Clinical application of a digital method to improve the accuracy of color perception in toluidine blue stained oral mucosal lesions

Pelin Guneri, PhD, DDS1/Joel B. Epstein, DMD, MSD, FRCD(C), FDS RCS(E)2/Betul Ilhan, PhD, DDS3/Aslihan Kaya, PhD, DDS4/Hayal Boyacioglu, PhD5

Objective: Color perception is an important variable in detecting and assessing oral conditions. The aim was to investigate clinicians’ perception of toluidine blue (Tblue) staining compared to digital color analysis, which may impact mucosal lesion detection, affect the decision to biopsy, and biopsy site selection. Method and Materials: Four oral lesions were stained with Tblue. Digital color analyses of eight areas on each image were completed and were considered as “gold standard” (GS). Twenty specialists ranked these areas according to their perceived intensity of blue stain in two sessions. Results: Consistency between GS and observers rankings was 0.8791. However, more than half of the observers inaccurately perceived the intermediate blue tones. Overall interobserver agreement was 0.8714; stability between two sessions decreased to 45% for intermediate tones. Conclusion: Assessing the equivocal blueness of an oral mucosal lesion in clinical settings may vary due to variation in visual perception. A digital method for objective color analysis in clinical practice may be used to eliminate this deficiency by implementing a mathematical formula. (Quintessence Int 2013;44:619–627; doi: 10.3290/j.qi.a29510)

Key words: color perception, oral cancer, oral diagnosis, oral medicine, toluidine blue

Toluidine blue (Tblue) is a cationic meta-chromatic dye that may selectively bind to sulfate, phosphate, and carboxylate radicals of large molecules.1 It may stain DNA and/or may be retained in intracellular spaces of oral squamous cell carcinoma (OSCC) and oral premalignant lesions (OPLs) that clinically appear as royal blue areas.2,3 Tblue has been shown to assist in detection of oral mucosal areas with molecular changes that are associated with OPLs or OSCC, in assessment of margins of OPLs/OSCC and biopsy sites selection,3,4 and in detection of second primary oral cancers or recurrences.5 Furthermore, Zhang et al6 and Guo et al7 have reported that the stained oral mucosal areas with benign histology and mild dysplasia and normal margins may harbor clonal changes with risk of progression to malignant lesions. Thus, Tblue is used in vivo as an adjunct to the clinical oral examination to provide supplementary clinical information related to prognosis of premalignant lesions or lesions that may represent OSCC at diagnosis.4,6 However, in addition to the difficulties in accurately assessing variations in the texture and red or white changes of the original lesion, inaccurate perception of the blueness of a Tblue stained suspicious oral mucosal lesion may lead to future hazards, such as delay in early diagnosis and immediate therapy of oral mucosal malignancies.7,8 It has been reported that in long-
Human color perception is considered subjective because alterations in perception of color can occur as a result of numerous factors. Color can be measured and is expressed in the coordinates of a color order system. The most complete color space that describes colors visible to the human eye is defined as CIELAB by International Commission on Illumination (Commission Internationale d’Eclairage). In this system, three coordinates constitute the color space: L* refers to the lightness coordinate (range from 0 for perfect black to 100 for perfect white), and a* and b* represent chromaticity coordinates in the red-green axis and yellow-blue axis, respectively. Positive a* values reflect the red color range and negative values indicate the green color range. Likewise, positive b* values indicate the yellow color range while negative values indicate the blue color range. This method was supported by the American Dental Association in measuring color variations on dental surfaces and dental materials. It is used to establish the color of dentition and clinical parameters of lesions and to quantitatively evaluate the lesion progression and response to treatment.

The primary purpose of this study was to assess color perception in a clinical setting to investigate whether clinicians perceived intensity of blue color of specified areas of four oral lesions after Tblue staining. Color perception that was assessed by objective digital color analysis (the “gold standard”, GS) was compared to clinical visual findings. Second, we evaluated a digital method that may have clinical application in cases in which it may be difficult to establish the need and site selection for tissue biopsy.

METHOD AND MATERIALS

Four patients who were referred with chronic oral mucosal lesions with malignant potential/nature were enrolled in the study. These cases were various mucosal entities that would provide a range of mucosal Tblue staining intensity. They were representative of malignancy potential that may present with varying characteristics, which may make detection easy or challenging and even when readily visualized present challenges in biopsy site selection. Following the interview and clinical examination, patients’ informed consents were obtained. Institutional review board approval was not required since this process was part of routine procedure required for suspicious oral mucosal lesions. Helsinki Declaration guidelines were followed in this investigation. All lesions were stained with Tblue by rinse application in order to examine the presence of dysplastic/malignant changes of the lesion and the adjacent area, and were reexamined. Tblue solution was prepared by compounding laboratory grade powder (1 g toluidine blue powder, 10 ml acetic acid, 4.19 ml absolute alcohol, and 86 ml distilled water). Acetic acid rinse (1%) was used before and after Tblue staining, and color images of the lesions were obtained by a professional photographer using a digital camera (6.31 mega pixel resolution, 23.7 × 15.5 mm sensor size, automatic focus function; Olympus Camera C-2500L). In order to standardize the lightness between the images, a 5-mm diameter white disk was placed adjacent to each lesion during photographing. The images were transferred to a personal computer and saved as tagged image file format (TIFF) files. One of the investigators who was experienced in Tblue staining (PG) determined and selected the areas with varying blue color intensities within each lesion that were used for clinicians’ color assessment. On each image, eight circular areas, each 50 × 50 pixels, were designated within the lesion using the elliptical marking tool of Adobe Photoshop CS2 version 9.0.2 (Adobe Systems). Additionally, two sample areas (one from the pale colored area [A], the other from the dark area [B]) of normal-appearing oral mucosa were selected as well (Figs 1 to 4). In order to assess the perceptual differences between any two colors in the L*, a*, b* system (ΔE*ab), the differences in the lightness and chromaticity coordinates (ΔL*, Δa*, Δb*).
Δb*) were determined. Then, overall color change was assessed using the formula:

$$\Delta E^*_{ab} = (\Delta L^*^2 + \Delta a^*^2 + \Delta b^*^2)^{1/2}.$$  

The L*, a*, and b* values of each selected circular area and unstained adjacent oral mucosa samples (A and B) and the white calibration sites were measured using the histogram function of the software, as reported previously. In CIELAB color space, the higher the L* value, the brighter the image. Similarly, higher a* value reveals more redness of the region, whereas lower b* value represents more blueness of the area. In each image, three measurements were performed for each parameter and the mean values were calculated. These values that were obtained with computer image analysis were accepted as the GS. In order to investigate the correspondence between clinical color interpretations with the GS, 20 oral medicine specialists who had experience in evaluation of oral mucosal lesions examined the images. They were asked to rate the eight identified areas on each image according to the intensity of blue stain. The age of the observers ranged between 26 and 58 years (mean 38.4 years ± 9.93). Their experience in oral mucosal lesion diagnosis varied between 2 and 34 years (mean 15.2 years ± 10.04). Each observer had a separate blinded session, and all evaluated the image using the same computer screen. To assess intraobserver accuracy, each observer reevaluated the images 2 weeks after their initial interpretations under the same conditions.

Statistical analyses were performed using SPSS 10.0 (SPSS) and Minitab 13 software (Minitab Statistical Software 2000; Minitab). Pearson’s coefficient of contin-

**Fig 1** Case 1. Histological diagnosis: verrucous carcinoma.

**Fig 2** Case 2. Histological diagnosis: squamous cell carcinoma.

**Fig 3** Case 3. Histological diagnosis: squamous cell carcinoma.

**Fig 4** Case 4. Histological diagnosis: erosive lichen planus with moderate dysplasia.
RESULTS

The findings reported in this study were relevant to the results of the cases presented, and due to the small number of cases the findings are interpreted cautiously and may not be generalized. The visual examination of the lesions before and after staining with Tblue rinse revealed no satellite lesions that required further investigation. The lesions were surgically excised with at least 4-mm margins and were histologically evaluated, although histological diagnostic examination may not have been performed for the specified eight areas in each lesion. The histological diagnoses were made following World Health Organization (WHO) diagnostic criteria\(^{20}\) and were verruocous carcinoma for Case 1, squamous cell carcinoma for Case 2 and Case 3, and erosive lichen planus with moderate dysplasia for Case 4.

Firstly, digital analyses of the pale and dark samples of normal-appearing mucosa were completed on each image. As observed in our previous case study,\(^{17}\) we noted that the visual differences regarding the paleness/darkness between two areas originated from the \(L^*\) values rather than the \(a^*\) and \(b^*\) values of these areas. Thus, the variations in pale or dark color perception were impeded by the whiteness component \((L^*)\) of the area. In order to debug the effect of lightness from the appearance of blue tones and to clarify the blueness of the sample, we provided an objective ranking criterion for accurate establishment of \(L^*\), \(a^*\), and \(b^*\) values of the samples in each lesion’s digital color image, and used a formula to adjust the \(b^*\) value for lightness \((L^*)\):

\[
b_i^* = b_i + AC(L_W - L_X)
\]

where \(b^*\) is \(b\) value adjusted for \(L^*\), \(W\) is white calibration paper, \(b\) is blueness, and \(AC\) is adjustment coefficient, which was defined as:

\[
AC = \frac{L_A - L_B}{b_A - b_B}
\]

Here, \(A\) represented the healthy pale oral mucosa on the image, whereas \(B\) defined the healthy dark oral mucosa on the image. The expanded formulation of \(AC\) is as follows:

\[
AC = \frac{(L_W - L_B) - (L_W - L_A)}{(b_W - b_B) - (b_W - b_A)} = \frac{L_A - L_B}{b_A - b_B}
\]

The methodology may be better revealed with an example: (for sample number 1 in Case 2):

\[
CC = \frac{L_A - L_B}{b_A - b_B} = \frac{L_W - L_B}{b_W - b_B} = \frac{171.15 - 112.20}{143.58 - 149.96} = \frac{58.95}{-6.38} = -9.24
\]

Then, \(b\) value adjusted for lightness \((L^*)\) was calculated as:

\[
\]

This formula was used for correction of the \(b^*\) values of all samples in each image, and in order to expose the blueness of the area stripped from the influence of the lightness. Objective ranking of the samples according to the “adjusted \(b^*\) values”, ranging from the lightest to darkest area, are presented in Table 1.

In the study, 20 observers ranked eight areas in each image of the four cases, then they reevaluated the same images after 2 weeks; resulting in a total of 1,280 rankings. The objective rankings depending upon the adjusted \(b^*\) values, and the observers total rankings are presented in Table 2. The observers accurately ranked the lightest and darkest samples (98.12% and 87.50%, respectively). However, intermediate intensities of blueness were associated with less accurate rankings by as much as 40% (Table 2).

The analysis of the rankings of the observers and the objective measurements (GS) revealed that \(\chi^2 = 4,355.8\).

Pearson’s contingency coefficient was determined by using the following formula:

\[
CC = \sqrt{\frac{X^2}{X^2 + N}} = \sqrt{\frac{4355.8}{5635.8}} = 0.8791
\]

The agreement between the objective rankings and observers rankings was 0.8791.

The level of consistency between the first and second rankings of the observers (intraobserver agreement) is presented in Table 3. The observers’ agreements for the lightest and darkest samples in two sessions were very high (98.75% and 90%, respectively). However, the consistency
Table 1  Objective ranking of the samples according to the adjusted b* values, ranging from the lightest to darkest area

<table>
<thead>
<tr>
<th>Objective ranking</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b_{adj} value</td>
<td>128.18</td>
<td>125.56</td>
<td>118.70</td>
<td>116.00</td>
<td>111.53</td>
<td>91.17</td>
<td>86.54</td>
<td>83.56</td>
</tr>
<tr>
<td>Sample no. on the image</td>
<td>8</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Case 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b_{adj} value</td>
<td>129.26</td>
<td>129.94</td>
<td>128.88</td>
<td>127.14</td>
<td>119.65</td>
<td>119.57</td>
<td>114.36</td>
<td>110.01</td>
</tr>
<tr>
<td>Sample no. on the image</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Case 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b_{adj} value</td>
<td>131.66</td>
<td>118.17</td>
<td>117.02</td>
<td>112.04</td>
<td>110.62</td>
<td>109.10</td>
<td>105.31</td>
<td>101.00</td>
</tr>
<tr>
<td>Sample no. on the image</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>8</td>
<td>3</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Case 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b_{adj} value</td>
<td>118.06</td>
<td>112.13</td>
<td>111.47</td>
<td>109.95</td>
<td>107.01</td>
<td>107.00</td>
<td>106.48</td>
<td>106.3</td>
</tr>
<tr>
<td>Sample no. on the image</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2  The consistency between the objective rankings and the observers’ rankings (ranging from the lightest to darkest)

<table>
<thead>
<tr>
<th>Objective ranking</th>
<th>Observers’ ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6 7 8</td>
</tr>
<tr>
<td>1</td>
<td>157 1 2</td>
</tr>
<tr>
<td>2</td>
<td>3 122 31 2 2</td>
</tr>
<tr>
<td>3</td>
<td>28 82 30 19 1</td>
</tr>
<tr>
<td>4</td>
<td>4 10 66 66 12 2</td>
</tr>
<tr>
<td>5</td>
<td>5 31 20 64 36 4</td>
</tr>
<tr>
<td>6</td>
<td>4 35 7 100 13 1</td>
</tr>
<tr>
<td>7</td>
<td>7 2 10 122 19</td>
</tr>
<tr>
<td>8</td>
<td>1 19 140</td>
</tr>
<tr>
<td>Total</td>
<td>160 160 160 160 160 160 160 1280 66.64</td>
</tr>
</tbody>
</table>

Table 3  The consistency between the first and second rankings of the observers (ranging from the lightest to darkest)

<table>
<thead>
<tr>
<th>First ranking</th>
<th>1 2 3 4 5 6 7 8</th>
<th>Total</th>
<th>Recording the same rank in 2 sessions (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>79 1</td>
<td>80</td>
<td>98.75</td>
</tr>
<tr>
<td>2</td>
<td>1 62 14 2</td>
<td>80</td>
<td>77.50</td>
</tr>
<tr>
<td>3</td>
<td>13 43 21 3</td>
<td>80</td>
<td>53.75</td>
</tr>
<tr>
<td>4</td>
<td>2 18 36 15 7 2</td>
<td>80</td>
<td>45.00</td>
</tr>
<tr>
<td>5</td>
<td>2 4 13 47 12 2</td>
<td>80</td>
<td>58.75</td>
</tr>
<tr>
<td>6</td>
<td>1 7 10 53 8 1</td>
<td>80</td>
<td>66.25</td>
</tr>
<tr>
<td>7</td>
<td>1 5 6 61 7 8</td>
<td>80</td>
<td>76.25</td>
</tr>
<tr>
<td>8</td>
<td>1 7 72 80</td>
<td>80</td>
<td>90.00</td>
</tr>
<tr>
<td>Total</td>
<td>80 80 80 80 80 80 80 80 80 640 70.78</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
between the first and second rankings of the observers decreased to 45% for intermediate tones (Table 3), indicating a limited ability in color perception among the observers to rank the intermediate color tones with the same order in two settings.

For the overall interobserver agreement, analysis of the rankings recorded in two sessions of the observers revealed that $\chi^2 = 2019.4$. The coefficient of contingency (CC) was calculated using the following formula:

$$CC = \sqrt{\frac{X^2}{X^2 + N}} = \sqrt{\frac{2019.4}{2659.4}} = 0.8714$$

The overall interobserver agreement was 0.8714, which was only very slightly lower than the consistency between the objective and observers rankings (0.8791).

**DISCUSSION**

Color perception is a complex process that involves optical, physiological, and psychological components. It is initiated by illumination of the object with light of varying wavelengths and intensities: depending on the frequency of the light waves, the object absorbs and reflects the light in different ways. The light is processed by the eye of the observer and finally is perceived as color.9 If the light’s dominant wavelength is in the upper boundary of the visible spectrum, it is perceived as red; if in the lower end, then the light is defined as blue. In addition to the light’s physical properties, the observer’s eye-brain interaction is another determinant in color assessment.21

The human visual system can discriminate almost 35,000 colors,21 but color perception may be affected by the size of the object,21 the type and intensity of the illuminant,22 and the color of the background.9 Furthermore, the reflective spectra are not identical under disparate forms of light; thus, in such conditions, different tri-stimulus values and different shades of color may arise.9 This phenomenon is defined as metamerism, and it may cause variations between clinicians’ color perceptions and definitions in different examination settings. In order to unravel this problem, a control material (a white photographic paper) is used in photography and image processing in order to calibrate the differences prior to digital data analysis.14,17 In our previous report, which was based upon one clinical case, dark and pale oral mucosal areas had different L* values when compared to their a* and b* values.17 Moreover, these areas’ lightness parameter (defined by the L* value) influenced the clinical impression of the intensity of blue color. In other words, observers’ blue color perception was influenced by darkness/lightness of the area, which may in part be related to the intraoral location of the lesion.17 The same observation was encountered in this investigation, and standardization of the lightness was required in order to eliminate the impact of light on the blue tones’ perceptibility, and to provide an accurate color determination.

In the present study, we used clinical images which had apparent malignant or suspicious characteristics in order to test the applicability of our model on the representative cases. In this study of oral medicine experts, we sought to assess concordance of clinician perception compared to digital detection, mimicking the impact upon biopsy site selection, rather than focusing on detection of lesions. Further studies could be prepared for cases that may be difficult to visualize in order to evaluate the impact of blue tones upon detection of lesions. We also planned to test the sensitivity and specificity of the model on future complicated cases, since the ones used in the present paper were already representative of malignancy potential.

In the literature, only “a royal-blue” intense Tblue stain was accepted as positive in some studies, while any staining was considered positive in others. Gandolfo et al23 reported that all OSCC stained dark blue with Tblue. However, Gray et al1 and Missmann et al2 accepted equivocal staining as “positive” and showed that the sensitivity of Tblue was as low as 40% and as high as 100% in these instances. If equivocal stained lesions were considered negative, the sensitivity rate varied from 100% to 81%.1 In another study, when equivocal staining was considered positive, the specificity of Tblue in OPLs ranged from 31% to 93%.1 Nonetheless, accepting pale blue staining as negative increased the specificity significantly.1 The value of Tblue staining
in preliminary evaluation of OPLs and OSCC is based on the color of the stained lesion coupled with clinical appearance, including the variation in color and surface texture of the lesion and its location. However, as observed in our previous study where color analysis was examined in a single case, this judgment was affected by stain intensity and color perception of the observer.

In the present study, after correcting the blueness of the Tblue stained sample areas using a basic mathematical formula, we observed that the clinicians categorized the darkest and lightest blue areas accurately, and the ranking order of the observers were in accordance with L*, a*, and b* values of the same areas. Also, the observers’ agreement was highest in darkest and lightest colored areas. However, the eye of the clinical observers failed to rank the intermediate color’s intensity accurately when compared to the GS. Additionally, the observers’ agreement was lower in the midtone areas when compared to that of the darkest and lightest colored areas. It is known that the human eye is more sensitive to green and least sensitive to blue, which may be an explanation of observers’ poorer performance in ranking the intermediate blue tones with respect to the GS. In this study, we investigated the blueness of the lesion because we used images after Tblue staining; however, variation in red and white as well as texture could also be critical variables in the clinical care of patients with mucosal change. Future studies regarding the impact of color (especially red and white changes) of unstained lesions on visual perception, and the influence of this process on the clinical management of oral mucosal lesions are required.

The association between visual color perception and spectrophotometric analysis in dental esthetics has been documented. In medicine, quantitative color analysis was used to describe the chromoendoscopic findings more objectively for early diagnosis of esophageal squamous cell carcinoma. Ishihara et al reported that for diagnosing high-grade intra-epithelial neoplasia and cancer, quantitative assessment of the pink-color sign (a marker of malignancy) resulted in a high sensitivity and specificity (88% and 95%, respectively). Jönsson et al showed that the values from image analysis of extravasated plasma albumin marked with Evans blue in burned tissue specimen correlated with invasive measurements from spectrophotometric analysis and subjective decisions made by the naked eye.

In order to assist the clinical decision process, Patel et al have suggested use of computer-assisted decision support technologies as an adjunct to conventional decision-making means. The present study investigated the place of digital color analysis in the clinical decision-making process and resulted in a number of clinical implications: the observers ranked the darkest and the palest samples visually in accordance to the rank order of digital color analysis (GS); however, they had more difficulty in identifying the blue intensity in the areas with intermediate blue tones. In the present study, we observed that the lightness or darkness of the area influenced clinicians’ blue tone perceptibility, which may affect clinical assessment of the malignancy potential of oral lesions and in choosing biopsy site selection. This may be impacted by clinical experience as well as visual perception. In our investigation, we have suggested a digital method to provide objective assessment of blue color intensity after Tblue staining via standardizing the color parameters. In order to achieve a GS/objective color establishment, b* values of Tblue stained oral lesions on digital color images can be “corrected” with a simple mathematical formula using the differences between L* values of the white calibration paper and oral lesion. By utilizing this method in clinical practice, visually problematic areas may be ranked according to their corrected b* values and more accurate blueness, which may help to determine biopsy site selection of Tblue stained oral mucosal lesions with malignant potential. Additionally, the areas that require further investigation may be revealed in the clinical setting. The objective method suggested for assessing intermediate blue tones does not disregard the necessity of histopathologic examination of a suspicious lesion. Rather, it may assist the clinician in detecting blue stain retentive sites of a particular...
lesion that would be the most appropriate site for biopsy and histopathologic analysis.

A systematic literature review recommended the use of TBlue in high risk patients by experienced providers. In the present study, we observed that the observers had poor agreement in the midtone blue areas when compared to that of the darkest and lightest blue colored sites. Provided that human color perception is “subjective” because of alterations in color perception caused by numerous uncontrollable factors, relying on the clinicians’ ability to accurately perceive the blue tone intensity after TBlue rinse application may reduce the utility of TBlue. Visualization of TBlue retention should be considered along with other clinical features of the lesions including homogeneity or nonhomogenous color, texture, induration, ulceration, and symptoms. While we did not evaluate red or white changes specifically in this study, this may affect clinical assessment of potentially malignant lesions due to variability in color perception, and should be examined in future studies. Determination of the lesion color using an objective method such as the one presented may provide additional ability to detect and to predict OPLs at risk of progression to cancer, since any staining with TBlue should elevate the index of suspicion and should prompt further evaluation. The methods employed were simple and can be performed in the clinic with a personal computer and a common graphics program. However, the utility and validity of the presented method require further analyses with more complicated cases in order to establish its application to evaluate the impact of color perception on detection of oral lesions and evaluation of color in detection and diagnosis of suspicious oral mucosal lesions.

ACKNOWLEDGMENTS

The authors thank all coworkers who participated in evaluation of the images, and Prof Dr Saim Kendir for his contributions to the analyses of data.

REFERENCES


