Validation and verification of the OPI 2.0 System

Richard Abelson1,2
Keith J Lane3
John Rodriguez3
Patrick Johnston3
Endri Angjeli3
George Ousler3
Douglas Montgomery1

1School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA
2Statistics and Data Corporation, Tempe, AZ, USA
3Ora, Inc, Andover, MA, USA

Purpose: The Ocular Protection Index (OPI) 2.0 System was developed to evaluate ocular surface protection under a natural blink pattern and normal visual conditions. The OPI 2.0 System implements fully automated software algorithms which provide a real-time measurement of corneal exposure (breakup area) for each interblink interval during a 1-minute video. Utilizing this method, the mean breakup area (MBA) and OPI 2.0 (MBA/interblink interval) were calculated and analyzed. The purpose of this study was to verify and validate the OPI 2.0 System for its ability to distinguish between dry eye and normal subjects, and to accurately identify breakup area.

Methods: In order to verify and validate the OPI 2.0 System, a series of artificial images and a series of still image frames captured during an actual clinical session using fluorescein staining videography were analyzed. Finally, a clinical validation process was completed to determine the effectiveness and clinical relevance of the OPI 2.0 System to differentiate between dry eye and normal subjects.

Results: Software analysis verification conducted in a set of artificially constructed images and in actual videos both saw minimal error rates. MBA and OPI 2.0 calculations were able to distinguish between the qualifying eyes of dry eye and normal subjects in a statistically significant fashion (P < 0.001 for both outcomes). As expected, dry eye subjects had a higher MBA and OPI 2.0 than normal subjects (0.232, dry eye; 0.040, normal and 0.039, dry eye; 0.006, normal, respectively). Results for the worst eyes and all qualifying analyses based on staining, forced-stare tear film breakup time, and MBA were numerically similar.

Conclusion: The OPI 2.0 System accurately identifies the degree of breakup area on the cornea and represents an efficient, clinically relevant measurement of the pathophysiology of the ocular surface.

Keywords: tear film breakup time, interblink interval, natural blink conditions, Ocular Protection Index

Introduction

Reduced tear film stability is a key driving factor in the development of dry eye. The measurement of tear film breakup time (TFBUT) using fluorescein with forced-stare is a well established clinical metric for evaluating the health of the precorneal tear film.1-4 More recently, the development of the Ocular Protection Index (OPI) was an important step in evaluating the interaction between blinking and TFBUT. This tool has been used in numerous observational studies and clinical trials and has been widely adopted by clinicians.5-12 However, an increased understanding of the complexities of blink physiology and tear film breakup suggests that this methodology has the potential to be improved upon. First, TFBUT and interblink interval (IBI) measurements are
performed at different times. Second, TFBUT is evaluated using the forced-stare technique, which is not representative of the physiological action of an unaltered blink pattern. Third, this methodology provides no information on what occurs on the ocular surface between actual tear film breakup and the next blink, which is the point of corneal affliction.

In order to address these shortcomings, the OPI 2.0 System was developed to evaluate ocular surface protection under normal visual conditions. The approach yields a real-time measurement of percent of cornea exposed (tear film breakup area) for each IBI frame during a 1-minute video. The system also provides a simultaneous measurement of TFBUT and IBI. Utilizing this method, the mean breakup area (MBA) and OPI 2.0, defined as MBA/IBI, are calculated and analyzed. Initially, a method of retrospective manual analysis of fluorescein staining video data was utilized with the OPI 2.0 System. In this method, which is referred to as video capture with manual analysis (VCMA), a panel of examiners evaluated the integrity of the tear film and determined IBI and TFBUT by manually stopping the video to note and confirm the time stamp, and record the time of each blink and the first appearance of a micelle within each IBI. This method of breakup area evaluation utilized a sectoral transect of the corneal surface. Grading was made based on a binary evaluation of breakup within each region. A given region was counted as fully broken if any breakup was observed in that area regardless of the actual extent of exposure. Results utilizing the VCMA method demonstrated successful differentiation between normal and dry eye subjects; however, this methodology required numerous technician hours to manually grade the area of corneal coverage and did not reach the desired level of precision.

To improve the efficiency of the analysis, a complex set of algorithms was developed in order to automate the analysis of video footage collected. The processing of a video consists of two stages. The first is an image segmentation stage during which the corneal image is extracted from the background of the video frame using a template matching algorithm. The second stage consists of measurement of the exposed area from the image sequence. The areas of exposure are summed pixel-by-pixel and divided by the mean corneal area over the entire video. This is to account for small variations in palpebral fissure width, and the calculation yields the average percent area of corneal exposure as a function of time.

The development of the software analysis had three goals: to calculate more precise values for the percent of corneal area exposed by way of computerized image analysis; to decrease human error (ie, error introduced by the use of a stopwatch in the technician’s calculation as there is an inherent delay between the time the doctor can detect a break and the time the stopwatch is stopped); and to increase the speed of analysis. Previous work on tear film breakup area has been conducted but it is uncertain how much validation has been completed on the procedures used. Advances in technology have prompted the use of video images to determine breakup area. While some techniques measure breakup area from the last video frame before the IBI, the OPI 2.0 System is designed to measure MBA, which is an average of the percent of cornea exposed over the entire video (Figure 1).

The goal of adding software analysis to the OPI 2.0 System
was to accurately measure the amount of breakup area on the cornea and provide an efficient, clinically relevant measurement of the pathophysiology of the ocular surface.

Methods
Verification
To calibrate the software analysis and demonstrate that it can correctly identify the area of exposure, a set of artificially constructed images was created to mimic the visual properties of images captured during an actual clinical session using fluorescein staining videography. The parameters were density, breakup dispersion, and image brightness. Breakup density represented the extent of tear film breakup as a percentage of the corneal surface area. Breakup dispersion represented the degree to which the exposed areas were distributed over the corneal surface, ie, the number of individual isolated regions of exposed cornea. Image brightness represented the pixel intensity level of the green channel of the image.

In addition to the eight images created to bracket the range of values of the three parameters (designed HHL for high dispersion, high density, and low brightness, etc), a middle image was created at the mean parameter values to create a total of nine images. To measure the effectiveness of the software, an image was output with the areas of detected simulated tear film breakup shown in red. For the purposes of this verification procedure, the artificially constructed images created to mimic the visual properties of images captured during an actual clinical session using fluorescein staining videography will be referred to as the “artificial” images. The software analysis output of the image with the areas of detected simulated tear film breakup will be referred to as the “detected” images. There are two types of incorrect detections of breakup area with regard to discrepancies seen between the number of pixels detected in real images and the detected images: false negatives and false positives. A false negative detection occurs when breakup in the real image is not observed by the software analysis in the detected image. A false positive detection occurs when the software analysis detects breakup in the detected image that is not considered breakup in the real image.

Validation
The second stage involved using video images collected from subjects during the clinical validation process. The image properties were selected to correspond to the range of image values similar to the artificial images. After the selection, the images were graded manually by an expert grader and areas of exposed cornea were marked using image editing software in blue. These images were used as ground truths to measure the effectiveness of the software. The software was used to output the original image indicating the areas of detected breakup in red. This allowed for a simple visual comparison between red (software detected) and blue (technician graded). The images were also compared with regard to pixel count.

Clinical validation
A single-center, one-visit study enrolling 29 dry eye subjects and 16 normal subjects was conducted. All subjects were enrolled based on qualifying eyes, meaning a subject could contribute one or two eyes. Qualifying eyes included 49 eyes from the 29 dry eye subjects and 29 eyes from the 16 normal subjects, for a total of 78 qualifying eyes. Qualifying eyes for the dry eye subjects were based on three inclusion criteria: a forced-stare TFBUT of ≤5 seconds in at least one eye; a corneal fluorescein staining score ≥2 (0–4 point Ora scale) in at least one region of the eye; and a reported history of dry eye disease or ocular symptomatology with the desire to use artificial tears. Normal subjects were excluded if they had a history of dry eye, irritation, or any other ocular problems, wore contact lenses or had laser-assisted in situ keratomileusis eye surgery, or habitually used artificial tears or tear substitutes. To ensure that normal subjects were largely free of keratitis, qualifying eyes for the normal subjects must also have had a ≤1.5 staining score in each region of both eyes (0–4 point Ora scale). A staining score of ≥2 in any region of any eye was exclusionary. In addition to forced-stare TFBUT and fluorescein staining evaluations, all enrolled subjects were measured by the OPI 2.0 System. Additionally, all dry eye subjects and a random sampling of four normal subjects underwent Schirmer’s test evaluations.

Three additional analyses were performed looking at worst eye only, meaning that each subject only contributed a single eye to each analysis. The eye was defined as “worst” using three separate criteria in three independent analyses; first by looking at total staining score, second by looking at forced-stare TFBUT, and finally by looking at MBA. Finally, dry eye and normal groups were compared with respect to variability. Ratios of standard deviation were used for Schirmer’s and staining, while ratios of coefficients of variation were used for forced-stare TFBUT, IBT, MBA, and OPI 2.0.

Forced-stare TFBUT was evaluated by instilling sodium fluorescein solution (5 µL, 2% preservative-free) into the inferior conjunctival cul-de-sac of each eye, and the subject was asked to blink several times to mix the fluorescein with...
their tear film. The subject was then asked to blink twice and then stare without blinking for as long as possible. The examiner monitored the integrity of the tear film through a slit-lamp biomicroscope (BQ900 slit lamp; Haag-Streit AG, Koniz, Switzerland) with an 8 mm scanning beam (using an excitation blue filter and a barrier Wratten #12 yellow filter), and measured the time from eye opening to the first appearance of micelles with a stopwatch. The eyes were evaluated sequentially (right eye, left eye). Two measurements were taken and averaged unless the two measurements were both less than 10 seconds and differed by more than 2 seconds, in which case a third measurement was taken and the two closest of the three were averaged.

Following the traditional clinical assessments, OPI 2.0 System measurements were taken. The examiner instilled sodium fluorescein solution (5 µL, 2% preservative-free) into the inferior conjunctival cul-de-sac of each eye and the subject was asked to blink several times to mix the fluorescein with the tear film. While the subject performed a standard visual task (watching a documentary on television from a 5-ft viewing distance), the eye was recorded using a digital video camera (EyeCap™ IM 900® camera; Haag-Streit AG, Koeniz, Switzerland) at 8× magnification at a rate of 15 frames per second through a slit-lamp biomicroscope using an excitation blue filter and a barrier Wratten #12 yellow filter. A minimum of 1 minute of continuous data was recorded for each eye with approximately 30 seconds between recordings of the two eyes. The eyes were recorded from right to left. Subsequently, a computer program analyzed the cornea on a frame-by-frame basis and provided breakup area for each IBI during the 1-minute video. From this analysis, MBA and OPI 2.0 were calculated and analyzed. The software also provided a measurement of TFBUT; however for the purposes of this paper, this data was not analyzed.

Statistical analysis
The comparison between independent dry eye and normal qualifying eyes was based on 78 eyes (dry eye = 49, normal = 29). Normal linear models estimated by generalized estimating equation methods were used for staining scores and Schirmer’s scores. Gamma multiplicative models, also estimated by generalized estimating equation methods, were used for MBA, IBI, OPI 2.0, and forced-stare TFBUT.

These models provided estimates for group means, differences of means for linear models, and ratios of means for multiplicative linear models. Corresponding 95% confidence intervals, and \( P \) values for tests of equality, were calculated. All models were fit using the GENMOD procedure of SAS version 9.2 (SAS Institute Inc, Cary, NC).\(^17\)

Results
Verification
The software analysis was able to correctly identify the area of exposure in a set of artificially constructed images created to mimic the visual properties of actual clinical images captured using fluorescein staining videography. For all nine images, out of 3,642,590 pixels, there was a total of 62 false errors, yielding a 99.9983% accuracy rate. Seven of the errors were false negatives while 55 were false positives. The OPI 2.0 System false positive and false negative errors were dependent on the given parameters (density, \( P = 0.004; \) dispersion, \( P = 0.038; \) and brightness, \( P < 0.001 \)) of the real images (Figure 2). In the artificial eye designated LLL (low density, low dispersion, and low brightness; Figure 2A), the OPI 2.0 System detected the greatest number of false positive and false negative pixels with a total of 18, zero of which were false negative and all 18 of which were false positive. In the artificial eyes designated HLH (high density, low dispersion, and high brightness; Figure 2B) and HHH (high density, high dispersion, and high brightness; Figure 2C), the OPI 2.0 System detected the least number of false positive and false negative pixels, both with a total of zero.

Validation
The software analysis was able to correctly identify the area of exposure in a set of video images collected (Figure 3). For all nine images, out of 3,165,062 pixels, there was a total of 38,728 false errors, yielding a 98.7764% accuracy rate. Of the errors, 14,050 were false negatives while 24,678 were false positives. In the technician-graded eye designated HHL (high density, high dispersion, and high brightness; Figure 3A), the OPI 2.0 System detected the greatest number of false positive and false negative pixels with a total of 12,857; of these, 5550 were false negatives and 7307 were false positives. While this error rate was the highest at 2.8948%, the discrepancy could possibly be attributed to the inaccuracy of the technician’s grading. In the technician-graded eye designated LLH (low density, low dispersion, and high brightness), the OPI 2.0 System detected the least number of false positive and false negative pixels with a total of zero.

Clinical validation
The mean ages for the dry eye (n = 29) and normal (n = 16) subjects with qualifying eyes were 59.08 and 34.03 years,
Figure 2 The OPI 2.0 System false positive and false negative errors and verification of the software analysis. Image of (A) low density, low dispersion, and low brightness; (B) high density, low dispersion, and high brightness; and (C) high density, high dispersion, and high brightness, with designated artificial eye on the left and OPI 2.0 System output with the areas of detected simulated tear film breakup in red on the right.

Note: The OPI 2.0 System false positive and false negative errors were dependent on the given parameters.

Abbreviation: OPI, Ocular Protection Index.
respectively. A total of 49 and 29 eyes qualified for the dry eye and normal subjects, respectively. Three additional analyses were performed looking at the worst eye of both dry eye and normal subjects using three separate criteria: worst eye based on staining, worst eye based on forced-stare TFBUT, and worst eye based on MBA. Each subject only contributed a single eye to each analysis, for a total of 45 eyes per analysis (dry eye = 29, normal = 16). For the variability analysis, ratios of standard deviation were used for staining, while ratios of coefficients of variation were used for forced-stare TFBUT, IBI, MBA, and OPI 2.0.

**All qualifying eyes**

The Schirmer’s score means for the dry eye qualifying eyes and for the four qualifying eyes of the randomly selected normal subjects were 11.938 mm and 21.000 mm, respectively, for a ratio of 0.568 (P = 0.330). The forced-stare TFBUT means for the dry eye and normal qualifying eyes were 2.599 seconds and 10.908 seconds, respectively, for a ratio of 0.238 (P < 0.001). Figure 4 shows histograms for both dry eye and normal qualifying eyes. The staining score means of the entire cornea for the dry eye and normal qualifying eyes were 1.983 and 0.241, respectively, for a ratio of 8.215 (P < 0.001). The staining score means of the superior region of the cornea for the dry eye and normal qualifying eyes were 1.878 and 0.207, respectively, for a ratio of 9.075 (P < 0.001). The staining score means of

<table>
<thead>
<tr>
<th>Density</th>
<th>Dispersion</th>
<th>Brightness</th>
<th>False negative</th>
<th>False positive</th>
<th>Total false</th>
<th>Total pixels</th>
<th>% error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>L</td>
<td>L</td>
<td>8</td>
<td>0</td>
<td>8</td>
<td>325,620</td>
<td>0.0025</td>
</tr>
<tr>
<td>L</td>
<td>L</td>
<td>H</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>396,360</td>
<td>0.0000</td>
</tr>
<tr>
<td>L</td>
<td>H</td>
<td>L</td>
<td>1852</td>
<td>983</td>
<td>2835</td>
<td>364,704</td>
<td>0.7773</td>
</tr>
<tr>
<td>L</td>
<td>H</td>
<td>H</td>
<td>1169</td>
<td>769</td>
<td>1938</td>
<td>256,432</td>
<td>0.7558</td>
</tr>
<tr>
<td>M</td>
<td>M</td>
<td>M</td>
<td>1654</td>
<td>1021</td>
<td>2675</td>
<td>444,730</td>
<td>0.6015</td>
</tr>
<tr>
<td>H</td>
<td>L</td>
<td>L</td>
<td>740</td>
<td>6168</td>
<td>6908</td>
<td>313,040</td>
<td>2.2067</td>
</tr>
<tr>
<td>H</td>
<td>L</td>
<td>H</td>
<td>1561</td>
<td>4044</td>
<td>5605</td>
<td>282,400</td>
<td>1.9848</td>
</tr>
<tr>
<td>H</td>
<td>H</td>
<td>L</td>
<td>5550</td>
<td>7307</td>
<td>12,857</td>
<td>444,136</td>
<td>2.8948</td>
</tr>
<tr>
<td>H</td>
<td>H</td>
<td>H</td>
<td>1516</td>
<td>4386</td>
<td>5902</td>
<td>337,640</td>
<td>1.7480</td>
</tr>
</tbody>
</table>
the central region of the cornea for the dry eye and normal qualifying eyes were 1.765 and 0.103, respectively, for a ratio of 17.065 (P < 0.001). The staining score means of the inferior region of the cornea for the dry eye and normal qualifying eyes were 2.306 and 0.414 respectively, for a ratio of 5.573 (P < 0.001).

The IBI means for the dry eye and normal qualifying eyes were 10.710 and 7.114 seconds, respectively, for a ratio of 1.506 (P = 0.098). The MBA (mean percent of the cornea exposed) of the entire cornea for the dry eye and normal qualifying eyes was 0.232 and 0.040, respectively, for a ratio of 5.882 (P < 0.001). The MBA of the central region of the cornea for the dry eye and normal qualifying eyes was 0.137 and 0.013, respectively, for a ratio of 10.730 (P = 0.001). The MBA of the superior region of the cornea for the dry eye and normal qualifying eyes was 0.043 and 0.006, respectively, for a ratio of 6.111 (P = 0.029). The MBA of the inferior region of the cornea for the dry eye and normal qualifying eyes was 0.052 and 0.004, respectively, for a ratio of 3.256 (P = 0.061). OPI 2.0 (in units of mean percentage cornea exposed/second) represents the fraction of corneal surface exposed per IBI. OPI 2.0 of the entire cornea for the dry eye and normal qualifying eyes was 0.023 and 0.001, respectively, for a ratio of 2.03 (P = 0.282). The coefficients of variations for MBA and IBI while the normal qualifying eyes were clustered along the X-axis. The coefficients of variations for MBA for the entire cornea of dry eye and normal eyes were statistically significant for dry eye and normal worst eyes based on staining, forced-stare TFBUT, and MBA (Table 1). The forced-stare TFBUT means were statistically significant for dry eye and normal worst eyes based on staining, forced-stare TFBUT, and MBA (Table 1).

The MBA of the entire and inferior cornea was statistically significant for dry eye and normal worst eyes based on staining, forced-stare TFBUT, and MBA (Table 2). OPI 2.0 of the entire and inferior cornea was also statistically significant for dry eye and normal worst eyes based on staining, forced-stare TFBUT, and MBA (Table 2).

Variability analysis
Generally, dry eye qualifying eyes showed greater variability than normal qualifying eyes; typically dry eyes were twice as variable (Table 3). This is also demonstrated graphically in Figure 7. The dry eye qualifying eyes had higher variability for MBA and IBI while the normal qualifying eyes were clustered along the X-axis. The coefficients of variations for MBA for the entire cornea of dry eye and normal qualifying eyes were 8.72 and 4.29, respectively, for a ratio of 2.03 (P = 0.282). The coefficients of variations for OPI 2.0 of entire cornea for dry eye and normal qualifying eyes were 4.08 and 1.59, respectively, for a ratio of 2.57 (P = 0.022).
Discussion

An enhanced understanding of the complexities involved with tear film breakup and blink physiology led to an alternative method for the evaluation of ocular surface protection under normal visual conditions. Although forced-stare TFBUT has been a standard diagnostic tool for over 40 years, it does not provide sensitive information about the overall health of the tear film, namely what occurs after the break in the tear film. In contrast, the OPI 2.0 System implements fully automated software algorithms which provide a real-time measurement of corneal exposure for each IBI during a 1-minute video. From this system, MBA and OPI 2.0 are calculated and analyzed to garner a more complete picture of ocular surface health. The retrospective manual analysis originally used, however, required numerous technician hours to manually grade the area of corneal coverage. The development of the software analysis allows for a frame-by-frame analysis of percent of corneal area exposed and utilizes computer programs to increase the speed of analysis. In addition, the computer program minimizes human error or bias and achieves the outcomes in a more precise manner.

The OPI 2.0 System was able to distinguish between a group of predefined dry eye and normal subjects by way of both MBA and OPI 2.0 in statistically significant fashions. Utilizing the software analysis allows for much more precise calculations of MBA and OPI 2.0 than the manual analysis. This can be attributed to the fact that grading for the manual analysis was made based on a binary evaluation of breakup within each region, where a given region was considered to have breakup in that area regardless of the actual extent of exposure. In contrast, the software analysis provides an actual pixel count of breakup area, which in turn affords a more precise assessment. Additionally, the manual analysis provides average measurements of the percent of cornea exposed for the 1-minute observation period at only time zero (immediately following the blink), at the time of tear film breakup, and at the maximum level of tear film breakup at the end of the IBI. The software analysis, however, analyzes the cornea on a frame-by-frame basis, accounting for individualized points of breakup area.

The evaluation on a region-by-region basis parallels other clinical assessments such as staining grading. On an
aggregate basis, the results of this study suggest that there may be a relationship between MBA and staining, as an increase in MBA of the dry eye population coincided with higher staining scores. The results of this study also indicate that certain regions of breakup, in particular the inferior region, may be important indicators of dry eye. The worst eye analysis confirmed the authors’ interest in the inferior region of the cornea as a key indicator of dry eye, although further research is warranted.

While the goals of this study were to verify and validate the software analysis, the OPI 2.0 System may also be used to classify dry eye patients into subgroups. Dry eye patients are largely variable, due in part to varying disease states, diurnal variations, extensive visual tasks, or environmental stressors that may exacerbate or influence dry eye signs and symptoms. In this study, dry eye patients were typically twice as variable as normal patients, which may be indicative of various subgroups of dry eye patients based on minimal or significant ocular surface exposure and IBI. These various subgroups may represent underlying variations in disease pathophysiology in addition to a distinct opportunity for advances in potential therapies.

It is evident that forced-stare TFBUT alone does not provide enough information to adequately categorize and assess dry eye patients. The OPI 2.0 System allows MBA and OPI 2.0 to be calculated and analyzed. MBA is a global way of assessing the percent of cornea exposed, while OPI 2.0 provides information on tear film stability by factoring the IBI to garner a more complete understanding of overall ocular surface health. Possible limitations of this study include the small normal population analyzed, the measurement of Schirmer’s on only four randomly selected normal subjects, and the conduct of the study without a therapeutic agent. Studies are underway employing the OPI 2.0 System to assess the therapeutic value of a study drug in a clinical trial setting. Further research to understand the relationship between MBA, OPI 2.0, and potential dry eye subgroups is necessary.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


7. D’Arienzo P, Ousler GW 3rd, Schindelar MR. A comparison of two marketed artificial tears in improvement of tear film stability as measured by tear film break-up time (TFBUT) and ocular protection index (OPI). Paper presented at: 5th International Conference on the Tear Film and Ocular Surface: Basic Science and Clinical Relevance; September 5–8, 2007; Taormina, Sicily, Italy.


