ACR Open Rheumatology
Vol. 1, No. 5, July 2019, pp 292–302
DOI 10.1002/acr2.1033

AMERICAN COLLEGE of RHEUMATOLOGY
Empowering Rheumatology Professionals

Reproducibility of Ocular Surface Staining in the Assessment of Sjögren Syndrome–Related Keratoconjunctivitis Sicca: Implications on Disease Classification

Astrid Rasmussen,[1] Donald U. Stone,[2] C. Erick Kaufman,[3] Kimberly S. Hefner,[4] Nicole R. Fram,[5] Rhea L. Siatkowski,[6] Andrew J. W. Huang,[7] James Chodosh,[8] Pablo T. Rasmussen,[9] Dustin A. Fife,[10] Nathan Pezant,[10] Kiely Grundahl,[10] Lida Radfar,[3] David M. Lewis,[3] Michael H. Weisman,[11] Swamy Venuturupalli,[11] Daniel J. Wallace,[11] Nelson L. Rhodus,[12] Michael T. Brennan,[13] Courtney G. Montgomery,[10] Christopher J. Lessard,[14] R. Hal Scofield,[15] and Kathy L. Sivils[14]

Objective. The objective of this study was to assess the performance and reproducibility of the two currently used ocular surface staining scores in the assessment of keratoconjunctivitis sicca in Sjögren syndrome (SS) research classification.

Methods. In a multidisciplinary clinic for the evaluation of sicca, we performed all tests for the American European Consensus Group (AECG) and the American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) classification criteria, including the van Bijsterveld score (vBS) and the Ocular Staining Score (OSS), in 994 participants with SS or with non-SS sicca. We analyzed the concordance between the scores, the diagnostic accuracy and correlation with clinical variables, and interrater and intrasubject reproducibility.

Results. A total of 308 (31.1%) participants had a discordant vBS and OSS that was due to extra corneal staining points in the OSS. The presence of one or more of the additional points was highly predictive of SS classification (odds ratio = 3.66; $P = 1.65 \times 10e\text{-}20$) and was associated with abnormal results of all measures of autoimmunity and glandular dysfunction. Receiver operating characteristic curves showed optimal cutoff values of four for the vBS (sensitivity = 0.62; specificity = 0.71; Youden's $J = 0.33$) and five for the OSS (sensitivity = 0.56; specificity = 0.75; Youden's $J = 0.31$). Notably, there was very poor consistency in interobserver mean scores and distributions ($P < 0.0001$) and in intrasubject scores after a median of 5.5 years (35% changed status of the ocular criterion).

Conclusion. Ocular surface staining scores are useful for SS research classification; however, they are subject to significant interrater and intrasubject variability, which could result in changes in classification in 5%-10% of all subjects. These results highlight the need for objective and reproducible markers of disease that have thus far remained elusive for SS.

INTRODUCTION

Sjögren syndrome (SS) is a complex autoimmune disorder characterized by xerostomia and xerophthalmia caused by exocrine gland dysfunction. The clinical manifestations and glandular damage are mediated by autoantibodies and lymphocytic infiltration of the salivary and lacrimal glands, which leads to significant dental decay and keratoconjunctivitis sicca (KCS). Furthermore, a significant proportion of patients develop extraglandular manifestations, including lymphadenopathy, arthritis, and respiratory,

Presented in part at the 2014 Annual Scientific Meeting of the American College of Rheumatology, Boston, MA, November 2014.

The contents of this article are the sole responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Supported by the NIH (grants AR-053483, AR-050782, DE-018209, DE-015223, AI-082714, GM-104938, AR-060804, AI-118787, AR-065953, AI-012717, RR-020143, and GM-103510). Additional funding was obtained from the Department of Veterans Affairs (VA Merit Review BX001451), the Oklahoma Medical Research Foundation, the Phileona Foundation, the Sjögren's Syndrome Foundation, and the Lupus Research Institute Novel Research Award.

[1]Astrid Rasmussen, MD, PhD: Oklahoma Medical Research Foundation, Oklahoma City, and Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, México City, México; [2]Donald U. Stone, MD: Johns Hopkins University, Baltimore, Maryland (current address: Dean McGee Eye Institute, University of Oklahoma Health Sciences Center, Oklahoma City); [3]C. Erick Kaufman, MD, Lida Radfar, DDS, MS, David M. Lewis, DDS: University of Oklahoma, Oklahoma City; [4]Kimberly S. Hefner, DO: Hefner Eye Care and Optical Center, Oklahoma City, Oklahoma; [5]Nicole R. Fram, MD: David Geffen

SIGNIFICANCE & INNOVATIONS

- Ocular surface staining scores are useful for assessing corneal and conjunctival damage in Sjögren syndrome (SS) and for disease classification.
- This study validates, in an external cohort, a revised cutoff level of five or more for the Ocular Staining Score (OSS) in the American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) SS research classification criteria.
- The scoring of the additional corneal fluorescein staining points, as described in the OSS, increases the predictive value of the test compared with the van Bijsterveld score.
- A major concern is the poor repeatability of the scores, which show significant interrater and intrasubject variability, even among trained and calibrated specialists.
- The low reproducibility of the tests reduces their usefulness as outcome measures for treatment and may impact patient classification.

renal, muscular, neurologic, and hematologic manifestations, as well as an increased risk of lymphoid malignancies (1–3).

The diagnosis of SS is difficult to establish because there is no single diagnostic gold standard test. Furthermore, in the clinical setting, diagnosis is mostly based on expert opinion but may require a multidisciplinary team and invasive procedures. For research purposes, several classification methods have been described, and the most widely used are the 2002 revised American European Consensus Group (AECG) classification criteria (4). These were revised anew with support from both the American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR), resulting in the updated 2016 ACR/EULAR classification criteria for SS (5). Critical to the development of effective criteria is the selection of items with the greatest demonstrated validity, sensitivity, specificity, and reproducibility and determining their optimal cutoff levels (6).

KCS is a core phenotypic feature of SS and was recognized by Sjögren (7), who described that damage to the cornea and interpalpebral conjunctiva was detectable by staining the ocular surface. The objective assessment of KCS by ocular staining has also gone through multiple grading systems, the basis of which was van Bijsterveld's (8) Rose Bengal–based semiquantitative score (van Bijsterveld score [vBS]), one of the AECG's SS classification criteria. A modification to the score to allow for the use of fluorescein to stain the cornea and for the use of lissamine green to stain the conjunctiva was introduced with the 2002 revised AECG criteria because Rose Bengal is not available in many countries and is associated with significant patient discomfort (4). Furthermore, several studies have demonstrated similar staining properties of the dyes (9), with the advantage of decreased toxicity and pain associated with lissamine green (10).

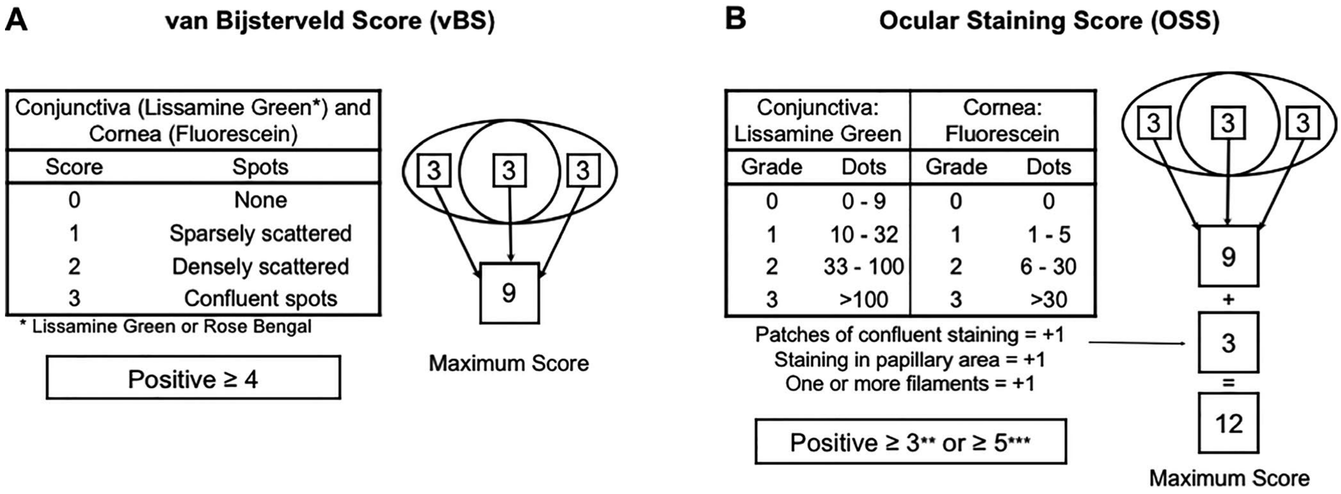
As part of the effort to develop standardized classification criteria for SS, the National Institutes of Health (NIH) funded the Sjögren's International Collaborative Clinical Alliance (SICCA) international registry (11). One of the objectives of SICCA was to develop objective measures of SS, among them being the modification of previous KCS grading systems to develop a new quantitative dry-eye grading scheme: the SICCA Ocular Staining Score (OSS) (12). The OSS uses lissamine green dye to stain the conjunctiva and fluorescein to stain the cornea and gives equal weight to corneal and conjunctival changes by adding three additional points to the corneal staining: patches of confluent staining, pupillary staining, and filaments. An abnormal OSS was originally defined as being a score of three or more in either eye (12) (Figure 1). The OSS correlated well with other objective measures of SS and showed good sensitivity but poor specificity when evaluated in the SICCA cohort (13). Furthermore, its performance was not directly compared with that of the vBS or other preexisting staining scores, nor was it validated in external cohorts; it has been suggested that the current cutoff level results in inadequate specificity of the test (14–16). This resulted in a revised cutoff level of five or more being introduced in the new ACR/EULAR criteria (5).

The objective of this study was to directly compare the performance of the OSS with that of the vBS in a cohort of subjects with sicca manifestations who were systematically evaluated for SS and sicca syndrome, with the goal of testing the diagnostic accuracy, optimal cutoff level, and reliability of the OSS in an independent cohort.

School of Medicine, University of California, Los Angeles; [6]Rhea L. Siatkowski, MD: University of Oklahoma and Dean McGee Eye Institute, Oklahoma City; [7]Andrew J. W. Huang, MD: School of Medicine, Washington University in St. Louis, St. Louis, Missouri (current address: University of Minnesota); [8]James Chodosh, MD, MPH: Massachusetts Eye and Ear and Harvard University, Boston (current address: Dean McGee Eye Institute, University of Oklahoma Health Sciences Center, Oklahoma City); [9]Pablo T. Rasmussen: Harvard University, Cambridge, Massachusetts; [10]Dustin A. Fife, PhD (current address: Oklahoma Medical Research Foundation, Oklahoma City), Nathan Pezant, MS, Kiely Grundahl, BS, Courtney G. Montgomery, PhD: Oklahoma Medical Research Foundation, Oklahoma City; [11]Michael H. Weisman, MD, Swamy Venuturupalli, MD, Daniel J. Wallace, MD: Cedars-Sinai Medical Center, Los Angeles, California; [12]Nelson L. Rhodus, DMD, MPH: School of Dentistry, University of Minnesota, Minneapolis; [13]Michael T. Brennan, DDS, MHS: Carolinas Medical Center, Charlotte, North Carolina; [14]Christopher J. Lessard, PhD, Kathy L. Sivils, PhD: Oklahoma Medical Research Foundation and University of Oklahoma, Oklahoma City; [15]R. Hal Scofield, MD: Oklahoma Medical Research Foundation, University of Oklahoma, and Department of Veterans Affairs Medical Center, Oklahoma City

Dr. Scofield has received consulting fees related to Sjögren syndrome from AbbVie Pharmaceuticals and an honorarium and travel funding from Eisai Korea (less than $10 000 each). Dr. Huang has received speaking fees from Mallinckrodt Pharmaceuticals (less than $10 000). No other disclosures relevant to this article were reported.

Address correspondence to Kathy L. Sivils, PhD, Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, 825 NE 13th Street, MS57, Oklahoma City, OK. E-mail: sivilsk@omrf.org



Conjunctiva (Lissamine Green*) and Cornea (Fluorescein)	
Score	Spots
0	None
1	Sparsely scattered
2	Densely scattered
3	Confluent spots

* Lissamine Green or Rose Bengal

Positive ≥ 4

Conjunctiva: Lissamine Green		Cornea: Fluorescein	
Grade	Dots	Grade	Dots
0	0 - 9	0	0
1	10 - 32	1	1 - 5
2	33 - 100	2	6 - 30
3	>100	3	>30

Patches of confluent staining = +1
Staining in papillary area = +1
One or more filaments = +1

Positive ≥ 3** or ≥ 5***

Figure 1. Diagram of scoring of the van Bijsterveld score (vBS) and the Ocular Staining Score (OSS). **A**, Semiquantitative method for determining the vBS as described originally by van Bijsterveld (8) and incorporated into the American European Consensus Group (AECG) classification criteria (4,34). However, for this study and to diminish interrater subjectivity, a dot-counting method based on the OSS was used to determine the vBS (14). **B**, The OSS with the three additional corneal staining points. The original cutoff level of the OSS was three or more (12); however, a cutoff level of five or more has been adopted for the American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) classification criteria (5).

METHODS

Participant recruitment and disease classification. The present study is the retrospective analysis of data collected between 2005 and 2017 from participants in a large cohort of patients with sicca who were evaluated in the Sjögren's Research Clinic at Oklahoma Medical Research Foundation or at similar clinics at the University of Minnesota and Cedars-Sinai Medical Center (14). All the tests necessary for classification by AECG (4) and ACR/EULAR (5) criteria were performed in addition to collection of detailed clinical and serological measures (14). We used the AECG classification criteria as the diagnostic gold standard for the study given that the ocular surface staining is solely based on the vBS, the score against which we compared the diagnostic properties of the OSS.

The institutional review board of each institution approved all procedures, and each participant provided written informed consent prior to entering the study. The study was conducted in accordance with current regulations protecting human subjects participating in research, the Health Insurance Portability and Accountability Act (HIPAA), and the Declaration of Helsinki.

Ocular assessment and calibration. The ocular specialists underwent standardized training and calibration exercises for vBS and OSS scoring with preexisting and externally validated videos and patient photographs (ie, the gold standard for this study) before patients were enrolled in the study or before any ocular evaluation occurred. Each specialist performed the complete eye examination, including both the vBS and OSS, using a standardized protocol derived from the methods described in the SICCA cohort (12). A summary is illustrated in Figure 1, and a detailed description of the methods is included in the Supplemental Materials. Every rater evaluated different participants in the study; the participants were assigned to the rater based on the clinic at which they were evaluated but otherwise in an unbiased manner. The clinical and sociodemographic features of each subset of participants were not significantly different (data not shown).

The vBS system relies on comparison with a representative drawing, with scores given from 0-3: 0 represents no staining, 1 represents few dots, 2 represents many dots, and 3 represents confluent dots (8) (Figure 1). To reduce the interobserver and intraobserver variability inherent to the descriptive quality of the vBS, the scoring for each section of the ocular surface was determined using the SICCA dot-counting method, as previously described, in this cohort (14). Furthermore, all examiners were trained on standardized photos to use the dot-counting method. As described in the AECG criteria, a vBS of 4 or more in either eye was considered positive (4).

The OSS was calculated for each eye as follows: for the corneal score, the number of punctate epithelial erosions was counted and tabulated as a score. If fluorescein staining was absent, the score was 0. If one to five erosions were noted, the corneal score was 1; if six to 30 erosions were noted, the score was 2, and for 30 or more erosions, the score was 3. An additional point was added to the corneal score for each of the following: erosions in the central 4 mm of the cornea (pupilar staining), the presence of one or more corneal filaments, or areas of confluent staining. Conjunctival lissamine green staining of the interpalpebral conjunctiva was scored similarly, with a grade of 0 for zero to nine lissamine dots, a grade of 1 for 10-32 dots, a grade of 2 for 33-100 dots, and a grade of 3 for

Table 1. Effect and predictive value of the Ocular Staining Score additional corneal points on classification for primary SS[a]

	SS (n = 445), n (%)	Non-SS Sicca (n = 549), n (%)[b]	PPV (95% CI)	NPV (95% CI)	Odds Ratio (95% CI)	*P*
Additional points (positive)	201 (45)	107 (20)	0.66 (0.61-0.71)	0.65 (0.61-0.69)	3.66 (2.76-4.85)	1.65 × 10e-20
Confluent staining (positive)	183 (41)	83 (15)	0.69 (0.63–0.74)	0.64 (0.60-0.67)	3.92 (2.9-5.29)	2.44 × 10e-20
Pupillary staining (positive)	138 (31)	59 (11)	0.70 (0.63-0.76)	0.61 (0.58-0.65)	3.37 (2.66-5.23)	1.25 × 10e-15
Filaments (positive)	35 (8)	16 (3)	0.69 (0.54-0.80)	0.57 (0.53-0.60)	2.84 (1.55-5.21)	0.0005

Abbreviation: AECG, American European Consensus Group; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; SS, Sjögren syndrome.

[a] SS classification is based on AECG criteria. We excluded all participants with secondary SS or overlap syndromes or those with incomplete ocular assessments (4).

[b] Participants with non-SS sicca are participants who have self-described sicca symptoms and meet at least one AECG criterion but do not meet enough criteria to be classified as having SS.

more than 100 dots. Each eye received a conjunctival score for the temporal and nasal region; therefore, the total ocular surface staining score for each eye was the addition of the corneal score, the nasal conjunctiva score, and the temporal conjunctiva score (Figure 1). At its inception and for the ACR criteria, an OSS of 3 or more in either eye was considered positive; however, the new ACR/EULAR criteria use a threshold of 5 or more for a positive OSS (5,12). The vBS and OSS were considered concordant if their numerical values were identical in the same subject and were considered discordant if the values were different (ie, discordant values were only present if the subject presented one or more of the additional three points included in the OSS).

Statistical analysis. The Wilcoxon rank sum test was used to compare the mean vBS and OSS between concordant and discordant individuals. Dichotomous variables were evaluated using χ^2 tests. Receiver operating characteristic (ROC) curves were used to compare the trade-off between sensitivity and specificity for the OSS and vBS predictor variables using the Youden *J* index criteria to define optimal cutoff values (17,18). The reliability analysis included the vBS and OSS of the eye specialists who evaluated 18 or more participants. The mean, the median with the interquartile range (IQR), and the median absolute dispersion (MAD) of the scores were determined. The difference of the medians was analyzed using the Brown-Forsythe and Welch analysis of variance tests, correcting for multiple comparisons using the Games-Howell statistical hypothesis test. For the 20 subjects whose eyes were evaluated twice, intraclass correlation coefficient (ICC) estimates were calculated based on an absolute agreement, two-way random-effects model for interrater reliability, and consistency of agreement, two-way mixed-effects model for intrarater reliability (19–22). The impact of scoring differences was evaluated by identifying subjects who might have been classified differently if their ocular score had been assessed by a different observer. Subjects with SS that could become non-SS sicca were defined as those with four classification criteria, a positive ocular surface staining score, and a $\text{vBS} - \text{MAD}_{\text{vBS}}$ of less than 4 or an $\text{OSS} - \text{MAD}_{\text{OSS}}$ of less than 5, whereas participants with non-SS sicca that would become SS had three classification criteria, a negative ocular surface staining score, and a $\text{vBS} + \text{MAD}_{\text{vBS}}$ of 4 or more or an $\text{OSS} + \text{MAD}_{\text{OSS}}$ of 5 or more. Hypothesis tests were performed in R version 3.2.0 (23), GraphPad Prism version 8.0.1 for Mac OS X (GraphPad Software), STATA Statistical Software: Release 15 (StataCorp LLC), and VassarStats (online at vassarstats.net) and ROC curves were generated using the pROC package (24) roc function. $P < 0.05$ was considered statistically significant.

RESULTS

Characteristics of study participants. We evaluated 1703 subjects for SS based on the AECG and ACR/EULAR classification criteria in multidisciplinary clinics for patients with sicca symptoms; included in our analysis are 994 patients who had both vBS and OSS results and met criteria for primary SS (n = 445) or for non-SS sicca (n = 549) according to the AECG criteria. Non-SS sicca was defined as sicca symptoms and at least one AECG criterion but less than four AECG criteria or the absence of objective autoimmunity (negative serology and histopathology test results). The excluded subjects included those with other non-SS autoimmune diseases or secondary SS and those with incomplete ocular evaluations (Supplemental Table 1).

Predictive value of the corneal staining points of the OSS. The nominal (positive vs. negative) concordance rate of the vBS and the OSS was significantly different, independent of the OSS cutoff level selected (McNemar's test of paired samples: positive OSS of 3 or more: $P < 0.1 \times 10e\text{-}06$; positive OSS of 4 or more: $P = 0.004$; positive OSS of 5 or more: $P < 0.1 \times 10e\text{-}06$), but still achieved a moderate to high agreement rate (positive OSS of 3 or more: $\kappa = 0.74$ [95% confidence interval (CI): 0.71-0.79]; positive OSS of 4 or more: $\kappa = 0.82$ [95% CI: 0.79-0.86]; positive OSS of 5 or more: $\kappa = 0.86$ [95% CI: 0.83-0.99]). The ordinal analysis (comparing the numerical scores) showed that 686 (69%) participants had an identical vBS and OSS, whereas 308 (31%) participants had different scores. The participants with discordant scores had more anti-Ro/SSA ($P = 1.85 \times 10e\text{-}14$) and anti-La/SSB antibodies ($P = 5.01 \times 10e\text{-}09$) and a focus score of one or more on a minor salivary gland biopsy ($P = 1.01 \times 10e\text{-}09$) compared with subjects with a concordant OSS and vBS (Supplemental Table 2). They also had a significantly higher vBS than the concordant group (Wilcoxon rank sum test: $P = 2.2 \times 10e\text{-}16$) (Supplemental Figure 1).

The central methodological difference between the vBS and OSS is the addition of the three corneal staining points (CSPs). Thus, we investigated the contribution of these points to the disease classification of SS and investigated their association with other measures of disease, as shown in Table 1. The additional CSPs were present in 308 of 994 (31%) participants, with the patches of confluent staining being the most common (266 of 994; 84%), followed by pupillary staining in 197 (64%) participants, and the filaments being the least common (51 of 308; 16%). The presence of any extra CSP was strongly predictive of SS classification (odds ratio [OR] = 3.66; 95% CI: 2.76-4.85; $P = 1.65 \times 10e\text{-}20$), and the same held true if each of the three points was analyzed separately. The strongest association was with the patches of confluent staining (OR = 3.92; 95% CI: 2.90-5.29; $P = 2.44 \times 10e\text{-}20$) (Table 1). A small number of subjects with a vBS of less than 4 (thus negative) had a positive OSS (5 or more) because of extra CSPs (10 of 994; 1%).

Every objective measure included in the classification criteria showed significant association with each of the three patterns of staining (ie, patches of confluent staining, pupillary staining, and filaments). Confluent staining showed the strongest association with positive anti-Ro/SSA test results, pupillary staining showed the strongest association with positive Schirmer's test results, and filaments showed the strongest association with positive anti-La/SSB and whole unstimulated salivary flow (WUSF) test results. An increasing number of CSPs were also significantly associated with most criteria (in particular, with WUSF and Schirmer's tests) but not with abnormal labial salivary gland biopsy results (Table 2).

A small but significant proportion of patients who did not meet criteria for SS presented with an abnormal vBS and extra CSPs on the OSS (n = 107; 20% of subjects with non-SS sicca); their median vBS and OSS were high (median = 5 [IQR: 4-7] and median = 7 [IQR: 5-9], respectively) and coexisted with positive Schirmer's test results in 43% of cases, suggesting significant lacrimal gland involvement in a subset of patients with non-SS sicca.

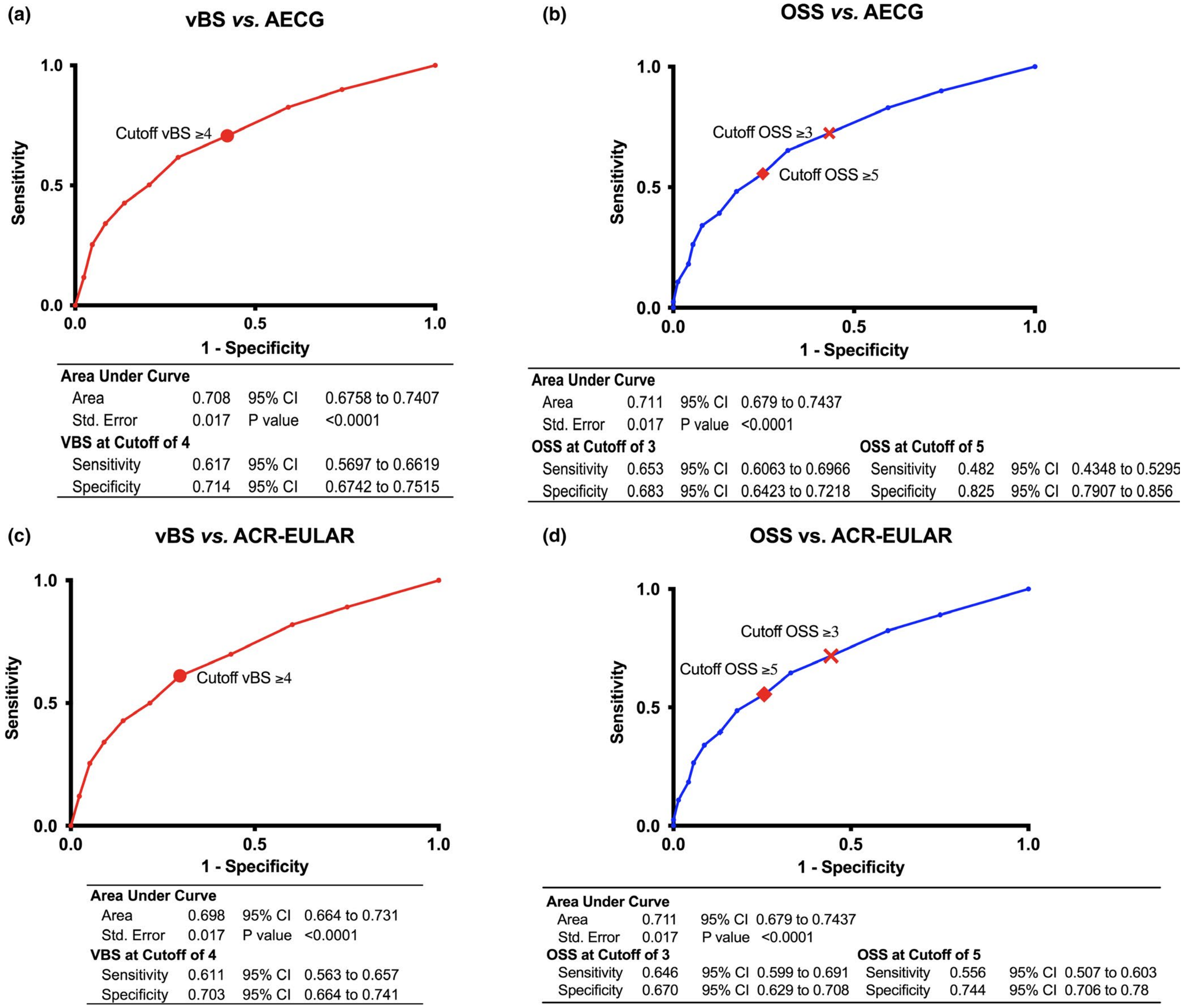
Optimal cutoff values for the OSS. To assess the performance of each of the two eye-staining scores on the same subject, ROC curves were generated, plotting sensitivity and 1-specificity against nominal SS or non-SS sicca classification based on the AECG criteria (Figure 2A) and the ACR/EULAR criteria (Figure 2B). The area under the curve (AUC) for the vBS was 0.708 and 0.698 for the AECG and ACR/EULAR classification criteria, respectively, whereas the AUC for the OSS was 0.711 and 0.702 for the AECG and ACR/EULAR classification criteria, respectively; the difference in AUCs was not significant ($P = 0.222$). The current cutoff value for the vBS of 4 or more had a sensitivity of 0.617 and 0.611 and a specificity of 0.714 and 0.703 for the AECG and ACR/EULAR classification criteria, respectively. The original cutoff value for the OSS of 3 or more had a sensitivity of 0.653 and 0.646, and a specificity of 0.683 and 0.670 for the AECG and ACR/EULAR classification

Table 2. Association of the Ocular Staining Score additional corneal points with classification criteria for SS[a]

	Confluent Staining	Pupillary Staining	Filaments	Three CSPs	Six CSPs
	OR (95% CI), *P*	OR (95% CI), *P*	OR (95% CI), *P*	OR (95% CI), *P*	OR (95% CI), *P*
Anti-Ro/SSA (positive)	3.11 (2.32-4.18), 6.34 × 10e-15	2.82 (2.05-3.89), 5.12 × 10e-11	2.53 (1.44-4.45), 5.16 × 10e-04	3.24 (1.5-6.64), 0.0038	1.63 (0.63-3.84), 0.363
Anti-La/SSB (positive)	2.54 (1.83-3.52), 6.87 × 10e-09	2.67 (1.88-3.79), 9.26 × 10e-09	3.07 (1.72-5.49), 4.05 × 10e-05	3.23 (1.42-6.95), 0.0045	3.13 (1.29-8.2), 0.018
Biopsy (positive)	2.73 (2.01-3.7), 2.53 × 10e-11	2.72 (1.93-3.81), 1.87 × 10e-09	1.91 (0.99-3.68), 0.029	2.37 (1.05-5.46), 0.038	1.09 (0.41-2.8), 1.0
Schirmer's test (positive)	2.53 (1.9-3.37), 6.54 × 10e-11	3.33 (2.42-4.58), 1.20 × 10e-14	2.09 (1.08-4.05), 0.015	3.29 (1.54-6.89), 0.003	3.66 (1.24-10.42), 0.032
WUSF (positive)	2.42 (1.79-3.25), 1.99 × 10e-09	2.88 (2.04-4.07), 3.26 × 10e-10	4.93 (2.29-10.6), 3.99 × 10e-06	3.81 (1.57-8.95), 0.003	8.49 (1.34-91.1), 0.014

Abbreviation: CI, confidence interval; CSP, corneal staining point; OR, odds ratio; SS, Sjögren syndrome; WUSF, whole unstimulated salivary flow.

[a]Based on analysis of 316 participants with extra points: n = 209 for SS and n = 107 for non-SS sicca. Presence (positive) (21) for each criterion is based on Vitali et al (4).



(a)

Area Under Curve			
Area	0.708	95% CI	0.6758 to 0.7407
Std. Error	0.017	P value	<0.0001
VBS at Cutoff of 4			
Sensitivity	0.617	95% CI	0.5697 to 0.6619
Specificity	0.714	95% CI	0.6742 to 0.7515

(b)

Area Under Curve							
Area	0.711	95% CI	0.679 to 0.7437				
Std. Error	0.017	P value	<0.0001				
OSS at Cutoff of 3				**OSS at Cutoff of 5**			
Sensitivity	0.653	95% CI	0.6063 to 0.6966	Sensitivity	0.482	95% CI	0.4348 to 0.5295
Specificity	0.683	95% CI	0.6423 to 0.7218	Specificity	0.825	95% CI	0.7907 to 0.856

(c)

Area Under Curve			
Area	0.698	95% CI	0.664 to 0.731
Std. Error	0.017	P value	<0.0001
VBS at Cutoff of 4			
Sensitivity	0.611	95% CI	0.563 to 0.657
Specificity	0.703	95% CI	0.664 to 0.741

(d)

Area Under Curve							
Area	0.711	95% CI	0.679 to 0.7437				
Std. Error	0.017	P value	<0.0001				
OSS at Cutoff of 3				**OSS at Cutoff of 5**			
Sensitivity	0.646	95% CI	0.599 to 0.691	Sensitivity	0.556	95% CI	0.507 to 0.603
Specificity	0.670	95% CI	0.629 to 0.708	Specificity	0.744	95% CI	0.706 to 0.78

Figure 2. Receiver operator curves for the performance of ocular surface staining scores versus Sjögren syndrome classification criteria. The table beneath each graph shows the area under the curve (AUC) and the sensitivity and specificity of the score at the present and past cutoff values for each classification system. **A**, the van Bijsterveld score (vBS) versus the American European Consensus Group (AECG) classification criteria. **B**, The Ocular Staining Score (OSS) versus the AECG classification criteria. **C**, the vBS versus the American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) classification criteria. **D**, the OSS versus the ACR/EULAR classification criteria. CI, confidence interval.

criteria, respectively. To achieve a better specificity, compared with that of the vBS, the OSS cutoff would have to be 5 or more (sensitivity = 0.482 and 0.556; specificity = 0.825 and 0.744) (Supplemental Tables 3-6).

Interrater and intrasubject variability of the vBS and the OSS. Although each rater evaluated different participants, the distribution of the median staining scores showed great variation across observers, with a MAD for vBS of 2 and a MAD for OSS of 2.5, as shown in Figure 3. The difference in the means across all observers was highly significant ($P < 0.0001$) both for the vBS and the OSS. The poor consistency across raters persisted even when analyzing only the three observers with the largest number of observations (n = 140, 212, and 236), all of whom were trained by the same eye specialist ($P < 0.0001$ for the vBS and the OSS; data not shown). A major caveat to this observation is the fact that each rater evaluated a different set of individuals; however, there were no systematic biases in the assignation of rater to subject.

To estimate how the rater could impact disease classification, we identified all subjects in whom a nominal (ie, positive to negative or negative to positive) change in ocular surface staining score would result in a different SS classification by the AECG classification criteria (details in Methods section) (25–27). These were either

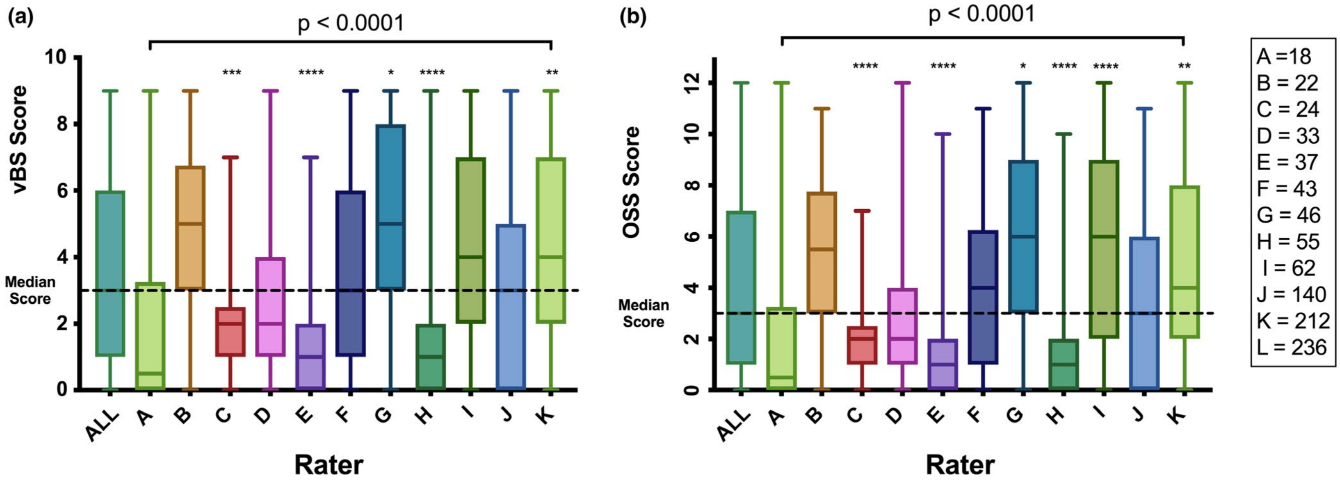


Figure 3. Interrater reproducibility of the ocular surface staining scores. The median, minimum, and maximum scores of all the evaluations performed by 11 raters (A-K) are shown. Each rater evaluated a different set and number of subjects, as shown in **A**. The difference of the medians was analyzed using the Brown-Forsythe and Welch analysis of variance tests, correcting for multiple comparisons using Games-Howell statistical hypothesis testing. The overall *P* value presented for raters A through K reflects the difference across all medians; the asterisks above each bar reflect the comparison of the individual rater versus the median of the scores of all raters. **A**, The median van Bijsterveld scores (vBS) by rater. **B**, The median Ocular Staining Scores (OSS) by rater.

subjects who were classified as having SS based on meeting three AECG classification criteria plus having a positive ocular surface staining score or subjects who were classified as having non-SS sicca based on meeting three AECG classification criteria and having a negative ocular surface staining score. In the first instance, a decrease of the vBS – MAD_{vBS} to less than 4 would result in 22 (4.9%) subjects with SS being classified as having non-SS sicca, whereas an increase of the vBS + MAD_{vBS} to 4 or more could result in 69 (12.6%) subjects with non-SS meeting SS criteria. When a similar exercise is done with the OSS, an OSS – MAD_{OSS} of less than 5 would result in 11 (2.5%) subjects with SS no longer meeting criteria, and an OSS + MAD_{OSS} of 5 or more would lead to 28 (5.1%) subjects with non-SS being classified as having SS.

In the context of the new weighted ACR/EULAR classification criteria, the percentage of subjects who are in the situation in which SS versus non-SS status is dependent on the ocular surface staining score is between approximately 5% and 10%; 54 of 532 (10.2%) subjects with non-SS sicca have either a positive biopsy result or a positive serology test result and a negative Schirmer's test result, a negative WUSF test result, and a negative OSS, whereas among those classified as having primary SS by these criteria, 24 of 432 (5.6%) only have a positive OSS beyond either a positive biopsy result or a positive serology test result. Thus, 78 of 994 (7.9%) subjects are at the boundary of ACR/EULAR classification criteria; they are in the hypothetical situation in which their SS classification by the new weighted criteria depends on the OSS. If the OSS of these subjects changed by the MAD, five subjects with non-SS would have SS (1% of the subjects with non-SS) and 14 subjects with SS (3.2%) would have non-SS.

As a pilot assessment of intrasubject variability, 20 subjects were invited to return to the clinic to have all study procedures, including the ocular criteria evaluation, performed again after a median of 5.5 years (Table 3). Nine subjects were reevaluated by the same eye specialist who did their initial evaluations, and 11 subjects were reevaluated by a different specialist; in all cases, the raters were blinded to the initial eye scores or to the identity of the rater at the first visit. Among the nine participants evaluated by the same clinician, a moderate concordance rate was observed in the categorical (positive vs. negative) scoring of the ocular staining ($\kappa = 0.571$ for both the vBS and the OSS), resulting in two of nine (22.2%) subjects changing from a positive to a negative vBS and OSS. The classification status of these two participants whose scores changed remained the same because they had sufficient additional features to meet SS classification. The intrarater agreement in the precise numerical score was moderate to good (ICC = 0.77 [95% CI: 0.26-0.94] for the vBS; ICC = 0.74 [95% CI: 0.20-0.94] for the OSS) (22).

In the case of the 11 participants evaluated by different specialists, the interrater categorical agreement was slight to none ($\kappa = 0.154$ for the vBS and equal to chance for the OSS), and the numerical agreement was poor (ICC = 0.55 [95% CI: 0.18-0.85] for the vBS; ICC = 0.38 [95% CI: −0.25 to 0.78] for the OSS) (22). Among these subjects, six of 11 (54.5%) had a changed nominal ocular score (four of 11 [36.4%] from positive to negative and two of 11 [18.2%] from negative to positive). The result of this change in the nominal score would be that one of the subjects (9%) would have a changed classification from primary SS at the first evaluation to non-SS sicca at the second time point. None of these changes were associated with the use of punctal plugs or prescription eye drops (data not shown). It is important to note that the intrarater variability analysis of this study was a pilot study with a very small sam-

Table 3. Intrasubject variability of the vBS and the OSS in 20 subjects evaluated at two time points with a median gap of 5.5 years

	Year of Evaluation			vBS (Call)		OSS (Call)		
Subject	First	Second	Time Gap, y	First	Second	First	Second	Classification Item Change
Same rater for both evaluations								
1	2010	2017	7	7 (+)	5 (+)	9 (+)	7 (+)	No change
2	2010	2017	7	2 (−)	1 (−)	4 (−)	1 (−)	No change
3	2012	2017	5	5 (+)	5 (+)	6 (+)	7 (+)	No change
4	2013	2017	4	3 (−)	3 (−)	4 (−)	3 (−)	No change
5	2013	2017	4	5 (+)	6 (+)	6 (+)	7 (+)	No change
6	2015	2017	2	0 (−)	0 (−)	0 (−)	0 (−)	No change
7	2015	2017	2	6 (+)	7 (+)	7 (+)	7 (+)	No change
8	2015	2017	2	4 (+)	0 (−)	5 (+)	0 (−)	Pos to neg
9	2015	2017	2	4 (+)	2 (−)	5 (+)	2 (−)	Pos to neg
Different rater for each evaluation								
1	2008	2017	9	3 (−)	1 (−)	4 (−)	2 (−)	No change
2	2008	2017	9	2 (−)	3 (−)	2 (−)	6 (+)	Neg to pos
3	2009	2017	8	0 (−)	1 (−)	0 (−)	1 (−)	No change
4	2009	2017	8	3 (−)	7 (+)	3 (−)	9 (+)	Neg to pos
5	2009	2017	8	5 (+)	1 (−)	6 (+)	2 (−)	Pos to neg
6	2009	2017	8	5 (+)	3 (−)	5 (+)	3 (−)	Pos to neg
7	2010	2017	7	9 (+)	9 (+)	11 (+)	10 (+)	No change
8	2011	2017	6	9 (+)	7 (+)	9 (+)	9 (+)	No change
9	2011	2017	6	9 (+)	7 (+)	9 (+)	9 (+)	No change
10	2014	2017	3	9 (+)	3 (−)	11 (+)	3 (−)	Pos to neg
11	2015	2017	2	7 (+)	3 (−)	9 (+)	3 (−)	Pos to neg

Abbreviation: ACR/EULAR, American College of Rheumatology/European League Against Rheumatism; neg, negative ocular staining; OSS, Ocular Staining Score; pos, positive ocular staining; vBS, van Bijsterveld score.
The cutoff values for the vBS and the OSS are based on the ACR/EULAR classification criteria (5).
Subjects whose ocular criterion changed between the two time points.
Items went from a positive (abnormal) score to a negative (normal) one.

ple size, and therefore generalizable conclusions should not be drawn. Rather, a properly designed larger prospective study would be important to replicate or discard these findings.

DISCUSSION

We compared the performance of two alternative ocular surface staining scores, the vBS and the OSS, in a large cohort of patients with sicca. The same observer evaluated each participant by both scoring systems. The concordance rate of the two scores was substantial ($\kappa = 0.78$) (28), but the specificity of the OSS for SS classification was lower than that of the vBS at the originally established cutoff level of 3 or more. These findings are in support of our previous observations that approximately 25% of our study participants had a positive OSS (cutoff of 3 or more) but a negative vBS, a highly significant difference ($P < 1 \times 10e\text{-}06$). The relevance of the divergent scores is highlighted by the fact that it accounted for 71% of discrepant AECG and ACR SS classifications in our cohort (13,14). We suggested at that point that a detailed comparison of the two scoring systems was necessary to determine a cutoff value for the OSS that had a better balance between sensitivity and specificity (14).

In the current analysis, our findings support the use of the OSS as a valuable tool, which correlates well with all the other domains of disease and final classification of SS. Based on ROC curve analyses, the newly proposed cutoff of the test to 5 or more significantly improves the specificity, both in comparison with a cutoff of 3 or more and in comparison with the vBS, without major sacrifice of the sensitivity of the test.

It has previously been noted that one of the limitations of the ocular surface staining scores in clinical practice is that, just as is the case with the minor salivary gland lip biopsy, these evaluations require additional medical specialists and sophisticated equipment (29–32). A very small percentage (approximately 1%) of vBS-negative subjects have a positive OSS based on the extra CSPs, so it could be argued that both scores are equivalent. However, the OSS is not different from the vBS in technical difficulty or requirements, so assessing the additional CSPs and scoring each domain on a quantitative basis should be a worthy pursuit. Although higher scores of either vBS or OSS have a positive correlation with all classification criteria for SS, the identification of additional CSPs in the OSS has a better positive predictive value for SS classification and correlates strongly with measures of objective autoimmunity, such as the presence of anti-Ro/SSA, anti-La/SSB autoantibodies, and positive labial salivary gland biopsies.

Conditions like rheumatic diseases are hard to diagnose because of heterogeneous and overlapping clinical manifestations and the lack of a gold standard diagnostic test. This is particularly evident when trying to conduct and evaluate the results of research studies that require recruitment of patients with a well-defined disease. The central role of classification criteria is to help distinguish patients with or without the disease of interest in a consistent and reproducible manner across diverse research projects. The most relevant characteristics of classification criteria are sensitivity (Do the criteria identify the subjects with the disease?), specificity (Do they identify those without the disease?), and criterion validity (Do they predict or correlate with a gold standard, most often expert opinion?) (6). In the case of participant selection for clinical trials, particularly those involving biological agents or immunosuppressants with the potential for serious side effects, the classification criteria must provide high specificity to avoid exposing subjects who are not affected to unacceptable toxicity (14,29). The last step in the design of classification criteria is their validation in external and independent cohorts to ensure that when applied under different circumstances, the criteria will yield comparable results (repeatability and reproducibility) (6).

As previously mentioned, a relevant consideration when designing any new classification system or diagnostic test is its reproducibility (interobserver reliability) (6,19,28). When the OSS was created, the authors noted that the sequence and intervals of each of the ocular tests—such as Schirmer's eye test, fluorescein instillation, tear break-up time, grading of the corneal staining, lissamine green dye application, and conjunctival examination—is crucial for accuracy and reproducibility (12). The SICCA consortium developed a careful protocol for the assessment of ocular dryness measures, and all participating ocular specialists were trained to perform the procedures consistently, which are the same procedures that were used in this study (SICCA website; online at http://sicca.ucsf.edu) (11,12). In the development article, internal reliability was confirmed by a low rate of intragrader/intrasubject (same patient, same observer) variability in the score of the left versus the right eye (12). More recently, the same group evaluated a subset of their cohort to assess the intergrader reliability of the OSS among SICCA-trained ophthalmologists examining the same patient on the same day and reported an excellent agreement (ICC = 90-91) (16). This is in contrast to a previous report of low to moderate intrarater agreement of corneal and conjunctival fluorescein and Rose Bengal staining (κ = 0.25 and 0.21, respectively) (33,34) and to our findings of significant variation across multiple observers. A significant limitation of our study is that each rater evaluated different participants; however, the poor agreement of the score distributions persisted even when including only the most experienced raters (ie, those with the largest number of subjects evaluated) in the analysis. Moreover, ocular surface staining varied significantly over time in the same subject, even when evaluated by the same clinician, leading to a change of the criterion from positive to negative or vice versa in more than one-third of the subjects who were evaluated twice.

The ultimate impact of these changes on disease classification was small but not insignificant. For AECG classification, in which every criterion carries the same weight, approximately 3%-13% of subjects would have reverse classification based on the ocular surface staining score, whereas approximately 1%-3% would have a changed status based on the ACR/EULAR criteria. The weighted ACR/EULAR classification system was more resilient to change because the ocular staining criterion is scored as one point, whereas the serology and minor salivary gland histopathology test results are scored as three points each. These findings highlight the influence of the observer and the subjectivity of the tests on classification and they underscore the relevance of proper training and calibration of the examiners. As a corollary, when the suspicion of SS in the clinical setting is high, borderline results by one observer should not suffice to rule out the diagnosis, for which expert opinion is still the gold standard.

Beyond the clinical setting, the poor interrater reproducibility and reliability of the ocular staining tests become a significant problem when classification criteria, which rely, in part, on these tests, determine inclusion or exclusion in research projects. The operator-dependent scoring of the ocular surface involvement may introduce bias in the homogeneity of patient selection, may allow for inclusion of subjects not meeting SS classification criteria and potentially expose them to unnecessary drugs, or may exclude patients with SS who could benefit from participation in clinical trials. Thus, it is crucial that the study design includes homogenization strategies, such as a detailed standardization of the methods, accurate training of the ocular specialists, or the use of a single scoring team to evaluate images of all study participants.

The weakness of not having multiple raters evaluate the same subject is, at the same time, an excellent reflection of a real-life scenario. In centers that have the infrastructure and expertise to perform the ocular staining tests as part of clinical care, any given patient may be seen by a different specialist, with potentially different outcomes. Furthermore, when the procedure is intended for participation in research, the same individual might be tested at multiple time points by different observers. Thus, unless training and measurement against a well-defined standard is implemented, the logical expectation is high variability. The possibility of a single rater evaluating the same individual on more than one occasion or the possibility of more than one rater assessing the same patient to attain reproducibility is a highly desirable but often unreal scenario.

The use of ocular surface staining as a measure of KCS in SS is valuable, particularly the OSS, which adds predictive diagnostic and prognostic power in SS evaluation without increasing technical burden. However, the potential of the misclassification of 5%-10% of patients, based on either rater variability or changes of the score over time, is less than ideal for a classification criterion and even more so when using ocular staining as a measure of response to treatment or as an outcome measure in clinical trials. These results strongly support the notion that we need better, objective, and reproducible biological markers of disease.

ACKNOWLEDGMENTS

We are grateful to all the individuals with SS and those serving as controls who participated in this study. We thank the following individuals at the Oklahoma Medical Research Foundation for their help in the collection and ascertainment of the samples used in this study: Judy Harris, Sharon Johnson, Sarah Cioli, Cherilyn Pritchett-Frazee, Laura Battiest, Wes Daniel, and Stuart Glenn.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual contact, and all authors approved the final version to be published.

Study conception and design. A. Rasmussen, Stone.

Acquisition of data. A. Rasmussen, Stone, Kaufman, Hefner, Fram, Siatkowski, Huang, Chodosh, Grundahl, Radfar, Lewis, Weisman, Venuturupalli, Wallace, Rhodus, Brennan, Lessard, Scofield.

Analysis and interpretation of data. A. Rasmussen, P. T. Rasmussen, Fife, Pezant, Montgomery, Scofield, Sivils.

REFERENCES

1. Brito-Zeron P, Baldini C, Bootsma H, Bowman SJ, Jonsson R, Mariette X, et al. Sjögren syndrome. Nat Rev Dis Primers 2016;2:16047.
2. Mariette X, Criswell LA. Primary Sjögren's syndrome. N Engl J Med 2018;378:931–9.
3. Ramos-Casals M, Brito-Zeron P, Seror R, Bootsma H, Bowman SJ, Dorner T, et al. Characterization of systemic disease in primary Sjögren's syndrome: EULAR-SS Task Force recommendations for articular, cutaneous, pulmonary and renal involvements [published erratum appears in Rheumatology (Oxford) 2017;56:1245]. Rheumatology (Oxford) 2015;54:2230–8.
4. 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.
5. Shiboski CH, Shiboski SC, Seror R, Criswell LA, Labetoulle M, Lietman TM, et al. 2016 American College of Rheumatology/European League Against Rheumatism classification criteria for primary sjögren's syndrome: a consensus and data-driven methodology involving three international patient cohorts. Arthritis Rheumatol 2017;69:35–45.
6. Singh JA, Solomon DH, Dougados M, Felson D, Hawker G, Katz P, et al. Development of classification and response criteria for rheumatic diseases. Arthritis Rheum 2006;55:348–52.
7. Sjögren H. Zur kenntis der keratoconjunctivitis sicca (keratitis filiformis bei hypofunktion der tranendrusen). Acta Ophthalmol 1933;11 Suppl 2:1–151.
8. Van Bijsterveld OP. Diagnostic tests in the Sicca syndrome. Arch Ophthalmol 1969;82:10–4.
9. Machado LM, Castro RS, Fontes BM. Staining patterns in dry eye syndrome: rose bengal versus lissamine green. Cornea 2009;28:732–4.
10. Bron AJ, Evans VE, Smith JA. Grading of corneal and conjunctival staining in the context of other dry eye tests. Cornea 2003;22:640–50.
11. Daniels TE, Criswell LA, Shiboski C, Shiboski S, Lanfranchi H, Dong Y, et al. An early view of the international Sjögren's syndrome registry. Arthritis Rheum 2009;61:711–4.
12. 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.
13. Shiboski SC, Shiboski CH, Criswell LA, Baer AN, Challacombe S, Lanfranchi H, et al, for the Sjögren's International Collaborative Clinical Alliance (SICCA) Research Groups. 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 Car Res (Hoboken) 2012;64:475–87.
14. 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.
15. Cornec D, Saraux A, Jousse-Joulin S, Pers JO, Boisrame-Gastrin S, Renaudineau Y, et al. The differential diagnosis of dry eyes, dry mouth, and parotidomegaly: a comprehensive review. Clin Rev Allergy Immunol 2015;49:278–87.
16. Rose-Nussbaumer J, Lietman TM, Shiboski CH, Shiboski SC, Bunya VY, Akpek EK, et al. Inter-grader agreement of the Ocular Staining Score in the Sjögren's International Clinical Collaborative Alliance (SICCA) Registry. Am J Ophthalmol 2015;160:1150–3.
17. Machado PM. Measurements, composite scores and the art of 'cutting-off'. Ann Rheum Dis 2016;75:787–90.
18. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol 2006;163:670–5.
19. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound Obstet Gynecol 2008;31:466–75.
20. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1:30–46.
21. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–8.
22. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155–63.
23. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.
24. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77.

25. Korb DR, Herman JP, Finnemore VM, Exford JM, Blackie CA. An evaluation of the efficacy of fluorescein, rose bengal, lissamine green, and a new dye mixture for ocular surface staining. Eye Contact Lens 2008;34:61–4.
26. Manning FJ, Wehrly SR, Foulks GN. Patient tolerance and ocular surface staining characteristics of lissamine green versus rose bengal. Ophthalmology 1995;102:1953–7.
27. Kim J, Foulks GN. Evaluation of the effect of lissamine green and rose bengal on human corneal epithelial cells. Cornea 1999;18:328–32.
28. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
29. Vitali C, Bootsma H, Bowman SJ, Dorner T, Gottenberg JE, Mariette X, et al. Classification criteria for Sjögren's syndrome: we actually need to definitively resolve the long debate on the issue. Ann Rheum Dis 2013;72:476–8.
30. Bootsma H, Spijkervet FK, Kroese FG, Vissink A. Toward new classification criteria for Sjögren's syndrome? Arthritis Rheum 2013;65:21–3.
31. Cornec D, Saraux A, Cochener B, Pers JO, Jousse-Joulin S, Renaudineau Y, et al. Level of agreement between 2002 American-European Consensus Group and 2012 American College of Rheumatology classification criteria for Sjögren's syndrome and reasons for discrepancies. Arthritis Res Ther 2014;16:R74.
32. Sankar V, Noll JL, Brennan MT. Diagnosis of Sjögren's syndrome: American-European and the American College of Rheumatology classification criteria. Oral Maxillofac Surg Clin North Am 2014;26:13–22.
33. Nichols KK, Mitchell GL, Zadnik K. The repeatability of clinical measurements of dry eye. Cornea 2004;23:272–85.
34. 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.