A Standardized Grading System for Scleritis

H. Nida Sen, MD, MHSc,[1] Amit A. Sangave, BS,[1] Debra A. Goldstein, MD,[2] Eric B. Suhler, MD, MPH,[3,4] Denise Cunningham, FOPS, MS,[1] Susan Vitale, PhD, MHS,[1] Robert B. Nussenblatt, MD, MPH[1]

Objective: This study evaluated the performance of a standardized grading system for scleritis using standard digital photographs.

Design: Cross-sectional interobserver agreement study.

Participants: Photo archives from the National Eye Institute.

Methods: Three uveitis specialists from 3 different centers graded 79 randomly arranged images of the sclera with various degrees of inflammation. Grading was done using standard screen resolution (1024×768 pixels) on a 0 to 4+ scale in 2 sessions: (1) without using reference photographs and (2) with reference to a set of standard photographs (proposed grading system). The graders were masked to the order of images, and the order of images was randomized. Interobserver agreement in grading the severity of inflammation with and without the use of grading system was evaluated.

Main Outcome Measures: Interobserver agreement.

Results: The proposed grading system for assessing activity in scleritis demonstrated a good interobserver agreement. Interobserver agreement (pooled κ) was poor (0.289) without photographic guidance and improved substantially when the "grading system" with standardized photographs was used ($\kappa = 0.603$).

Conclusions: This system of standardized images for scleritis grading provides significantly more consistent grading of scleral inflammation in this study and has clear applications in clinical settings and clinical research.

Financial Disclosure(s): Proprietary or commercial disclosure may be found after the references. *Ophthalmology 2011;118:768–771 © 2011 by the American Academy of Ophthalmology.*

Assessment and standardization of the degree of activity in ocular inflammation are important for both patient care and clinical research. A quantitative, standardized grading of the severity of inflammation could serve as an inclusion criterion or outcome measure. The comparison of clinical research data requires reproducibility of such measures. It also requires that the methodology be practical for the clinical setting. Scleritis is a chronic, painful, and destructive inflammatory disease of the sclera frequently associated with an infection or underlying systemic disease and is one of the most challenging conditions to manage in ophthalmology.[1] The Standardization of Uveitis Nomenclature (SUN) Working Group published standards for grading the location and degree of activity of intraocular inflammation, including endorsement of standardized photographs for grading vitreous haze;[2] however, no such system has been established for scleritis. To assess the performance of a new reference photograph-based scleritis grading system, we studied the level of agreement among uveitis specialists from different centers in grading the severity of scleral inflammation with and without the aid of photographic guidance.

Materials and Methods

All photographs used were high-resolution (2544×1696 pixels) color images of sclera captured with the Canon EOS 20D digital camera (Tokyo, Japan) mounted on a Haag-Streit BX900 slit lamp (Bern, Switzerland) after a 10% phenylephrine instillation according to a standardized protocol at the National Eye Institute. These were then saved in a secure digital database (OIS WinStation 4000SL, Sacramento, CA) and used in accordance with Declaration of Helsinki guidelines. Seventy-nine digital photographs of the sclera with various degrees of inflammation were selected from this digital photo-archive and downloaded in JPEG file format (1124×742 pixels). Three uveitis specialists from 3 different centers graded the photos on a 0 to 4+ scale, using a standard screen resolution (1024×768 pixels). The grading was done twice: (1) without using photographic references ("session 1") and (2) guided by standard photographic references ("session 2"). The same 79 photographs were used for each session; however, the graders were masked to the order of images and the order of images was randomly assorted in both parts of the study.

Each grader was instructed on how to grade the photos in each session. For session 1, the graders were asked to grade each digital image on a 0 to 4+ (with a 0.5+ grade between 0 and 1+) scale, where 0 represented no scleral inflammation and 4+ represented the most severe form of scleral inflammation, necrotizing scleritis. For session 2, the graders were asked to compare each digital photograph with a similar one on the "grading system poster" provided to them at the end of the first session (Fig 1). This poster was developed by the investigators, who were not involved in grading, by choosing photos from the National Eye Institute's digital photo archive. Scleral inflammation in this poster was similarly graded with an ordinal scale of 0 (no scleral inflammation with complete blanching of vessels), 0.5+ (trace inflammation with minimally dilated deep episcleral vessels), 1+ (mild scleral inflammation with diffuse mild dilation of deep episcleral vessels), 2+ (moderate scleral inflammation with tortuous and engorged deep episcleral vessels), 3+ (severe scleral inflammation with diffuse significant redness of sclera ± obscuration of deep epis-

Published by Elsevier Inc.

ISSN 0161-6420/11/$–see front matter
doi:10.1016/j.ophtha.2010.08.027

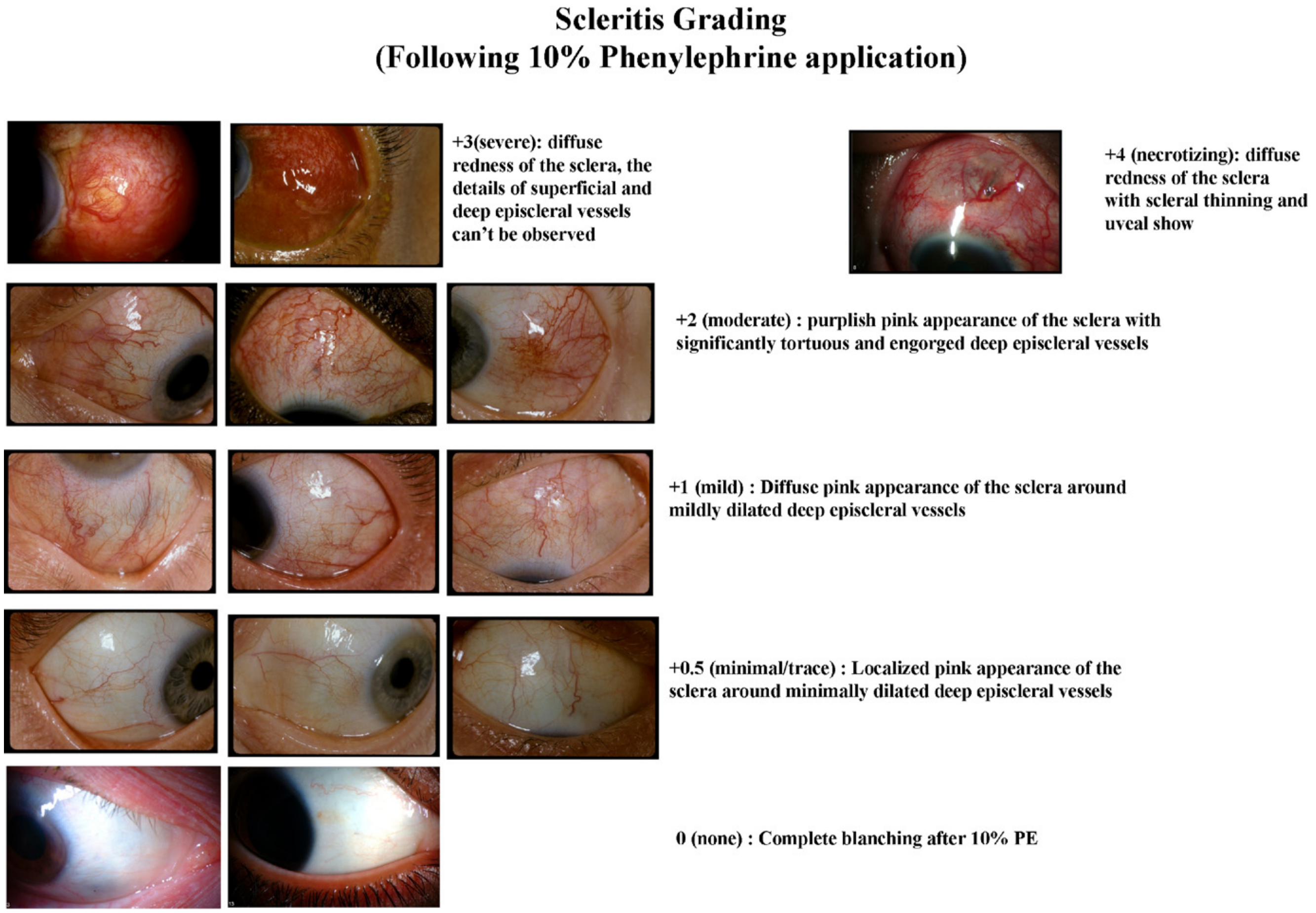
Scleritis Grading
(Following 10% Phenylephrine application)
+3(severe): diffuse redness of the sclera, the details of superficial and deep episcleral vessels can't be observed
+4 (necrotizing): diffuse redness of the sclera with scleral thinning and uveal show
+2 (moderate) : purplish pink appearance of the sclera with significantly tortuous and engorged deep episcleral vessels
+1 (mild) : Diffuse pink appearance of the sclera around mildly dilated deep episcleral vessels
+0.5 (minimal/trace) : Localized pink appearance of the sclera around minimally dilated deep episcleral vessels
0 (none) : Complete blanching after 10% PE

Figure 1. Scleritis grading system. Standardized digital photographs of scleritis of varying severity. Graders used this illustration as reference photographs in session 2. PE = phenylephrine.

cleral vessels with edema and erythema), and 4+ (necrotizing scleritis with or without uveal show) (Fig 1).

Interobserver agreement was assessed between each pair of graders (graders 1 and 2, graders 2 and 3, and graders 1 and 3), separately for session 1 and session 2. We also calculated the overall agreement among the 3 graders, separately for session 1 and session 2, by using pooled Kappa (κ). The kappa (κ) statistic is an index that compares the agreement against that which might be expected by chance. We computed kappa values using SAS 9.0 software (SAS Inc., Cary, NC). Possible values for Kappa can range from +1 (perfect agreement) via 0 (no agreement above that expected by chance) to −1 (complete disagreement). The convention is that values of kappa from 0.41 to 0.6 represent moderate agreement and values of 0.61 to 0.8 represent good or substantial agreement.[3] In addition, to compare the 3-way agreement between session 1 and session 2, the number of agreeing grades (among the 3 graders) was evaluated.

Results

Results of the grading are shown in Table 1 (available at: http://aaojournal.org). The overall distributions of grader 1's grades were similar between sessions 1 and 2. Grader 2 was less likely to assign grade 0 and more likely to assign grade 1 in session 2 than in session 1. Grader 3 did not use the 0.5 grade at all in session 1 but did use it in session 2.

Because grader 3 did not use the 0.5 category at all in session 1, and kappa requires a "square" table (i.e., the same number of categories must be used for the 2 entities whose agreement is being assessed), we combined the grade of 0.5 with the grade of 0 to compute the kappa statistics. The agreement of each pair of graders was assessed separately for sessions 1 and 2 (Table 2). Graders 1 and 2 had the closest agreement in session 1 (possibly because of grader 3's non-use of the 0.5 category). Pairwise kappas for session 1 ranged from 0.09 to 0.44. The pooled kappa for session

Table 2. Interobserver Agreement

	Grader 1 vs. 2	Grader 1 vs. 3	Grader 2 vs. 3	Pooled Kappa (Graders 1, 2, and 3)
Unweighted (only exact agreement is considered)				
Session 1	0.442	0.348	0.086	0.2896
Session 2	0.587	0.672	0.576	0.6034
Weighted (agreement within ±1 category is considered)				
Session 1	0.638	0.554	0.342	NA
Session 2	0.726	0.797	0.708	NA

NA = not applicable (represents where a pooled kappa is not applicable). Agreement among graders for sessions 1 and 2 when categories 0 and 0.5 are combined (kappa statistic) is shown. Note the improvement in weighted kappa between sessions 1 and 2.

1 was 0.29. Agreement was higher in session 2 than in session 1 for all pairwise comparisons, with kappas ranging from 0.58 to 0.67. The pooled kappa for session 2 was 0.60, which is a substantial improvement.

Because the values of kappa for sessions 1 and 2 are not independent, we were not able to perform statistical testing to see whether agreement differed significantly between sessions. We therefore examined agreement in the following way: For each of the 79 photographs, there are 3 session 1 grades (from graders 1, 2, and 3) and 3 session 2 grades (from graders 1, 2, and 3). For each photograph in session 1, we recorded the number of graders who agreed exactly among the 3 graders (possible values were 0, 2, or 3). We did the same for each photograph in session 2 and cross-tabulated the values (Fig 2A, available at http://aaojournal.org). Of the 11 photographs for which there was no agreement in session 1, 7 had agreement of all 3 graders and 3 had agreement of 2 graders in session 2. Of the 50 photographs for which 2 graders agreed in session 1, 29 had agreement of all 3 graders in session 2 and none had 0 agreeing graders. To perform statistical testing to see if the level of agreement tended to be higher in session 2 than in session 1, we collapsed the table in Figure 2A to create a 2×2 table and applied McNemar's test (which uses the discordant data to test the hypothesis that changes between sessions 1 and 2 occurred randomly, i.e., half the discordant cases should fall in the no agreement–agreement category and half should fall into the agreement–no agreement category). If categories 2 and 3 were grouped (i.e., any agreement vs. no agreement) (Fig 2B, available at http://aaojournal.org), the probability that changes in the discordant cases occurred purely due to chance was 0.001, that is, agreement was significantly more likely in session 2 than in session 1. If categories 0 and 2 were grouped (i.e., total agreement vs. less-than-total agreement) (Fig 2C, available at http://aaojournal.org), the probability that changes in the discordant cases occurred purely due to chance was $<$0.001, that is, agreement was significantly more likely in session 2 than in session 1.

Discussion

Scleritis has been classified on the basis of anatomic location (anterior, posterior) and nature of involvement (diffuse, nodular, necrotizing).[4] Although this classification is helpful in the clinical setting and has implications for prognosis, within each category the severity varies. The SUN working group addressed standardization of grading of intraocular inflammation, including anterior chamber cells, flare, and vitreous haze, with the latter defined by means of a photographic standard.[2] However, there is currently no standardized grading system for assessing severity or activity in scleritis. McCluskey and Wakefield[5] proposed a system for scoring the extent and severity of scleritis, based on common clinical signs. Their scoring system takes into account clinical features, such as the area of inflammation, pain, corneal involvement, and associated anterior chamber inflammation. Although the system grades pain and the area of inflammation, it does not specifically address the degree or severity of inflammation. The detailed nature of this system makes it reflective of the disease course rather than the inflammatory activity at one visit. In addition, the system requires the grading of 8 components, each of which can be assigned subjective values from 0 to 2 or 0 to 4, which is likely to require more time from the grader.

In the absence of a standardized grading system, a clinician typically grades the overall scleritis activity on a none, mild, moderate, severe scale (or from 0 to 4), as is done with most other grading systems in ocular inflammatory disorders. Determination of the degree of inflammation is a critical step for the management of patients with scleritis and allows for accurate assessment of response to therapy. Accurate, consistent assessment of the degree of inflammation is particularly important when patients are seen by different physicians at different visits, in comparing treatment effects between trials, and in interpreting published results from different groups. To the best of our knowledge, this is the first scleritis grading system using photographic standards and focusing on the severity of scleral inflammation that has been formally evaluated for its utility. A large number of standardized photos were graded by 3 graders from 3 different centers with varying years of experience in the field of ocular inflammatory diseases. Our proposed grading system for scleritis is simple and easy to implement with a 0 to 4+ grading scheme similar to that used for anterior chamber cell grading. A similar standard photographic grading system has been published by Nussenblatt et al[6] for vitreous haze. Agreement using this system was tested in a pilot study with favorable results, and since then it has been widely used and is now accepted by the SUN working group as the standard for grading vitreous haze. In addition, agreement using this scale has since been reevaluated and was found to be moderate to substantial.[7]

Most studies grade scleritis subjectively as active or inactive.[8] Although this may be satisfactory for patient care in some settings, a more consistent and less subjective method is desirable. The grading system proposed herein improved agreement between observers; the improvement in agreement was more notable when there was a greater difference between years of experience of graders. A scleritis grading system similar to the current one was used by our group in a pilot trial[9] and found to be helpful in adjudicating the degree of inflammation by different examiners.

As with every grading system, ours has limitations. The grading system was tested for agreement on the basis of clinical photographs and not actual patients. It could be argued that comparing photos with photos may have inflated the level of agreement. However, this testing method may have avoided bias introduced by physicians' knowledge of a particular patient's disease course if the testing of the system were to be done in a clinical setting. In addition, this may serve as an advantage for clinical trials, allowing standard photographs to be graded independently by a reading center, eliminating individual investigator bias. These scleral photographs were taken 15 to 20 minutes after instillation of 10% phenylephrine, and each photograph reflected only 1 quadrant of the eye. In the future, further development of this grading system may include use of a composite score for each eye (based on, e.g., grades 0 to 4 being assigned to each quadrant) and for each patient. In addition, the current grading system requires further validation studies using a clinical outcome.

The graders in this study were practicing, uveitis-trained ophthalmologists, which may have overestimated the interobserver agreement, but these results are probably applicable to uveitis specialists who participate in clinical trials, particularly given the graders were from different centers with different years of experience.

In conclusion, we believe that this grading system, although far from perfect, provides ophthalmologists with a reproducible, quantitative system and allows for more reliable and consistent assessment of scleral inflammation regardless of the number of years of experience. The system is simple and practical and can easily be implemented in the clinical setting and in clinical trials.

References

1. Watson PG. Doyne Memorial Lecture, 1982. The nature and the treatment of scleral inflammation. Trans Ophthalmol Soc U K 1982;102:257–81.
2. Standardization of Uveitis Nomenclature (SUN) Working Group. Standardization of uveitis nomenclature for reporting clinical data: results of the First International Workshop. Am J Ophthalmol 2005;140:509–16.
3. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
4. McCluskey PJ, Wakefield D. Scleritis and episcleritis. In: Pepose JS, Holland GN, Wilhelmus KR, eds. Ocular Infection and Immunity. St. Louis, MO: Mosby; 1996:642–62.
5. McCluskey PJ, Wakefield D. Prediction of response to treatment in patients with scleritis using a standardized scoring system. Aust N Z J Ophthalmology 1991;19:211–5.
6. Nussenblatt RB, Palestine AG, Chan C, Roberge F. Standardization of vitreal inflammatory activity in intermediate and posterior uveitis. Ophthalmology 1985;92:467–71.
7. Kempen JH, Ganesh SK, Sangwan VS, Rathinam SR. Interobserver agreement in grading activity and site of inflammation in eyes of patients with uveitis. Am J Ophthalmol 2008; 146:813–8.
8. Taylor SR, Salama AD, Joshi L, et al. Rituximab is effective in the treatment of refractory ophthalmic Wegener's granulomatosis. Arthritis Rheum 2009;60:1540–7.
9. Sen HN, Sangave A, Hammel K, et al. Infliximab for the treatment of active scleritis [report online]. Can J Ophthalmol 2009;44:e9-12. Available at: http://article.pubs.nrc-cnrc.gc.ca/RPAS/rpv?hm=HInit&calyLang=eng&journal=cjo&volume=44&afpf=i09-061.pdf. Accessed July 30, 2010.

Footnotes and Financial Disclosures

Originally received: April 18, 2010.
Final revision: August 11, 2010.
Accepted: August 12, 2010.
Available online: November 20, 2010. Manuscript no. 2010-564.

[1] National Eye Institute, National Institutes of Health, Bethesda, Maryland.

[2] University of Illinois at Chicago, Chicago, Illinois.

[3] Oregon Health Sciences University, Portland, Oregon.

[4] Portland VA Medical Center, Portland, Oregon.

Financial Disclosure(s):
Dr Eric Suhler receives financial support from Celgene, Abbott, Genentech, Novartis, and the Department of Veterans' Affairs, and institutional support from Research to Prevent Blindness.

Supported by the National Eye Institute Intramural research program support (HNS, RBN, AS, SV).

Correspondence:
H. Nida Sen, MD, MHSc, National Eye Institute, 10 Center Drive, Building: 10 Room:10N112, Bethesda, MD 20892. E-mail: senh@nei.nih.gov.