

Automated Grading System for Evaluation of Superficial Punctate Keratitis Associated With Dry Eye

John D. Rodriguez,¹ Keith J. Lane,¹ George W. Ousler III,¹ Endri Angjeli,¹ Lisa M. Smith,¹ and Mark B. Abelson^{1,2}

¹Ora, Inc., Andover, Massachusetts, United States

²Harvard Medical School, Boston, Massachusetts, United States

Correspondence: John D. Rodriguez, Ora, Inc., 300 Brickstone Square, Andover, MA 01810, USA; jrodriguez@oraclinical.com.

Submitted: July 24, 2014

Accepted: February 19, 2015

Citation: Rodriguez JD, Lane KJ, Ousler GW III, Angjeli E, Smith LM, Abelson MB. Automated grading system for evaluation of superficial punctate keratitis associated with dry eye. *Invest Ophthalmol Vis Sci*. 2015;56:2340–2347. DOI:10.1167/iov.14-15318

PURPOSE. To develop an automated method of grading fluorescein staining that accurately reproduces the clinical grading system currently in use.

METHODS. From the slit lamp photograph of the fluorescein-stained cornea, the region of interest was selected and punctate dot number calculated using software developed with the OpenCV computer vision library. Images ($n = 229$) were then divided into six incremental severity categories based on computed scores. The final selection of 54 photographs represented the full range of scores: nine images from each of six categories. These were then evaluated by three investigators using a clinical 0 to 4 corneal staining scale. Pearson correlations were calculated to compare investigator scores, and mean investigator and automated scores. Lin's Concordance Correlation Coefficients (CCC) and Bland-Altman plots were used to assess agreement between methods and between investigators.

RESULTS. Pearson's correlation between investigators was 0.914; mean CCC between investigators was 0.882. Bland-Altman analysis indicated that scores assessed by investigator 3 were significantly higher than those of investigators 1 and 2 (paired t -test). The predicted grade was calculated to be: $G_{pred} = 1.48\log(N_{dots}) - 0.206$. The two-point Pearson's correlation coefficient between the methods was 0.927 ($P < 0.0001$). The CCC between predicted automated score G_{pred} and mean investigator score was 0.929, 95% confidence interval (0.884–0.957). Bland-Altman analysis did not indicate bias. The difference in SD between clinical and automated methods was 0.398.

CONCLUSIONS. An objective, automated analysis of corneal staining provides a quality assurance tool to be used to substantiate clinical grading of key corneal staining endpoints in multicentered clinical trials of dry eye.

Keywords: dry eyes, corneal epithelium, image analysis

Punctate keratitis is a feature of many ocular surface diseases of infectious,¹ traumatic,² and inflammatory origin.³ Punctate keratitis presents differently for different diseases, yet is not pathognomonic to one disease, and a critical part of patient management is following the progression of ocular surface disease through changes in keratitis. Different patterns of keratitis provide clues to the underlying disease, whereas different amounts are indicative of severity.

Since the end of the 19th century, clinicians have visualized punctate keratitis with the aid of dyes, such as sodium fluorescein.^{4,5} When a drop of sodium fluorescein dye is instilled into the tear fluid, under cobalt-blue filter illumination, surface regions of staining become visible in a time-dependent fashion. These regions, commonly referred to as micro- or macro-punctate dots, superficial punctate keratitis, or SPK, have been interpreted in various ways, such as pooling in cavities resulting from desquamation of the corneal epithelium or penetration into intercellular spaces.^{6,7} However, recent investigations have questioned some of these hypotheses⁸ and instead have indicated a role of cellular uptake; it was shown in in vitro human and ex vivo rabbit models that damaged or dying cells take up fluorescein at a higher rate than healthy cells, and dead cells or surface gap pooling are probably not the cause of

staining.⁹ This relationship to cell turnover also would explain why healthy subjects have staining,^{6,10} and how pathological staining might reflect an incremental increase in cellular fluorescein uptake by dying cells.

Grading by the clinician normally involves a qualitative estimation of punctate dots in various corneal regions.^{11–14} A plethora of scoring systems has been developed to better describe what the clinician sees at the slit lamp, and these are the current standard in clinical research. These clinical grading scales use an ordinal scale based on punctate dot estimation with grade levels defined by qualitative descriptors, drawn or photographic imagery, or a combination. The cornea is typically divided into several regions (e.g., inferior, superior, central) with each region graded separately. Some of the more commonly used scales are the Baylor scale,¹¹ the National Eye Institute/Industry Workshop grading system,¹² the Oxford scale,¹³ and the van Bijsterveld scale.¹⁴ For contact lens-associated staining, the Efron scales are widely used.^{15,16} In the context of contact lens studies, additional parameters, such as staining extent or depth, may be included.^{15,16} Such scales are well accepted in all types of clinical studies and treatment-based clinical trials in dry eye.

We use a modified scale specific to dry eye (Ora Calibra Staining Scale^{17–19}), which is properly anchored with descrip-

tions and photographs at each level, and with half-unit increments of grading allowing for greater granularity. In the clinic, staining is assessed in the inferior, central, and superior regions. However, inferior corneal staining has recently been shown to be the key corneal zone required to predict any subtype of dry eye using receiver operating characteristic analysis of single, paired, and summed corneal zones as predictors of various types of dry eye.²⁰

There are several drawbacks that limit the effectiveness of clinical grading systems in the drug-approval process, which requires greater sensitivity of measure. Practical limitations of estimating punctate dot number, as well as subjectivity in interpretation of qualitative descriptors, can hinder objectivity, precision, and reproducibility.²¹ This becomes a particular challenge when assessing treatment efficacy in multicenter clinical trials with more than 35 participating investigators. Excessive variability in the data must then be offset by increasing the subject population, adding expense and complexity to the trial. Recent efforts to improve clinical grading involved modifying slit lamp hardware to minimize the artifact of nonuniform illumination.²² Other researchers have presented new scale concepts based on a further quantification of staining scores through the use of fluorophotometry at the slit lamp, relating results to permeability of the corneal surface.²³

The increments in current scales designed to measure punctate keratitis are nonlinear,¹³ and total punctate dot number can vary greatly at higher grades. Consequently, significant reductions of staining may still fall within the same clinical grade. Although precision is improved by a finer granularity in scales, this only partially resolves the problem. In fact, near-continuous scales with 0- to 100-point increments may exceed limits of human perceptual resolution.²⁴

The automated assessment of keratitis presented here is a simple method to quantify punctate dots in images taken at the slit lamp. This system can be generalized to complement and corroborate any clinical grading system, but as described here is specific to the investigation of scales based on punctate dot enumeration in dry eye disease. The slit lamp evaluation of keratitis specifies that the corneal surface be divided into several subregions that the investigator must estimate visually. Using quantitative digital image analysis, these regions can be extracted from the image more accurately and consistently. Furthermore, this alternative provides an effectively continuous objective scale that allows for greatly enhanced precision.

Because staining scores are adopted as precise endpoints in Food and Drug Administration clinical trials of new drug effects, known to be quite subtle in dry eye disease, quality assessment/control protocols must be in place for both the method of staining and the method of grading staining. Digital image analysis provides a useful tool to corroborate the quality of the grading that occurs at the slit lamp, fully recognizing that the gold standard in clinical trials is clinical grading and that these automated methods serve a complementary function of quality assurance across multicenters.

The objective of the present study, therefore, was to define the statistical correlation and agreement between the dataset derived from three experienced investigators grading a series of images as they would at the slit lamp, and comparing agreement to the dataset generated by the automated software.

METHODS

Subjects

The informed consent and study protocol were approved by a properly constituted institutional review board (Alpha IRB, San Clemente, CA, USA), and the study was conducted in

accordance with the ethical principles of the Declaration of Helsinki.

Subjects had to have a history of use of artificial tear substitutes for symptoms of dry eye within the previous 6 months and a patient- or investigator- reported history of dry eye in both eyes. Subjects were excluded who had (1) any clinically significant slit lamp findings, including active blepharitis, meibomian gland dysfunction, ocular infection, or active ocular or lid margin inflammation that required therapeutic treatment; (2) worn contact lenses; (3) a history of laser-assisted in situ keratomileusis or similar type of corneal refractive surgery within 12 months before the first visit; (4) used any of the prohibited medications within the appropriate prestudy washout period; or (5) a condition that the investigator felt may have put the subject at significant risk, may have confounded study results, or may have interfered significantly with the subject's participation in the study.

Image Capture Technique

Staining was evaluated after instillation of sodium fluorescein solution (5 μ L 2% preservative-free) into the inferior conjunctival cul-de-sac of both eyes of each subject. The subject was asked to blink several times to ensure proper mixing of the fluorescein dye throughout the tear film. No rinsing step was used. Because residence time is known to influence fluorescence, all images were taken 5 minutes after fluorescein instillation.

At the slit lamp, clinical grading of staining would have occurred for all five corneal regions. However, for the purposes of this agreement study, we focused on the inferior cornea, as this is the region most commonly and highly stained in dry eye,²⁰ as well as being the primary efficacy endpoint in many clinical trials.^{17-19,25}

Images of the inferior cornea were acquired with a 12-megapixel digital camera (Canon 5D Mark I DSLR; Canon USA, Inc., Melville, NY, USA) attached to a Haag Steit BX 900 (Bern, Switzerland) slit lamp system. An excitation blue filter (Haag Streit, Bern, Switzerland) and a barrier Wratten #12 yellow filter (Yellow Wratten #12; Eastman Kodak Company, Rochester, NY, USA) were used to optimize visibility of punctate dots. Tear layer reflection was minimized by using a diffused flash system. Images were captured in RAW format to maximize the information from each subject and converted to RGB, 8-bit jpeg format for processing. Shutter speed, aperture size, ISO, and other camera parameters were constant across all subjects.

Selection of Region of Interest

A sample image is shown in Figure 1a. The region of interest is the geometrically defined area of the inferior cornea (Fig. 1b). The first step of the analysis process is the segmentation of this portion from the remainder of the image. Due to variations in intersubject corneal diameter, eyelid shape, and other geometric parameters, this step cannot be fully automated.

From the original image, three points along the edge of the inferior cornea are selected with the mouse. The final segmentation of the inferior region is then done automatically. The OpenCV computer vision library,²⁶ an open source library written in C++, was used to develop software for segmenting and cropping the image, as well as for detection and enumeration of total punctate dot number.

Calculation of Punctate Dot Number (N_{dots})

After the initial cropping step to isolate the region of interest, the blue and red channels are discarded and analysis of the green channel proceeds using a process of image normalization with background smoothing (Fig. 1c).

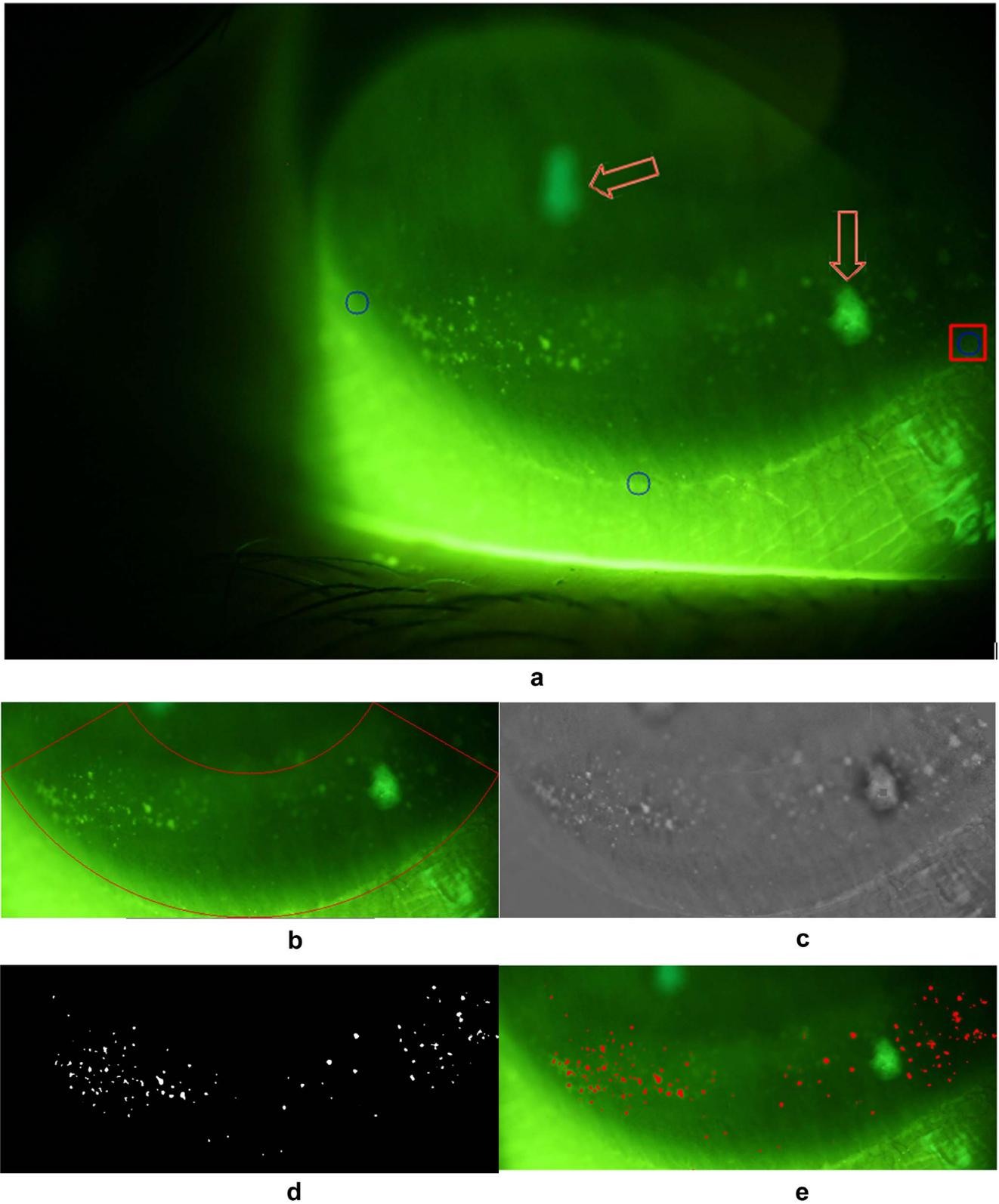


FIGURE 1. (a) Digital image of the inferior cornea (arrows indicate flash artifacts). User defines the inferior region by selecting three points along the periphery of the cornea (blue circle). (b) Cropping of the image based on the region of interest (in red outline). (c) Processing the image (green channel) with background smoothing and image normalization. (d) Binary image after thresholding. (e) Detected punctate dots are in red.

TABLE 1. Repeatability Over Three Trials of Automated System

	Trial 1 vs. Trial 2	Trial 1 vs. Trial 3	Trial 2 vs. Trial 3
Pearson's correlation coefficient (95% CIs)	0.996 (0.993–0.998)	0.996 (0.993–0.998)	0.996 (0.993–0.998)
Bland-Altman mean difference (<i>P</i> value)	–0.00482 (0.624)	0.001367 (0.899)	0.006186 (0.533)
Difference SD	0.072101	0.079307	0.072592
CCC (95% CIs)	0.996 (0.994–0.998)	0.996 (0.994–0.998)	0.996 (0.993–0.997)

After extraction of the green RGB image channel, image brightness is normalized to mean pixel intensity of 100. The image background is next smoothed to remove the effects of slowly varying image intensities due to the spherical geometry of the corneal surface. Punctate dots are isolated by thresholding the image by pixel intensity level of approximately 20% above the image mean intensity.

To remove noise in the image, the images were next converted to binary format and a blob detection algorithm²⁶ was used to filter connected regions of white pixels below the typical size of a corneal epithelial cell (assumed to be 36 μm).²⁷ The minimum blob size in pixels is then determined by the apparent diameter of the cornea (in pixels) in the image based on assumed actual diameter of 10 mm.

The blob filtering is also used to remove flash artifact from the image by filtering blobs above 500 pixels. Finally, blobs not falling within the geometric area of interest (here the inferior cornea) are also discarded. The remaining blobs (Fig. 1d) are then counted by the algorithm to obtain the punctate dot number, N_{dots} .

Detected punctate dots are highlighted in red (Fig. 1e).

Image Database

Images selected for this study were required to be free of defects that prevented clear visibility of staining in the region of interest including the following: overexposure, too weak a fluorescein stain intensity to distinguish staining from background, limited visibility, or poor focus in the region of interest, and tear film interference with staining visibility. Images failing these criteria were rejected.

A total of 229 images, from 115 subjects, were first graded by the software. The scores were then sorted into bins defined by the minimum and maximum log punctate dot number over the entire population partitioned into six equal ranges. Nine images were then selected at random from each severity grade level.

The final database consisted of fifty-four images. Each image was presented to each investigator in a random sequence with no information on the initial assessment.

Clinical Grading System

The final 54 images were graded by three experienced clinical research investigators (162 clinical grades comprised this dataset) using the clinical grading system conventionally used at the slit lamp. This scale uses a 0 to 4 grading system (Ora Calibra Fluorescein Staining Scale) with grade levels based on number of punctate dots as specified by qualitative descriptive language, where 0 = none and 4 = severe confluent staining. The cornea is divided into regions that are graded separately. This averaged investigator grade was adopted as the “clinical grade” to be then compared to the software grade. The grading of images, as opposed to actual grading at the slit lamp, does constitute an additional layer of separation from the clinic. However, we felt that the comparison of images evaluated by both investigator and software represents a critical first step to understanding actual clinical grading.

Statistical Methods

Correlation and Agreement Among Investigators. The pairwise correlations among the three investigators were first determined using Pearson's correlation coefficient. This provided information regarding correlation, but not agreement. To calculate agreement, Lin's Concordance Correlation Coefficient²⁸ (CCC) was used. This analysis was supplemented by Bland-Altman²⁹ analysis comparing mean difference and SD of mean difference between each pairwise investigator combination.

Agreement Between Clinical and Automated Predicted Score. The mean score of the three investigators for each image was taken as best representing the true clinical score. The primary analysis was then a comparison of the mean investigator-graded score to the automated predicted score.

An estimator for the predicted score was derived by linear regression fit to all 162 investigator scores against the automated score. Using the results of this analysis, the best approximation to the clinical score may be written as follows:

$$G_{pred} = C_1 \log(N_{dots}) + C_2, \quad (1)$$

where G_{pred} is the automated predicted keratitis score and N_{dots} is the computed punctate dot number.

The agreement between the automated predicted keratitis score based on the regression fit, and the clinical score assigned to the image by the investigator, was again determined using Lin's CCC as well as Bland-Altman analysis in the same manner as comparison between investigators.

RESULTS

Repeatability of Automated Method

Because the grading method requires manual input, it is important to consider intra-user variability under repeated trials. To test this requirement, a single software user supplied manual input for each image in the dataset in a series of three trials. Bland-Altman analysis of the resulting predicted score G_{pred} is shown in Table 1. The maximum between-trial grade difference over all images was 0.245. The mean concordance between trials was 0.996 and no statistical difference was found between any of the trials.

Correlation and Agreement Between Investigators

A summary of the correlation and agreement statistics comparing all clinical investigators is shown in Table 2. The mean Pearson's correlation between investigators was 0.914. The mean CCC between investigators was 0.882. Bland-Altman analysis indicated that scores assessed by investigator 3 tended to be significantly higher than those assessed by both investigators 1 and 2 based on a paired *t*-test. Mean difference in assessed scores was greatest between investigator 1 and investigator 3 (–0.370) ($P < 0.0001$). The mean of the SDs of the staining score differences between investigators was (0.497) or approximately one-half a clinical grade point.

TABLE 2. Agreement Among Investigators Using Clinical Grading System

Clinical Score	Investigator 1 vs. Investigator 2	Investigator 1 vs. Investigator 3	Investigator 2 vs. Investigator 3
Pearson's correlation coefficient (CIs)	0.905 (0.840-0.944)	0.903 (0.838-0.943)	0.934 (0.888-0.961)
Bland-Altman mean difference (<i>P</i> value)	-0.111 (0.135)	-0.370 (<0.0001)*	-0.259 (0.0002)*
Difference SD	0.538	0.4687	0.483
CCC (CIs)	0.895 (0.828-0.936)	0.874 (0.806-0.919)	0.876 (0.800-0.925)

* Difference between investigators was significantly greater than zero, $P < 0.05$.

Correlation and Agreement Between “Clinical” and Automated Predicted Scores

The logarithmic nature of punctate dot number has been previously observed; that is, the Oxford grading scale follows a logarithmic progression from panel to panel.¹³ In Figure 2, the investigator-assigned score for all images is plotted against the computed punctate dot number, N_{dots} in logarithmic scale. The number of computed punctate dots (N_{dots}) is potentially unlimited, although no images have been observed to date with N_{dots} greater than approximately 1000; that is, $\log(N_{dots}) = 3$.

A linear regression calculation of all investigator scores versus the automated punctate dot number for each image determined $C_1 = 1.48$ and $C_2 = -0.206$ ($R = 0.890$) with two-sided confidence intervals (CIs) (0.853-0.918). The automated predicted grade was then estimated from this regression fit as

$$G_{pred} = 1.48\log(N_{dots}) - 0.206(\text{for } N_{dots} > 0) \quad (2)$$

and

$$G_{pred} = 0 \quad \text{for } N_{dots} = 0. \quad (3)$$

Both clinical and predicted scores are plotted in Figure 3 for each of the 54 images.

Table 3 summarizes the statistical agreement and correlation between the clinical and automated predicted scores. The

two-point Pearson's correlation coefficient between the methods was 0.927 ($P < 0.0001$).

The CCC between predicted automated score and mean investigator score was 0.929.

Figure 4 shows the Bland-Altman plot, with score differences $I(n) - G_{pred}(n)$ on the y -axis, where $I(n)$ is the mean score for all investigators for each image and $G_{pred}(n)$ is the predicted score for each image ($1 < n < 54$). The average grade $[I(n) + G_{pred}(n)]/2$ is shown on the x -axis. The mean difference between methods for all images was -0.0153 . This value was not significantly different from zero ($P = 0.779$, based on a paired t -test), and hence did not indicate bias. The range of the 95% CI was $(-0.124$ to $0.0933)$.

The SD of mean differences between clinical and automated methods was 0.398.

DISCUSSION

The use of imaging techniques and the associated software analysis as a complement or substitute for clinical scoring is already widespread in many medical specialties, such as radiology and retinal diseases.³⁰ These methods provide increased efficiency for processing clinical data, as well as the benefit of standardizing grading across multicentered clinical trials. Standardization of grading in turn provides a more statistically powerful dataset, as well as greater precision

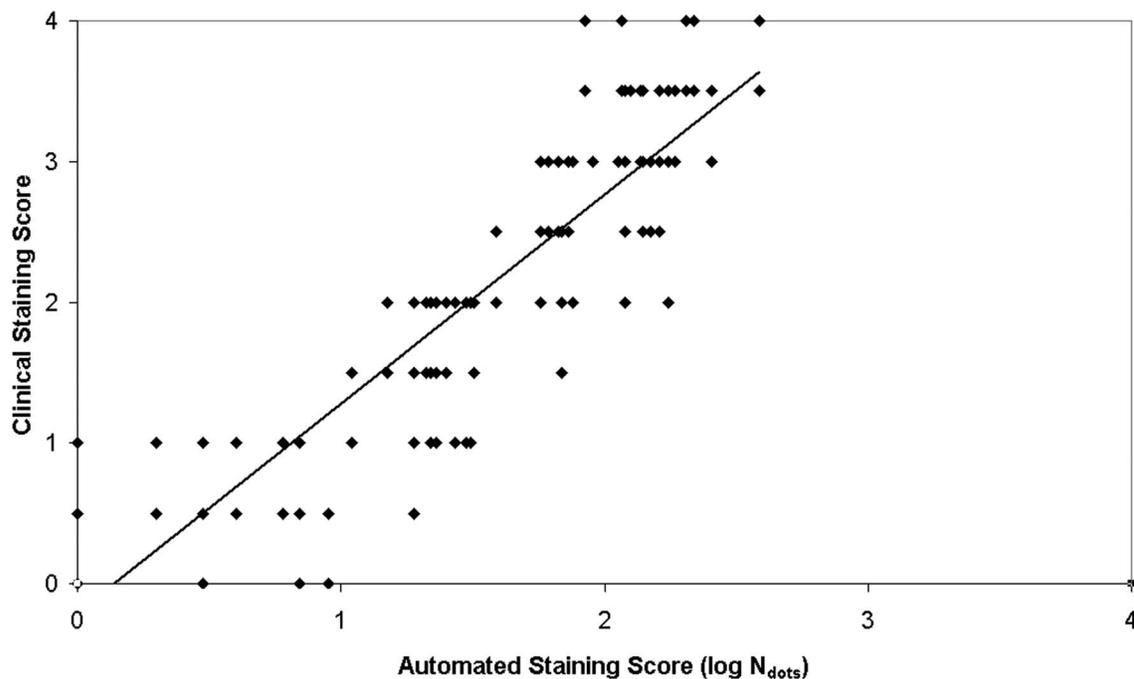


FIGURE 2. Individual clinical staining scores assigned by the three investigators for all 54 images (162 data points) versus software-derived punctate dot number (N_{dots}); per linear regression $G_{pred} = 1.48\log(N_{dots}) - 0.206$ ($R = 0.890$, CI 0.853-0.918).

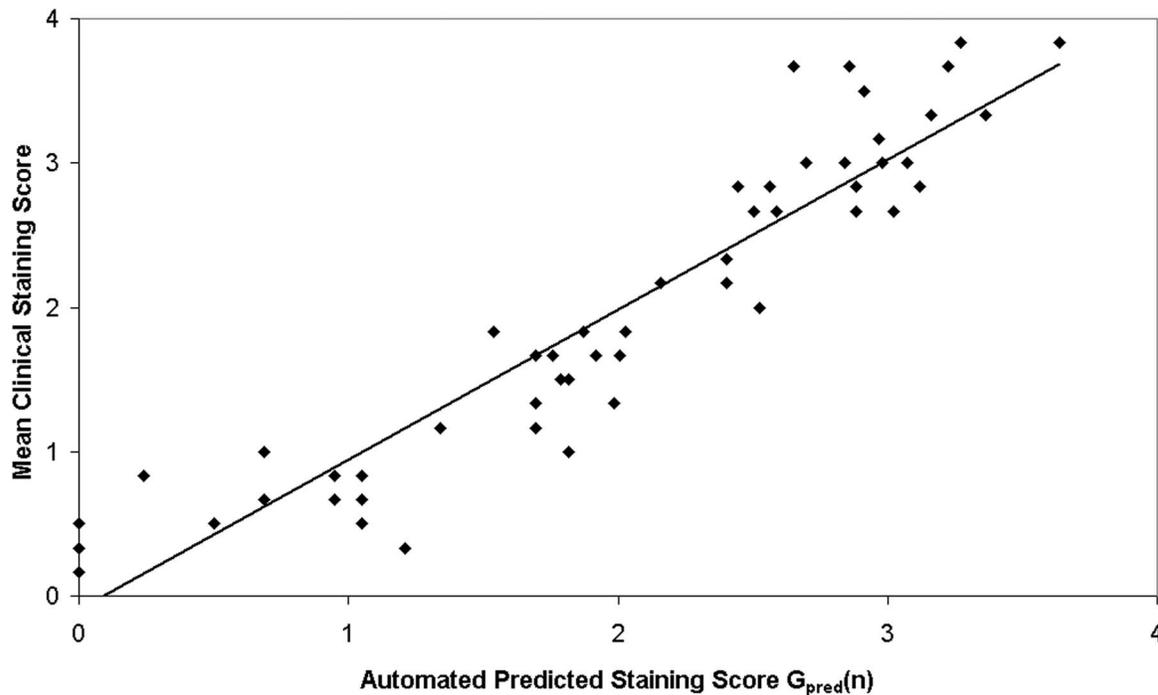


FIGURE 3. Mean clinical staining score (of three investigators) for each of the 54 images [$I^m(n)$] ($n = 1-54$) versus corresponding automated predicted staining score $G_{pred}(n)$.

when assessing treatment effects. In addition, imaging data allow for more detailed analyses of the effects of treatment than standard techniques, and allow for retrospective reanalysis, as a permanent database is created.

In dry eye clinical trials, the acquisition of image data is relatively less common. In a previous study,³¹ we investigated the practicality of automated grading of images of conjunctival hyperemia. In that study, the morphological structure of the vascular structure was shown to be relevant as a diagnostic tool in the context of dry eye. In other applications in ophthalmology, the use of imaging data of the conjunctiva has been explored extensively in studies of contact lens wear.^{15,16}

Software analysis of digital images of corneal staining with sodium fluorescein^{22,32,33} has yet to be explored as extensively as conjunctival hyperemia.³⁴⁻³⁸ On one level, automated analysis of hyperemia is more complex because both clinical and automated scoring incorporate two parameters in the score: redness intensity and vessel morphology/geometry.^{24,36} Such assessments have been shown to be particularly difficult for a human investigator. In contrast, corneal staining, particularly in dry eye disease, and as defined by existing clinical scales, is dependent on evaluating the punctate number in a given corneal region. This one-parameter approach to grading may help reduce inconsistencies among investigators in a clinical trial. Although it is not possible to

ascertain the “true” number of stained cells, the close agreement of the qualitative and quantitative methods strengthens the validity of both. Furthermore, the repeatability of the software method appears to be excellent. For a complete comparison of clinical and automated grading, aspects of repeatability, such as that of the photographic technique and possible variation between the image grade and slit lamp grade, must be addressed, as well as whether sufficient image-acquisition quality in an actual clinical trial can be maintained. Answers to these questions will be investigated with a more extensive data set.

Assessment of confluent staining poses challenges for both manual and automated grading because, at times, staining presents as small isolated patches of confluent regions within a relatively clear cornea. In these cases, the clinician must then rely on his or her clinical acumen to grade the severity of staining accordingly. For the data set considered in this study, automated analysis based on punctate dot count does successfully reproduce the gestalt manual grade over the entirety of the grading range of the Ora Calibra Staining Scale. However, subtleties in the grading of confluence at the upper end of the scale may require further refinement of program parameters, based on clinical input, including the possible addition of a parameter to measure total staining area as well as the consideration of blobs of larger size. This will be examined in future studies.

In the present study, the average SD of interinvestigator differences was 0.497, or less than one-half clinical grade. The corresponding agreement between predicted score and mean investigator score was 0.398. The results from the Bland-Altman analysis show that there was no statistically significant bias between predicted and mean investigator score. For the investigators, despite close agreement of differences, the scores assessed by Investigator 3 were significantly higher than those assessed by the other investigators. Finally, recent results from clinical studies have shown that the evaluation of treatment efficacy in dry eye is extremely subtle, and

TABLE 3. Agreement Statistics Between Clinical and Automated Staining Scores

	Clinical Score vs. Automated Score
Pearson's correlation coefficient (CI)	0.927 (0.877-0.957)
Bland-Altman mean difference (P value)	-0.0153 (0.760)
Difference SD: clinical score vs. automated punctate dot number	0.398
CCC (CI)	0.929 (0.884-0.957)

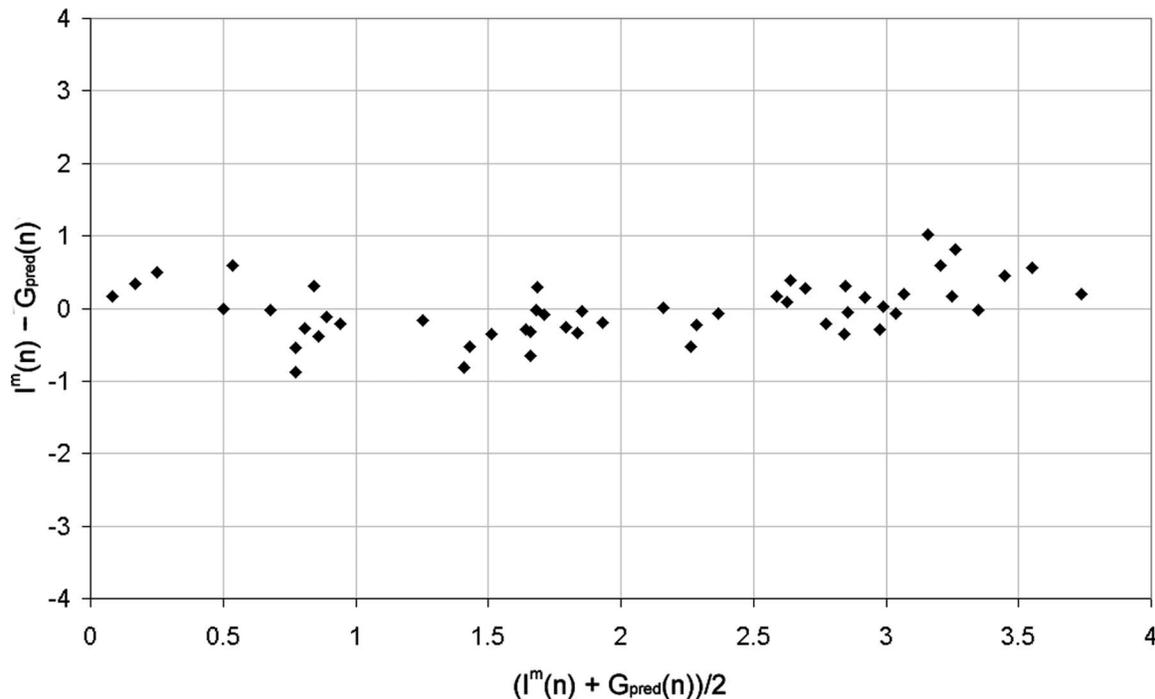


FIGURE 4. Bland-Altman comparison between mean investigator clinical staining score (I^m) and automated predicted staining score (G_{pred}). Average score (x-axis) versus score difference (y-axis) for all images ($n = 1-54$). Mean score difference: -0.0153 ($P = 0.760$), SD: 0.398 .

measurements must be able to detect small changes.^{17-19,25} Mean significant differences in corneal staining after treatment can be less than a 0.3 clinical grade between vehicle and active subject groups.⁹

Results of this agreement study show that, although remaining a work in progress, automated assessment of corneal staining, based on digital image analysis, provides increased objectivity and decreased scoring variability. Although we expect that clinical grading will continue to remain the gold standard in clinical trials, image analysis holds promise to ensure the quality of clinical staining score outcomes across multicenters, and could thereby improve outcomes in clinical studies of dry eye.

Acknowledgments

Supported by Ora, Inc.

Disclosure: **J.D. Rodriguez**, Ora, Inc. (E, R); **K.J. Lane**, Ora, Inc. (E, R); **G.W. Ousler III**, Ora, Inc. (E, R); **E. Angjeli**, Ora, Inc. (E, R); **L.M. Smith**, Ora, Inc. (E); **M.B. Abelson**, Ora, Inc. (S)

References

- Hill GM, Ku ES, Dwarakanathan S. Herpes simplex keratitis. *Dis Mon.* 2014;60:239-246.
- Matthews TD, Frazer DG, Minassian DC, Radford CF, Dart JKG. Risks of keratitis and patterns of use with disposable contact lens. *Arch Ophthalmol.* 1992;110:1559-1562.
- (No authors listed). Management and therapy of dry eye disease: report of the management and therapy subcommittee of the International Dry Eye Workshop. *Ocul Surf.* 2007;5:163-178.
- Joyce PD. Corneal vital staining. *Ir J Med Sci.* 1967;6:359-367.
- Kim J. The use of vital dyes in corneal disease. *Curr Opin Ophthalmol.* 2000;11:241-247.
- Norn MS. Micropunctate fluorescein vital staining of the cornea. *Acta Ophthalmol (Copenh).* 1970;48:108-118.
- Kikkawa Y. Normal corneal staining with fluorescein. *Exp Eye Res.* 1972;14:13-20.
- Bandamwar KL, Garrett Q, Papas EB. Mechanisms of superficial micropunctate corneal staining with sodium fluorescein: the contribution of pooling. *Cont Lens Anterior Eye.* 2012;35:81-84.
- Bandamwar K, Papas EB, Garrett Q. Fluorescein staining and physiological state of corneal epithelial cells. *Cont Lens Anterior Eye.* 2014;37:213-223.
- Korb DR, Korb JM. Corneal staining prior to contact lens wearing. *J Am Optom Assoc.* 1070;41:228-232.
- DePaiva CS, Pflugfelder SC. Corneal epitheliopathy of dry eye hyperesthesia to mechanical air jet stimulation. *Am J Ophthalmol.* 2004;137:109-115.
- Lemp MP. Report of the National Eye Institute/Industry Workshop on Clinical Trials in Dry Eyes. *CLAO J.* 1995;21:221-232.
- Bron AJ, Evans VE, Smith JA. Grading of corneal and conjunctival staining in the context of other dry eye tests. *Cornea.* 2003;22:640-650.
- van Bijsterveld OP. Diagnostic tests in the Sicca syndrome. *Arch Ophthalmol.* 1969;82:10-14.
- Efron N, Morgan PB, Katsara SS. Validation of grading scales for contact lens complications. *Ophthalmic Physiol Opt.* 2001;21:17-29.
- Efron N. Grading scales for contact lens complications. *Ophthalmic Physiol Opt.* 1998;18:182-186.
- Meerovitch K, Torkildsen G, Lonsdale J, et al. Safety and efficacy of MIM-D3 ophthalmic solutions in a randomized, placebo-controlled Phase 2 clinical trial in patients with dry eye. *Clin Ophthalmol.* 2013;7:1275-1285.
- Semba CP, Torkildsen GL, Lonsdale JD, et al. A Phase 2 randomized, double-masked, placebo-controlled study of a novel integrin antagonist (SAR 1118) for the treatment of dry eye. *Am J Ophthalmol.* 2012;153:1050-1060.

19. Sheppard JD, Torkildsen GL, Lonsdale JD, et al. Lifitegrast ophthalmic solution 5.0% for treatment of dry eye disease. Results of the OPUS-1 Phase 3 Study. *Ophthalmology*. 2014; 121:475-483.
20. Fenner BJ, Tong L. Corneal staining characteristics in limited zones compared with whole cornea documentation for the detection of dry eye subtypes. *Invest Ophthalmol Vis Sci*. 2013;54:8013-8019.
21. Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effects of scaling. *Invest Ophthalmol Vis Sci*. 1991;32:422-432.
22. 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;39:239-246.
23. 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;121:1537-1539.
24. Fieguth P, Simpson T. Automated measurement of bulbar redness. *Invest Ophthalmol Vis Sci*. 2002;43:340-347.
25. Patane MA, Cohen A, From S, et al. Ocular iontophoresis of EGP-437 (dexamethasone phosphate) in dry eye patients: results of a randomized clinical trial. *Clin Ophthalmol*. 2011;5: 633-643.
26. Dr. Dobb's Journal of Software Tools. The Open CV Library. Available at: <http://www.drdoobs.com/open-source/the-opencv-library/184404319>. Accessed April 29, 2013.
27. Romano AC, Espana EM, Yoo SH, Budak MT, Wolosin JM, Tseng SC. Different cell sizes in human limbal and central corneal basal epithelia measured by confocal microscopy and flow cytometry. *Invest Ophthalmol Vis Sci*. 2003;44:5125-5129.
28. Lin L. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45:225-268.
29. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician*. 1983;32:307-317.
30. Wolbarst AB, Capasso P, Wyant AR. *Medical Imaging: Essentials for Physicians*. New York: Wiley-Blackwell; 2013.
31. Rodriguez JD, Johnston PR, Ousler GW, Smith LM, Abelson MB. Automated grading system for evaluation of ocular redness associated with dry eye. *Clin Ophthalmol*. 2013;7: 1-8.
32. Peterson RC, Wolffsohn JS. Objective grading of the anterior eye. *Optom Vis Sci*. 2009;86:273-278.
33. 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.
34. Owen CG, Fitzke FW, Woodward EG. A new computer assisted objective method for quantifying vascular changes of the bulbar conjunctivae. *Ophthalmic Physiol Opt*. 1996;16:430-437.
35. Willingham FF, Cohen KL, Coggins JM, Tripoli NK, Ogle JW, Goldstein GM. Automatic quantitative measurement of ocular hyperemia. *Curr Eye Res*. 1995;14:1101-1108.
36. Papas EB. Key factors in the subjective and objective assessment of conjunctival erythema. *Invest Ophthalmol Vis Sci*. 2000;41:687-691.
37. Wolffsohn JS, Purslow C. Clinical monitoring of ocular physiology using digital image analysis. *Cont Lens Anterior Eye*. 2003;26:27-35.
38. Yoneda T, Sumi T, Takahashi A, et al. Automated hyperemia software analysis: reliability and reproducibility in healthy subjects. *Jpn J Ophthalmol*. 2012;56:1-7.