Published in final edited form as:

Ocul Surf. 2018 October ; 16(4): 436–440. doi:10.1016/j.jtos.2018.06.003.

A New Scale for the Assessment of Conjunctival Bulbar Redness

Ilaria Macchi, MD^{1,2}, Vatinée Y. Bunya, MD¹, Mina Massaro-Giordano, MD¹, Richard A. Stone, MD¹, Maureen G. Maguire, PhD¹, Yuanjie Zheng, PhD^{1,3,4}, Min Chen, PhD⁵, James Gee, PhD^{5,6}, Eli Smith, BA¹, and Ebenezer Daniel, MBBS PhD¹

¹Scheie Eye Institute, University of Pennsylvania, Philadelphia, PA

²Department of Ophthalmology, University Campus Biomedico, Rome

³School of Information Science and Engineering, Shandong Normal University

⁴Shandong Normal University, School of Information Science and Engineering, Key Lab of Intelligent Computing & Information Security in Universities of Shandong, Institute of Life Sciences, Shandong Provincial Key Laboratory for Distributed Computer Software Novel Technology, and Key Lab of Intelligent Information Processing, Jinan, 250358, China

⁵Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

⁶Institute of Translational Medicine and Therapeutics, University of Pennsylvania, Philadelphia, PA

Abstract

Purpose: Current scales for assessment of bulbar conjunctival redness have limitations for evaluating digital images. We developed a scale suited for evaluating digital images and compared it to the Validated Bulbar Redness (VBR) scale.

Methods: From a digital image database of 4889 color corrected bulbar conjunctival images, we identified 20 images with varied degrees of redness. These images, ten each of nasal and temporal views, constitute the Digital Bulbar Redness (DBR) scale. The chromaticity of these images was assessed with an established image processing algorithm. Using 100 unique, randomly selected images from the database, three trained, non-physician graders applied the DBR scale and printed VBR scale. Agreement was assessed with weighted Kappa statistics (K_w).

Results: The DBR scale scores provide linear increments of 10 from 10–100 when redness is measured objectively with an established image processing algorithm. Exact agreement of all graders was 38% and agreement with no more than a difference of ten units between graders was 91%. K_w for agreement between any two graders ranged from 0.57 to 0.73 for the DBR scale and from 0.38 to 0.66 for the VBR scale. The DBR scale allowed direct comparison of digital to digital images, could be used in dim lighting, had both temporal and nasal conjunctival reference images, and permitted viewing reference and test images at the same magnification.

Corresponding Author: Ebenezer Daniel, MBBS, MS, MPH, PhD, Ophthalmology Reading Center, Department of Ophthalmology, University of Pennsylvania, 3535 Market St, Ste 700, Philadelphia, PA 19104 (ebdaniel@mail.med.upenn.edu). Ph: 215-615-1523. Fax: 215-615-1533.

Conflict of interest: None of the authors have any potential conflicts of interest.

Conclusion: The novel DBR scale, with its objective linear chromatic steps, demonstrated improved reproducibility, fewer visualization artifacts and improved ease of use over the VBR scale for assessing conjunctival redness.

Keywords

Conjunctival redness; Digital image-based scale

Introduction

Bulbar redness is a common clinical sign in many ocular conditions such as allergic and infectious conjunctivitis, dry eye disease, ocular rosacea, contact lens wear, reactions to topical medications.^{1–5} Accurate and consistent evaluation of bulbar redness is important not only in clinical settings to assess progression or improvement over time, but also in clinical trials of topical ocular agents that may induce or alleviate redness.

Historically, the estimation of bulbar redness has been achieved using simple, arbitrary and subjective ordinal scales that are easy to use but difficult to replicate and have several limitations.^{6–9} In many clinical trials, standard sets of printed images of the bulbar conjunctiva with different degrees of redness have been used to assess bulbar redness. Frequently used photographic scales for estimating bulbar redness include the McMonnies and Chapman-Davies scale (M-CD scale),¹⁰ Institute for Eye Research (IER) scale (previously known as CCLRU scale),¹¹ the Efron scale,^{12,13} and the Validated Bulbar Redness (VBR) grading scale.¹⁴ There are many differences among these scales including the number of reference images, the range of redness, the linearity of the scores as a measure of redness, and the conjunctival region displayed in the reference images. For example, the Efron scale was developed from artist rendered paintings while the CCLRU scale was constructed from a montage of photographs;¹⁵ additionally the VBR scale had 10 reference images, the Efron scale had 5, the M-CD scale had 6 and the IER scale had only 4.^{10–16} The reproducibility of grading using these scales has been inconsistent,^{7–9} and cross-calibrating these clinical grading scales to allow comparison of grading results among the scales has also been difficult.^{16,17} These printed scales were intended primarily to be used during the clinical examination of patients and were not designed for comparison with digital images.

Photographic imaging of the eye allows for the accurate and reproducible assessment of morphological changes and has been the standard for both patient management and clinical research in several ocular diseases including diabetic retinopathy, macular degeneration, glaucoma, and retinopathy of prematurity.^{18,19}

Evaluation of the images at reading centers reduces the variability in grading seen across clinicians. The availability and increased use of digital photography of the bulbar conjunctiva allows for the advantages of remote assessment at a centralized reading facility, and yet it also raises some challenges when trying to apply printed scales to digital images viewed on a computer monitor or other electronic display screens. As a result, a different scale is needed that is designed for use when grading digital images displayed on electronic screens.

In this study, we describe the development of a new digital image based scale for the assessment of bulbar redness on digital images, and compare the accuracy and ease of use of this new scale relative to the VBR, a well-established clinical grading scale.

MATERIALS AND METHODS:

Digital Images of the Bulbar Conjunctiva

A database of 4889 nasal and temporal digital images of eyes with varying degrees of bulbar redness was available at the repository of the Scheie Image Reading Center at the University of Pennsylvania. These de-identified images were obtained from participants in a clinical trial of a topical agent that induced bulbar redness. The study was approved by the Institutional Review Board (IRB) at the University of Pennsylvania.

Images were acquired by project-certified photographers according to a previously published photography protocol that was modified to image the bulbar conjunctiva.²⁰ Briefly, the images were acquired using a Canon EOS Rebel EF-S T2i camera fitted with a Canon EF 100mm f/2.8L IS Macro Lens. A white index card was positioned a few millimeters away from the lateral canthal area so that it appeared in the periphery of each image. The subject was positioned in a chair and with the head stabilized by a disposable styrofoam cup placed between the head and the wall, three photographs for each eye were taken beginning with the right eye; an overall view of the eye with the subject looking straight ahead; a temporal view of the conjunctiva with subject looking as far as possible in the nasal direction with the eyes open wide; and a nasal view of the conjunctiva with the subject looking as far as possible in the temporal direction with the eyes wide open. The photographic field of the two bulbar conjunctiva extended from the outer (temporal conjunctiva) or inner (nasal conjunctiva) canthus to at least the midpoint of the cornea. Each session also included a photograph of a mini (2.25 × 3.25 in) Gretag Macbeth Color Checker Chart. Images of the color checker and white card were used to normalize the color and luminance of the images from different imaging sessions and/or clinical centers. Images were uploaded through a secure web-based system and processed for color and luminance normalization.

Development of the Digital Bulbar Redness (DBR) scale

Reference images for the DBR scale were selected from the digital image database of nasal and temporal images of the conjunctiva, normalized for color and luminance. Conjunctival images (separately for nasal and temporal images) with the lowest and highest levels of bulbar redness were selected as anchors for the scale and were assigned scores of 10 (least redness) and 100 (most redness) (Figure 1). Next, images with eight different intermediate levels of progressively increasing bulbar redness were identified and included in the scale in 10 point increments from 20–90.

This draft set of reference images was then evaluated in two ways. First, the 10 images for the nasal and temporal conjunctival views were ordered from least to most redness by two masked trained graders from the Scheie Image Reading Center. Second, the linearity of the scale was assessed using an established image processing algorithm presented by Fieguth et al. to calculate the average measurement of redness per unit area of the bulbar conjunctiva.⁶

The algorithm performs this measurement by first evaluating the difference at each pixel between the value of the red color channel and the sum of the values of the green and blue color channels of the image. It then divides this difference by the sum of all three color channels at that pixel. Then, it averages this ratio across all pixels in the bulbar conjunctiva to produce the final objective redness measurement for each reference image.

Based on these evaluations, it was determined that the objective redness of one of the original reference images was too close to that of an adjacent image on the scale, based on the algorithm output. Replacing that image with a different image improved the linearity of the resultant subjective scale in relation to the algorithm output (Figure 2). The images were cropped so that only the bulbar conjunctival region was visible to allow focusing the attention of the grader on only the area between the canthus and limbus. Reference images were selected from both eyes. The images were then digitally oriented so that the scale representing each conjunctival region had the identical images but reversed between left and right eyes. The digital images were assembled into a collage to form the final Digital Bulbar Redness (DBR) grading scale (Figure 3).

Assessment of Reproducibility in Grading

A random sample of 100 unique images, (temporal or nasal views, not necessarily from the same eye or subject) was selected from the dataset of 4889 images. These images were different from those included as reference images in the DBR scale. To assess bulbar redness in these images, graders used calibrated dual monitors. The study images were displayed in Photoshop (Adobe Systems, Inc.) on one of the calibrated monitors with the magnification of the image adjusted to 25% in Photoshop. The scale reference images of the DBR scale were opened with the Windows Photo viewer on an adjacent monitor and enlarged so that each reference image was matched to the size of the study image to be graded. Lighting in the grading room was dimmed to minimize incident light on the monitors

Each of three trained graders examined both the study and the reference images and selected the best match from the reference images and recorded the score (10–100). Each of the graders independently evaluated the 100 images. The integers (10, 20, 30, etc.) corresponding to the closest image was recorded; no interpolation between reference images was performed. The trained graders used the VBR scale to evaluate the same 100 images, recording a score for each image from 10 to 100. The integers (10, 20, 30, etc.) corresponding to the closest image was recorded; no interpolation between reference images was performed. The 10-level VBR scale was selected for comparison because it spans the widest range of redness with the most reference images among the 4 commonly used scales; and the scores also lie along a linear scale of an objective redness measure.^{21,22}

Statistical Analysis

Agreement among graders was assessed with weighted Kappa statistics using Cicchetti-Allison weights. McNemar's test for paired proportions was used to assess the difference between the DBR scale and the VBR scale in the percentage with exact agreement among the three graders.

RESULTS:

The values obtained using the algorithm to calculate a global measurement of redness per unit area of image of the bulbar conjunctiva for the 10 temporal and 10 nasal conjunctival reference images for the final DBR scale are displayed in Figure 2. For both nasal and temporal reference images, the magnitude of change in the objective measure of redness between adjacent images was nearly equal, indicating a linear relationship between the DBR scale scores and the objective redness measure.

The scores of the 100 test images demonstrated moderate to substantial agreement among the pairs of graders for both scales, with weighted kappa values of 0.57 to 0.73 for the DBR scale and 0.38 to 0.66 for the VBR scale (Table).²³ The 3 trained readers agreed exactly on more of the images using the DBR scale (38%) than using the VBR scale (7%; $p < 0.05$).

DISCUSSION

We developed a digital image-based scale for evaluating digital images of bulbar redness. While there are several printed subjective scales currently available for clinical grading of bulbar redness, they have several disadvantages for the assessment of digital images. For example, the low number of reference images for the IER scale (4 images) and the Efron scale (5 images) provides a coarse categorization that reduces sensitivity to detect change.²⁴ The McMonnies scale has 6 reference images but they span a narrower range of redness than the DBR scale.^{21,22} In addition, the increase in redness between adjacent reference images for these three scales is not constant so that the scores are not on a linear scale. Linearity between scores permits the use of the scores as interval scale measurements and summary statistics such as mean and standard deviation. Precision of the scoring and the ability to detect change can be increased by assigning score values between two standard image scores when the redness fall between them; e.g., assigning a score of 43 to an image that has redness that exceeds that in reference image 40 but is closer to reference image 40 than to reference image 50.²⁴

The VBR scale has 10 reference images that span a range of redness at least as wide as the other 3 scales, allows for a finer categorization, and is linear. However, the VBR scale also has certain limitations. The reference images for the highest degrees of redness are the same image modified using Adobe Photoshop to artificially create more severe grades of bulbar redness; this approach may not accurately reflect the in vivo appearance. In addition, the VBR scale includes only temporal conjunctival views, complicating the comparison with the nasal bulbar view.

The DBR scale was designed to minimize the disadvantages present with other subjective redness scales and optimize use with digital images. The DBR scale includes 10 reference images per view that span a wide range of redness with scores on a linear scale of redness. Thus, these measurement properties have the same advantages as the VBR scale. However, the DBR scale has additional favorable properties that facilitate its use for assessing digital images. All of the DBR reference images are of patients without any artificially generated redness as done to create the more intensely red images in the VBR scale. Both temporal and

nasal views are available and all extraneous parts of the images (e.g. cornea, eyelids) have been cropped. Importantly, the reference images and images to be assessed can be scaled on a monitor to the same size. These properties simplify the process of grading by eliminating the need for the grader to mentally transform the image under evaluation to the pattern of redness, the orientation (nasal to temporal), and, the size to the reference image. Also, assessments of digital images are best performed either in the dark or with as little incident light as possible so that the colors displayed on calibrated monitors are not affected by incident light from room lighting or from outdoor light. Unlike the DBR scale that is designed for use in a darkened room, viewing the photographic image reference scales such as VBR and IER require ambient lighting. These favorable features of the DBR scale that facilitate the grading task are likely responsible for the slightly better reproducibility of the DBR scale gradings compared to the VBR scale gradings.

However, there are some limitations to the DBR scale. The scale was derived from images accumulated as part of a clinical trial of a topical medication that caused bulbar redness as a side effect. The range of bulbar redness due to causes such as contact lens use, dry eye disease, and allergic conjunctivitis may be greater. However, the range of the DBR scale is very similar to the range of the VBR scale and exceeds the range of the Efron, IER and M-CD scales, which have been applied to studies of contact lens wear.^{15, 25–27} Finally, because the DBR scale is optimized for assessment of digital images, it may be more cumbersome for routine clinical use in evaluating bulbar redness when compared to an easily portable 4–5 image photo-card scale.

In recent years, there has been increased interest in the use of non-invasive imaging techniques for the evaluation of the ocular surface²⁸. For example, there are a number of objective algorithms have been developed for evaluating bulbar redness that have the advantage of providing highly reproducible results.^{29–32} Many of these algorithms require manual or semi-manual segmentation of the conjunctiva prior to image analysis. There is currently a commercially available instrument, the Oculus Keratograph 5M (Oculus, Arlington, WA) that allows for the quantification of bulbar redness and has been used to study redness in different subgroups of patients.^{33,34} However, as these new instruments are developed, their results must still be validated using subjective gradings of images by clinicians to use as the “gold standard” and therefore there continues to be a need for better image-based scales.³⁵ In addition, subjective scales are still useful in that clinically acquired images can differ in quality and human graders may be able to better compensate for the variability in photographic field and focus than a fully automated computer algorithm.

In summary, we have developed a digital image-based scale to assess bulbar redness. This scale has several advantages over the presently available subjective scales in evaluating bulbar redness on digital images. The DBR scale is especially well suited for evaluation of bulbar redness in clinical trials.

Acknowledgments

Source of Funding: Vatinnee Y. Bunya-National Eye Institute R01 EY026972; Yuanzie Zheng - Natural Science Foundation of China (NSFC) (61572300); Natural Science Foundation of Shandong Province in China (ZR2014FM001); Taishan Scholar Program of Shandong Province in China (TSHW201502038); Richard A. Stone

- Paul and Evanina Bell Mackall Foundation Trust; Research to Prevent Blindness ; Vision Research Center Core grant (NEI P30 EY001583).

References:

1. Lanier BQ, Tremblay N, Smith JP, deFaller JM. A double-masked comparison of ocular decongestants as therapy for allergic conjunctivitis. *Ann Allergy*.1983;50:174–7. [PubMed: 6338768]
2. Deinema LA, Vingrys AJ, Wong CY, et al. A Randomized, Double-Masked, Placebo-Controlled Clinical Trial of Two Forms of Omega-3 Supplements for Treating Dry Eye Disease. *Ophthalmology*. 2017;124:43–52. [PubMed: 27817918]
3. Stapleton F, Ramachandran L, Sweeney DF, Rao G, Holden BA. Altered conjunctival response after contact lens-related corneal inflammation. *Cornea*.2003;22:443–7. [PubMed: 12827050]
4. Trzeciecka A, Paterno JJ, Toropainen E, et al. Long-term topical application of preservative-free prostaglandin analogues evokes macrophage infiltration in the ocular adnexa. *Eur J Pharmacol*. 2016;788:12–20. [PubMed: 27288881]
5. Quarterman MJ, Johnson DW, Abele DC, Leshner JL Jr, Hull DS, Davis LS. Ocular rosacea. Signs, symptoms, and tear studies before and after treatment with doxycycline. *Arch Dermatol*. 1997;133:49–54. [PubMed: 9006372]
6. Fieguth P, Simpson T. Automated measurement of bulbar redness. *Invest Ophthalmol Vis Sci*. 2002;43:340–7. [PubMed: 11818375]
7. Chong T, Simpson T, Fonn D. The repeatability of discrete and continuous anterior segment grading scales. *Optom Vis Sci*. 2000;77:244–251. [PubMed: 10831214]
8. Papas EB. Key factors in the subjective and objective assessment of conjunctival erythema. *Invest Ophthalmol Vis Sci*. 2000;41:687–691. [PubMed: 10711682]
9. Peterson RC Wolffsohn JS. Sensitivity and reliability of objective image analysis compared to subjective grading of bulbar hyperaemia. *Br J Ophthalmol*. 2007;91:1464–66. [PubMed: 17475716]
10. McMonnies C, Chapman-Davies A. Assessment of conjunctival hyperemia in contact lens wearers. *Am J Optom Vis Sci*. 1987;64:246–250.
11. IER. IER Grading Scales. Institute for Eye Research: Sydney, Australia; 2007.
12. Efron N Clinical application of grading scales for contact lens complications. *Optician*. 1997;213:26–35.
13. Efron N. Grading scales. *Optician*. 2000;219:44–45.
14. Schulze M Jones D Simpson T The development of validated bulbar redness grading scales. *Optom Vis Sci*. 2007;84:976–983. [PubMed: 18049363]
15. Efron N, Morgan PB, Katsara SS. Validation of grading scales for contact lens complications. *Ophthalmic Physiol Opt*. 2001;21:17–29. [PubMed: 11220037]
16. Schulze MM, Hutchings N, Simpson TL. Grading bulbar redness using cross-calibrated clinical grading scales. *Invest Ophthalmol Vis Sci*. 2011;52:5812–7. [PubMed: 21474775]
17. Schulze MM, Hutchings N, Simpson TL. The use of fractal analysis and photometry to estimate the accuracy of bulbar redness grading scales. *Invest Ophthalmol Vis Sci*. 2008;49:1398–406. [PubMed: 18385056]
18. Daniel E, Quinn GE, Hildebrand PL, et al. Validated System for Centralized Grading of Retinopathy of Prematurity: Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity (e-ROP) Study. *JAMA Ophthalmol*. 2015;133:675–82. [PubMed: 25811772]
19. Grunwald JE, Daniel E, Ying GS, et al. Photographic assessment of baseline fundus morphologic features in the Comparison of Age-Related Macular Degeneration Treatments Trials. *Ophthalmology*. 2012;119:1634–41. [PubMed: 22512984]
20. Bunya VY, Brainard DH, Daniel E, et al. Assessment of signs of anterior blepharitis using standardized color photographs. *Cornea* 2013; 32:1475–1482. [PubMed: 24055901]
21. Schulze MM, Hutchings N, Simpson TL. The perceived bulbar redness of clinical grading scales. *Optom Vis Sci*. 2009;86:E1250–8. [PubMed: 19770812]

22. Baudouin C, Barton K, Cucherat M, Traverso C. The measurement of bulbar hyperemia: challenges and pitfalls. *Eur J Ophthalmol*. 2015;25:273–9. [PubMed: 26069220]
23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–74. [PubMed: 843571]
24. Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effects of scaling. *Invest Ophthalmol Vis Sci*. 1991;32:422–32. [PubMed: 1993595]
25. McMonnies C, Chapman-Davies A. Assessment of conjunctival hyperemia in contact lens wearers. *Am J Optom Vis Sci*. 1987;64: 246–250.
26. Efron N, Morgan PB, Jagpal R. Validation of computer morphs for grading contact lens complications. *Ophthalmic Physiol Opt* 2002;22:341–9. [PubMed: 12162486]
27. Efron N, McCubbin S. Grading contact lens complications under time constraints. *Optom Vis Sci*. 2007;84:1082–6. [PubMed: 18091306]
28. Qazi Y, Aggarwal S, Hamrah P. Image-guided evaluation and monitoring of treatment response in patients with dry eye disease. *Graefes Arch Clin Exp Ophthalmol*. 2014;252:857–872. [PubMed: 24696045]
29. Peterson RC Wolffsohn JS. Sensitivity and reliability of objective image analysis compared to subjective grading of bulbar hyperaemia. *Br J Ophthalmol*. 2007;91:1464–1466. [PubMed: 17475716]
30. Amparo F, Wang H, Emami-Naeini P, Karimian P, Dana R. The Ocular Redness Index: a novel automated method for measuring ocular injection. *Invest Ophthalmol Vis Sci*. 2013;54:4821–6. [PubMed: 23766472]
31. Sorbara L, Simpson T, Duench S, Schulze M, Fonn D. Comparison of an objective method of measuring bulbar redness to the use of traditional grading scales. *Cont Lens Anterior Eye*. 2007;30:53–9. [PubMed: 17254834]
32. Park IK, Chun YS, Kim KG, Yang HK, Hwang JM. New clinical grading scales and objective measurement for conjunctival injection. *Invest Ophthalmol Vis Sci*. 2013;54:5249–57. [PubMed: 23833063]
33. Pérez Bartolomé F, Martínez de la Casa JM, Arriola Villalobos P, Fernández Pérez C, Polo V, Sánchez Jean R, García Feijoó J Ocular Redness Measured with the Keratograph 5M in Patients Using Anti-Glaucoma Eye Drops. *Semin Ophthalmol*. 2017;16:1–8.
34. Wu S, Hong J, Tian L, Cui X, Sun X, Xu J. Assessment of Bulbar Redness with a Newly Developed Keratograph. *Optom Vis Sci*. 2015;92:892–9 [PubMed: 26099055]
35. Downie LE, Keller PR, Vingrys AJ. Assessing ocular bulbar redness: a comparison of methods. *Ophthalmic Physiol Opt*. 2016;36:132–9 [PubMed: 26890702]

Figure 1:
Representative bulbar conjunctival images used to create Digital Bulbar Redness (DBR) scale. Color images of the temporal bulbar conjunctiva showing the least amount of redness (A) and the most amount of redness (B) in the DBR scale.

Figure 2:
Graph showing the linearity of the Digital Bulbar Redness Scale with the algorithm redness score.

Figure 3:
The Digital Bulbar (DBR) scale showing the redness scale in all four areas of the bulbar conjunctiva

Table:

Agreement on the Digital Bulbar Redness (DBR) scale and the Validated Bulbar Redness (VBR) scale scores between three Trained Readers.

Scale	Exact % agreement (195% CI)	Weighted Kappa* (95% CI)
Digital Bulbar Redness Scale		
Grader 1 vs 2	68% (58%, 77%)	0.73 (0.65, 0.82)
Grader 2 vs 3	52% (42%, 62%)	0.60 (0.48, 0.71)
Grader 1 vs 3	48% (38%, 58%)	0.57 (0.46, 0.68)
Validated Bulbar Redness Scale		
Grader 1 vs 2	18% (11%, 27%)	0.38 (0.29, 0.47)
Grader 2 vs 3	55% (45%, 65%)	0.66 (0.57, 0.75)
Grader 1 vs 3	25% (17%, 35%)	0.46 (0.36, 0.56)

* Cicchetti-Allison weights