The Ocular Surface
A JOURNAL OF REVIEW, LIVING LABORATORY SCIENCE, CLINICAL SCIENCE, AND CLINICAL PRACTICE

Optimising subjective grading of corneal staining in Sjögren's syndrome dry eye disease

James S. Wolffsohn, Alberto Recchioni, Olivia A. Hunt, Sònia Travé Huarte, Giuseppe Giannaccare, Marco Pellegrini, Marc Labetoulle

PII: S1542-0124(24)00035-1

DOI: https://doi.org/10.1016/j.jtos.2024.03.005

Reference: JTOS 957

To appear in: *Ocular Surface*

Received Date: 3 December 2023

Revised Date: 13 February 2024

Accepted Date: 12 March 2024

Please cite this article as: Wolffsohn JS, Recchioni A, Hunt OA, Travé Huarte Sò, Giannaccare G, Pellegrini M, Labetoulle M, Optimising subjective grading of corneal staining in Sjögren's syndrome dry eye disease, *Ocular Surface* (2024), doi: https://doi.org/10.1016/j.jtos.2024.03.005.

Optimising subjective grading of corneal staining in Sjögren's syndrome dry eye disease

James S. Wolffsohn,[1] Alberto Recchioni,[1,2,3] Olivia A. Hunt,[1] Sònia Travé Huarte,[1] Giuseppe Giannaccare,[4] Marco Pellegrini,[5] Marc Labetoulle.[6]

1) School of Optometry, Health and Life Sciences, Aston University, Birmingham, UK
2) Academic Unit of Ophthalmology, Institute of Inflammation and Ageing, University of Birmingham, UK
3) Birmingham and Midland Eye Centre, Sandwell & West Birmingham NHS Trust, Birmingham, UK
4) Eye Clinic, Department of Surgical Sciences, University of Cagliari, Cagliari, Italy
5) Department of Translational Medicine, University of Ferrara, Ferrara, Italy
6) Ophthalmology Département, Hôpitaux Universitaires Paris-Saclay, APHP, Université Paris-Saclay, IDMIT Infrastructure, Fontenay-aux-Roses Cedex, France

Funding: This study was funded by MC2 Therapeutics Ltd, Hoersholm, Denmark

Abstract:

Aim: To assess whether smaller increment and regionalised subjective grading improves the repeatability of corneal fluorescein staining, and to determine the neurological approach adopted for subjective grading by practitioners.

Methods: Experienced eye-care practitioners (n=28, aged 45 ± 12 years), graded 20 full corneal staining images of patients with mild to severe Sjögren's syndrome with the Oxford grading scheme (both in 0.5 and 1.0 increments, globally and in 5 regions), expanded National Eye Institute (NEI) and SICCA Ocular Staining Score (OSS) grading scales in randomised order. This was repeated after 7-10 days. The digital images were also analysed using ImageJ for comparison.

Results: The Oxford grading scheme was similar with whole and half unit grading (2.77vs2.81,$p=0.145$), but the variability was reduced (0.14vs0.12,$p<0.001$). Regional grade was lower ($p<0.001$) and more variable ($p<0.001$) than global image grading (1.86±0.44 for whole increment grading and 1.90±0.39 for half unit increments). The correlation with global grading was high for both whole ($r=0.928$,$p<0.001$) and half increment ($r=0.934$,$p<0.001$) grading. Average grading across participants was associated with particle number and vertical position, with 74.4-80.4% of the linear variance accounted for by the digital image analysis.

Conclusions: Using half unit increments with the Oxford grading scheme improve its sensitivity and repeatability in recording corneal staining. Regional grading doesn't give a comparable score and increased variability. The key neurally extracted features in assigning a subjective staining grade by clinicians were the number of discrete staining locations (particles) and how close to the vertical centre was their spread, across all three scales.

Keywords: corneal staining; subjective grading; objective grading; Sjögren's syndrome; dry eye disease.

Background:

Corneal staining with fluorescein dye has been long recognised as a biomarker of ocular surface disease [1, 2]. The Tear Film and Ocular Surface Society (TFOS) Dry Eye Workshops (DEWS) included ocular surface staining as a marker of a loss of homeostasis of the tear film, which together with symptomology, constitutes one of the criteria for the diagnosis of dry eye disease [3]. The ODISSEY European Consensus Group agreed that following diagnosis, symptom-based assessment and corneal fluorescein staining are sufficient to determine the severity of dry eye disease in the majority of patients [4]. The Asia Dry Eye Society's stated definition of dry eye: "Dry eye is a multifactorial disease characterized by unstable tear film causing a variety of symptoms and/or visual impairment, potentially accompanied by ocular surface damage" also emphasises the importance of fluorescein staining [5]. Additionally, both the American–European Consensus Group (AECG) criteria [6] and the 2016 American College of Rheumatology/European League Against Rheumatism (ACR–EULAR) criteria[7], that are the most widely accepted classification criteria for primary Sjögren's syndrome, include fluorescein staining assessment.

It is important for follow-on care that damage to the ocular surface is accurately assessed and recorded. Grading scales, with broad increments, were developed for ocular conditions such as corneal damage in the 1990´s to provide reference images against which observed damage could be recorded in a easy and straight forward way. These scales, such as the Oxford grading scheme [8], are well accepted in clinical practice and have been used by some international eye-care specialists for over 30 years. It has been proposed, following modelling, that the sensitivity of grading can be improved by interpolating to 0.1 unit steps between grade images, rather than reporting 1 unit steps [9]; however, sub-unit grading is rarely adopted by practitioners [10]. It has recently been demonstrated for the grading of ocular redness, that half unit sub-increments can increase sensitivitiy at least as much as using 0.1 unit steps [11], but this approach has not been investigated for corneal staining.

Scales with a limited number of steps typically have good repeatability, but lack sensitivity [12]. Dividing the ocular surface into regions could aid in relating the staining to clinical impact such as symptomology [13]; however no study to date has explored how a global score relate to

regional grading beyond anecdotal reporting of differences between three clinicians [14]. It has also been suggested that zonal grading can help in the differential diagnosis of ocular surface disease, with more temporal conjunctival staining found in Sjögren's syndrome than other forms of keratoconjunctivitis sicca [15].

The lack of a single, widely accepted, “gold standard” staining scale [13], has an important impact on the endpoints of clinical trials of ocular surface treatments. Of the most commonly adopted scales, the National Eye Institute (NEI)/Industry scale [16] adopts the approach of grading 5 corneal zones and scoring the zones by the density of stained dots on a 0–3 scale. The Oxford grading scheme [8] also grades the density of stained dots within the cornea and nasal and temporal conjunctiva, but introduced the concept of log unit increases in the number of stained dots between grades. The Sjögren's International Collaborative Clinical Alliance (SICCA) Ocular Staining Score (OSS) scale [17] includes this feature of coalescence by adding a single grade point for each of the following features present on the cornea: confluent staining, filaments or staining in the pupillary area. The OSS also advocated using fluorescein dye to stain the cornea and lissamine green to stain the conjunctiva, with the scores from each equally weighted in the overall score, although no scientific evidence was provided to justify this approach [17] and interobserver consistency was poor [18].

The aim of this study was to determine whether subjective grading to smaller increments and regionalised grading with established scales improves the sensitivity and repeatability of corneal staining recording. The study also compared grading between the expanded National Eye Institute / Industry Workshop Corneal Fluorescein Staining scale (expanded NEI), Oxford grading scheme and corneal part of the SICCA Ocular Staining Score (SICCA OSS) to investigate their comparability and repeatability. Finally, the approach adopted for subjective grading by practitioners was identified by correlating investigator ratings with objective image analysis of staining dot counting, staining area, intensity and location.

Method

The study was given a favourable opinion by the Aston University Research Ethics Committee and followed the tenets of the Declaration of Helsinki. Participants were experienced eye-care practitioners (n=28, aged 45 ± 12 years, 6 female, qualified for 19 ± 11 years, 15 ophthalmologists and 13 optometrists, examining 237 ± 360 ocular surface patients a month [median 95, range 15 to 1,600]), involved in corneal staining as part of their practice, recruited from professional body lists (Tear Film and Ocular Surface Society and European Dry Eye Society), who gave written informed consent after the nature and risks of the study had been explained to them. Training was provided in the form of sample images to grade using the electronic format followed by discussion on how they differed from a group of five experienced graders (non-participants in the study), repeated with a second set of images. They were provided with an electronic file with a series of 20 randomly sequenced full corneal images of patients with mild to severe dry eye disease owing to Sjögren syndrome with positive fluorescein corneal staining imaged with blue light and a yellow observation filter. They were asked to view them for around 30s each and to grade them with the Oxford grading scheme, expanded NEI and SICCA OSS scales in randomised order. For the Oxford grading scheme they were required to report the image with the nearest whole number increment from the grading scale reference images to the global amount of staining, and in central, superior, inferior, nasal and temporal regions (see Figure 1a). They also graded the resequenced images again (altered in a Latin square approach) with the Oxford grading scheme to the nearest half unit increment, in a randomized sequence (Figure 1b). The eye care practitioners then repeated the complete exercise a second time 7-10 days later in the opposite questionnaire order, but with the image sequence again randomized. One image of the 20 was repeated to allow intrasession repeatability to be assessed; reviewing of previous scores was not permitted.

Figure 1: Grading report form examples.

Image Analysis was performed using ImageJ (v1.53t http://imagej.nih.gov/ij). Pixel to millimeter calibration was achieved by imaging a ruler with the same slit lamp and settings as the image was captured with. Color thresholding was applied to sample the green pixels in HSB color space and Huang thresholding was applied, with a saturation and brightness in the range 20-80% found

to best highlight the area of observed staining. The cornea was manually segmented and particle analysis applied to identify the number of particles, the average size (mm^2), the proportion $>0.1mm^2$, the proportion of total staining area consisting of particles $>0.1mm^2$, the proportion of corneal area covered by staining, the average intensity (8-bit green percentage), average horizontal position of the centroid of staining (with 100% being on the inferior limbus) and distribution (the average distance between particles).

Data Analysis

Based on a 0.4 SD for subjective grading [19], a sample size of 24 clinicians was required to allow the detection of a 0.25 difference in mean with 80% power ($p<0.01$ significance level) (G*Power, National Institute for Health) [20]. As corneal staining subjective grading scales are ordinal in nature, non-parametric related-sample Wilcoxon signed rank test and Spearman's rank correlations were conducted with $p<0.05$ taken as significant. Multivariate analysis was conducted to determine the contribution of objectively extracted staining features to subjective grading using stepwise and enter methods (SPSS Statistics v29.01, IBM, USA). Spearman rank correlations were also performed for an individual graderbetween each of the grading scales.

Results

Despite initial training, one experienced grader used a 4 or 5 for all images with the Oxford grading scheme except one at both visits, resulting in an average score 20% higher than the next highest grader and therefore their results were excluded from the analysis.

Grading Increment

The average grade with the Oxford grading scheme was similar with whole and half unit grading (2.77 vs. 2.81, p=0.145), but the variability with the former was reduced (average standard deviation 0.14 vs. 0.12, p<0.001). When the grading was repeated 7-10 days later, the average staining grade was 0.08 grade units lower with a 95% confidence interval of 0.19 when grading to whole units, whereas the second repeat was almost identical (0.01 higher) with a 95% confidence interval of 0.17 when grading to 0.5 increments, with a significant difference between them (p=0.007, Figure 2). The intrasession repeatability was -0.09 ± 0.05 (mean ± 95% confidence interval) for whole unit grading (p=0.006), but reduced to 0.02 ± 0.01 for half unit grading (p=0.824).

Figure 2: Bland-Altman plot of mean versus difference in repeated grading with 0.5 of 1.0 increment units of 20 corneal staining images of patient with Sjogren's syndrome.

Regional grading

The average regional grade was lower (p<0.001) and the variability higher (p<0.001) than global image grading (1.86 ± 0.44 for whole increment grading and 1.90 ± 0.39 for half unit increments). The correlation with global grading was high for both whole unit (r=0.928, p<0.001) and half increment (r=0.934, p<0.001) grading. Regional grading (1.0 increments) increased the intersession repeatability to ±1.06 units (95% confidence interval), which was larger as a proportion of the scale, to global grading (5.3% versus 3.9%).

Comparison between scales

The Oxford grading scheme (1.0 increments) average grade for all the participants for each image was strongly associated with that of the OSS (r=0.802, p<0.001) and NEI (r=0.912, p<0.001). The OSS and NEI were also strongly correlated (r=0.888, p<0.001). However, for an

individual grader, the correlations between scales was much more variable (Oxford vs NEI: r=0.070 to 0.668; Oxford vs OSS: r=0.050 to 0.546; NEI vs OSS: r=0.019 to 0.726). The repeatability as a percentage of the scale range was greatest for the OSS (16.6%) which was higher than the NEI scale (13.4%; 1.0 increments; p=0.022) and lowest with the NEI (9.4%, p<0.001). The 0.5 increment Oxford grading scheme (11.9%) was also more variable than the NEI (p=0.015).

The intrasession repeatability was -0.09 ± 0.05 units (mean ± 95% confidence interval) for the Oxford grading scheme (1.0 increments), -0.07 ± 0.59 units for the OSS and -0.04 ± 1.56 units for the NEI scale. The intersession repeatability was 0.19 units (95% confidence interval) for the Oxford grading scheme (1.0 increments), 0.37 units for the OSS and 0.74 units for the NEI scale.

Effect of Experience

The years of qualification was generally negatively associated with absolute mean difference from the mean with each image for the Oxford scale (1.0 increments: r=-0.382, p=0.049; 0.5 increments: r=-0.476, p=0.012), NEI scale (r=-0.262, p=0.186) and OSS (r=-0.354, p=0.070). However, the number of gradings performed per month was not associated with absolute mean difference from the mean with each image for the Oxford scale (1.0 increments: r=-0.230, p=0.248; 0.5 increments: r=-0.143, p=0.477), NEI scale (r=-0.202, p=0.311) and OSS (r=-0.057, p=0.777).

Features associated with subjective grading

The correlation between each of the objective staining metrics and mean subjective grading score (average of both completitons) are presented in table 1. Average grading across participants was associated with particle number (accounting for 47.1/48.9% of the variance) and vertical position (accounting for a further 17.2/16.2%) with a total of 75.4/78.7% of the linear variance accounted for by the digital image analysis for the Oxford (0.5/1.0 increments) grading scheme. Average grading across participants was associated with vertical position (accounting for 45.0% of the variance) and particle number (accounting for a further 13.3%) with a total of 74.4% of the linear variance accounted for by the digital image analysis for the OSS scale. Average grading across participants was associated with particle number (accounting for 49.0% of the variance) and

vertical position (accounting for a further 14.3%) with a total of 80.4% of the linear variance accounted for by the digital image analysis for the OSS scale.

Table 1: Means and correlations of objectively analysed features influencing eye care practitioner subjective grading of corneal staining images Note vertical position scaled from 0 (superior limbus) to 100 (inferior limbus). * $p<0.05$; ** $p<0.01$, *** $p<0.001$

Discussion

Water-soluble dyes are excluded from the normal epithelium by tight junctions, the plasma membranes and the surface glycocalyx. Shed cells, or those with a compromised glycocalyx barrier, have been hypothesized to 'stain' through transcellular entry and diffusion across defective tight junctions [21]. Due to its low molecular weight compared to other ocular dyes, fluorescein can spread from initial sites of punctate staining initially by a paracellular route and then by transcellular diffusion [21]. This can be minimised by reducing the amount of fluorescein applied [22]. Fluorescein staining is best visualised following the minimum application of dye, illuminated with a blue light with a peak around 495nm, observed through a yellow filter with a sharp cut off around 500nm, between 20-160s after instillation [2].

The first aim of this study was to determine whether subjective grading to smaller increments and regionalised grading with established scales improves the repeatability of corneal staining recording. While the average grade with the Oxford grading scheme was similar with whole and half unit grading, allowing studies that use either approach to be directly compared, the variability among observers within a visit and across two visits was statistically reduced with half unit grading. This supports a previous study on other types of ocular physiological feature grading, that grading to half increments is more repeatable than whole unit grading [11]. While the difference may not be considered clinically significant, the overall benefits of half increment grading outweigh any disadvantages. Dividing the ocular surface into regions has been adopted by many clinical studies as a potentially more accurate way to grade ocular physiological features such as staining [13]; the present study was unique in systematically assessing how a global score relates to regional grading. Interestingly, the diameter of the central zone has only been specified (beyond stating zones should be of similar size [23, 24]) by Woods and colleagues [12], who stated the central zone was to have a diameter of half that of the cornea. The assigned average regional grade was lower than the global image approach for both whole and half unit grading. This would suggest a tendency for clinicians to base their overall grade on the intensity of staining in a localized area, rather than as a percentage of the whole ocular surface. The correlation between global and regional grading was strong, accounting for around 86% of the variance for both whole and half unit grading. However, the 95% confidence interval was statistically higher for regional grading, even when scaled for the higher range of scores

generated, which will require a larger sample size to be powered to detect differences between groups by adopting regionalized grading. Hence while the differences many not be considered clinically significant, the disadvantages of this approach seem to outweigh any advantages.

The present study also compared three commonly used staining grading scales. Grading scales for ocular surface staining adopt different approaches to what defines severity. The expanded NEI [16] and SICCA OSS [17] scale grades increase with the number of dots and the actual numbers for each grade are stipulated in the SICCA OSS Scale. The authors of the Oxford grading scheme [8] do not recommend counting punctate staining dots, but the number of dots in each grade increase in a logarithmic nature as the grade increases; drawings depict the increasing density of dots with each grade, unevenly distributed within each zone, clustering and eventually coalescing (Grade IV) around the limbus across the interpalpebral zone. Coalescent rather than punctate staining is seen in DED with more conjunctival damage and with lower reflex tear volume as found in Sjogren syndrome patients [25]. Mucus plaques (containing mucus, epithelial cell and proteinaceous and lipoidal material) of varying size and shape, attached to the corneal epithelium, which stain with fluorescein dye have been described in patients with accompanying system disease such as Sjogren's Syndrome. This sign is more common when filaments are present [26]. The possible mechanisms responsible for the manifestation of coalescent patches of staining are the increase in MUC16 concentration in tears due to inflammation induced increased shedding, the accumulation of mucins due to delayed tear clearance, the reduction in repulsive forces from the corneal surface due to both of these factors and the increased friction due to reflex tear deficiency [25]. The terms "confluence" or "coalescence" of stained dots are included in several scales. In the CCLRU scale [23], coalescence is a category of stain, while in the SICCA OSS Scale [17], a point is added for confluent staining of the cornea. Therefore, it is clear that the local density of staining, which may be so dense as to be coalesced or confluent, is considered an important aspect of grading scales for dry eye and other ocular surface conditions [17, 23, 27].

The expanded NEI scale is not linear as grades 0.5 to 1.5 are attributed to a non-linear increase in micropunctate staining spots, 2.0 and 2.5 to moderate macropunctate area, 3.0 and 3.5 to clumped macropunctate area and 4.0 to diffuse macropunctate stain. A pharmaceutical company

has created another modified version of the NEI scale in 0.5 grade incredments with a linear increase in punctate dots up to grade 3, but still with coalesced areas a requirement of grades 3.5 and 4.0; however, the reliability and repeatability was no better than the previous expanded NEI scale [28]. The CORE scale [12] aimed to generate continuous data to facilitate parametric analysis, but still attributed a type of staining (micropunctate, macropunctate, coalescent and patch staining) as anchors to point values; staining type (1-100), extent (1-100) and depth (1–4, based on the timing and extent of stromal glow) are graded and multiplied together (max 40,000). This is repeated in 5 zones to create Zone Staining Scores. However, the 15 separate grades are time consuming to score and is likely to decrease inter-grader concordance. In practice, the modified Oxford grading scheme has been shown to be subjective and observer dependent, besides being susceptible to poor reproducibility and high inter-observer and intra-observer variability in contrast to computer-assisted, objective digital analysis [29-31].

Due to these differences in scoring range and approach, staining grading scales cannot be directly compared. However, the average grading score correlation between the group of clinicians was strong (ranging from r^2= 0.65 to 0.83). However, for an individual grader, the correlations between scales was much more variable (from r^2= 0.01 to 0.53), which would be statistically significant (80% power) with the number of graders involved [32]. This could, in part, have been due to differing amounts of grader experience with the individual scales, although consistently those with more years of experience were closer to the mean score for each image with each scale. In addition, this result was calculated after one clinician's grades were removed due to their very different approach, thus highlighting that individual clinician's can interpret grading scoring guidance very differently even after training. A limitation of the study was the time the clinician took to make their grading decision was not monitored. When assessing repeatability as a percentage of the scale range, the NEI was the most repeatable and the OSS the least repeatable. If the NEI reflects the findings with the Oxford grading scheme, its regional grading approach will have reduced the average score and hence the variability between measures would be expected to be lower (although this was unexpectantly not the case with the Oxford grading scheme analysis). The additional grades that can be added to the OSS on the presence of certain

features being noted (excluding filaments in this study due to the static nature of the images being graded) is likely to have resulted in the poorer repeatability, as proposed previously [18].

Finally, the approach adopted for subjective grading by practitioners was identified by correlating investigator ratings with objective image analysis of staining. Techniques for objective analysis of corneal staining have been developed and tested using: edge detection and color extraction [33, 34]; an observer-dependent thresholding technique [35]; luminance correction across the image [36]; green channel isolation and thresholding, along with size thresholds for particles [37]; intensity green thresholding [30]; green channel isolation and automated contrast enhancement, convoluted background subtraction, auto-threshold "triangle-white" following manual corneal selection with size and circularity thresholds for particles identified applied by an ImageJ macro [31]; and a combination of the difference of Gaussians (DoG), edge detection for morphologic properties of corneal erosions, and the red-green-blue (RGB) systems and hue-saturation-value (HSV) color model for detection of colour [38]. The effect of prior image enhancement with a median filter, Otsu thresholding, and a contrast-limited adaptive histogram equalization has been investigated [38], but the correlation to subjective grading using a number of different scales remained strong (r=0.85 to 0.92). The expanded NEI scale correlated slightly more strongly with objective measurement (r=0.90) than the Oxford grading scheme (r=0.85), but the subjective grading of the two scales was not compared directly [38]. The corneal staining index (the ratio between the staining and total corneal area) has been found to be strongly correlated with the expanded NEI and Oxford (accounting for 60 and 68% of the variance) and showed good interobserver reliability; the circularity and roundness of staining spots (manually traced and quantified objectively) were significantly higher in patients with ocular graft versus host disease compared to those diagnosed with Sjogren's Syndrome, with a distinguishing sensitivity and specificity of 65% and 60% respectively for circularity and 80% and 70% for roundness [29]. However, while objective grading of staining has advantages, it relies on high quality image capture which can be influenced by practitioner skill, instrumentation as well as the iris colour and features.

Chun and colleagues acknowledged that despite a strong correlation between their objective punctate staining count and the subjective grading by two experienced ophthalmologists, their

342 objective strategy “could not account for the human eye’s detailed perception of corneal staining
343 morphology characteristics, such as coalescence and dispersion”[38]. Therefore the objective
344 analysis conducted in this study chose to analyse not only the number of particles detected, but
345 also their average size, intensity of fluorescence, the covered area (in relation to the corneal
346 area), the proportion and relative area covered by coalescence (defined as a detected area of
347 staining greater than $0.1mm^2$, based on the average punctate dot being 15-27μm [39]), vertical
348 centration of the staining within the cornea and spread across the cornea. With all of the
349 subjective scales, the average clinical subjective grade related principally on the number of
350 particles (accounting for 43.5 to 74.8% of the variance), vertical centration (accounting for 40.1
351 to 63.2% of the variance), fluorescent intensity (accounting for 39.7 to 46.9% of the variance)
352 and corneal coverage (accounting for 22.5 to 43.2% of the variance). However, these metrics are
353 inter-related, such as more particles and greater coalesence will be related to the corneal area
354 covered by staining, and as the staining is more centred within the cornea the distribution is
355 likely to increase. Hence linear multivariate analysis identified that the main neurally extracted
356 features in assigning a subjective staining grade were the number of discrete staining locations
357 (particles) and how close to the vertical centre was their spread, across all three scales. As the
358 images had a wide range of punctate and coalescent staining between them, this might suggest
359 that separate scoring criteria for coalescence may not be required, allowing the scale grade
360 decriptions to be more linear. The overall variance accounted for was similar in this study to that
361 reported by Chun and colleagues for the Oxford grading scheme (75.4% versus 72.3%) and NEI
362 scale (80.4% versus 81.5%, both finding the NEI subjective grading to be slightly more strongly
363 associated with objective staining analysis [38].
364
365 In conclusion, using half unit increments with the Oxford grading scheme improves its
366 repeatability in recording corneal staining, whereas regional grading increased variability. The
367 three commonly used staining grading scales (the Oxford grading scheme, SICCA OSS and
368 expanded NE)I have different scale ranges, so their mean scores are not comparable; however,
369 the mean score of a group of clinicians with each of the scales are strongly correlated. Individual
370 clinician approaches to grading with each of the scales are quite variable and therefore it is
371 important to use multiple subjective graders in clinical trials. Finally, despite the limitations of
372 applying objective image analysis to complex staining patterns, the correlation with subjective

grading is strong and demonstrates that the key features extracted in assigning a subjective staining grade by clinicians were the number of discrete staining locations and how close to the vertical centre was their spread; this novel finding may inform more linear grading scale design in the future.

Disclosures

JSW received funding from MC2 Therapeutics Ltd for this investigator initiated study. None of the other authors have any declarations of interest to declare.

References

[1] Barabino S, Aragona P, di Zazzo A, Rolando M. Updated definition and classification of dry eye disease: Renewed proposals using the nominal group and Delphi techniques. European journal of ophthalmology. 2021;31:42-8.
[2] Peterson RC, Wolffsohn JS, Fowler CW. Optimization of anterior eye fluorescein viewing. American journal of ophthalmology. 2006;142:572-5.
[3] Wolffsohn JS, Arita R, Chalmers R, Djalilian A, Dogru M, Dumbleton K, et al. TFOS DEWS II Diagnostic methodology report. Ocular Surface. 2017;15:539-74.
[4] Baudouin C, Aragona P, Van Setten G, Rolando M, Irkec M, Benitez del Castillo J, et al. Diagnosing the severity of dry eye: a clear and practical algorithm. The British journal of ophthalmology. 2014;98:1168-76.
[5] Tsubota K, Yokoi N, Shimazaki J, Watanabe H, Dogru M, Yamada M, et al. New Perspectives on Dry Eye Definition and Diagnosis: A Consensus Report by the Asia Dry Eye Society. The ocular surface. 2017;15:65-76.
[6] Vitali C, Bombardieri S, Jonsson R, Moutsopoulos HM, Alexander EL, Carsons SE, et al. Classification criteria for Sjogren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. Ann Rheum Dis. 2002;61:554-8.
[7] 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 Sjogren's Syndrome: A Consensus and Data-Driven Methodology Involving Three International Patient Cohorts. Arthritis Rheumatol. 2017;69:35-45.
[8] 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.
[9] Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effects of scaling. Investigative ophthalmology & visual science. 1991;32:422-32.
[10] Wolffsohn JS, Naroo SA, Christie C, Morris J, Conway R, Maldonado-Codina C. Anterior eye health recording. Cont Lens Anterior Eye. 2015;38:266-71.
[11] Vianya-Estopa M, Nagra M, Cochrane A, Retallic N, Dunning D, Terry L, et al. Optimising subjective anterior eye grading precision. Cont Lens Anterior Eye. 2020;43:489-92.
[12] Woods J, Varikooty J, Fonn D, Jones LW. A novel scale for describing corneal staining. Clinical ophthalmology. 2018;12:2369-75.
[13] Begley C, Caffery B, Chalmers R, Situ P, Simpson T, Nelson JD. Review and analysis of grading scales for ocular surface staining. The ocular surface. 2019;17:208-20.
[14] Begley CG, Barr JT, Edrington TB, Long WD, McKenney CD, Chalmers RL. Characteristics of corneal staining in hydrogel contact lens wearers. Optom Vis Sci. 1996;73:193-200.
[15] Caffery B, Simpson T, Wang S, Bailey D, McComb J, Rutka J, et al. Rose bengal staining of the temporal conjunctiva differentiates Sjögren's syndrome from keratoconjunctivitis sicca. Investigative ophthalmology & visual science. 2010;51:2381-7.
[16] Lemp MA. Report of the National eye institute/ industry workshop on clinical trials in dry eyes. CLAO J. 1995;21:221-32.
[17] Whitcher JP, Shiboski CH, Shiboski SC, Heidenreich AM, Kitagawa K, Zhang S, et al. A simplified quantitative method for assessing keratoconjunctivitis sicca from the Sjogren's Syndrome International Registry. American journal of ophthalmology. 2010;149:405-15.
[18] Rasmussen A, Stone DU, Kaufman CE, Hefner KS, Fram NR, Siatkowski RL, et al. Reproducibility of Ocular Surface Staining in the Assessment of Sjögren Syndrome-Related

Keratoconjunctivitis Sicca: Implications on Disease Classification. ACR open rheumatology. 2019;1:292-302.
[19] Efron N. Grading scales for contact lens complications. Ophthalmic Physiol Opt. 1998;18:182-6.
[20] Erdfelder E, Faul F, Buchner A. GPOWER: A general power analysis program. Behavior Research Methods, Instruments, & Computers. 1996;28:1-11.
[21] Bron AJ, Argueso P, Irkec M, Bright FV. Clinical staining of the ocular surface: mechanisms and interpretations. Progress in retinal and eye research. 2015;44:36-61.
[22] Courrier E, Renault D, Kaspi M, Marcon A, Lambert V, Garcin T, et al. Micro-instillation of fluorescein with an inoculation loop for ocular surface staining in dry eye syndrome. Acta Ophthalmol. 2018;96:e140-e6.
[23] Terry RL, Schnider CM, Holden BA, Cornish R, Grant T, Sweeney D, et al. CCLRU standards for success of daily and extended wear contact lenses. Optom Vis Sci. 1993;70:234-43.
[24] Jalbert I, Sweeney DF, Holden BA. The characteristics of corneal staining in successful daily and extended disposable contact lens wearers. Clin Exp Optom. 1999;82:4-10.
[25] Komai S, Yokoi N, Kato H, Komuro A, Sonomura Y, Kinoshita S, et al. Clinical Implication of Patchy Pattern Corneal Staining in Dry Eye Disease. Diagnostics (Basel). 2021;11.
[26] Fraunfelder FT, Wright P, Tripathi RC. Corneal mucus plaques. American journal of ophthalmology. 1977;83:191-7.
[27] Barr JT, Schechtman KB, Fink BA, Pierce GE, Pensyl CD, Zadnik K, et al. Corneal scarring in the Collaborative Longitudinal Evaluation of Keratoconus (CLEK) Study: baseline prevalence and repeatability of detection. Cornea. 1999;18:34-46.
[28] Sall K, Foulks GN, Pucker AD, Ice KL, Zink RC, Magrath G. Validation of a Modified National Eye Institute Grading Scale for Corneal Fluorescein Staining. Clinical ophthalmology. 2023;17:757-67.
[29] Pellegrini M, Bernabei F, Moscardelli F, Vagge A, Scotto R, Bovone C, et al. Assessment of Corneal Fluorescein Staining in Different Dry Eye Subtypes Using Digital Image Analysis. Translational vision science & technology. 2019;8:34.
[30] Amparo F, Wang H, Yin J, Marmalidou A, Dana R. Evaluating Corneal Fluorescein Staining Using a Novel Automated Method. Investigative ophthalmology & visual science. 2017;58:Bio168-bio73.
[31] Kourukmas R, Roth M, Geerling G. Automated vs. human evaluation of corneal staining. Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie. 2022;260:2605-12.
[32] Hulley SB, Cummings SR, Browner WS, Grady D, Newman TB. Designing clinical research : an epidemiologic approach. 4th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2015.
[33] Wolffsohn JS, Purslow C. Clinical monitoring of ocular physiology using digital image analysis. Cont Lens Anterior Eye. 2003;26:27-35.
[34] Peterson RC, Wolffsohn JS. Objective grading of the anterior eye. Optom Vis Sci. 2009;86:273-8.
[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;26:3-9.
[36] 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;39:239-46.

[37] Rodriguez JD, Lane KJ, Ousler GW, 3rd, Angjeli E, Smith LM, Abelson MB. Automated Grading System for Evaluation of Superficial Punctate Keratitis Associated With Dry Eye. Investigative ophthalmology & visual science. 2015;56:2340-7.
[38] Chun YS, Yoon WB, Kim KG, Park IK. Objective assessment of corneal staining using digital image analysis. Investigative ophthalmology & visual science. 2014;55:7896-903.
[39] Courrier E, Lépine T, Hor G, Fournier C, He Z, Chikh M, et al. Size of the Lesions of Superficial Punctate Keratitis in Dry Eye Syndrome Observed With a Slit Lamp. Cornea. 2016;35:1004-7.

Metric	Range across images	Assocation with Oxford grading scheme (0.5 increments)	Assocation with Oxford grading scheme (1.0 increments)	Assocation with NEI scale	Assocation with OSS scale
N^{o} of particles,	8-4232	.865***	.851***	.652**	.765***
Average size of particles (mm^2)	0.01-0.16	-.094	-.020	.146	.022
Proportion >0.1mm^2 (%)	0.0-91.7	-.213	-.187	.057	-.007
Proportion coalesced (%)	0.0-98.7	.176	.160	.350	.362
Corneal coverage (%)	0.1-37.1	.657**	.657**	.475*	.587**
Average intensity (%)	8.1-26.4	-.647**	-.685***	-.630**	-.636**
Vertical position	31.7-88.9	-.795***	-.775***	-.640**	-.714***
Distribution (mm)	0.1-0.4	.435	.507*	.430	.553*

Table 1: Means and correlations of objectively analysed features influencing eye care practitioner subjective grading of corneal staining images Note vertical position scaled from 0 (superior limbus) to 100 (inferior limbus). * $p<0.05$; ** $p<0.01$, *** $p<0.001$