tvst

Article

Canonical Grading Scales of Corneal and Conjunctival Staining Based on Psychophysical and Physical Attributes

Trefford Simpson[1], Carolyn G. Begley[2], Ping Situ[2], Yunwei Feng[1], J. Daniel Nelson[3], Barbara Caffery[4], Clark Springs[5], and Sara Butterworth Connell[6]

[1] University of Waterloo, School of Optometry and Vision Science, Waterloo, Ontario, Canada
[2] Indiana University School of Optometry, Bloomington, IN, USA
[3] HealthPartners Institute, Bloomington, MN, USA
[4] Toronto Eye Care, Toronto, Ontario, Canada
[5] Indiana University Department of Ophthalmology, Indianapolis, IN, USA
[6] Clinical Trial Consultant, Iowa City, IA, USA

Correspondence: Trefford Simpson, University of Waterloo, School of Optometry and Vision Science, Waterloo, ON N2L3G1, Canada.
e-mail: trefford.simpson@uwaterloo.ca

Received: August 7, 2020
Accepted: July 7, 2021
Published: August 17, 2021

Keywords: grading scale; fluorescein staining; lissamine green staining; cornea; conjunctiva; staining; psychophysical scaling; spectrophotometry

Citation: Simpson T, Begley CG, Situ P, Feng Y, Nelson JD, Caffery B, Springs C, Connell SB. Canonical grading scales of corneal and conjunctival staining based on psychophysical and physical attributes. Transl Vis Sci Technol. 2021;10(9):17, https://doi.org/10.1167/tvst.10.9.17

Purpose: In this study, we apply psychophysical scaling principles based on physical (photometric) attributes of images to better understand the factors involved in clinician judgement of ocular surface staining and, using that knowledge, to develop photographic scales for the assessment of staining for dry eye (DE) and related conditions.

Methods: Subjects with noninfectious ocular surface staining were enrolled at five clinical sites. Following instillation of fluorescein, photographs of corneal staining were taken every 30 seconds for at least 5 minutes. The same procedure was followed for conjunctival staining after instillation of 2 µl of 1% lissamine green. A subset of the best corneal and bulbar conjunctival staining images were anonymized and a spectroradiometer measured photometric attributes (luminance and chromaticity). The images were scaled psychophysically by study investigators, who participated in constructing grading scales based on physical and psychophysical analyses. The final grading scales were refined following consultation with outside DE experts.

Results: Photographs were collected from 142 subjects (81% women), with an average age of 58 ± 17 years; 89% were diagnosed with DE. There was a monotonic relationship between between physical measurements and psychophysically scaled staining of both corneal (fluorescein) and bulbar (lissamine green) staining. Michelson contrast and u′ (chromaticity) accounted for 66% and 64% of the variability in the psychophysically scaled images of fluorescein corneal and lissamine green conjunctival staining, respectively.

Translational Relevance: This paper provides examples of the first ever clinically usable ocular surface staining scales validated using psychophysical scaling and the physical attributes (luminance and chromaticity) of the staining itself. In addition, it provides a generalizable method for the development of other clinical scales of ocular appearance.

Introduction

Vital stains have been used to evaluate ocular surface disease and damage since the late 19th and early 20th century.[1–3] Fluorescein dye is typically used to assess corneal staining, whereas rose bengal, and more recently, lissamine green[4–6] are most often used for evaluation of conjunctival staining. A number of scales have been developed using these dyes to assess dry eye (DE),[7–12] contact lens complications,[13–15] graft-versus-host disease,[16] and keratoconus.[17] However, despite the plethora of scales, there is unfortunately no widely accepted "gold-standard" ocular surface staining scale.

Current ocular surface staining scales use a variety of methods to judge the level of surface disease or damage. Some scales are based on assessing the intensity or density or both of stained areas,[7,9,14,17,18] whereas others evaluate the area of staining or quantify stained dots.[8,10,11,13] Corneal staining scales may also

add points to account for more severe DE disease.[10,11] For example, one scale increases the grade if central staining is present,[10] whereas another uses more than one scale to judge several aspects of staining, such as coalescence.[14] All of these scales have been created using clinical expertise to drive the attributes of the scale and to assign step-sizes between grades.

Another approach to devising biometric scales, such as ocular surface grading scales, is to utilize measurement theory to understand the clinical assessment of disease severity and to assign step sizes. The assignment of the number of centimeters to the height of an object is a common example of measurement. Sometimes these measures are more complex because a standardized reference is unavailable, so a simple measurement is, out of necessity, indirect, and is termed a soft measurement.[19–22] Judging the appearance of a clinical presentation, such as corneal staining, falls into this soft measurement realm and is considered complex because of the many characteristics of staining, such as area, depth, and intensity.[23–27] Psychophysics is the examination of the relationship between one's internal mental representation, such as the clinical grading of corneal staining, with external physical attributes, such as measurement of the luminance or area of staining. To our knowledge, the bulbar redness scale is the only ocular surface clinical scale that has been developed using these psychophysical principles to relate measured physical attributes to the scale.[28–31]

In this study, we applied psychophysical scaling principles to better understand the clinical gestalt underlying the grading of corneal staining and, using that knowledge, to develop scales for the assessment of ocular surface staining in DE and related conditions. Clinical investigators with expertise in grading corneal and conjunctival staining were used to provide and select initial reference images, to assist in developing initial scale attributes, to scale images using linear separation from no stain to worst staining, to select scale step number and exemplar images, and to verify/adjust the scale to optimize its utility. The scales reported in this paper are image-based to provide an accurate depiction of clinical staining references.

Methods

Figure 1 shows a flowchart of the procedures in this study.

Collection of Exemplar Images

The first step in this investigation involved setting up photographic and staining protocols for study investigators and training personnel. The five study investigators, (Carolyn Begley, OD, MS, Bloomington, IN, USA; Barbara Caffery, OD, Toronto, ON, CA, Clark Springs, MD, Indianapolis, IN, USA; Joseph Tauber, MD, Kansas City, MO, USA; and JD. Nelson, St Paul, MN, USA). chosen for their experience in grading ocular surface staining, participated in this phase of the study. Four of the sites were in private practice and one (Indiana University, Bloomington, IN, USA) was in a university setting. The Coordinating Center for the study was located at Indiana University, which also served as a clinical site. Following data collection, study investigators had access only to de-identified photographs and data.

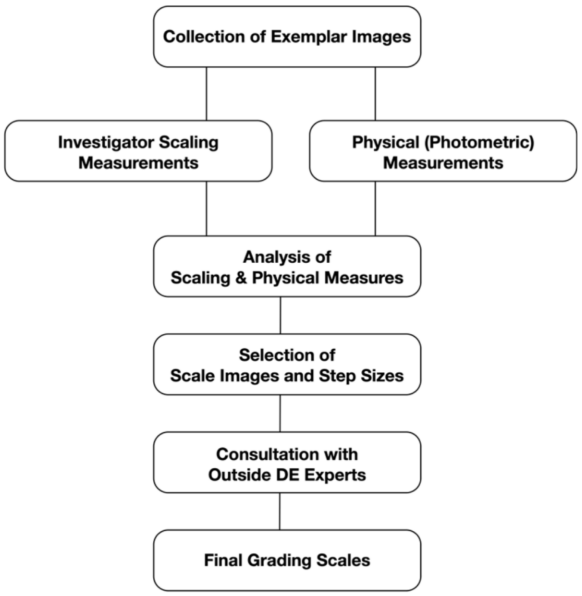

Figure 1. Flow chart for the procedures in this study.

Photography Procedures

The study investigator team met at Indiana University to set photographic protocols and train personnel, in order to provide uniformity in photography procedures. To further ensure uniformity of photography, all sites were equipped with identical slit lamp biomicroscopes (Topcon SL-D701 LED) and photography equipment (Eye Photo Systems EC 100 Imaging System) for use in the study.

Images of corneal staining were taken at 16× magnification at an angle of 25 to 30 degrees temporally with a cobalt blue filter over the illumination system and a Wratten #12 filter over the observation port of the slit lamp biomicroscope. Five microliters of 2% liquid sodium fluorescein dye (Premier Pharmacy Labs, Weeki Wachee, FL, USA) were instilled in the right eye and photographs were taken every 30 seconds for a period of 5 minutes or until the fluorescein dye was

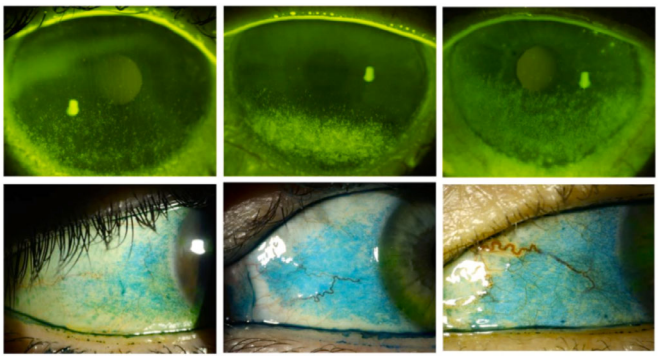

Figure 2. Representative photographs of corneal fluorescein staining (*top row*) and conjunctival lissamine green staining (*bottom row*) from this study.

visibly washed out. The same procedure was repeated for the left eye.

After an additional 5 minutes, to allow fluorescein dye to wash out of the eyes, 5 µl of 1% liquid lissamine green (Premier Pharmacy Labs) was instilled into the right eye. Photographs of the temporal conjunctiva and then nasal conjunctiva were taken at 10× magnification at an angle of 25 to 30 degrees every 30 seconds for 5 minutes or until the dye visibly washed out of the eye. No filters or diffusers were used. Identical procedures were repeated for the left eye. Sequential photographs were taken to address the uncertainty of the timing required to obtain the best photographs of staining.

Selection of Subjects for Photographs of Staining

Ethics approval was received by each site, two from institutional review boards (IRBs) and three from centralized IRBs. The study followed the tenets of the Declaration of Helsinki. Informed consent was obtained for each subject at all sites.

In order to acquire a wide range of corneal and conjunctival staining images, subjects with DE and other ocular surface conditions with corneal and/or conjunctival staining were included in the study at each clinical site. Subjects with corneal or conjunctival staining due to infectious diseases were excluded. Some subjects with no staining or minimal levels of staining were included to ensure that all levels of staining were represented in the final staining scales.

Images from each site were uploaded to a cloud-based central repository (Indiana University, IUBox.com) and renamed using a custom MATLAB program designed to assign a numerical name to each file that protected the identity of the subject. The site investigators were unaware of the naming procedure for each image file to avoid bias in later decisions on grading images or choosing images to include in the scale. The images (see examples in Fig. 2) were captured as high-resolution (5184 × 3456 horizontal and vertical pixels, respectively) jpeg images and stored in a de-identified manner on the Indiana University cloud server (IUBox).

Objective Photometric Protocol for Staining Evaluation

Contrast (Weber and Michelson), integrated luminance and chromaticity (u') were outcome measures. All images were measured using PR650 spectroradiometer (Photo Research Inc. USA). This instrument measures the radiometric, photometric, and colorimetric quantity of light. It was mounted on a tripod at a fixed measurement distance of 45 cm, which when centered (and orthogonal to screen surface for each image), was triggered manually. The measured images were displayed on a 17-inch Acer AL716 computer monitor (1280 × 1024 active matrix color TFT LCD monitor; 500:1 contrast ratio; luminance midpoint 250 DC/m2), each 33 × 27 cm (horizontal and vertical length, respectively), surrounded by a gray strip, and the computer monitor was covered above and on the left and right sides. At the set distance, the integration area of the measuring circle of the spectroradiometer was 21.6 mm^2.

For each corneal and conjunctival image, the luminance was measured in an area with the highest estimated staining and a nonstained area. The relative intensity, or difference between the luminance of the area of highest corneal and conjunctival staining compared to nonstained regions, was calculated using Weber and Michelson contrasts. An additional analysis was conducted on the conjunctival images. The chromaticity (u'), was measured using Lu'v' values to give the coordinates of color in the CIE Lu'v' color space. The measurements and position of the recordings were noted for each image.

Weber and Michelson contrasts were calculated from the stained and unstained luminance measurements of both corneal and conjunctival images. For each session, the luminance of the middle of the upper part of the surrounding grey strip was measured in order to monitor and control for the variation in screen luminance during the 2-month period over which measurements were made.

In addition, a subset of the corneal fluorescein images was measured (while displayed on the same monitor) in a way that provided an integrated luminance. Integrated luminance provides an overall spatial average of luminance over a large area. The spectroradiometer was moved to seven meters to allow the measuring area on the instrument to contain a large enough area to include a zone from just below the lower

pupil margin (excluding pupil autofluorescence) to just above the lower limbus (excluding the tear meniscus fluorescence).

Psychophysical Scaling Protocol for Corneal and Conjunctival Staining

A representative subset of corneal and conjunctival images (50) was selected (based on general photographic qualities of focus, coverage, and range of fluorescence) to represent a range of staining and 9 identical sets of 18.6 × 12.4 cm (horizontal and vertical, respectively) prints on medium gloss paper were made using a Kodak APEX 7000 dye-sublimation printer. All images were numbered randomly and all information as to subject or site removed. Each chosen image was printed nine times in succession. These were cut and a set distributed to each of the study investigators.

Study investigators, who were working independently, were instructed to set up 2 separate meter sticks end-to-end to form a 2-meter continuous number line. They were instructed to place all of the corneal staining images in a pile and then systematically place the pictures along the meter stick number line, from least severe to most severe staining, according to their own clinical judgments of staining severity. Once that was complete, the study investigators measured the position of each image's bottom left-hand corner relative to the meter stick number line and recorded its position in centimeters (from 0 to 200) on the back of the photograph. The photographs were then returned to the Coordinating Center and the position data were entered into a spreadsheet.

This procedure was repeated for conjunctival lissamine green staining photographs that included both nasal and temporal images.

Scale Development and Refinement/Tuning

Analysis of Photometric Measurements and Investigator Scaling

Study investigator scalings were returned to Indiana University where correlations between objective photometric measures of luminance and chromaticity, Michelson and Weber contrast and the investigator distance representations of staining were performed. This exploratory data analysis was carried out using R[32] and JASP[33] statistical software. Exploratory analyses comprised exploratory and confirmatory factor analyses of corneal and conjunctival psychophysically (practitioner judgment) scaled setting and physical measurements. Once it was determined how the variables were related to each other, additional correlations between untransformed and transformed variables were used to identify those attributes with strongest associations. Ordinary and weighted least squares were used to derive the relationships between the selected photometric predictor and psychophysically scaled outcome variables.

Selection of Scale Images and Step Size

Based on the results of the investigator scaling and the plot of the objective photometric measurements, potential corneal and conjunctival grading scales with a wide range of grading steps (ranging from 3 to 11) were prepared. The representative images were selected along the horizontal line from the ordinate (as in Figs. 3, 4) at the appropriate step. For example, a 3-point scale would be made up of images along the horizontal lines extending from 0, 100, and 200 (as rated by each investigator on the 2-meter stick number line) on the ordinate.

One month after the final set of scaling data were obtained, the study investigators reconvened at Indiana University to determine which images would be used for the scale and to determine the number of step sizes that represented clinical significance in their judgment.

Review of Scales by Outside DE Experts

Following the development of the initial grading scales, advice and comment was sought from outside DE experts, who were experienced in grading ocular surface staining. (Penny Asbell, MD, Memphis, TN, Deborah Jacobs MD, MS, Boston, MA, Brett King, OD, Bloomington, IN, Jerry Paugh, OD, PhD, Fullerton, CA, Stephen Pflugfelder, MD, Houston, TX) The outside DE experts were sent copies of the initial grading scales along with the same corneal and conjunctival images used at the investigator meeting. The outside DE experts were instructed to independently grade the corneal and conjunctival images according to the proposed photographic grading scales and then to use the scales clinically to assess five patients with DE. When completed, a 1-hour interview with each outside DE expert was conducted (and recorded), querying both the ease of use of the scales on the images of corneal and conjunctival staining, as well as its ease of use clinically. Critical commentary and suggestions were also invited.

Refinement of Corneal and Conjunctival Scales

The comments and suggestions of the outside DE experts were reviewed by the study investigator team and presented for further comment in videoconferences. Investigator comments and consensus were then used to select final images for the corneal and

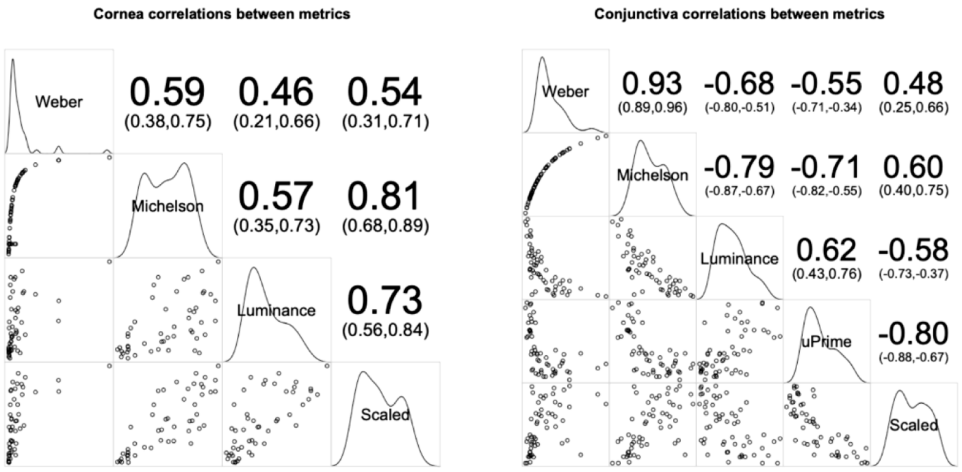

Figure 3. Correlograms showing relationships between Weber contrast, Michelson contrast, luminance, chromaticity (u') and mean study investigator scaled severity position. The left panel is for fluorescein stained corneal images and right panel is lissamine green stained conjunctival images. Legend: uPrime is chromaticity (u'), scaled is the placement of the images on the 200 meter scale, luminance is integrated luminance.

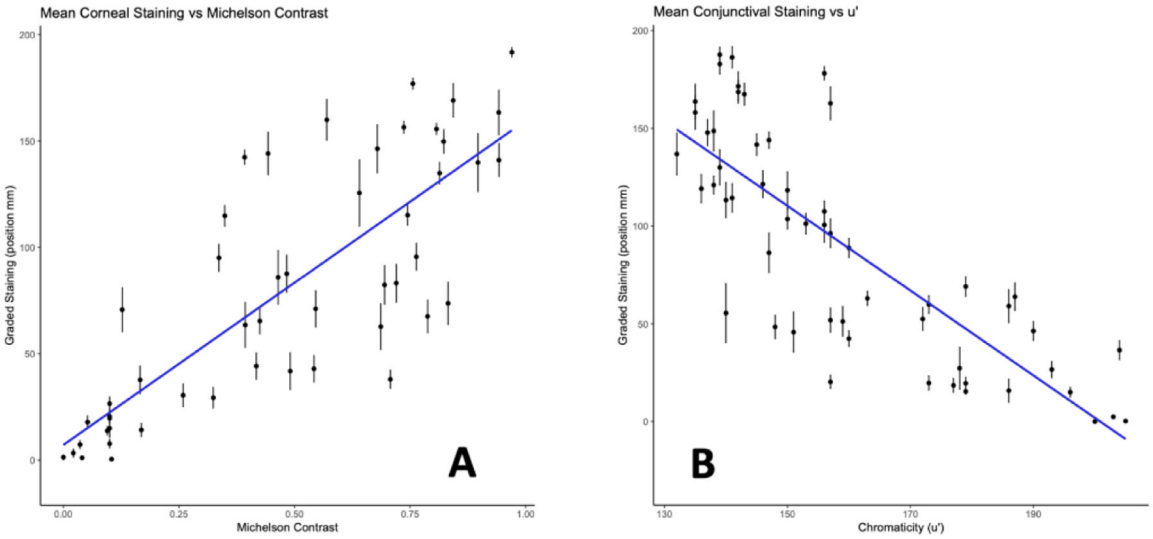

Figure 4. (**A**) Mean (± standard error) image position scaled to represent corneal fluorescein stain study investigator scaling (0–200, ordinate) versus Michelson contrast (0–1.0). (**B**) Mean (± standard error) image position scaled to represent conjunctival lissamine green stain study investigator scaling (0–200, ordinate) versus chromaticity (u') (0.130–0.20). For both graphs, the blue line is the ordinary least squares linear regression fit to the data and the gray band is the 95% confidence interval.

conjunctival staining scales based on both clinical judgment and photometric measures.

Results

Subjects

A total of 142 subjects were enrolled at the 5 clinical sites. The demographic information and clinical diagnoses are listed in the Table. The majority of subjects in this study were white women over the age of 40 years, with 80.2% having a primary ocular diagnosis of DE disease or Sjögren's syndrome. Another 8.4% had DE associated with other conditions, as listed in the Table. Sixteen subjects without a DE diagnosis and no ocular pathology were included to provide images with no or minimal ocular surface staining.

Photography

A total of 9358 corneal and conjunctival photographs were collected from five clinical sites.

A representative subset of 50 corneal and 50 conjunctival images was selected based on the selection criteria outlined in the Methods section.

Table. Demographics and Diagnoses of Subjects in This Study

Subject Demographics		
Age (mean ± standard deviation)	57.6 ± 16.7 years	
Sex	Male	27 (19%)
	Female	115 (81%)
Ethnicity	Asian	9 (5.6%)
	Black/African American	6 (4.2%)
	Hispanic/Latino	2 (1.4%)
	White	119 (83.8%)
	Other	6 (4.2%)
Primary ocular diagnoses	DE disease	81 (57%)
	Sjögren's syndrome	33 (23.2%)
	Lacrimal gland removal	3 (2.1%)
	DE disease secondary to refractive or other corneal surgery, pterygia	6 (4.2%)
	Glaucoma associated DE disease	3 (2.1%)
	Other	16 (11.3%)

Corneal and Conjunctival Photography

Figure 2 shows an example of selected images from a series of corneal fluorescein and conjunctival lissamine green staining from subjects in this study.

Photometric Measurements and Psychophysical Scaling

The correlograms in Figure 3 show how the physical measures (Weber and Michelson contrast), integrated luminance and chromaticity (u') were correlated to each other, as well as to the mean study investigator scaled severity position for each corneal and conjunctival image (left and right panels, respectively). As is illustrated, the highest correlations for corneal staining images were between Michelson contrast and corneal fluorescein study investigator severity scaling (r = 0.81, 95% confidence interval [CI] = 0.68 to 0.89) and between chromaticity (u') and conjunctival lissamine green study investigator intensity scaling (r = −0.80, 95% CI = −0.86, to −0.67).

Figure 4A shows the relationship between the objective photometric contrast assessment of corneal fluorescein staining and the scaled position of the images, as set by the study investigators on a 200 cm scale. As is apparent, there is a linear relationship between the two variables, Pearson r = 0.81, implying that 66% of the variability (r^2) in the study investigator scaled settings can be accounted for by the Michelson contrast of the image.

Figure 4B shows the negative relationship between the objective assessment of bulbar conjunctival lissamine green staining chromaticity (u') and the scaled position of the images, as set by the study investigators on a 200 cm scale. As is apparent, there is a monotonic relationship, Pearson r = −0.80, implying that 64% of the variability (r^2) in the study investigator scaled settings can be accounted for by the chromaticity (u') metric.

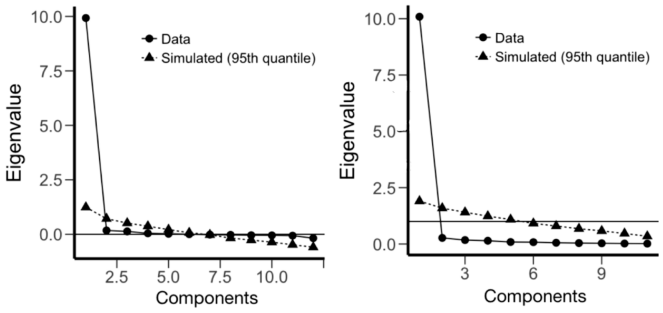

Figure 5. Scree plots for the corneal and conjunctival principal component analyses (left and right panels, respectively). These show that the vast majority of the variance can be accounted for by the first principal component and that none of the eigenvalues of the other components is greater than 1, again, pointing to one dimension accounting for the variance/covariance in the data.

Psychometric Properties of the Scale

An exploratory analysis of scaling by the study investigator group (Fig. 5) showed that the scaling of the images (both corneal and conjunctival) was essentially unidimensional (i.e. based on one objective feature of the individual stains). Scatterplot matrices highlighted the physical measurements most strongly associated with the study investigator scaling. Thus, the photometric analyses indicated that the psychophysical scaling data were well described by the predictors

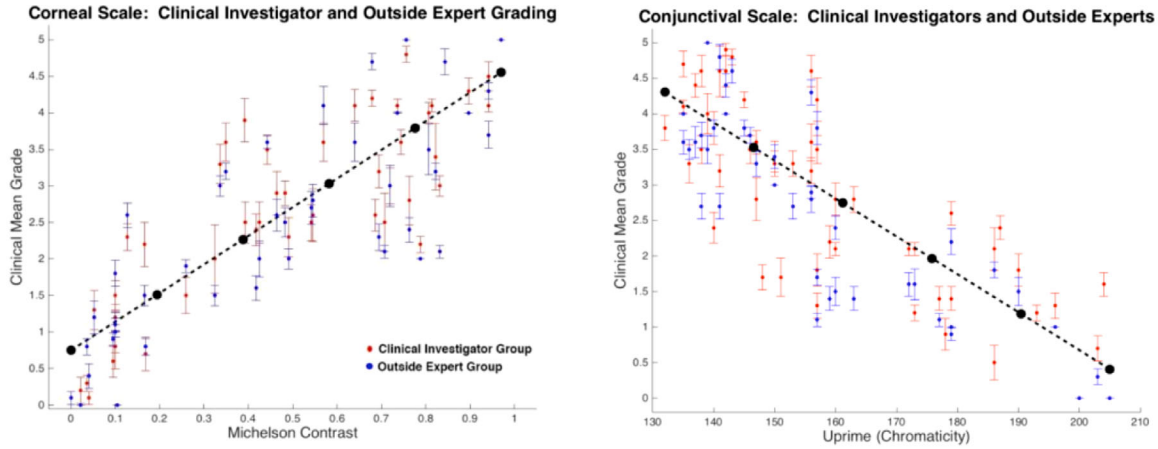

Figure 6. (**A**) Plot of final clinical grading of photographs of corneal staining by the study investigators (*red*) and the outside DE experts (*blue*) versus objective measurements. (**B**) Plot of clinical grading of photographs of conjunctival staining by the study investigators and the outside DE experts versus objective measurements. Error bars show the standard error of grading by each group. The black dots on the dashed line represent equally spaced points in 0 to 5 grading scales.

shown in Figure 4, inasmuch as approximately two thirds of the scaled image variability was accounted for by the contrast and the chromaticity predictor for corneal and conjunctival staining, respectively. The high association with Michelson contrast on the corneal scale illustrated that the relative intensity of fluorescein staining was strongly correlated with corneal grading, whereas the chromaticity (the quality of color independent of intensity) of the lissamine green, was strongly correlated with conjunctival grading.

Selection of Scale Images and Step Size

One month after the final set of scaling data were obtained and the analyses of photometric measurements and study investigator scaling were completed, the study investigators regrouped to discuss and determine which images would be used for the final staining scales and to decide upon the number of step sizes that represented clinical significance in their judgement. Because the psychophysical grading analysis (see below) demonstrated that study investigator staining judgements of severity followed a single objective quality of the staining, this unidimensionality of the grading suggested that one scale for cornea and one for conjunctiva would be indicated.

During the meeting, a 0 to 5 scale was selected for both the cornea and conjunctiva by study investigator group consensus. The final images for these scales were selected by the study investigators along the horizontal line from the ordinate axes of Michelson contrast for cornea and chromaticity (u') for conjunctiva (as in Figs. 4A, 4B), equally spaced to obtain 0 to 5 grading steps. At the same meeting, after the photographs were selected for each scale, the study investigators used the newly derived scales and regraded the same set of corneal and conjunctival images.

Consultation With Outside DE Experts

Copies of the initial grading scales along with the same corneal and conjunctival images that were previously scaled and photometrically measured were then sent to the five outside DE experts who were not part of the study. They used the newly derived scales and graded the same set of corneal and conjunctival images. They were given simple instructions to "grade the photographs of corneal and conjunctival staining using the scales provided." In addition, they gave feedback on the scales and their clinical use.

The outside DE experts made a number of comments concerning the utility of the scales. Only the comments made by at least two of the outside DE experts will be reported here. Most found the corneal and conjunctival scales straightforward and simple to use in a clinical setting. Several suggested that the scales should consider the area of staining and that corneal staining over the pupil was more important than other locations. It was also pointed out that the superior cornea was not visible in the photographs. Some of the individual photographs in the scale were criticized as too difficult to see or too similar to the next grade. Showing the photographs in the scale at a larger magnification was suggested. During the interviews, which were conducted after the photographic grading, all of the outside DE experts were highly interested in our findings that the relative intensity (contrast) of

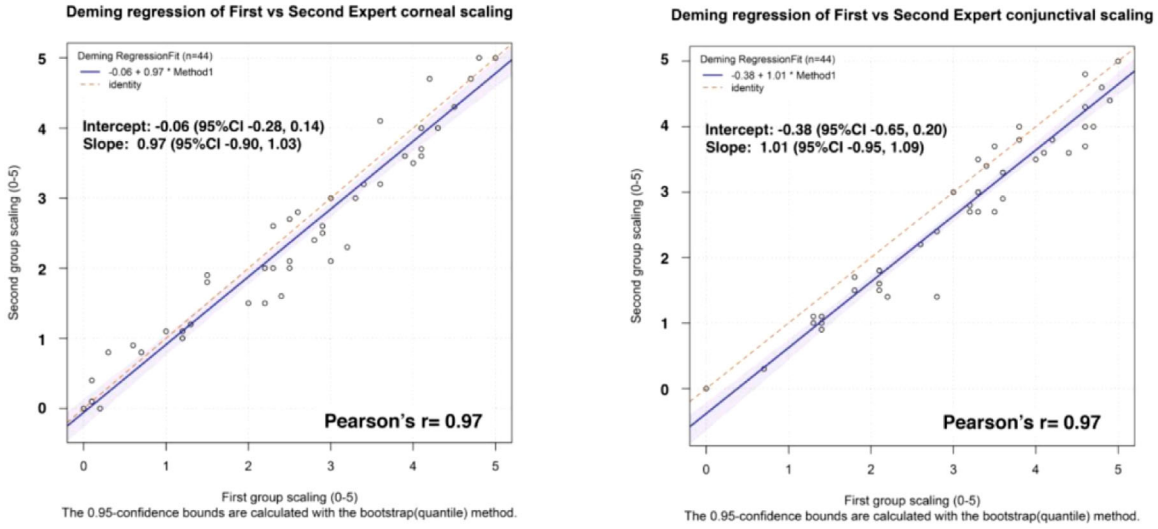

Figure 7. (**A**) Deming regression between corneal staining grading of study investigators and outside DE experts. (**B**) Deming regression between conjunctival staining grading of study investigators and outside DE experts. The 95% confidence bounds were calculated with the bootstrap (quantile) method.

corneal staining and color of conjunctival staining were most strongly correlated with the attributes used for grading. Several suggested that, if necessary, the same scale could also be used in corneal zones[7,13] if knowing the exact area and location of staining was critical.

The gradings of both the study investigators and the outside DE experts, using the initial scale, was compared in Figures 6A and 6B. The dashed line with black dots represents equally spaced grades along the objective scales. It is clear from Figure 6 that the grades from both groups cluster along these lines. Figures 7A and 7B shows the Deming regression between study investigators and outside experts of their clinical gradings of corneal and conjunctival staining, respectively.

Refinement/Fine Tuning of the Scale

After the interviews with the outside DE experts, the study investigator group reconvened over video conference and discussed all points raised by the outside DE experts. In particular, the critiques of individual photographs in each scale were considered. For example, the study investigators agreed with the outside DE experts that staining in the original grades 1 and 2 corneal scale was too inferior and difficult to see. These photographs were replaced with photographs chosen by the study investigator group that had similar photometric measures (Michelson contrast), but better visibility. For the conjunctival scale, only one photograph was replaced (grade 4) in response to comments by the outside DE experts that grades 4 and 5 conjunctival staining looked too similar. The new photograph was chosen by the study investigator group and was very close to the original photograph's chromaticity value.

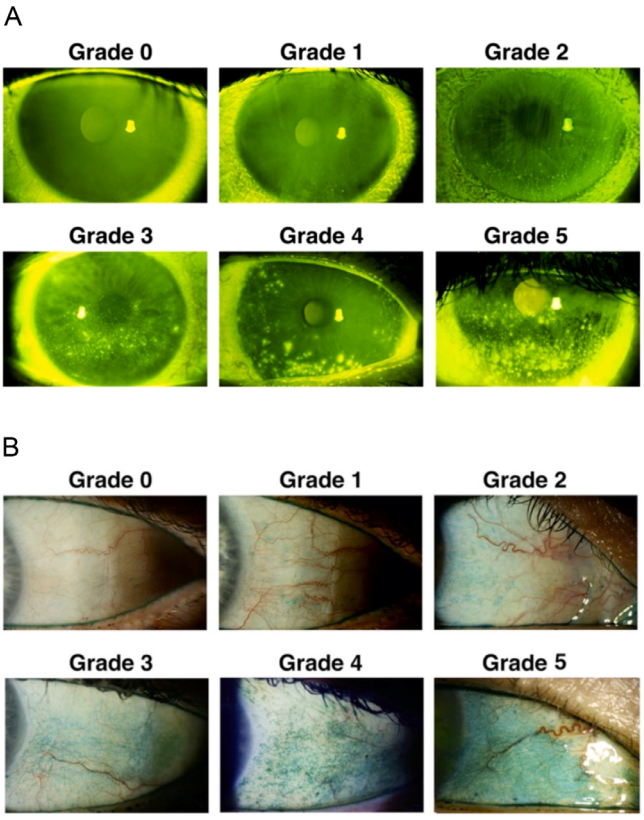

Figure 8. (**A**) The final corneal grading scale. (**B**) The final conjunctival grading scale.

The corneal and conjunctival scales are shown in Figure 8. The image investigator scaling and

photometric measure associations were determined again. Corneal staining grade was linearly associated with Michelson contrast, Pearson r = 0.96, whereas the conjunctival staining grade was linearly negatively associated with chromaticity (u'), Pearson r = −0.93.

Discussion

From the development of our two scales there emerged important empirical "epiresults." We showed that both corneal and conjunctival images were correlated with physical (photometric) measurements. This provides support for the notion that the clinical estimations of these outcomes could be superseded by physical measurements, although more study of this concept is warranted. We also demonstrated that staining can be psychophysically scaled repeatably between observers and directly related to physical measurements. This psychophysical approach can be used to better understand the clinical gestalt underlying the judgment of the grading of corneal and conjunctival staining. The utility of this method has been demonstrated previously in the development of a psychophysically and photometrically derived scale of bulbar conjunctival redness.[28–31]

We used psychophysical scaling methods[34] to understand how clinicians judged the separation of physically measured corneal and bulbar conjunctival staining images to develop grading scales for ocular surface staining. This method showed that the clinicians' psychophysically scaled staining grade was most strongly correlated with contrast of staining (r = −0.81) for fluorescein corneal staining and chromaticity (r = 0.80) for lissamine green staining of the conjunctiva in their clinical judgments of the severity of staining. One interpretation of these empirical findings is that a single (different) dimension may account for the scaling of corneal and conjunctival staining. This putative unidimensionality of clinical grading decision supports our development of single grading scales, rather than assessing multiple factors,[9–11,14] for corneal and conjunctival staining. These are the first photographic ocular surface grading scales based on physical attributes developed using psychophysical methods.

Exactly how the psychophysically scaled staining intensities were related to the contrast (for the cornea) and the chromaticity (for the conjunctiva) is speculative. We do not believe that observers are directly calculating Michelson contrast, or computing u' chromaticity. The chromaticity of an unstained conjunctiva derives from the underlying sclera and as more staining occurs, so the chromaticity will systematically change in the direction of the chromaticity of the blueish dye. As staining increases, chromaticity moves in the direction of more saturated blue. This is exactly what is illustrated in the right panel of Figure 4. It, of course, is unclear what judgment is being made clinically, because it is unlikely that absolute chromaticity is being quantified by the clinicians. Rather, we assume, the extent of the staining is numerically similar to the physical measurement of chromaticity, using the transformation/scale we selected (Lu'v'). The psychophysical relationship between corneal staining and contrast is less clear, but the data suggest that a relative judgment of darkness to lightness (captured by the Michelson contrast metric) is either used by or related to the psychophysical scaling judgments made by the group of experts in this experiment.

There are many previous scales that have been developed to grade ocular surface staining based on clinician assessment of severity. Some use drawings or paintings of the staining,[7,8,10,15] some are textual descriptions,[12,14,18] and others use digital methods to measure the area of staining over the cornea.[35–41] The development of our scales differ in that it provides a photometric, quantifiable basis for the scale and also incorporates the relationship between the physical measures and clinician grading. In this study, we found that a single photometric measurement accounts for approximately two thirds of the variability in clinician scaling. For corneal staining, it is the Michelson contrast measured between the most intensely stained region and the unstained background, whereas for the conjunctiva, the psychophysical variability is accounted for by the chromaticity metric u'. Other aspects of grading corneal staining may be explored in future studies.

The unidimensionality in the corneal scale is perhaps somewhat surprising inasmuch as other scales of corneal staining use additional vectors to quantify the staining (e.g. type, extent, location, and depth)[14] or add points for additional severity measures, such as staining over the visual axis or filaments.[10] Miyata et al.[9] use both the area and intensity of corneal staining as equal factors in their staining grid. However, with the simple instruction to grade the severity of staining, our investigators were able to independently position photographs, according to staining severity, along a two-meter ruler-scale with high interobserver correlations. This unidimensionality was supported by the principal component analyses (of investigator scaling and photometric measures) as seen in Figure 5 as well as the R-squared of the relationships between investigator scaling and the contrast or chromaticity predictors, accounting

for approximately two thirds of scaling variability (Figs. 4A, 4B).

These results demonstrate that this method may be used to develop clinical scales, and that the accuracy of these scales can be enhanced by their direct relationships to physical measures. The "validity" of these scales is automatic because they can be directly calibrated to simple, physical measurements. Therefore, complex validation procedures, commonly necessary for novel clinical scales, are less important. The clinical utility of these scales still, of course, does require empirical demonstration in future studies, as do some other aspects of their performance. For example, we showed that the application of the novel scales produces high interobserver reliability. That demonstration of repeatability along with the accuracy (how the scaling relates to the physical measures) indicates that accuracy and repeatability assessment is technically possible and is perhaps something to be aimed for in clinical scale development.

Another aspect of this scale development is that physical measures, such as the Michelson contrast and chromaticity or u' found in this study, could be used as a tool to quantify ocular surface staining from photographs in clinical trials. Other digital staining methods have been proposed for the cornea, but all are based on measuring the area of staining, not the intensity.[35–41] The results of this study suggest that the relative intensity of fluorescence and the chromaticity of the blue-green color of lissamine green may be substantial components used by clinicians to grade ocular surface staining. Thus, in digitally captured images, measurement of relative intensity over the cornea and chromaticity over the conjunctiva could potentially be used to grade or measure ocular surface staining.

However, use of photography to digitally measure and quantify ocular surface staining has some pitfalls, most of which are connected to capturing good quality photographs that adequately demonstrate the physical qualities of the staining. Part of this project was to formalize those methods and study them, in addition to devising grading scales. We found that many aspects of capturing photographs of corneal and conjunctival staining were problematic. One issue was fluorescein dye-concentration quenching, in which higher dye concentrations takes longer to reach maximum tear fluorescence.[42] It is important to note that a Wratten filter should be used for best photographic results. Another issue that occurred was fading of conjunctival staining lissamine green over time.[43,44] Although we selected the best corneal staining images from subjects for this study, some images in each subject's series showed less intense corneal staining due to quenching of fluorescein dye.[45,46] It is possible that use of darker images, affected by quenching or loss of fluorescein dye, may have been graded differently by clinicians, thus affecting the correlations between grading and the physical measures found in this study. Issues such as these could affect grading scales that are based on the luminance or chromaticity of ocular surface staining.

The scales developed in this study can be used together or independently, as the clinical need arises, and may be used for ocular surface conditions other than DE. The physical basis of these scales may reduce the problems associated with clinical trials where ocular surface staining is an important end point. However, it is important to note that this study can be regarded as a first step in developing ocular surface staining scales based on psychophysical principles and physical attributes, perhaps similar to the process used in the first steps of questionnaire development.[46] A large scale clinical trial, including more investigators than in this study, will be an important step to test the methods and scales developed in this study against existing ocular surface grading scales.

In this study, we developed ocular surface grading scales based on psychophysical and photometric measures using rigorous photography procedures. The scales developed in this project incorporate an understanding of the underlying basis of clinician grading of the severity of ocular surface disease using vital dyes. The first ocular staining scale, developed by van Bijsterveld[18] in 1969 was also based on grading the intensity of staining, although the dye he used was rose bengal. Thus, the concept of intensity as an important metric in grading corneal and conjunctival staining is not a new one. It is perhaps ironic that in some ways we appear to have come full circle in understanding parameters important in assessment and grading of vital dye staining of the ocular surface.

Acknowledgments

The authors thank Joseph Tauber for his early participation in this study and Penny Aspbell, Deborah Jacobs, Steven Pflugfelder, Jerry Paugh, and Brett King for their assistance in refinement of the scales. Additional thanks go to Eye Photo Systems for their custom design and assistance in the use of the EC 100 Imaging System.

Supported by Santen Pharmaceutical Co., Ltd.

Disclosure: **T. Simpson**, None; **C.G. Begley**, None; **P. Situ**, None; **Y. Feng**, None; **J.D. Nelson**, None; **B. Caffery**, None; **C. Springs**, None; **S.B. Connell**, None

References

1. Pflüger, Zu Ernahrung der cornea. *Klin Monatsbl Augenheilk*. 1882;20:69–81.
2. Straub M. Fluoresceinlösung al ein diagnostiches Hilfmittel fur Hornhauterkrankungen. *Cent F Augenheilk*. 1888;12:75–77.
3. Schirmer O. Studien zur Physiologie und Pathologie der Tränenabsonderung und Tränenabfuhr. *Albr Graefes Arch für Klin Exp Ophthalmol*. 1903;56:197–291.
4. Amaki S, Ogata T, Konishi M, Shimizu K, Yamada M, Mashima Y. Lissamine green B staining in the evaluating of keratoconjunctivitis sicca. *Folia Ophthalmologica Japonica*. 1999;50(7):536–539.
5. 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(1):61–64.
6. Machado LM, Castro RS, Fontes BM. Staining patterns in dry eye syndrome: rose bengal versus lissamine green. *Cornea*. 2009;28(7):732–734.
7. Lemp MA. Report of the National Eye Institute/Industry workshop on Clinical Trials in Dry Eyes. *CLAO J*. 1995;21(4):221–232.
8. Bron AJ, Evans VE, Smith JA. Grading of corneal and conjunctival staining in the context of other dry eye tests. *Cornea*. 2003;22(7):640–650.
9. Miyata K, Amano S, Sawa M, Nishida T. A novel grading method for superficial punctate keratopathy magnitude and its correlation with corneal epithelial permeability. *Arch Ophthalmol*. 2003;121(11):1537–1539.
10. Whitcher JP, Shiboski CH, Shiboski SC, et al. A simplified quantitative method for assessing keratoconjunctivitis sicca from the Sjögren's Syndrome International Registry. *Am J Ophthalmol*. 2010;149(3):405–415.
11. De Paiva CS, Pflugfelder SC. Corneal epitheliopathy of dry eye induces hyperesthesia to mechanical air jet stimulation. *Am J Ophthalmol*. 2004;137(1):109–115.
12. Abelson MB, Rimmer DW, Ousler G. Diagnosing dry eye: It's now a fine art. *Ophthalmol Manag*. 2016;20(April):48–50.
13. Caffery BE, Josephson JE. Corneal staining after sequential instillations of fluorescein over 30 days. *Optom Vis Sci*. 1991;68(6):467–469.
14. Terry RL, Schnider CM, Holden BA, et al. Cclru standards for success of daily and extended wear contact-lenses. *Optom Vis Sci*. 1993;70(3):234–243.
15. Efron N. Grading scales for contact lens complications. *Ophthal Physiol Opt*. 1998;18(2):182–186.
16. Ogawa Y, Kim SK, Dana R, et al. International Chronic Ocular Graft-vs-Host-Disease (GVHD) Consensus Group: Proposed Diagnostic Criteria for Chronic GVHD (Part I). *Scientific Reports*. 2013;3:3419–3419.
17. Barr JT, Schechtman KB, Fink BA, et al. Corneal scarring in the Collaborative Longitudinal Evaluation of Keratoconus (CLEK) Study: baseline prevalence and repeatability of detection. *Cornea*. 1999;18(1):34–46.
18. van Bijsterveld OP. Diagnostic tests in the Sicca syndrome. *Arch Ophthalmol*. 1969;82(1):10–14.
19. Rossi GB. A probabilistic theory of measurement. *Measurement*. 2006;39(1):34–50.
20. Rossi GB. Measurability. *Measurement*. 2007;40(6):545–562.
21. Rossi GB. Cross-disciplinary concepts and terms in measurement. *Measurement*. 2009;42(9):1288–1296.
22. Mari L, Ugazio E. Preliminary analysis of validation of measurement in soft systems. *Journal of Physics: Conference Series*. 2010;238(1):012026.
23. Mari L, Lazzarotti V, Manzini R. Measurement in soft systems: Epistemiological framework and a case study. *Measurement*. 2009;42(2):241–253.
24. Finkelstein L. Widely, strongly and weakly defined measurement. *Measurement*. 2003;34:39–48.
25. Finkelstein L. Problems of measurement in soft systems. *Measurement*. 2005;38:267–274.
26. Finkelstein L. Widely-defined measurement-An analysis of challenges. *Measurement*. 2009;42(9):1270–1277.
27. Sparrow JM, Bron AJ, Brown NA, Ayliffe W, Hill AR. The Oxford Clinical Cataract Classification and Grading System. *Int Ophthalmol*. 1986;9(4):207–25.
28. Schulze MM, Hutchings N, Simpson TL. The perceived bulbar redness of clinical grading scales. *Optom Vis Sci*. 2009;86(11):E1250–E1258.
29. Schulze MM, Jones DA, Simpson TL. The development of validated bulbar redness grading scales. *Optom Vis Sci*. 2007;84(10):976–83.
30. Schulze MM, Hutchings N, Simpson TL. The conversion of bulbar redness grades using psychophysical scaling. *Optom Vis Sci*. 2010;87(3):159–167.
31. Macchi I, Bunya VY, Massaro-Giordano M, et al. A new scale for the assessment of conjunctival bulbar redness. *Ocul Surf*. 2018;16(4):436–440.
32. R_Development_Core_Team. *R: A Language and Environment for Statistical Computing*. Vienna,

Austria: R Foundation for Statistical Computing: 2010.
33. JASP Team (2020). JASP (Version 0.14) [Computer software].
34. Gescheider GA. *Psychophysics: the fundamentals*. 3rd ed. Mahwah, NJ: L. Erlbaum Associates; 1997.
35. Pritchard N, Young G, Coleman S, Hunt C. Subjective and objective measures of corneal staining related to multipurpose care systems. *Cont Lens Anterior Eye*. 2003;26(1):3–9.
36. Wolffsohn JS, Purslow C. Clinical monitoring of ocular physiology using digital image analysis. *Cont Lens Anterior Eye*. 2003;26(1):27–35.
37. Tan B, Zhou Y, Svitova T, Lin MC. Objective Quantification of Fluorescence Intensity on the Corneal Surface Using a Modified Slit-lamp Technique. *Eye Contact Lens*. 2013;39(3):239–246.
38. Chun YS, Yoon WB, Kim KG, Park IK. Objective assessment of corneal staining using digital image analysis. *Invest Ophthalmol Vis Sci*. 2014;55(12):7896–7903.
39. Rodriguez JD, Lane K, Ousler GW, Angjeli E, Smith L, Abelson M. Automated Grading System for Evaluation of Corneal Superficial Punctate Keratitis Associated with Dry Eye. *Invest Ophthalmol Vis Sci*. 2015;56(4):2340–2347.
40. Amparo F, Wang H, Yin J, Marmalidou A, Dana R. Evaluating Corneal Fluorescein Staining Using a Novel Automated Method. *Invest Ophthalmol Vis Sci*. 2017;58(6):BIO168–BIO173.
41. Guillon M, Dumbleton K, Theodoratos P, Patel T, Karkkainen T, Moody K. Objective Assessment of Ocular Surface Response to Contact Lens Wear in Presbyopic Contact Lens Wearers of Asian Descent. *Eye Contact Lens*. 2018;44(3):182–189.
42. Webber WR, Jones DP. Continuous fluorophotometric method of measuring tear turnover rate in humans and analysis of factors affecting accuracy. *Med Biol Eng Comput*. 1986;24(4):386–392.
43. Hamrah P, Alipour F, Jiang S, Sohn JH, Foulks GN. Optimizing evaluation of Lissamine Green parameters for ocular surface staining. *Eye (Lond)*. 2011;25(11):1429–1434.
44. Foulks GN. Challenges and pitfalls in clinical trials of treatments for dry eye. *Ocul Surf*. 2003;1(1):20–30.
45. Begley C, Caffery B, Chalmers R, Situ P, Simpson T, Nelson JD. Review and analysis of grading scales for ocular surface staining. *Ocul Surf*. 2019;17:208–220.
46. Abetz L, Rajagopalan K, Mertzanis P, Begley C, Barnes R, Chalmers R, et al. Development and validation of the impact of dry eye on everyday life (IDEEL) questionnaire, a patient-reported outcomes (PRO) measure for the assessment of the burden of dry eye on patients. *Health Qual Life Outcomes*. 2011;9:111.