/or stock options in Isis Pharmaceuticals and Schering-Plough. Dr. J. S. Greenspan has received consultant fees (less than $10,000) from GlaxoSmithKline.

Address correspondence to S. C. Shiboski, PhD, University of California, San Francisco, Department of Epidemiology and Biostatistics, Division of Biostatistics, 185 Berry Street, Lobby 5, Suite 5700, San Francisco, CA 94107. E-mail: steve@biostat.ucsf.edu.

Submitted for publication March 31, 2011; accepted in revised form December 20, 2011.

Phase 1: expert panel member selection and item generation. *Expert panel member selection.* We first identified panel members who were experts in the relevant clinical specialties (rheumatology, ophthalmology, or oral medicine) and constituted a heterogeneous group with respect to geographic area, sex, and seniority/level of expertise. The panel included 20 experts: 7 rheumatologists, 6 ophthalmologists, and 7 experts in oral medicine. Nine members (45%) were from the US, and the rest were from 4 countries on 3 continents. All of the panel experts practiced their specialty within a university-affiliated medical center. Sixteen (80%) were senior investigators at the professor level, 20% were at the associate professor level, and 40% were women. All of the investigators had been selected for their experience with SS within their clinical specialty and for geographic representation.

Expert panel first face-to-face meeting: review of evidence-based literature and item generation. In February 2004, the panel of experts gathered for a face-to-face meeting moderated by a statistician (SCS) and an epidemiologist (CHS). The goal of this meeting was to obtain consensus (at least 80%) on the target population to whom the classification criteria would apply, and the initial list of variables or criteria items that would be collected as part of SICCA. The meeting began with presentation of a comprehensive literature review by one of the senior investigators (TED) of the 11 previous classification and diagnostic criteria for SS that had been published in the past 40 years, none of which had been endorsed by the ACR or EULAR.

There was consensus among the panel that the criteria should apply to the population of patients who may be referred to a specialist because of signs and/or symptoms possibly suggesting SS. Recruitment strategies and eligibility criteria are described below. The rationale for selecting this target population is that a given patient would not be evaluated for SS unless she/he had signs or symptoms suggesting this diagnosis. There was also consensus that if asked to select cases and controls for validation of new classification criteria, panel members would use objective tests (e.g., specific serum measures of autoimmunity, ocular staining reflecting lacrimal hypofunction, and LSG biopsy reflecting FLS) that would likely be part of the new classification criteria, leading to circularity. Therefore, it was agreed that no diagnostic labels would be used for enrollment, and that all of the participants would undergo the same set of standardized objective tests and questionnaires capturing various signs and symptoms.

The panel agreed upon the examinations and tests used to assess ocular and oral signs and symptoms, tear and salivary function, LSG biopsy results, and various serum measures of autoimmunity. The list created was based both on published results and on the clinical experience of panel members. There was discussion among the rheumatologists regarding which extraglandular manifestations possibly associated with SS should be captured, and consensus was achieved regarding a list of signs/symptoms that would be measured through a targeted rheumatologic examination, review of systems, careful medical history, and serologic laboratory measures. Similarly, the oral medicine specialists agreed on a list of tests measuring salivary function (both stimulated parotid and UWS flow rates) and salivary gland expression of autoimmunity through biopsy of LSG, examining them for the presence of FLS, and measuring FS accordingly as described in detail elsewhere (15). The ophthalmologists agreed on tests evaluating participants for the presence of keratoconjunctivitis sicca (KCS). There was consensus that, while rose bengal had been widely used for grading conjunctival and corneal damage in patients with KCS, it is inherently toxic to epithelial cells and very painful for patients. Therefore, fluorescein was selected to grade the cornea and lissamine green was selected to grade the bulbar conjunctiva. Effectiveness for grading KCS is established for both (16). They agreed on a standardized quantitative grading system that would be easily reproducible and could be used in clinical practice in the future (17). The ocular staining score (OSS) is the sum of a 0–6 score for fluorescein staining of the cornea and a 0–3 score for lissamine green staining of both the nasal and temporal bulbar conjunctivae, yielding a total score ranging from 0–12. Alternative established tests for dryness used in prior criteria, such as tear breakup time (TBUT) and unanesthetized Schirmer's test, were also included.

The final list of criteria items that was agreed upon by the end of the first meeting included nearly all those previously reported in the relevant literature. It has been described previously (12) and is available online at http://sicca.ucsf.edu.

Phase 2: item reduction. *Expert panel second face-to-face meeting: review of preliminary SICCA data analyses.* Following 2 years of standardized data collection, including the criteria components selected in phase 1, another face-to-face panel meeting was convened in April 2006. Data analysis summaries were presented to the group by the epidemiologist and statistician who moderated the initial meeting, and the panel was divided into small specialty-specific focus groups to review evidence-based results for each clinical specialty. The goals of these exploratory analyses included understanding the relationship between variables representing the oral/salivary, ocular, and systemic features of the disease; defining cutoff values for particular tests that could be used as components of the classification criteria; and assessing the value of tests that could serve as surrogates for primary objective tests in alternate criteria sets. Methods used included frequency tables, binary regression, and classification trees. We also used area-proportional Venn diagrams to visualize the overlapping relationships of 3 variables simultaneously, each representing one of the primary disease features (18).

Phase 3: candidate SS criteria. *Expert panel third face-to-face meeting: data-driven, consensus-based selection of preliminary classification criteria for SS.* The panel met in May 2009 to decide on preliminary classification criteria. Additional analyses were presented, including longitudinal assessment of the stability of criteria components over time (based on results from scheduled 2-year followup visits that mirrored baseline assessments). Results from a statistical classification based on latent class analysis (LCA) (19) were presented (approach further described

below). The data presented to the panel represented a subset of participants (n = 1,107) consecutively enrolled as of April 1, 2009.

Following presentation of results by the statistician, a discussion among panel members was moderated by the epidemiologist and statistician. The goal of the discussion was for members to select, based on their understanding of the data analyses presented and on their clinical experience, which objective test(s) they believed was/were the most specific for SS within their own specialty. Furthermore, various preliminary classification criteria were discussed, and panel members were asked to select which criteria they believed would best classify patients with SS. Following this third face-to-face meeting, a report summarizing the data analyses was circulated among the panel members and a questionnaire was distributed by e-mail. The questionnaire was designed to assess consensus among the expert panel members regarding the preliminary classification criteria. It also measured the level of consensus within each specialty regarding which criteria component was thought to have the highest level of face validity within that specialty. The response to each question was on a scale of 1–5, with 1 indicating strong disagreement and 5 indicating strong agreement.

SICCA registry cohort. The participants in the SICCA cohort have been enrolled since 2004 in 5 collaborating academically-based research groups, located in Argentina, China, Denmark, Japan, and the US, and directed from the University of California, San Francisco (12) (Table 1). Subsequently, additional research groups joined the SICCA project: in 2007 from the UK and in 2009 from India and 2 additional sites in the US.

To be eligible for the SICCA registry, participants must be at least 21 years of age and have at least 1 of the following: symptoms of dry eyes or dry mouth; a previous suspicion or diagnosis of SS; elevated serum antinuclear antibody (ANA) titer, positive rheumatoid factor (RF), or anti-SSA and/or anti-SSB; bilateral parotid enlargement in a clinical setting of SS; a recent increase in dental caries; or diagnoses of rheumatoid arthritis (RA) or systemic lupus erythematosus (SLE) and any of the above. The rationale for these eligibility criteria is that only patients with such characteristics would be evaluated for SS or considered for enrollment in a clinical trial designed to evaluate a potential therapeutic agent for SS. Therefore, our classification criteria target individuals with signs and symptoms that may be suggestive of SS, not the general population.

Participants are recruited through local or national SS patient support groups, health care providers, public media, and populations served by all 9 SICCA research groups. Exclusion criteria include known diagnoses of hepatitis C, human immunodeficiency virus, sarcoidosis, amyloidosis, active tuberculosis, graft versus host disease, or autoimmune connective tissue diseases other than RA or SLE; past head and neck radiation treatment; current treatment with daily eye drops for glaucoma; corneal surgery in the last 5 years to correct vision; cosmetic eyelid surgery in the last 5 years; or a physical or mental condition interfering with successful participation in the study. Contact lens wearers are asked to discontinue wear for 7 days before the SICCA examination. We do not exclude participants taking prescription drugs that may affect salivary or lacrimal secretion, but record their use and all other medications currently taken.

Phase 4: criteria validation. In the absence of a gold standard diagnostic test for SS, conventional methods of validation based on direct estimation of quantities such as sensitivity and specificity are not directly applicable. The practice of defining a gold standard based on a series of cases and controls identified by expert clinicians is also not practical for SS because diagnosis must rely on 3 clinical specialties. Further, since such diagnoses rely on the same tests that form the basis of the proposed criteria, estimates of sensitivity and specificity will be inherently biased.

Acknowledging these difficulties, we based the initial evaluation and validation of preliminary criteria primarily on a data-based approach, including: 1) comparison to alternate versions of the preliminary criteria, each defined by substituting alternate tests for those used in the criteria definition; 2) comparison to a "gold standard diagnosis" derived from a statistical model fitted to data from a range of diagnostic tests; 3) comparison to versions of the AECG criteria defined using results of selected component diagnostic tests; and 4) assessment of the stability of classifications produced by the proposed criteria, both in participants distinct from those used for criteria development and over a 2-year period within a subsample of participants for whom scheduled followup data were available.

Results reported here are based on complete participant data on key diagnostic features from 6 SICCA sites collected through March 2010. In addition, for external validation we utilized data from approximately 300 participants not included in the data set used for criteria development and representing 10 months of additional recruitment. We excluded participants with RA, SLE, scleroderma, or other connective tissue diseases from these analyses since there were only 87 such participants (6%). Therefore, the proposed classification criteria apply to a target population of individuals who do not have SLE, RA, or other connective tissue diseases. The methodologic approaches for each of the steps are outlined below.

Validation through alternate criteria sets. We considered alternate sets of criteria, based on substituting simpler and/or less invasive tests for the preliminary criteria. These included: 1) substitution of UWS flow rate for the LSG biopsy with FLS and FS $\geq$1 focus/4 mm^2; 2) substitutions of TBUT $<$10 seconds or unanesthetized Schirmer's test $\leq$5 mm/5 minutes for an OSS $\geq$3; and 3) positive RF, ANA titer $\geq$1:320, positive serum anti-SSA and/or anti-SSB, and each of the 3 used individually to represent the serologic component of the disease. Performance was assessed via sensitivity and specificity estimated by taking the preliminary criteria as a "gold standard," and summarizing the results with exact binomial 95% confidence intervals (95% CIs). The results of these analyses were used to evaluate possible effects of such substitutions on classification performance of the preliminary criteria.

Table 1. Demographic and SS-related phenotypic characteristics among 1,618 participants enrolled in the Sjögren's International Collaborative Clinical Alliance registry as of March 8, 2010*

	Value
Sources of complete baseline enrollments	
Argentina	280 (17)
China	236 (15)
Denmark	318 (20)
Japan	249 (15)
UK (since May 2007)	109 (7)
US	426 (26)
Women	1,493 (93)
SS-related characteristics	
Symptoms of:	
Dry mouth	1,459 (90)
Dry eyes	1,366 (85)
Both dry mouth and eyes	1,274 (79)
Positive serum	
Anti-SSA/Ro	573 (37)
Anti-SSB/La	368 (24)
Anti-SSA and SSB	354 (23)
Rheumatoid factor	590 (38)
ANA titer ≥1:40	1,054 (67)
ANA titer ≥1:320	632 (39)
Anti–hepatitis C	18 (1)
Hypergammaglobulinemia (IgG >1,445 mg/dl)	575 (37)
Unanesthetized Schirmer's test ≤5 mm/5 minutes	509 (32)
Tear breakup time <10 seconds	1,318 (82)
Ocular staining score ≥3 (max of left and right)†	1,166 (72)
UWS flow <0.1 ml/minute	898 (56)
Parotid enlargement	
Unilateral	239 (15)
Bilateral	171 (11)
Labial salivary gland biopsy diagnosis results‡	
Nonspecific/sclerosing chronic sialadenitis	578 (36)
Granulomatous inflammation/within normal limits	19 (1)
Mucosa-associated lymphoid tissue lymphoma	1 (0)
Inadequate specimen	28 (2)
Focal lymphocytic sialadenitis (FS assessable on n = 962)	992 (62)
FS >1	636 (66)
FS 1	30 (3)
FS <1	296 (31)
Continuous variables	
Age, median (range) years	54 (21–90)
UWS flow rate, median (25th, 75th percentiles) ml/minute	0.08 (0.03, 0.18)
Stimulated parotid flow, median (25th, 75th percentiles) ml/minute	0.12 (0.03, 0.26)

* Values are the number (percentage) unless otherwise indicated. Denominators may vary due to missing observations (<3%) for some variables. SS = Sjögren's syndrome; ANA = antinuclear antibody; UWS = unstimulated whole salivary; FS = focus score.
† Ocular staining score is assessed by fluorescein staining of the cornea and lissamine green staining of the interpalpebral conjunctivae and scored by a system in which ≥3 represents the presence of keratoconjunctivitis sicca. Details and results of these examinations are published elsewhere (17).
‡ Details of the histopathologic examination and further analyses of the labial salivary gland biopsy specimens are published elsewhere (15).

Model-based validation. An assessment of plausible levels of sensitivity and specificity of the preliminary and alternate criteria was provided using LCA (19). LCA provides a model-based clustering of individuals into a specified number of "disease" classes based on the observed patterns of a series of binary predictor variables representing the presence or absence of important diagnostic features. The resulting classes can then be related to disease status based on the class-specific patterns of diagnostic features. Because LCA methods rely on the restrictive assumption that component test results are independent conditional on the true classification (20), we applied a variant of LCA that relaxes this assumption (21). We fitted a series of models, presuming as few as 1 and as many as 4 disease classes. Models were based on 10 predictor variables encompassing the major ocular, oral/salivary, and systemic features of the disease. We also used a standard multivariate clustering procedure known as K-means (22). We used the results of LCA and clustering analyses to provide alternate model-based disease classifications to which the preliminary criteria (and alternate versions of these criteria) could be compared. This allowed estimation

of sensitivity and specificity using the model-based classification as a "gold standard" (23,24). In the absence of knowledge of the true disease classification, the accuracy of these estimates cannot be assessed. However, comparison of results between alternate versions of criteria allows an assessment of a plausible range of sensitivity and specificity of the proposed criteria. Further, consistency of results between alternate methods of deriving model-based standards helps establish stability of conclusions and reveals possible dependence of conclusions on assumptions inherent in the models.

Supplementary analyses included use of random forest classification as a means of assessing the importance of individual tests in predicting the model-based gold standard. The random forest approach is a generalization of standard classification trees (25). It is applied to a collection of predictor variables measured on individuals with known outcome classification to build a nonparametric classification rule that predicts the outcome as accurately as possible. One of the outputs of this analysis is a variable importance ranking for the predictors. We applied this approach to the classification produced by the latent class model, using the same predictor variables as inputs.

Validation against AECG criteria. To investigate how preliminary criteria compare to previous criteria, we classified participants in the validation sample using both the preliminary criteria and the AECG criteria (11). The 2002 AECG criteria for SS (11) are a modification of the 1993 European criteria (9) based on reanalysis of 180 cases selected from the original data set. It applies 6 types of clinical signs or tests: ocular symptoms, oral symptoms, ocular signs (Schirmer's test or ocular staining), histopathology (FLS in minor salivary glands), salivary involvement (reduced UWS flow, parotid sialography, or salivary scintigraphy), and serum autoantibodies (anti-SSA, anti-SSB, or both). Primary SS is indicated in the presence of any 4 of these 6 as long as either the histopathology or serology is present and none of 7 exclusions are present. The AECG criteria were defined for SICCA participants using the specified oral/salivary, ocular, and systemic components; substituting the SICCA OSS for rose bengal staining; and using a definition of participant-reported ocular and oral symptoms based on questions most closely matching the corresponding questions used in the AECG criteria. Because of the flexibility inherent in the definition of the AECG criteria, we considered alternate classifications using: 1) all available tests, 2) restricting the ocular test to be based on Schirmer's test only, 3) restricting the oral/salivary test to be based on the UWS only, and 4) restricting both the salivary and ocular tests to be the UWS and Schirmer's test, respectively. Consistency of the 2 approaches is summarized by estimated proportions of agreement and disagreement with 95% CIs, and using the kappa statistic.

Validation in an external set of "cases" and "controls." Validation of proposed criteria with a set of expert-defined disease cases and disease-free controls collected from sources external to the study population used for criteria development is a common means of validating classification criteria. Despite the potential for circularity in the expert assessments arising from use of the diagnostic variables comprising the proposed criteria, this type of validation can potentially yield complementary information to the other approaches just described. To provide a preliminary assessment of this type, we obtained a series of disease cases from 2 sites recently added to the registry. The directors of the Johns Hopkins University (JHU) site (ANB) and of the University of Pennsylvania site (FV) were asked to identify patients that they (or their rheumatology faculty practice colleagues) had diagnosed as having SS, using standard clinic procedures, prior to entry into the SICCA registry. We could not use clinical diagnosis to identify controls, since only people with suggested signs/symptoms of SS are referred to the SICCA registry (as in real clinical practice, only those with suggested signs/symptoms of SS would receive an evaluation to confirm/rule out the disease). Therefore, controls were selected among the participants observed to be negative according to the AECG criteria (described above) and recruited subsequent to the final date for inclusion in the sample of participants considered for criteria development. We compared the case/control classification to that obtained using the preliminary criteria, taking the former as the "gold standard" for the purpose of estimating sensitivity and specificity.

Testing criteria stability over time. To examine temporal stability of the preliminary classification criteria, we compared individual classifications made using test results from enrollment visits with classifications made on 2-year followup visits. Results are summarized by estimated proportions of agreement and disagreement with 95% CIs.

RESULTS

A total of 1,618 participants were enrolled in the SICCA registry as of March 8, 2010. A summary of the demographic characteristics and phenotypic features of SS is shown in Table 1.

Item generation and reduction. *Symptoms of dry mouth or dry eyes.* High proportions of participants in the SICCA registry experienced symptoms of dry mouth, dry eyes, or both (Table 1). However, as reported previously, dry eye/mouth symptoms did not show a statistically significant association with the presence of FLS, serum anti-SSA and/or anti-SSB, or an OSS $\geq$3 (12,26). Numbers of participants not experiencing dry eyes, dry mouth, or either were 247, 154, and 62, respectively. While these individuals represent no more than 15% of the cohort, 39–49% of these asymptomatic patients had positive anti-SSA and/or anti-SSB, 36–41% had LSG biopsies with FLS and FS $\geq$1, and 63–73% had an OSS $\geq$3.

Interrelationship of various phenotypic features of SS. Analyses investigating the associations between phenotypic features of SS found that the odds of having positive anti-SSA and/or anti-SSB serology was 12 times higher among those with an FS $\geq$1 than among those with an FS $<$1 or without FLS (95% CI 9.3–15.5). Those with FLS and FS $\geq$1 were 4 times more likely to have an OSS $\geq$3 than those with an FS $<$1 or without FLS (95% CI 3.1–5.3). The

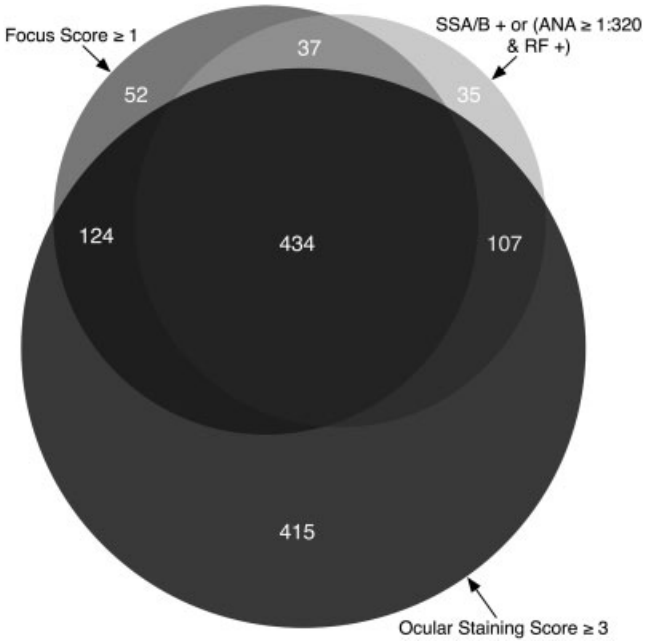

Figure 1. Area-proportional Venn diagram visualizing the interrelationships between abnormal ocular staining score, labial salivary gland focus scores ≥1, and positive anti-SSA and/or anti-SSB antibodies (SSA/B) or (positive rheumatoid factor [RF+] and antinuclear antibody [ANA] titer ≥1:320). The diagram is based on 1,507 individuals with complete data on the variables (3 objective tests) represented. The 303 individuals not included in the shaded regions did not possess any of the 3 defining characteristics.

association between an OSS ≥3 and positive anti-SSA and/or anti-SSB serology was also strong (odds ratio [OR] 4.8, 95% CI 3.6–6.4), but much less so than the association between FLS with FS ≥1 and positive anti-SSA and/or anti-SSB serology. The relationships among these 3 measures, as depicted by Figure 1, defined a large group of participants who had KCS without other components of SS (KCS only) representing a clinical entity distinct from the KCS associated with SS (17).

Extraglandular manifestations. Diagnostic confirmation of participants' histories of thyroid, liver, or kidney diseases or lymphoma was sought from the diagnosing physician and obtained for 78% of those reported histories. The prevalence of confirmed thyroid, liver, and kidney extraglandular manifestations included 18 diagnoses of Graves' disease, 43 diagnoses of Hashimoto thyroiditis, 15 diagnoses of primary biliary cirrhosis, 3 diagnoses of renal tubular acidosis, 2 diagnoses of interstitial nephritis, and 5 diagnoses of lymphoma. We found strong associations between phenotypic features of SS and serologic characteristics of autoimmunity. For example, participants with FLS and FS ≥1 (compared to those with FS <1 or without FLS) were 9 times more likely to be RF positive (95% CI 7.0–11.4), to have higher ANA titers (≥1:320; 95% CI 7–11.3), were 14 times more likely to have hypergammaglobulinemia with IgG >2,013 mg/dl (95% CI 9.3–21.1), and 2.4 times more likely to have hypocomplementemia with C4 <16 mg/dl (95% CI 1.7–3.3). Similarly, participants with an OSS ≥3 (compared to those with an OSS <3) were 4 times more likely to be RF positive (95% CI 3.0–5.3), 5 times more likely to have an ANA titer ≥1:320 (95% CI 3.5–6), 7 times more likely to have hypergammaglobulinemia (95% CI 4–11.3), and twice as likely to have hypocomplementemia (95% CI 1.2–2.7). A detailed description of extraglandular manifestations is published elsewhere (27).

Proposed preliminary classification criteria for SS and level of consensus of expert panel members: candidate SS criteria. As part of phases 2 and 3 of our consensus methodology, earlier versions of the analyses summarized above were presented and discussed, in addition to classification tree analyses and various iterations of the Venn diagram shown in Figure 1. These demonstrated the strong interrelationship between the 3 main serologic, ocular, and oral/salivary phenotypic features of SS measured by objective tests (anti-SSA and/or anti-SSB–positive serology, FLS with FS ≥1, and an OSS ≥3). The rheumatologists discussed potential roles for RF and ANA titer in the absence of anti-SSA and/or anti-SSB–positive serology in the classification criteria. The consensus was that a positive RF or ANA titer in the absence of positive anti-SSA and/or anti-SSB would be too nonspecific. However, there was strong support for substituting both positive RF and high ANA titers in the absence of anti-SSA and/or anti-SSB as a way to capture participants who have negative anti-SSA and/or anti-SSB serology, but a strong expression of autoimmunity from these 2 other tests. The relative importance of each of the 3 main serologic, ocular, and oral/salivary phenotypic features of SS measured by objective tests was discussed within each subgroup of the panel (rheumatology, ophthalmology, and oral medicine). Furthermore, various combinations of the 3 main phenotypic features of SS measured by objective tests, such as at least 1 of 3, 2 of 3, or 3 of 3, were discussed by the entire panel.

Results from a questionnaire administered following phase 3 revealed high consensus among each of the clinical specialties. More specifically, 6 (86%) of the 7 rheumatologists either agreed or strongly agreed that positive anti-SSA and/or anti-SSB serology represented the most specific serologic marker of SS, and 86% also thought that positive RF and ANA titer ≥1:320 represented a satisfactory substitute for a negative anti-SSA and/or anti-SSB serology. All 6 ophthalmologists either agreed or strongly agreed that an OSS ≥3 (using lissamine green and fluorescein) represented the most specific way to diagnose the ocular component of SS. Only 1 ophthalmologist (17%) agreed that TBUT represented the next best substitute, and none agreed with the use of Schirmer's test as a specific measure of the ocular component of SS. All 7 oral medicine specialists strongly agreed that the presence of FLS in an LSG biopsy sample with an FS ≥1 was the most specific test to determine the presence of the salivary component of SS. There was also 100% consensus in that group that neither UWS nor stimulated parotid flow rate would represent specific measures of the salivary component of SS. Among the entire panel, 86% agreed or strongly agreed that the preliminary criteria for SS should be at least 2 of 3 of the following objective tests: 1) positive serum anti-SSA and/or anti-SSB or (positive RF and ANA titer ≥1:320), 2) OSS ≥3 (using lissamine green and fluorescein) to

Table 2. Sensitivity and specificity of alternate classification criteria sets (each 2 of 3) compared to SICCA preliminary criteria*

Alternate sets	Sensitivity (95% CI), %	Specificity (95% CI), %
UWS flow rate <0.1 ml/minute replacing [FLS with FS ≥1]	89.8 (87.2–92.0)	74.3 (71.0–77.5)
TBUT <10 seconds replacing OSS ≥3†	94.8 (92.8–96.4)	94.4 (92.4–96.0)
Schirmer's test ≤5 mm/5 minutes replacing OSS ≥3†	74.8 (71.3–78.1)	98.9 (97.8–99.5)

* Sjögren's International Collaborative Clinical Alliance (SICCA) preliminary criteria defined as at least 2 of 3 of the following 3 objective tests: labial salivary gland with FLS and FS ≥1 focus/4 mm^2; OSS ≥3; or positive anti-SSA and/or SSB serology or (positive rheumatoid factor and antinuclear antibody titer ≥1:320). 95% CI = 95% confidence interval; UWS = unstimulated whole salivary; FLS = focal lymphocytic sialadenitis (in labial salivary gland biopsy); FS = focus score; TBUT = tear breakup time; OSS = ocular staining score.
† The OSS is the sum of a 0–6 score for fluorescein staining of the cornea and a 0–3 score for lissamine green staining of the conjunctiva (17).

diagnose KCS (17), and 3) presence of FLS in an LSG biopsy sample with an FS ≥1 (15). Thirteen panel members (62%) either disagreed or strongly disagreed, while 4 (19%) agreed that the preliminary criteria for SS should be 3 of 3 of the above objective tests. There was 100% consensus that preliminary classification criteria for SS could not be limited to only 1 of the 3 objective tests.

Exclusion criteria included those initially defined in the Methods. It was also agreed that IgG4-related disease would be among the exclusion criteria. IgG4-related disease is a relatively new clinical entity characterized by increased serum IgG4 (>135 mg/dl) and marked infiltration of IgG4-positive plasma cells in various organs, especially the pancreas (so-called autoimmune pancreatitis) and lacrimal, submandibular, and parotid glands (28).

Criteria validation. *Alternate criteria in relation to preliminary classification criteria.* For consistency, results for this and subsequent validation analyses are based on a subset of 1,362 participants with complete data for 10 individual tests listed in the section on validation using LCA (i.e., participants with pending test results due to batch shipping of specimens from international sites were not included in these analyses). These tests were selected based on our preliminary analyses to represent the range of oral/salivary, ocular, and systemic features that characterize the disease, and also because they encompass characteristics used in previously developed criteria. The cases and controls defined according to the preliminary criteria were first used to explore possible sensitivity and specificity of alternate sets of criteria, each defined by substituting one component with an alternate test (Table 2).

Based on preliminary analyses, UWS flow was the only alternate oral/salivary measure considered that demonstrated a strong association with other objective tests such as positive anti-SSA and/or anti-SSB serology or FS ≥1. Classification based on substituting this for the LSG biopsy had a sensitivity of 89.8% (95% CI 87.2–92.0%), but a low specificity of 74.3% (95% CI 71.0–77.5%). Stimulated parotid flow rate was found to have a high number of missing observations (mostly because of technical difficulty encountered by examiners across multiple sites). We therefore did not include this variable in our analyses. Alternate classifications obtained by substituting TBUT or Schirmer's test for the OSS yielded high sensitivity and specificity (94.8% and 94.4%, respectively) for the former, and low sensitivity (74.8%) for the latter.

Estimates of sensitivity and specificity for specific tests based on LCA. A series of LCA models was fitted to the results of 10 diagnostic variables representing a wide range of ocular, oral, and systemic features of the disease for the 1,362 participants in the validation sample. Results indicated that a model with 2 latent classes fit adequately, with no significant improvement observed with the addition of a third class. Assignment of the disease "case" and "control" status was based on examination of observed patterns of results from the 10 component tests used as predictors in model fitting. Cases had clearly higher observed prevalence of positive results for the majority of these tests. Table 3 lists the estimated sensitivity and specificity val-

Table 3. Test-specific estimates of sensitivity and specificity in a latent class analysis (18) model*

Test	Sensitivity (95% CI), %	Specificity (95% CI), %
FLS with FS ≥1	83.5 (79.1–88.2)	82.3 (78.1–85.8)
UWS flow rate <0.1 ml/minute	64.6 (59.8–68.9)	49.7 (46.1–53.6)
Symptom of dry mouth	87.0 (83.7–90.1)	6.6 (4.8–8.6)
OSS ≥3	89.7 (86.4–92.7)	37.8 (34.2–41.2)
TBUT <10 seconds	90.5 (87.9–93.0)	21.4 (17.9–24.3)
Schirmer's test ≤5 mm/5 minutes	42.7 (37.8–47.6)	75.1 (71.7–78.2)
Symptom of dry eyes	80.0 (76.6–84.0)	11.9 (9.6–14.6)
Positive anti-SSA and/or anti-SSB serology	83.7 (78.0–89.3)	91.5 (87.8–94.9)
Positive rheumatoid factor	72.3 (67.7–77.6)	86.4 (83.2–89.9)
ANA titer ≥1:320	72.8 (67.5–77.7)	80.4 (76.9–84.0)

* 95% CI = 95% confidence interval; FLS = focal lymphocytic sialadenitis (in labial salivary gland biopsy); FS = focus score; UWS = unstimulated whole salivary; OSS = ocular staining score; TBUT = tear breakup time; ANA = antinuclear antibody.
† The OSS is the sum of a 0–6 score for fluorescein staining of the cornea and a 0–3 score for lissamine green staining of the conjunctiva (17).

Table 4. Sensitivity and specificity of alternate classification criteria sets (each 2 of 3) compared to latent class analysis classification*

Alternate sets	Sensitivity (95% CI), %	Specificity (95% CI), %
UWS flow rate <0.1 ml/minute replacing FS ≥1	91.0 (88.2–93.3)	65.7 (62.4–69.0)
TBUT <10 seconds replacing OSS ≥3	95.7 (93.6–97.2)	81.9 (79.1–84.5)
Schirmer's test ≤5 mm/5 minutes replacing OSS ≥3	83.3 (79.9–86.4)	93.6 (91.7–95.2)
SICCA preliminary classification criteria†	96.3 (94.3–97.7)	83.0 (80.3–85.5)

* 95% CI = 95% confidence interval; UWS = unstimulated whole salivary; FS = focus score of focal lymphocytic sialadenitis in labial salivary gland biopsy (15); TBUT = tear breakup time; OSS = ocular staining score (17).
† Sjögren's International Collaborative Clinical Alliance (SICCA) preliminary criteria defined as at least 2 of 3 of the following 3 objective tests: labial salivary gland with focal lymphocytic sialadenitis and FS ≥1 focus/4 mm^2; OSS ≥3; or positive anti-SSA and/or anti-SSB serology or (positive rheumatoid factor and antinuclear antibody titer ≥1:320).

ues from 10 component tests used for fitting the random-effects LCA. These estimates provide an indication of the importance of individual test results in predicting the overall disease classification provided by the model. Results indicate that an FS of at least 1 and positive serology for anti-SSA and/or anti-SSB provide the best overall combinations of sensitivity and specificity. The OSS with a cutoff of 3 yielded relatively high sensitivity but low specificity, and the alternate measure based on TBUT performed similarly. Indicators of the presence of ocular or oral symptoms were very nonspecific, whereas alternate systemic measures based on ANA titer and RF performed similarly to anti-SSA and/or anti-SSB, with somewhat lower values for sensitivity and specificity. A companion analysis using random forest classification (25) ranked the component tests in the following order of importance in determining the LCA results: FLS with FS ≥1, positive serum anti-SSA and/or anti-SSB, ANA titer ≥1:320, positive RF, OSS ≥3, UWS <0.1 ml/minute, Schirmer's test ≤5 mm/5 minutes, TBUT <10 seconds, symptoms of dry mouth, and symptoms of dry eyes. Restricting component tests to exclude symptoms had no discernible effect on results of the LCA.

Table 4 shows the estimated sensitivity and specificity values for the alternate sets listed in Table 2 and the preliminary criteria, compared to LCA classification. Results indicate that the preliminary criteria provide the best overall levels of both sensitivity (96.3%; 95% CI 94.3–97.7%) and specificity (83%; 95% CI 80.3–85.5%) relative to alternate sets. Alternate model-based classification approaches, including conventional LCA and K-means clustering, yielded similar estimates (not shown) to those displayed in Table 4.

Comparison of results obtained with AECG criteria. In Table 5 we compare 4 versions of the AECG classification against the preliminary SICCA criteria, taking the latter as the "gold standard." In the case where all diagnostic tests are available, classification by the AECG depends more heavily on the results of the LSG and anti-SSA and/or anti-SSB status. In this situation, we would expect results comparable to the preliminary SICCA criteria. This is confirmed by the results for sensitivity, specificity, and overall agreement (as measured by the kappa statistic) in Table 5. The level of agreement decreases for alternate versions of the AECG criteria defined by substituting alternate tests for the ocular and oral/salivary components of the disease. As noted previously, requiring the presence of dry eye/mouth symptoms will exclude some asymptomatic patients. The estimated sensitivity and specificity for the full AECG criteria for predicting the "gold standard" classification based on LCA were 88.6% (95% CI 85.6–91.1%) and 81.8% (95% CI 79.2–84.4%), respectively (Table 6). These results indicate somewhat lower sensitivity than the preliminary criteria, but similar specificity. Analogous results for alternate AECG criteria showed overall less agreement.

Since AECG criteria were published in 2002, they are likely the most commonly used in practice. As a result, expert clinician selection of SS cases and controls would

Table 5. Sensitivity and specificity of alternate AECG classification criteria sets compared to preliminary classification*

Alternate sets	Sensitivity (95% CI), %	Specificity (95% CI), %	Kappa†
AECG‡	92.1 (89.8–94.0)	95.5 (93.8–96.9)	0.88
AECG: Schirmer's test§	79.2 (75.8–82.2)	95.5 (93.8–96.9)	0.75
AECG: UWS flow¶	66.4 (62.6–70.0)	98.9 (97.3–99.2)	0.66
AECG: UWS flow and Schirmer's test#	52.4 (48.5–56.2)	98.4 (97.3–99.2)	0.52

* Sjögren's International Collaborative Clinical Alliance preliminary criteria defined as at least 2 of 3 of focus score ≥1; ocular staining score ≥3; or positive anti-SSA and/or SSB serology or (positive rheumatoid factor and antinuclear antibody titer ≥1:320). AECG = American–European Consensus Group; 95% CI = 95% confidence interval; UWS = unstimulated whole salivary.
† Kappa measure of agreement with preliminary classification.
‡ AECG defined using all available tests.
§ AECG defined using only Schirmer's test ≤5 mm/5 minutes to represent the ocular component.
¶ AECG defined using only UWS flow to represent the oral/salivary component.
# AECG defined using UWS flow and Schirmer's test ≤5 mm/5 minutes to represent the oral/salivary and ocular components.

Table 6. Sensitivity and specificity of preliminary and alternative AECG classification criteria sets compared to LCA classification*

Alternate sets	Sensitivity (95% CI), %	Specificity (95% CI), %	Kappa†
Preliminary criteria‡	96.3 (94.3–97.7)	83.0 (80.3–85.5)	0.76
AECG§	88.6 (85.6–91.1)	81.8 (79.2–84.4)	0.68
AECG: Schirmer's test¶	79.2 (75.5–82.5)	85.8 (83.3–88.0)	0.65
AECG: UWS flow#	74.4 (70.5–78.0)	94.9 (93.3–96.3)	0.72
AECG: UWS flow and Schirmer's test**	59.1 (54.8–63.3)	96.0 (94.5–97.2)	0.59

* AECG = American–European Consensus Group; LCA = latent class analysis; 95% CI = 95% confidence interval; UWS = unstimulated whole salivary.
† Kappa measure of agreement with LCA classification.
‡ Sjögren's International Collaborative Clinical Alliance preliminary criteria defined as at least 2 of 3 of focus score ≥1; ocular staining score ≥3; or positive anti-SSA and/or SSB serology or (positive RF and antinuclear antibody titer ≥320).
§ AECG defined using all available tests.
¶ AECG defined using only Schirmer's test ≤5 mm/5 minutes to represent the ocular component.
# AECG defined using only UWS flow to represent the oral/salivary component.
** AECG defined using UWS flow and Schirmer's test ≤5 mm/5 minutes to represent the oral/salivary and ocular components.

almost certainly involve their use. Therefore, we also explored the sensitivity and specificity of the SICCA preliminary criteria as compared to the AECG criteria used as the "gold standard." When the AECG criteria were applied using all available tests, the sensitivity and specificity of SICCA preliminary criteria were high, at 94.7% (95% CI 92.6–96.3%) and 93.3% (95% CI 91.3–95.0%), respectively.

Validation of preliminary criteria using an external set of cases and controls. When using an external data set obtained from 2 sites recently added to the registry whose participants were not included in the data set used to develop the preliminary classification criteria, we identified 40 participants who had been diagnosed as having SS by a JHU or University of Pennsylvania rheumatologist prior to, and independently of, entry into the SICCA registry. These clinical diagnoses were made by university-based rheumatologists (mainly ANB and FV) with expertise in SS prior to their involvement with SICCA. We also identified 263 controls defined as such, as they did not satisfy the AECG criteria as described in the Methods. In this external data set of 303 participants, we found the SICCA classification criteria to have a sensitivity of 92.5% (95% CI 80–98.4%) and a specificity of 95.4% (95% CI 92.2–97.6%).

Evaluation of criteria stability within the registry. To investigate the stability of the preliminary criteria over time, we classified 236 participants who had completed 2-year followup visits at both enrollment and followup. Results were concordant in 92% of participants. Among the 8% with discordant results (20 participants), 12 (60%) showed signs of progression from a disease-free classification at enrollment to classification as diseased at followup. The remaining 8 (40%) exhibited the reverse pattern. Among these, 2 reported taking a corticosteroid medication and 1 reported taking a tumor necrosis factor α inhibitor at baseline. However, none of the 8 participants was taking these or any other immunomodulating medication at the 2-year recall visit. These results indicate the general stability of disease status over a 2-year period. Additional analyses based on comparing the above validation results in participants recruited prior to September 8, 2009, with those recruited between September 8, 2009, and March 8,

Table 7. Proposed classification criteria for SS*

The classification of SS, which applies to individuals with signs/symptoms that may be suggestive of SS, will be met in patients who have at least 2 of the following 3 objective features:

1. Positive serum anti-SSA/Ro and/or anti-SSB/La or (positive rheumatoid factor and ANA titer ≥1:320)
2. Labial salivary gland biopsy exhibiting focal lymphocytic sialadenitis with a focus score ≥1 focus/4 mm^2†
3. Keratoconjunctivitis sicca with ocular staining score ≥3 (assuming that individual is not currently using daily eye drops for glaucoma and has not had corneal surgery or cosmetic eyelid surgery in the last 5 years)‡

Prior diagnosis of any of the following conditions would exclude participation in SS studies or therapeutic trials because of overlapping clinical features or interference with criteria tests:

History of head and neck radiation treatment
Hepatitis C infection
Acquired immunodeficiency syndrome
Sarcoidosis
Amyloidosis
Graft versus host disease
IgG4-related disease

* We excluded participants with rheumatoid arthritis, systemic lupus erythematosus, scleroderma, or other connective tissue disease from these analyses since there were only 87 (6%) such participants. SS = Sjögren's syndrome; ANA = antinuclear antibody.
† Using histopathologic definitions and focus score assessment methods as previously described (15).
‡ Using ocular staining score as previously described (17).

2010, yielded remarkably similar results for all comparisons and are not reproduced here.

Proposed classification criteria for SS. Based on the validation results, we propose the classification criteria shown in Table 7.

DISCUSSION

The SICCA registry represents a unique resource for establishing universally acceptable classification criteria based on: 1) the large size and international nature of the cohort with accordingly diverse ethnic backgrounds; 2) the international and multidisciplinary team of experts, including the 3 clinical specialties involved in the management of SS, epidemiologists, and statisticians; and 3) the standardized data collection procedures combining questionnaires, clinical examinations, and specimen collections performed by calibrated investigators. Using a consensus methodology derived from the nominal group technique among 20 experts and analyses involving 1,362 participants with complete data on 10 individual tests, we first developed preliminary classification criteria for SS. These relied on a combination of objective tests that assess the 3 main components of SS (serologic, ocular, and salivary). Accompanying analyses showed that symptoms of dry mouth and dry eyes had poor specificity due to a lack of association with objective phenotypic features. We found strong associations between the main objective phenotypic features, in particular FLS with FS $\geq$1 and positive anti-SSA and/or anti-SSB. ORs measuring the association between an OSS $\geq$3 and each of these 2 features were less than one-half the magnitude of the corresponding ORs measuring the association between them. A proportional Venn diagram (Figure 1) provides a good illustration of the lower specificity of OSS in relation to the salivary and serologic components.

Using a data-driven, consensus-based approach, we defined preliminary criteria for SS as at least 2 of 3 objective tests. We then performed a series of validation analyses. The first assessed sensitivity and specificity of alternate sets of criteria, each defined by substituting 1 item with an alternate test. The inclusion of alternate tests is important because classification criteria should be applicable in a wide variety of settings and some tests may not be available in certain settings, and some tests like the LSG biopsy may be perceived as invasive and cannot easily be performed by clinicians from specialties outside of oral medicine/surgery. However, the results did not identify any suitable alternate tests for the salivary and ocular phenotypic features of SS. While UWS $<$0.1 ml/minute had good sensitivity, it had low specificity compared to the LSG biopsy to measure FLS with FS $\geq$1. It also had both low sensitivity and specificity in comparison to model-based LCA validation results. While TBUT $<$10 seconds was found to have high sensitivity and specificity when substituted for an OSS $\geq$3, it was found to have very low specificity in LCA results. Although the specificity of an OSS $\geq$3 was also low in the LCA comparison, it was almost twice as high as the TBUT. Finally, the ophthalmologists in the panel all agreed that TBUT is decreased in many diseases with tear surface abnormality, thus supporting a lack of specificity for SS. Furthermore, it also requires the use of fluorescein and a slit lamp; therefore, it is not thought to be easier to administer than the OSS. With respect to serologic tests, positive anti-SSA and/or anti-SSB had the highest sensitivity and specificity based on the LCA comparison. Positive RF and ANA titer $\geq$1:320 had reasonable specificity but lower sensitivity, suggesting that either test alone would not be a good substitute for anti-SSA and/or anti-SSB. Although we did not identify any suitable alternate tests for the salivary and ocular phenotypic features of SS to be used in our proposed classification criteria for SS, UWS $<$0.1 ml/minute and TBUT $<$10 seconds may be suitable alternatives for diagnostic criteria. While classification criteria need to be stringent to prevent any misclassification because they are used to select participants for entry into clinical trials, diagnostic criteria that are used in clinical practice may allow for more flexibility.

Our analysis comparing the AECG criteria and proposed SICCA criteria revealed that if Schirmer's test was used in place of the OSS and if the UWS rate was used in place of the LSG biopsy, the level of agreement between both criteria was low (52%). However, when all objective tests were available to define the AECG criteria, the level of agreement between the SICCA classification criteria and the AECG criteria was high (88%). Also, if the AECG criteria were used as the "gold standard," which is a likely scenario if experts were asked at this time to select SS cases and controls, and all objective tests were available, the sensitivity and specificity of our SICCA criteria would be very high. In reality, because the AECG allows for substitution of criteria components, it is almost never applied with all objective tests only, which is one of its inherent weaknesses. Finally, in an external data set of 303 participants who were not included in the data set used to develop the preliminary classification criteria, we found the SICCA criteria to have a sensitivity of 93% and a specificity of 95%.

Until recently, since few therapeutic agents were being considered in the systemic management of SS, the development of classification criteria was mainly for the purpose of epidemiologic studies to estimate the prevalence of the disease. However, the development of new biologic immunomodulating agents that are being considered in the treatment of SS increases the need and importance of developing stringent classification criteria that can be used in the context of clinical trials. The consequence of misclassifying someone without SS as a case would be serious given the potentially toxic side effects of these agents. The results of the various validation analyses described herein indicate that the preliminary classification criteria we initially developed using a consensus methodology constitute a set of criteria that are stringent enough to be used as entry criteria into clinical trials. The SICCA classification criteria were found to perform very similarly to the AECG criteria when all objective tests are available for the AECG. However, the SICCA criteria do not have the weakness inherent to the AECG criteria that allows for the use of alternate tests like Schirmer's test, or reported symptoms

of dry mouth and/or eyes that we have shown to have poor specificity. Not only do the proposed classification criteria rely on a combination of objective tests, but they also require evidence of autoimmunity by serologic and/or histopathologic measures. Histopathologic examination of an LSG sample provides high disease specificity, wide availability, prediction of non-Hodgkin's lymphoma development with the presence of lymphoid germinal centers in the glands (29), and unparalleled insights into the autoimmune disease–active cells within an SS target organ. LSG biopsy has been criticized as being invasive and difficult to apply in all settings. However, the performance and analysis of nearly 1,400 biopsy samples as part of the SICCA protocol suggest otherwise. When the LSG biopsy is skillfully and conservatively performed, it is a minimally invasive 15-minute procedure that yields unique information about the extent and nature of the disease process.

The distinction between primary and secondary forms of SS is based on an early definition of the disease and may now be obsolete. The initial definition and diagnostic criteria for SS were the presence of "keratoconjunctivitis sicca ('dry eyes'), xerostomia ('dry mouth') and rheumatoid arthritis or other connective tissue disease" and "two of the three are generally considered sufficient for the diagnosis" (1). Patients who developed the dry eye/mouth components of SS without developing RA were initially labeled as having the "sicca syndrome" and later "primary SS," while those with RA who usually developed the dry eye/mouth components after onset of their joint disease were labeled "secondary SS" (30). Subsequently, objective measures were adopted for assessing lacrimal and salivary hypofunction, systemic components of primary SS were identified, and we learned that various organ-specific (e.g., thyroid, liver, kidneys, and lungs) autoimmune conditions can occur in SS patients, in other autoimmune connective tissue diseases, and independently. While the details of autoimmune pathogenesis remain elusive, many diseases have now been identified as having autoimmune mechanisms, mostly distinguished by the target organ(s) affected, and genetic causes or susceptibilities are emerging. It has also become clear that some individuals with one autoimmune disease have enhanced susceptibility to develop others. Therefore, it seems of little use and risks potential confusion to distinguish in a given patient one autoimmune disease as secondary to another. Accordingly, the diagnosis of SS should be given to all who fulfill these criteria while also diagnosing any concurrent organ-specific or multiorgan autoimmune diseases, without distinguishing as primary or secondary.

In summary, the SICCA classification criteria developed from registry data collected using standardized measures are easy to apply even though they may require the involvement of at least 2 clinical specialties and are based entirely on objective tests. A series of validation exercises indicates improved classification performance relative to existing alternatives, making them more suitable for application in situations where misclassification may present health risks.

ACKNOWLEDGMENTS

We would like to also express our gratitude to all of the participants in the SICCA and to Drs. Pamela McInnes, Jane Atkinson, and Xavier Mariette for their review of the manuscript and valuable input.

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. S. C. Shiboski, C. H. Shiboski, Criswell, Lanfranchi, Schiødt, Umehara, Zhao, Dong, D. Greenspan, Heidenreich, Helin, Kitagawa, Li, Sack, Sugai, Vollenweider, Whitcher, Wu, S. Zhang, W. Zhang, J. S. Greenspan, Daniels.
Acquisition of data. Baer, Challacombe, Lanfranchi, Schiødt, Umehara, Vivino, Zhao, Dong, Heidenreich, Helin, Kirkham, Kitagawa, Larkin, Li, Lietman, Lindegaard, McNamara, Shirlaw, Sugai, Vollenweider, Whitcher, Wu, S. Zhang, W. Zhang, J. S. Greenspan, Daniels.
Analysis and interpretation of data. S. C. Shiboski, C. H. Shiboski, Criswell, Lietman, Whitcher, J. S. Greenspan, Daniels.

REFERENCES

1. Bloch KJ, Buchanan WW, Wohl MJ, Bunim JJ. Sjögren's syndrome: a clinical, pathological, and serological study of sixty-two cases. Medicine (Baltimore) 1965;44:187–231.
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 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. Japan: Japanese Ministry of Health; 1977.
5. Homma M, Tojo T, Akizuki M, Yamagata H. Criteria for Sjögren's syndrome in Japan. Scand J Rheumatol Suppl 1986; 61:26–7.
6. Manthorpe R, Frost-Larsen K, Isager H, Prause JU. Sjögren's syndrome: a review with emphasis on immunological features. Allergy 1981;36:139–53.
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. Revised diagnostic criteria for Sjögren's syndrome. Rheumatology (Oxford) 2000;24:421–8.
11. Vitali C, Bombardieri S, Jonsson R, Moutsopoulos HM, Alexander EL, Carsons SE, et al, and the European Study Group on Classification Criteria for Sjögren's Syndrome. 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. Daniels TE, Criswell LA, Shiboski C, Shiboski S, Lanfranchi H, Dong Y, et al, for the Sjögren's International Collaborative Clinical Alliance Research Groups. An early view of the international Sjögren's syndrome registry. Arthritis Rheum 2009;61:711–4.
13. Classification and Response Criteria Subcommittee of the American College of Rheumatology Committee on Quality Measures. Development of classification and response criteria

for rheumatic diseases [editorial]. Arthritis Rheum 2006;55: 348–52.
14. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. Am J Public Health 1984;74:979–83.
15. Daniels TE, Cox D, Shiboski CH, Schiodt M, Wu A, Lanfranchi H, et al, for the Sjögren's International Collaborative Clinical Alliance Research Groups. Associations between salivary gland histopathologic diagnoses and phenotypic features of Sjögren's syndrome among 1,726 registry participants. Arthritis Rheum 2011;63:2021–30.
16. Feenstra RP, Tseng SC. Comparison of fluorescein and rose bengal staining. Ophthalmology 1992;99:605–17.
17. 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 2009;149:405–15.
18. Chow S, Ruskey F. Drawing area-proportional Venn and Euler diagrams: 11th International Symposium on Graph Drawing, Perugia, Italy. Lect Notes Comput Sci 2003;2912:466–77.
19. Goodman LA. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika 1974;61: 215–31.
20. Pepe MS, Janes H. Insights into latent class analysis of diagnostic test performance. Biostatistics 2007;8:474–84.
21. Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. Biometrics 1996;52:797–810.
22. Hartigan JA. Clustering algorithms. New York: John Wiley & Sons; 1975.
23. Baughman AL, Bisgard KM, Cortese MM, Thompson WW, Sanden GN, Strebel PM. Utility of composite reference standards and latent class analysis in evaluating the clinical accuracy of diagnostic tests for pertussis. Clin Vaccine Immunol 2008;15:106–14.
24. See CW, Alemayehu W, Melese M, Zhou Z, Porco TC, Shiboski S, et al. How reliable are tests for trachoma? A latent class approach. Invest Ophthalmol Vis Sci 2011;52:6133–7.
25. Breiman L. Random forests. Mach Learn 2001;45:5–32.
26. Daniels T, Greenspan JS, Cox D, Criswell LA, DeSouza Y, Dong Y. Objective measures in Sjögren's syndrome associated with each other but not with sicca symptoms: analysis of 564 enrollees in the SICCA International Registry and Repository [abstract]. Arthritis Rheum 2007;56 Suppl:S446.
27. Malladi AS, Sack KE, Shiboski S, Shiboski C, Baer AN, Banushree R, et al. Primary Sjögren's syndrome as a systemic disease: a study of participants enrolled in an international Sjögren's syndrome registry. Arthritis Care Res (Hoboken) 2012. E-pub ahead of print.
28. Masaki Y, Dong L, Kurose N, Kitagawa K, Morikawa Y, Yamamoto M, et al. Proposal for a new clinical entity, IgG4-positive multiorgan lymphoproliferative syndrome: analysis of 64 cases of IgG4-related disorders. Ann Rheum Dis 2009; 68:1310–5.
29. Theander E, Vasaitis L, Baecklund E, Nordmark G, Warfvinge G, Liedholm R, et al. Lymphoid organisation in labial salivary gland biopsies is a possible predictor for the development of malignant lymphoma in primary Sjögren's syndrome. Ann Rheum Dis 2011;70:1363–8.
30. Moutsopoulos HM, Webber BL, Vlagopoulos TP, Chused TM, Decker JL. Differences in the clinical manifestations of sicca syndrome in the presence and absence of rheumatoid arthritis. Am J Med 1979;66:733–6.

APPENDIX A: COLLABORATORS OF THE SJÖGREN'S INTERNATIONAL COLLABORATIVE CLINICAL ALLIANCE COHORT

Collaborators of the Sjögren's International Collaborative Clinical Alliance are as follows: D. Cox, R. Jordan (oral pathology), D. Lee (rheumatology), Y. DeSouza (operations director), D. Drury (clinical coordinator/phlebotomy), A. Do (clinical coordinator), L. Scott (clinical assistant), M. Lam (statistician/programmer), J. Nespeco (data manager), J. Whiteford (finance director), M. Margaret (administrative assistant): University of California, San Francisco; I. Adler, A. C. Smith, A. M. Bisio, M. S. Gandolfo (stomatology), A. M. Chirife, A. Keszler (oral pathology), S. Daverio (specimen processing), V. Kambo (group coordinator): University of Buenos Aires and German Hospital, Buenos Aires, Argentina; Y. Jiang, D. Xu, J. Su (rheumatology), D. Du (stomatology/pathology), H. Wang, Z. Li, J. Xiao (stomatology/labial salivary gland biopsies), Q. Wu (specimens/rheumatology), C. Zhang, W. Meng (phlebotomy), J. Zhang (project assistant): Peking Union Medical College Hospital, Beijing, China; S. Johansen, S. Hamann (ophthalmology), J. Schiødt, H. Holm (oral medicine), P. Ibsen (oral pathology), A. M. Manniche, S. P. Kreutzmann, J. Villadsen (group coordinators/specimen handling), Rigshospitalet, Copenhagen, Denmark; Y. Masaki, T. Sakai (rheumatology), N. Shibata (ophthalmology), M. Honjo (stomatology), N. Kurose, T. Nojima (oral pathology), T. Kawanami (specimen processing), T. Sawaki (hematology/immunology), K. Fujimoto (group coordinator): Kanazawa Medical University, Ishikawa, Japan; E. Odell, P. Morgan (pathology), L. Fernandes-Naglik (specimen processing), B. Varghese-Jacob, S. Ali (oral medicine), M. Adamson (project coordinator): King's College London, London, UK; S. Seghal, R. Mishra (rheumatology), V. Bunya, M. Massaro-Giordano (ophthalmology), S. K. Abboud (otolaryngology), A. Pinto, Y. W. Sia (oral medicine), K. Dow (group coordinator): University of Pennsylvania, Philadelphia; E. Akpek, S. Ingrodi (ophthalmology), W. Henderson (oral medicine), C. Gourin (otolaryngology), A. Keyes (group coordinator): Johns Hopkins University, Baltimore, Maryland; M. Srinivasan (group director), J. Mascarenhas, M. Das, A. Kumar (co-directors), P. Joshi (ophthalmology), R. Banushree (physician), U. Kim (surgeon), B. Babu (oral medicine), A. Ram, R. Saravanan, K. N. Kannappan (administration), N. Kalyani (group coordinator): Aravind Eye Hospital, Madurai, India.