Key Factors in the Subjective and Objective Assessment of Conjunctival Erythema

Eric B. Papas

**Purpose.** To establish objectively measurable characteristics of the conjunctival vasculature that correspond with the judgment of erythema by human observers.

**Methods.** Color images of bulbar conjunctiva from 21 subjects were digitally analyzed to extract the following variables characteristic of the scene: vessel width (W), number of vessels (V), proportion of area occupied by vessels (PA), relative redness both in vessels (RRV) and in the whole image (RRI), red-green difference both in vessels (RGB) and in the whole image (RGI), red-blue difference both in vessels (RBV) and in the whole image (RBI), and red hue value (RHIV). These data were compared with subjective judgments by a panel of seven trained observers who independently rated erythema in the same images, using a 0 to 4 scale with decimal interpolation between grades.

**Results.** Correlation analysis indicated significant associations (P < 0.05) between the mean response of the human observers and all the objective variables except RHIV. Associations with the morphometric variables PA (R² = 0.93) and V (R² = 0.90) were markedly stronger than for the best colorimetric variable RBV (R² = 0.62).

**Conclusions.** Judgments of erythema made by human observers do not rely primarily on color but can be closely approximated by a univariate, linear model involving only the proportion of the scene occupied by vessels. Under the conditions of this study, grading of erythema by trained observers can be considered to constitute measurement to at least an interval level. (Invest Ophthalmol Vis Sci. 2000;41:687–691)

Observing erythema (i.e., the appearance of redness), is a valuable clinical procedure, because changes in blood flow within the conjunctiva or sclera accompany a wide range of ocular conditions. In some circumstances, such changes are sufficiently gross that their observation requires little expertise. There are, however, many occasions when the subtlety of the event demands greater sensitivity. To assist in the process of assessment, clinicians and researchers have often resorted to the use of grading scales. Thus, a given presentation is gauged relative to a predetermined set of criteria chosen to represent different degrees of the condition of interest. Such scales vary in their design and may be either descriptive, artistically rendered, photographic, or computer generated.

As a clinical aid, grading scales have been convenient and useful; however, their inherent subjectivity is a source of some concern. Repeated viewing of the same scene, whether by different observers or by the same individual on separate occasions, typically produces a range of responses.

Presumably with this as well as other factors in mind, several groups have applied objective methods to the problem of measuring ocular surface vasculature. A variety of parameters have been used in this body of work in attempts to describe the changes associated with vascular activity. These include vessel caliber, vessel area, percentage vessel area, relative redness, number of vessel segments, intervessel spacing, and vessel length or area. Faced with such diversity, it seems natural to wonder about the nature of the relationship between the array of morphologic and colorimetric factors and the view of erythema arrived at by subjective means. Establishing the relative importance of individual objective measures by comparing them with subjective responses would seem a logical step therefore.

Some efforts in this direction have been made previously, although apparently with little success. Willingham et al. used an objective system to measure the relative redness and percentage of vessel coverage in each of the six reference photographs of an independently produced scale. After correlating the objective results with the scale integers, they claimed good agreement for both variables measured. However, because their data were all derived from just one eye, the validity of this claim is questionable. Owen et al. made objective and subjective measurements in a group of subjects but, unfortunately, the observed region of conjunctiva was not the same in both cases. Perhaps as a result, they did not compare their data over a range of responses. Finally, Guillon and Shah collected both objective and subjective data from 129 individuals, assessing two vascular parameters, vessel width and percentage of coverage. They were unable to demonstrate any consistent linkages, however, probably because of the limited range of conjunctival response present in their sample.

In the absence of other pertinent literature, the present study was undertaken with the purpose of identifying associations between individual, objectively measurable characteris-
tics of the ocular surface vasculature and typical subjective judgments of conjunctival erythema. Knowledge of these factors and their relative importance during the process of subjective judgment would be valuable, not only to designers of objective measurement systems, but also to those either using, or teaching the use of, subjective methods of assessment.

Methods

Image Preparation

The approach taken in this study was to allow the same scene, consisting of ocular surface vascular detail, to be evaluated independently by both human observers and an objective system. To achieve this, a series of images was prepared in the following manner.

After explaining the nature of the procedure and obtaining their informed consent, the bulbar conjunctiva of 21 individuals was imaged using a three-chip, charge-coupled device (CCD; model DXC 930P Sony, Tokyo, Japan) attached to a biomicroscope (30 SLM; Carl Zeiss, Jena, Germany). Still frames were acquired by means of a frame grabber (Flashpoint 128; Integral Technologies, Indianapolis, IN) interfaced with a computer where they were stored as tagged image file format (TIFF) files. Biomicroscope magnification was fixed at ×20 through the 500 × 200 pixels. Subsequent to acquisition, the images were cropped to a size of 500 horizontal picture elements (pixels) by 200 pixels vertically. This corresponded to 3.47 × 1.15 mm on the ocular surface. All potential reference points, such as the corneal periphery, canthi, or lid margins were excluded. Selection of images to be included was made from a range similar to that represented by the Cornea and Contact Lens Research Unit (CCLRU) Clinical Grading Scale. The tenets of the Declaration of Helsinki were observed throughout.

Human Assessment

The 21 images were separately displayed on a monitor (Multiscan 200ES, Sony). Seven observers independently viewed and assessed each scene. All individuals were optometrists who had previously been trained in the use of the CCLRU scale, were accustomed to assessing erythema with this scale, and had been doing so regularly and routinely for at least 1 year. Order of presentation was random and altered for each observer. Judgments were made using a 0 to 4 scale with their informed consent, the bulbar conjunctiva of 21 individuals was imaged using a three-chip, charge-coupled device (CCD; model DXC 930P Sony, Tokyo, Japan) attached to a biomicroscope (30 SLM; Carl Zeiss, Jena, Germany). Still frames were acquired by means of a frame grabber (Flashpoint 128; Integral Technologies, Indianapolis, IN) interfaced with a computer where they were stored as tagged image file format (TIFF) files. Biomicroscope magnification was fixed at ×20 throughout. Subsequent to acquisition, the images were cropped to a size of 500 horizontal picture elements (pixels) by 200 pixels vertically. This corresponded to 3.47 × 1.15 mm on the ocular surface. All potential reference points, such as the corneal periphery, canthi, or lid margins were excluded. Selection of images to be included was made from a range similar to that represented by the Cornea and Contact Lens Research Unit (CCLRU) Clinical Grading Scale. The tenets of the Declaration of Helsinki were observed throughout.

Objective Assessment

Digital processing of the same 21 images was performed in several ways to yield numerical parameters that would potentially be characteristic of the scene. Image manipulations were made within the framework of a commercially available software package (Image Pro Plus; Media Cybernetics, Silver Spring, MD). The parameters measured were as follows.

Relative Redness of Image. In the red-green-blue (RGB) system of representing color, each pixel in the image is associated with three values corresponding to the intensities of the colors red, green, and blue. These values range from 0 to 255. Relative redness for the image (RRI) was calculated as:

\[ RRI = \sum_{i=1}^{500} \sum_{j=1}^{200} \left( \frac{R_{ij}}{R_{ij} + G_{ij} + B_{ij}} \right) \]

where \( R_{ij} \), \( G_{ij} \), and \( B_{ij} \) are the red, green, and blue intensities, respectively, for a pixel at position \( i, j \) in the image array. The summation is taken over the entire image array of 500 × 200 pixels.

Relative Redness of Image. Relative redness of vessels (RRV) was calculated in a fashion similar to RRI, except that nonvascular background was excluded.

Red-Green Difference in Entire Image. The differences between red and green intensities (RGI) for each pixel were calculated and summed over the entire image:

\[ RGI = \frac{\sum_{i=1}^{500} \sum_{j=1}^{200} (R_{ij} - G_{ij})}{N} \]

Note that \( N = 500 \times 200 = 10^5 \) for all images.

Red-Green Difference in Vessels. Red-green difference in vessels (RGV) was calculated in a fashion similar to RGI, except that nonvascular background was excluded.

Red-Blue Difference in Entire Image. The difference between red and blue intensities (RBI) for each pixel were calculated and summed over the entire image:

\[ RBI = \frac{\sum_{i=1}^{500} \sum_{j=1}^{200} (R_{ij} - B_{ij})}{N} \]

Red-Blue Difference in Vessels. Red-blue difference in vessels (RBV) was calculated in a fashion similar to RBI, except that nonvascular background was excluded.

Red Hue Value. An alternative to the RGB approach to representing color is the hue, saturation, and intensity (HSI) system. Here, information relating to the “coloredness,” or hue, of each pixel is contained in the hue parameter (\( H \)), again with values ranging from 0 to 255. Pure red is centered at an \( H \) value of zero. The red hue value (RHV) was calculated as follows:

\[ RHV = \frac{\sum_{i=1}^{500} \sum_{j=1}^{200} |128 - H_{ij}|}{N} \]

Thus, RHV ranged from zero (turquoise) to 128 (pure red).

Number of Vessels. The original color image was first converted to 8-bit gray scale. Contrast was then enhanced using a local equalization technique similar to that described by Chen et al. The result was binarized so that the final image contained only white pixels corresponding to vessels in the original image and black pixels elsewhere.

To evaluate the number of vessels (\( V \)), 10 strips with dimensions 500 × 2 pixels were digitally imposed on the binarized image. All were placed parallel with the image’s long axis and evenly spaced at 20-pixel intervals along the short axis. Within each strip the number of vessels crossing both boundaries was counted. The average of this value over all 10 strips was taken as the number of vessels in the image (\( V \)).
Vessel Width. The mean width of vessels \((W)\) in the binarized image.

Proportion of Area Occupied by Vessels. To arrive at this parameter, the number of pixels representing vessels in the binarized image \((PA)\) was divided by the total number of pixels \((10^5)\).

**RESULTS**

Preliminary checks indicated that all data were approximately normally distributed (Kolmogorov–Smirnov; \(P > 0.05\)). The mean scores from all seven observers for each image together with the associated SD are shown in Figure 1. Overall, on this set of images, the average SD of judgments was 0.33 grades. This implies that 95% of observations made by these observers on the same image would be within a range of \(\pm 0.8\) grades.

Taking mean observed grade \((OG)\) as ordinate in all cases, a series of scatterplots was constructed, each having one of the measured variables as its abscissa. These plots are shown in Figure 2. In several instances, evidence of a linear relationship between the plotted variables was detected on visual inspection. Further confirmation was made by calculating bivariate Pearson correlation coefficients as shown in Table 1.

Overall, the strongest relationships found were those with \(V\) and \(PA\). These variables accounted for 90% and 93%, respectively, of the variance in \(OG\). Vessel width, showed only a relatively weak association with \(OG\) \((R^2 = 0.39)\).

Among the colorimetric variables, all except \(RHV\) were significantly correlated with \(OG\) and had fairly similar coefficients of determination. The association with the \(RBV\) was the strongest within this group \((R^2 = 0.62)\), although it was notably weaker than that for \(V\) and \(PA\). Attempts to improve the fit still further by combining colorimetric and morphometric information in a multivariate model were ineffective.

**DISCUSSION**

Given that the purpose of this study was to discover which observable characteristics of the conjunctival surface contribute to the perception of its redness, it seemed reasonable at the outset to expect that some aspect of the color information contained within the image would be an important factor. For this reason, several data constructs were incorporated into the analysis that, in one sense or another, encoded for redness. All these, with the exception of \(RHV\), produced reasonably good fits to the observers’ data. However, even the most successful measure \((RBV)\), fell considerably short of the conformity shown by the two morphologic variables \(V\) and \(PA\). That these measures were individually capable of accounting for at least 90% of the variance in the observers’ ratings, suggests that the process whereby this group made their judgments primarily involved gauging the extent or quantity of vessels in the scene rather than their color. Both these variables can be thought of as indicators of vascular density and were highly correlated with one another \((R^2 = 0.95)\).

The finding that assessments of conjunctival erythema can be modeled morphometrically is an important one so far as the design of an objective measurement system is concerned. The ability to ignore color permits monochromatic image acquisition, which is relatively inexpensive and offers advantages in CCD size and sensitivity. Furthermore, the amount of data involved is reduced by a factor of three, allowing processing and storage to be simpler, faster, and less demanding of computer resources.

The strong association between \(PA\) and \(OG\) has the intriguing consequence that good, subjective erythema judgments should be possible even from a monochromatic scene. Because the suggestion that the “redness” of a black and white image could be judged may be conceptually difficult, a supplementary experiment was performed to test its accuracy. Five of the original observers, who were available, were asked to re-view the test images. On this occasion, however, color information was removed before viewing by converting the images from their original RGB format to 8-bit gray scale. Because approximately 2 months had elapsed since the first viewing session, recall of previous scores was deemed to be unlikely. Nevertheless, the order of presentation was randomized for all observers. Correlating the mean grading scores from the monochromatic viewing with those from the original colored session yielded a coefficient of determination \((R^2)\) of 0.97 \((P < 0.0001)\). Mindful of the criticisms leveled at correlation as a method of indicating agreement, the difference between the gradings made on color and monochromatic images was plotted against their mean, as shown in Figure 3. No evidence of a relationship with measurement size was indicated \((R^2 = 0.1, P = 0.16)\), and the mean difference between color and monochromatic gradings was \(0.13 \pm 0.15\) (SD).

Overall, across the range of measurement, this result indicates a slight tendency for monochromatic judgments to underestimate those for color, but only by a small amount. The limits of agreement suggest that 95% of color-monochrome differences would fall between approximately 0.5 and \(-0.2\) grades. Neither the magnitude of the bias, nor its associated error are large compared with the random variability inherent in subjective grading, and it appears reasonable to conclude that removal of color information from this kind of grading task did not materially alter its outcome. This tends to confirm the suggestion that, for this group of observers at least, erythema judgment is essentially morphometric.

Apart from its methodologic implications, this study may also bear on the way graded data are treated analytically. Until now, information of this type has commonly been regarded as ordinal in nature. The reasons for this and its ramifications have been comprehensively discussed elsewhere, but briefly, ordinal scales simply require that some meaningful order, or
Figure 2. Mean observed erythema grade versus measured image characteristics. Error bars, 1 SD.
Correlation of Human Observations with Each Listed Variable

**TABLE 1.** Coefficients of Determination and *P* from Pearson Correlation of Human Observations with Each Listed Variable

<table>
<thead>
<tr>
<th>Variable</th>
<th><em>R</em>^2</th>
<th><em>P</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative redness image (RRI)</td>
<td>0.49</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Relative redness vessels (RRV)</td>
<td>0.55</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Red-green difference image (RGI)</td>
<td>0.49</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Red-green difference vessels (RGV)</td>
<td>0.52</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Red-blue difference image (RBRI)</td>
<td>0.62</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Red-blue difference vessels (RBRV)</td>
<td>0.59</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Red hue value (RHV)</td>
<td>0.17</td>
<td>0.06</td>
</tr>
<tr>
<td>Number of vessels (V)</td>
<td>0.90</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Vessel width (W)</td>
<td>0.39</td>
<td>0.0002</td>
</tr>
<tr>
<td>Percentage vessel area (PA)</td>
<td>0.95</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Achieving the next higher level (i.e., interval measurement) requires that successive scale increments be equally spaced. Clearly, this presents a problem for most grading scales because of difficulty in obtaining a suitable standard for comparison. However, an inspection of Figure 2J plainly shows that the relationship between OG (a graded score), and PA (a continuous variable) is closely linear across the grading range used. A line fitted to the data has the equation \( OG = 9.0 \times PA + 1.3 \) (\( R^2 = 0.93; P < 0.0001 \)). Thus, an increase of one erythema grade corresponded to a 0.11 increase in PA, irrespective of whether the change was from Grade 1 to 2, from 3 to 4, or from 2.7 to 3.7. This appears to satisfy the requirement for interval scale measurement. As such, the practice of avoiding using statistics such as the mean and SD, or parametric analytical techniques such as analysis of variance, simply on order of measurement grounds does not seem justified.

An interesting further observation along these lines is that PA goes to zero when OG is approximately 1.3, implying that grades below 1 are redundant. Making an axis translation to account for this results in a scale the zero point of which could be said to have a real physical meaning, (i.e., it corresponds to complete absence of vascular detail; \( PA = 0 \)). Interestingly, having a meaningful zero point is a property required of ratio scale measurement that constitutes the next, and highest, level of scaling.

In summary, this work has demonstrated that subjective erythema judgment can be closely modeled using a linear, univariate, morphometric approach and that, under the conditions of the study, clinical grading displays at least interval level measurement characteristics.

**Acknowledgments**

The author thanks Isabelle Jalbert, Lisa Keay, Edward Lum, Padmaja Sankaridurg, Cheryl Skotnitsky, Andrew Stephenson, and Rob Terry for supplying their clinical skills.

**References**