Inter-grader Agreement of the Ocular Staining Score in the Sjögren's International Clinical Collaborative Alliance (SICCA) Registry

JENNIFER ROSE-NUSSBAUMER, THOMAS M. LIETMAN, CAROLINE H. SHIBOSKI, STEPHEN C. SHIBOSKI, VATINEE Y. BUNYA, ESEN K. AKPEK, MUTHIAH SRINIVASAN, JEENA MASCARENHAS, GIACOMINA MASSARO-GIORDANO, NANCY A. MCNAMARA, JOHN P. WHITCHER, AND BRUCE D. GAYNOR, FOR THE SJÖGREN'S INTERNATIONAL COLLABORATIVE CLINICAL ALLIANCE RESEARCH GROUPS

• PURPOSE: To determine the intra-observer and inter-observer reliability of a novel ocular staining score among trained ophthalmologists.

• DESIGN: Reliability analysis within a prospective, observational, multicenter cohort study.

• METHODS: Those enrolled in the National Institutes of Health–funded Sjögren's International Collaborative Clinical Alliance (SICCA) who presented for follow-up at the University of California San Francisco, Aravind Eye Hospital, Johns Hopkins University, and the University of Pennsylvania were included. Study participants were graded using the ocular staining score by at least 2 masked SICCA-trained ophthalmologists. The primary outcome for this study was the intraclass correlation coefficient (ICC) for the total ocular staining score. ICCs were also calculated for tear break-up time (TBUT) and conjunctival and corneal staining.

• RESULTS: Total ocular staining score had an ICC of 0.91 for the right eye (95% confidence interval [CI] 0.85–0.96) and 0.90 for the left eye (95% CI 0.83–0.97). Corneal staining (right eye 0.86, 95% CI 0.76–0.93, left eye 0.90, 95% CI 0.81–0.95) and conjunctival staining (right eye 0.87, 95% CI 0.80–0.93, left eye 0.85, 95% CI 0.75–0.93) demonstrated excellent agreement. The ICC for TBUT was slightly lower (right eye 0.77, 95% CI 0.64–0.89; left eye 0.81, 95% CI 0.68–0.90).

• CONCLUSIONS: Previous studies have shown that the ocular staining score is correlated with other diagnostic components of Sjögren syndrome. In this study, we demonstrate high reliability in grading among trained ophthalmologists, completing the validation of this test. (Am J Ophthalmol 2015;160(6):1150–1153. © 2015 by Elsevier Inc. All rights reserved.)

SJÖGREN SYNDROME IS A CHRONIC AUTOIMMUNE INflammatory disorder characterized by decreased exocrine function of the salivary and lacrimal glands.[1,2] It can be primary or secondary, when it is related to other autoimmune conditions such as systemic lupus erythematosus or rheumatoid arthritis.[3] Owing to diagnostic challenges and different diagnostic criteria, the exact prevalence of Sjögren syndrome has not been well established. One study, using the American European Consensus Group criteria for diagnosis, found an annual incidence of 4 per 100 000; 70% of these cases were primary.[4] The American College of Rheumatology currently recommends that a classification of Sjögren syndrome be made based on meeting 2 of the following criteria: (1) positive serology for anti-SSA and/or anti-SSB; (2) ocular staining score of ≥3; (3) presence of focal lymphocytic sialadenitis with focus score ≥1 focus/4 mm^2.[2,5]

The Sjögren's International Collaborative Clinical Alliance (SICCA) is an NIH-funded group of investigators who are exploring the etiology, diagnosis, epidemiology, and treatment of Sjögren syndrome in a large prospective cohort. In order to characterize keratoconjunctivitis (KCS) associated with Sjögren syndrome, a new quantitative ocular staining score was developed and tested in this cohort. In a prior study in this cohort, abnormal ocular staining score was strongly associated with other features of Sjögren syndrome.[6] The reproducibility of dry eye measurements has been debated. In one study of a single grader who repeated Schirmer test, tear break-up time, and cotton-thread test measurements on different days, the repeatability was poor.[7]

In this study, we evaluate the inter-grader and intra-grader reliability of the ocular staining score among SICCA-trained ophthalmologists examining the same patient on the same day. If the ocular staining score is shown to be reliable, it may serve as a useful diagnostic tool in the evaluation of patients with both Sjögren syndrome–related and non–Sjögren syndrome–related KCS.

AJO.com Supplemental Material available at AJO.com.
Accepted for publication Aug 14, 2015.

From the Francis I. Proctor Foundation (J.R.-N., T.M.L., N.A.M., J.P.W., B.D.G.), and the Departments of Ophthalmology (J.R.-N., T.M.L., J.P.W., B.D.G.), Dentistry (C.H.S.), and Epidemiology and Biostatistics (T.M.L., S.C.S., J.P.W.), University of California San Francisco, San Francisco, California; Department of Optometry (J.R.-N., N.A.M.), University of California, Berkeley, Berkeley, California; Department of Ophthalmology (V.Y.B., G.M.-G.), University of Pennsylvania, Philadelphia, Pennsylvania; Department of Ophthalmology (E.K.A.), Johns Hopkins University, Baltimore, Maryland; and Aravind Eye Care System at Madurai, Madurai, India (M.S., J.M.).

Inquiries to Bruce D. Gaynor, 513 Parnassus Ave S334, San Francisco, CA 94122; e-mail: Bruce.Gaynor@ucsf.edu

0002-9394/$36.00
http://dx.doi.org/10.1016/j.ajo.2015.08.021

METHODS

STUDY PARTICIPANTS WHO PRESENTED FOR THEIR Sjögren's International Collaborative Clinical Alliance (NIH N01 DE32636) study visit at University of California San Francisco, Aravind Eye Hospital, Johns Hopkins University, or University of Pennsylvania were included. The methods of the cohort study, as well as the development of the ocular staining score and standardized training of all involved ophthalmologists, have been outlined previously.[6] Five ophthalmologists participated in the reliability analysis. In accordance with the SICCA study, participants were asked not to apply their routine drops for the 12 hours preceding their visit and to discontinue contact lens wear for at least 7 days prior to examination. Two to 3 ophthalmologists scored each patient and were masked to each other's ocular staining score grading. Institutional Review Board approval was obtained from all participating institutions for the Sjögren's International Collaborative Clinical Alliance prospective cohort study, including this reliability analysis.

During the examination, study participants had tests and examination in the following order: unanesthetized Schirmer test, slit-lamp examination assessing tear break-up time (TBUT) with fluorescein, and punctate epithelial erosions (PEEs) of the cornea with fluorescein and conjunctival staining patterns with lissamine green. Lid, conjunctival, and corneal abnormalities were also noted. A subset of participants also had osmolarity using TearLab (TearLab Corporation, San Diego, California, USA). Each parameter was graded sequentially by the different investigators, who remained masked to each others' grading.

Schirmer test I was performed prior to all other tests and before any drops were instilled in the eye. As soon as both Schirmer strips were in place the strips remained in place a maximum of 5 minutes or until completely saturated. Fluorescein (0.5% preservative free, Leiter's Pharmacy, San Jose, California, USA) was instilled immediately after removing the Schirmer strips. The cornea was examined at the slit lamp using 10× magnification and illumination set on "high" with the cobalt blue filter between 4 and 8 minutes after instillation. TBUT was defined as the time in seconds between the patient's last blink and the first appearance of a dry spot on the corneal surface. It was measured 3 times and the mean value was recorded. A value of 10 or greater was considered normal.[8] Participants were then given a score of 0 if there were no PEEs, 1 if there were 1–5 PEEs, 2 if there were 6–30 PEEs, and 3 if there were more than 30. They were then given an additional point each if they had patches of confluent staining, staining in the pupillary area, or 1 or more filaments, for a maximum possible corneal staining score of 6.

After detailed assessment for lid, conjunctival, and corneal abnormalities, lissamine green (1 drop of 1% lissamine green dye; Leiter's Pharmacy) was instilled in the eye. The conjunctival staining was then immediately assessed with 10× magnification with white illumination through a neutral-density filter. Interpalpebral nasal and temporal conjunctiva were assigned a grade of 0 for no staining, 1 for 10–32 dots, 2 for 33–100 dots, and 3 for greater than 100 dots. These values were then summed for a maximum possible score of 6 for conjunctival staining. The total ocular staining score was determined by adding the conjunctival and corneal staining with scores ranging from 0 to 12 for each eye (Supplemental Figure, available at AJO.com).

Inter-rater agreement was assessed using the intraclass correlation coefficient (ICC), estimated using analysis of variance methods for unbalanced data. Separate ICCs were estimated for the total ocular staining score, ocular staining score between the 2 eyes of an individual patient, and individual parts of the examination including TBUT, conjunctival lissamine staining pattern, and corneal fluorescein staining pattern. Precision of ICC estimates were summarized using percentile-based 95% bootstrap percentile confidence intervals based on resampling (n = 999 resamples) drawn at the level of the patient to account for clustering of responses. Calculations were performed in Mathematica 10.0 (Wolfram Research, Champaign, Illinois, USA).

TABLE 1. Baseline Characteristics of Study Participants Graded With Ocular Staining Score in the Sjögren's International Collaborative Clinical Alliance (n = 49)

Characteristic	Number	%	Median (IQR)
Age, y			57 (51–67)
Sex, female	41	84	
Serology positive for anti-SSA/SSB	14	29	
RF positivity	10	20	
Focus score ≥1	10	26	
Meets ACR criteria for Sjögren syndrome	16	33	
Schirmer mm/5 min, right eye			9 (5–16)
Shirmer mm/5 min, left eye			8 (5–18)
TBUT <10 s, both eyes	15	30	5 (4–7)
Ocular staining score (abnormal ≥3), right eye			5 (2–8)
Ocular staining score (abnormal ≥3), left eye			5 (3–8)

ACR = American College of Rheumatology; IQR = interquartile range; RF = rheumatoid factor; TBUT = tear break-up time.

RESULTS

NINETY-EIGHT EYES OF 49 STUDY PARTICIPANTS WERE EVALUated and scored at the University of California San Francisco, Aravind Eye Hospital, Johns Hopkins University, or University of Pennsylvania by at least 2 SICCA ophthalmologists between March 2011 and September 2012. One eye of 1 patient was excluded owing to missing data. Table 1 outlines

TABLE 2. Intraclass Correlation Coefficient for the Ocular Staining Score in the Sjögren's International Collaborative Clinical Alliance

Examination Element	Mean ICC (95% CI)		
	Right	Left	Both Eyes
Total OSS score	0.91 (0.85–0.96)	0.90 (0.83–0.97)	
Max OSS score			0.90 (0.83–0.96)
OSS score between eyes			0.95 (0.93–0.96)
Tear break-up time[a]	0.77 (0.64–0.89)	0.81 (0.68–0.90)	0.81 (0.68–0.91)
Conjunctival staining[b]	0.87 (0.80–0.93)	0.85 (0.75–0.93)	
Corneal staining[c]	0.86 (0.76–0.93)	0.90 (0.81–0.95)	

CI = confidence interval; ICC = intraclass correlation coefficient; OSS, ocular staining score.
[a]Measured in seconds, average of 3 measurements. Both eyes represents average tear break-up time between the 2 eyes.
[b]Calculated by adding nasal and temporal staining.
[c]Calculated by adding corneal stain plus 3 bonus points for patches of confluent staining, staining in pupillary area, and presence of filaments.

the characteristics of the study participants included. The majority of participants were female (n = 41, 84%) with a median age of 57 (interquartile range [IQR] 51–67). Sixteen (33%) met criteria for a diagnosis of Sjögren syndrome using criteria defined by the American College of Rheumatology.[5] Twenty-nine percent (n = 14) were positive for anti-Sjögren syndrome–related antigen A (anti-SSA) or anti–Sjögren syndrome–related antigen B (anti-SSB), while 20% (n = 10) were positive for rheumatoid factor (RF).

Table 2 shows the ICCs and confidence intervals (CI) for total ocular staining score, TBUT, conjunctival staining, and corneal staining. Total ocular staining score had an ICC of 0.91 for the right eye (95% CI 0.85–0.96) and 0.90 for the left eye (95% CI 0.83–0.97). The ICC between eyes for the same grader was 0.95 (95% CI 0.93–0.96). Cornea (right eye 0.86, 95% CI 0.76–0.93; left eye 0.90, 95% CI 0.81–0.95) and conjunctival (right eye 0.87, 95% CI 0.80–0.93; left eye 0.85, 95% CI 0.75–0.93) staining also demonstrated excellent agreement. ICC for TBUT was slightly lower (right eye 0.77, 95% CI 0.64–0.89; left eye 0.81, 95% CI 0.68–0.90), but improved when the TBUT in both eyes was averaged (0.81, 95% CI 0.68–0.91).

DISCUSSION

IN THIS STUDY, WE DEMONSTRATE HIGH INTRA- AND INTER-grader agreement in ocular staining score among trained ophthalmologists. In a prior study, abnormal ocular staining score, defined as a score of 3 or greater, was associated with other findings of Sjögren syndrome such as a focal lymphocytic sialadenitis with a focus score of greater than 1, and positive serologic results for anti-SSA or anti-SSB antibodies.[6] Although a score of 3 or greater was considered abnormal in that study, higher thresholds for an abnormal ocular staining score may be more appropriate. Determining reliability is another important component of validating a new diagnostic test. Our study demonstrates that only a small amount of the variance in ocular staining score is due to the examiner. Prior studies of the repeatability of dry eye measurements typically repeated the test on different days; therefore it is unknown whether the lack of repeatability was due to intra- or inter-observer variation or to variability in the measurement over time.[7] Although use of the ocular staining score for dry eye grading may improve reliability of dry eye measurements conducted on the same day, this study did not evaluate whether there are significant fluctuations in ocular staining score on different days. Given its high ICC, the ocular staining score has the potential to improve the internal validity of future dry eye studies as well as comparability between studies.

Agreement statistics have been calculated for determining the degree of clinical activity for other ocular diseases including trachoma,[9,10] as well as anterior[11,12] and posterior uveitis.[13] There has been discussion about the magnitude of ICC or kappa that constitutes adequate agreement. Many things must be considered when interpreting agreement statistics, making it difficult to define an acceptable level of agreement for all tests. Recommendations have been published—for example, those by Landis and Koch, who described values varying from 0–0.20, representing "slight" agreement, to 0.81–1, as "almost perfect" agreement for kappa.[14] Although this serves as a rough guide, there were no data supporting these recommendations, and they are not widely accepted. Intraclass correlation coefficient estimates the proportion of the variance that is due to the patient and not the observer. Cohen clarifies that kappa is an approximation of ICC, and may at times be identical.[15] Therefore, if these guidelines were accepted, the ocular staining score would have outstanding intra-observer and inter-observer reliability.

Individual portions of the examination including conjunctival and corneal staining pattern also demonstrated excellent agreement. However, a larger proportion

of the variance in tear break-up time was due to the examiner than of the variance in the total score. Taking the mean TBUT between the 2 eyes improved this marginally. Although there does not appear to be an advantage to using the maximum ocular staining score, it may be sufficient to calculate the score for only 1 eye, since the ICC between 2 eyes is high. Only about one third of our patients met the American College of Rheumatology criteria for a diagnosis of Sjögren syndrome. Twenty-eight percent of SICCA registry patients had abnormal ocular staining score without other evidence of Sjögren syndrome.[6] These data suggest that the ocular staining score is appropriate to evaluate Sjögren and non–Sjögren syndrome KCS patients.

There are several limitations to our study. The patients available for repeated testing may not be representative of the entire SICCA registry or of KCS patients. However, the goal of this study was to measure agreement between examiners; therefore this is not likely to be an important consideration. The intensive training we performed in this study likely improves ICC; therefore these results may be most applicable to research environments or academic centers.

Previously the ocular staining score has been shown to correlate with other indices of Sjögren syndrome; here we complete the validation of the ocular staining score by demonstrating its repeatability among different examiners. The ocular staining score for dry eye grading may improve reliability of dry eye measurements, which is of particular value in clinical trials where this may be an outcome measure. These results also suggest that the ocular staining score may be useful in the diagnosis and treatment of patients with aqueous tear deficiency from causes other than Sjögren syndrome.

FUNDING/SUPPORT: THE RESEARCH PORTION OF THIS WORK IS FINANCIALLY SUPPORTED BY CONTRACT N01-DE-32636 FROM the National Institutes of Health, Bethesda Maryland. Individual support is provided by contract #K12-EY-015398 (V.Y.B.) and #K12-EY-017269 (J.R.-N.) from the National Eye Institute, Bethesda Maryland and an unrestricted grant from the Peierls Foundation, Golden Colorado (J.R.-N.). Financial disclosures: Vatinee Y. Bunya has a sponsored research agreement with Amakem, Diepenbeek, Belgium. Financial disclosures: The following authors have no financial disclosures: Jennifer Rose-Nussbaumer, Thomas M. Lietman, Caroline H. Shiboski, Stephen C. Shiboski, Esen K. Akpek, Muthiah Srinivasan, Jeena Mascarenhas, Giacomina Massaro-Giordano, Nancy A. Mcnamara, Johnp.Whitcher, and Bruce D. Gaynor. All authors attest that they meet the current ICMJE criteria for authorship.
Members of the SICCA research group are available in the following website https://sicca-online.ucsf.edu.

REFERENCES

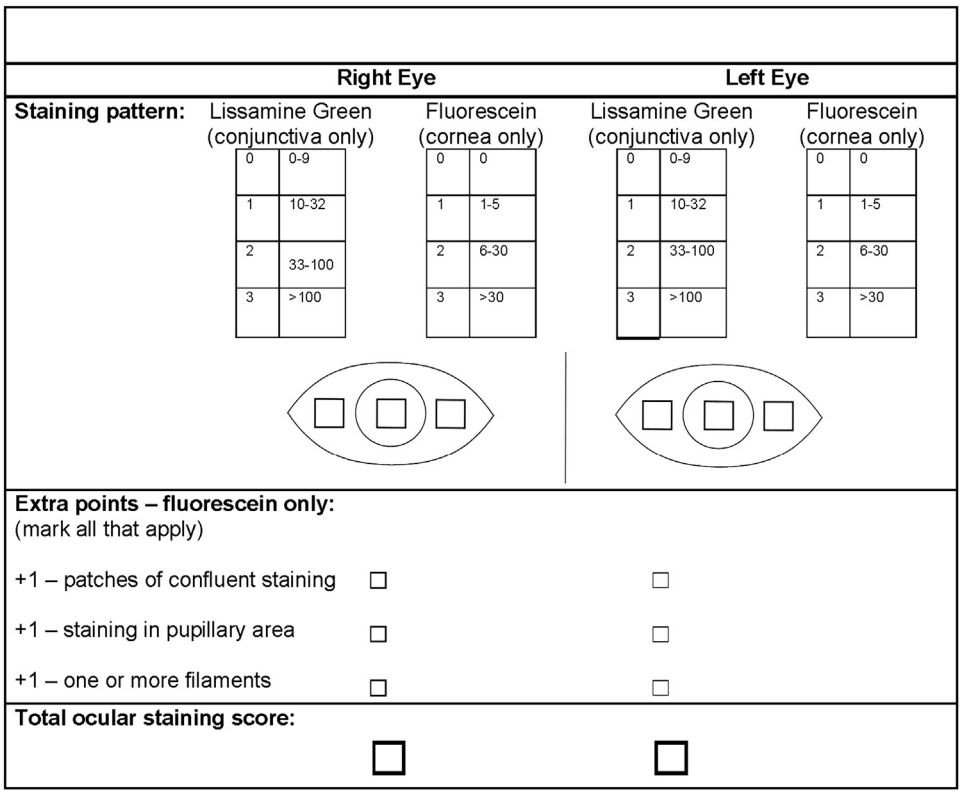
1. Sjogren H. Keratoconjunctivitis sicca and chronic polyarthritis. *Acta Med Scand* 1948;130(5):484–488.
2. Holm S, Sjogren H. Studies on keratoconjunctivitis sicca, based on examination of 500 subjects affected with rheumatism and an equally large control material. *Acta Ophthalmol* 1948;26(2):269–273.
3. 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(5):733–736.
4. Pillemer SR, Matteson EL, Jacobsson LT, et al. Incidence of physician-diagnosed primary Sjogren syndrome in residents of Olmsted County, Minnesota. *Mayo Clin Proc* 2001;76(6):593–599.
5. Shiboski SC, Shiboski CH, Criswell L, et al. American College of Rheumatology classification criteria for Sjogren's syndrome: a data-driven, expert consensus approach in the Sjogren's International Collaborative Clinical Alliance cohort. *Arthritis Care Res* 2012;64(4):475–487.
6. Whitcher JP, Shiboski CH, Shiboski SC, et al. A simplified quantitative method for assessing keratoconjunctivitis sicca from the Sjogren's Syndrome International Registry. *Am J Ophthalmol* 2010;149(3):405–415.
7. Nichols KK, Mitchell GL, Zadnik K. The repeatability of clinical measurements of dry eye. *Cornea* 2004;23(3):272–285.
8. Lemp MA. Report of the National Eye Institute/Industry workshop on Clinical Trials in Dry Eyes. *CLAO J* 1995;21(4):221–232.
9. Amza A, Kadri B, Nassirou B, et al. Community risk factors for ocular Chlamydia infection in Niger: pre-treatment results from a cluster-randomized trachoma trial. *PLoS Negl Trop Dis* 2012;6(4):e1586.
10. Solomon AW, Bowman RJ, Yorston D, et al. Operational evaluation of the use of photographs for grading active trachoma. *Am J Trop Med Hyg* 2006;74(3):505–508.
11. Kempen JH, Ganesh SK, Sangwan VS, Rathinam SR. Interobserver agreement in grading activity and site of inflammation in eyes of patients with uveitis. *Am J Ophthalmol* 2008;146(6):813–818.e1.
12. Jabs DA, Nussenblatt RB, Rosenbaum JT. Standardization of uveitis nomenclature for reporting clinical data. Results of the First International Workshop. *Am J Ophthalmol* 2005;140(3):509–516.
13. Nussenblatt RB, Palestine AG, Chan CC, et al. Standardization of vitreal inflammatory activity in intermediate and posterior uveitis. *Ophthalmology* 1985;92(4):467–471.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159–174.
15. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20(3):37–46.

Biosketch

Dr Jennifer Rose-Nussbaumer specializes in corneal transplantation including lamellar keratoplasty and Boston keratoprosthesis. In addition to her clinical work, she is an NIH-funded clinical researcher with a current focus on corneal ulcer treatment in India and Nepal with the Proctor Foundation International Research Team. She is the co-principle investigator of Descemet Endothelial Thickness Comparison Trial, an ongoing surgical randomized controlled trial to evaluate outcomes of Ultrathin Descemet Stripping Endothelial Keratoplasty versus Descemet Membrane Endothelial Keratoplasty.

Biosketch

Dr Bruce D. Gaynor, board certified in both internal medicine and ophthalmology, has broad experience in the treatment of infectious and inflammatory diseases of the eye. He has collaborated on the design, data collection and analysis of the ophthalmology portion of the Sjögren's International Clinical Collaborative Alliance (SICCA) study since 2010.



SUPPLEMENTAL FIGURE. Sjögren's International Collaborative Clinical Alliance ocular staining score form.