Published in final edited form as:

Am J Ophthalmol. 2010 November ; 150(5): 637–641.e1. doi:10.1016/j.ajo.2010.05.036.

Scale for Photographic Grading of Vitreous Haze in Uveitis

Janet L. Davis, Brian Madow, Jessica Cornett, Rick Stratton, Ditte Hess, Vittorio Porciatti, and William J. Feuer

University of Miami Miller School of Medicine, Bascom Palmer Eye Institute, Miami, Florida (J.L.D., B.M., J.C., R.S., D.H., V.P., W.J.F.).

Abstract

PURPOSE—To validate a scale for grading vitreous haze in uveitis using digitized photographs and standardized scoring.

DESIGN—Evaluation of clinical research methodology.

METHODS—Calibrated Bangerter diffusion filters inducing incremental decrements of spatial contrast were placed in front of the camera lens while photographing a normal eye to simulate vitreous haze. The photographs were digitized and an ordinal scale was created from 0 (none) to 8 (highest level of opacification at which fundus details could be seen). The scale steps correspond approximately to decimal Snellen visual acuities of 1.0, 0.8, 0.4, 0.2, 0.1, 0.04, 0.02, 0.01, and 0.002, with approximately 0.3 log step between each step. For validation, digitized fundus photographs of uveitis patients were displayed on a computer monitor for comparison with the standard photos. Three observers graded the test set twice under standard conditions. Interobserver and intraobserver variability and κ values for agreement greater than chance were calculated.

RESULTS—Variance component analysis determined that 87.7% of the variance in grades was attributable to the test item rather than to grader or session. The intraclass correlation between graders and grading sessions varied from 0.84 to 0.91. Simple agreement within 1 grade between graders and sessions occurred in $90 \pm 5.5\%$ of gradings. κ values averaged 0.91, which is considered near perfect.

CONCLUSIONS—A 9-step photographic scale was designed to standardize the grading of vitreous haze in uveitis patients using fundus photographs. The scale is potentially adaptable to clinical trials in uveitis.

Grading of intraocular inflammation using cell counts in the aqueous humor and anterior vitreous and a subjective assessment of anterior chamber flare was described first in 1959.¹ These measures recently have been standardized in a consensus document.² Vitreous haze is the obscuration of fundus details by vitreous cells and protein exudation. It impacts vision more profoundly than anterior inflammation and therefore used commonly as an outcome measure for clinical trials involving intermediate, posterior, or panuveitis.

Kimura and associates proposed a 5-level scale of vitreous haze based on the ophthalmoscopic clarity of the ocular fundus in 1959.³ To improve grading reproducibility for clinical trials, Nussenblatt and associates published clinical photographs illustrating 6 grades of vitreous haze in 1985.⁴ Visibility of the optic nerve head, retinal vessels, and nerve fiber layer defined the scale grades. The Nussenblatt scale has been used widely, and the Standardization of Uveitis Nomenclature Working Group consensus document on uveitis nomenclature accepted the scale for grading vitreous haze in 2005, with the exception that

the grade of trace was converted to 0.5+ to allow numeric scoring.² It is easy to use, reproducible,⁵ and currently accepted by the Food and Drug Administration for use in clinical trials.

There are disadvantages of the current scale. Intervals between grades are not quantified. Demonstrating a 2-step change in grade in a clinical trial can be difficult, given the relatively small number of grades of vitreous haze and the relatively poor discrimination of lower levels of haze, which are more common in uveitis than the higher grades of haze. Finally, the scale is not validated as a standard for grading fundus photographs. We propose a method of photographic grading of vitreous haze using a 9-level scale created by photographing a normal fundus through Bangerter occlusion filters to create increasing optical blur, simulating increasing vitreous haze.

METHODS

SCALE CREATION

Bangerter occlusion filters or foils are used in the treatment of amblyopia as an alternative to atropine or patching. They degrade visual acuity and contrast sensitivity stepwise and are calibrated to Snellen visual acuities.⁶ The foils are made of plastic and are cut easily for mounting. Bangerter Filters (Haag-Streit, Koeniz, Switzerland) were used singly or in combination to create a 9-point scale separated by approximately 0.3 log units of spatial resolution, equivalent to a doubling of the visual angle. Grade 0 had no superimposed filter. Haze grades 1, 2, 3, and 4 used 0.8, 0.4, 0.2, and 0.1 filters, respectively. Haze grades 5, 6, and 7 were achieved by overlaying the 0.1 filter on the 0.4, 0.2, and 0.1 filters, respectively. Grade 8 was an overlay of two 0.1 filters and the 0.2 filter. The 4 composite filters were estimated to correspond to visual acuities of 20/400, 20/1000, 20/2000, and 20/5000 (Table 1). Visual acuities through the actual filter combinations used were not verified.

Grade 0 photographs of the normal fundus of a 20-year-old volunteer were obtained without a filter with a 30-degree Zeiss fundus camera model FF4 (Carl Zeiss Meditec, Inc, Pleasanton, California, USA) with a Nikon film camera (Nikon Instruments Inc, Melville, New York, USA) loaded with Fuji-chrome 100 ASA film. Bangerter occlusion filter single and combined filters of increasing density then were suspended before the front aperture of the camera and the best photographs were selected. The film slides subsequently were digitized with a Nikon film scanner at 24-bit color and resolution of 300 dpi. Images were saved as tagged image file format (TIF). A set of 9 standard photographs with incremental optical blur was created (Figure 1). Level 0 corresponds to no haze and level 8 to maximum haze. A theoretical level 9 corresponds to haze greater than 8, creating a 10-point scale.

SCALE VALIDATION

Ninety-seven color film fundus photographs of patients with uveitis were selected from the photographic archives. Patients with significant cataracts and prominent fundus pathologic features were excluded. Levels of vitreous haze ranging from clear to total opacification were included in the set. The color slides were digitized as described above.

Preliminary validation was performed with the images projected in an auditorium with 25 graders at different levels of ophthalmologic training who were asked to compare the pictures to a hard copy print of the standards and to record the grade on a grade sheet. The standard photos were projected simultaneously one by one in dual projection adjacent to the test photograph. Each test photograph was scored on a scale from 0 through 8. Grading took approximately 40 minutes.

Subsequently, an interactive graphical user interface using a web browser was designed to allow simultaneous online comparison of the 9 standard photographs with the same set of 97 test photographs and refined as a Windows application (Microsoft, Redmond, Washington, USA). Three retina fellows graded the set twice at a 2-month interval on the same computer monitor under constant lighting conditions after reading a brief paragraph about how to use the grader. The score was recorded online during the grading. Grading took approximately 20 minutes per session.

STATISTICAL EVALUATION

Intraclass and interclass correlation coefficients were calculated to assess overall agreement for all graders and then for the 3 graders over 2 grading sessions. Variance component analysis was performed to identify the sources of variability.

Agreement between graders also was evaluated with a distribution of their differences in grading scores, and agreement within graders was evaluated with a distribution of differences between each grader's first and second grading. κ statistics⁷ estimated the excess chance of both exact agreement and agreement within 1 grade. Because the grading scale is ordinal, a weighted κ value was used, which awards credit for exact or nonexact matches according to the formula:

$$1 - ((\text{differences between ratings}) / (\text{maximum difference between ratings})).^7$$

Therefore, an exact match is weighted as 1.0 and a match within 1 grade is weighted as 0.75. κ values of less than .40 represent poor agreement beyond chance, values between .40 and .75 represent fair to good agreement, and values greater than .75 represent excellent agreement.

RESULTS

The variance estimate between the photographs was 3.75, with a good intraclass correlation coefficient of 69%. Between all graders, the variance estimate was 0.24 with a correlation coefficient of 4%. A third source of variance accounted for 1.44 of the total variance estimate; this tentatively was attributed to variation in lighting and other distractions during the auditorium grading. Variance component analysis of only the computer-based grading indicated an improved intraclass correlation of 87.7%. Systematic differences between graders accounted for 10.6% of variance, and differences between grading sessions accounted for 1.7% (Table 2).

Simple agreement between the 3 graders during 2 grading sessions and κ statistics are displayed in Table 3. Exact agreement occurred in $48 \pm 7.6\%$ of items. Agreement within 1 grade occurred in $90 \pm 5.5\%$ of items. Intraclass correlation between the 3 graders for the first grading was 0.91, and that for the second grading was 0.84. The intraclass correlation between the first and second gradings by the same grader was 0.88. κ statistics were in the range considered near perfect for each comparison between graders and gradings (Table 3).

Figure 2 shows the average score per group of test items grouped by haze and the distribution of the tested images among the different grading levels. Most test items were grouped in the 4 lowest categories of haze. The standard deviations ranged from 0.1 to 0.3.

DISCUSSION

Quantification of vitreous inflammation is difficult yet critical for the performance of clinical research in intermediate, posterior, and panuveitis. The Nussenblatt scale was created in the early 1980s by selecting fundus photographs from uveitis patients that

depicted varying degrees of clarity of the optic nerve and retina. Validation consisted of a single live grading of 5 uveitic eyes by 3 ophthalmologists with indirect ophthalmoscopy. Three of 5 eyes were graded identically by the 3 observers and the remaining 2 were graded identically by 2 of 3 observers. This was compared with grading of vitreous cells by the method of Kimura and associates by the same observers and was believed to be more reliable.³ The consensus haze grades of the uveitic eyes were trace or 1+ in 2 eyes each and 3+ in 1 eye. Photographic grading was considered but not regarded as feasible because of expense and possible artifact related to photographic technique.⁴

More recently, Kempen and associates evaluated inter-observer reliability of live clinical grading by 2 uveitis specialists using the Nussenblatt vitreous haze scale.⁵ Simple agreement values within 1 grade were 87.3%, similar to the 90% agreement with photographic grading we achieved. The κ value for agreement within 1 grade was 0.75, versus the average κ value of 0.91 for the 3 graders across 2 gradings in our study, indicating substantially better agreement for the photographic grading in the simulated reading center environment.

Recent clinical trials (NCT00404612, NCT00685399, and NCT01095250 at clinicaltrials.gov) have specified outcomes as a 2-level change in the Nussenblatt vitreous haze based on the lack of other validated scales and the acceptance of the Nussenblatt scale by the Standardization of Uveitis Nomenclature Working Group.² The currently used scale has inherent problems that would benefit from revision. It is an analog grading scale with unequal steps between the gradations. Validation of the scale was not statistically based, although it was recently revalidated by Kempen and associates in a clinical setting.⁵ The original vitreous haze scale was designed for live clinical grading, rendering it unsuitable for objective reading center evaluations in clinical trials. Furthermore, the steps are broad, which makes the scale inefficient in grading the lower ranges of vitreous haze that are more commonly seen in uveitis patients.⁵ Because a 2-step change in grade is the usual outcome measure in clinical trials, the predicted effect of a vitreous haze scale with fewer grades and less sensitivity in the range of haze most commonly seen in uveitis patients would be either to exclude most patients from clinical research or to encourage overestimation of haze, with grading up to 2+ vitreous haze at enrollment.

We capitalized on advances in digital technology to design a reproducible scale that can be used to grade vitreous haze objectively in fundus photographs using a viewer designed as a stand-alone Windows application. Because the separately validated Bangerter foils were used to create blur, the levels correlate with Snellen visual acuity at approximately equal 0.3 log increments, similar to the Early Treatment Diabetic Retinopathy Study chart. In comparison with the 6-level Nussenblatt grading scheme, our scale has 9 gradations that are clearer than 4+ haze in the Nussenblatt scale and effectively doubles the resolution in the critical range in which most uveitis patients fall. In addition, the scale underwent a more robust validation process.

There are several caveats in the use of a photographic system to grade haze. Standardization is required when images are acquired digitally and displayed to assure reproducibility. In clinical trials, this translates into certification of photographers who follow a specific protocol. In the reading center, uniform image resolution, brightness, contrast, and color balance will be essential, together with calibration of the viewing monitors and equalization of the ambient lighting environment. It is assumed that going forward, all clinical trials will use digitally acquired fundus images and that the additional step of scanning 35-mm slides will not be required.

Even with standardization of technique, other difficulties remain. Estimating vitreous haze would be relatively straightforward if the sole factor causing the image blur were vitreous

opacification; however, image quality can deteriorate from anterior opacities, especially the lens. Whereas clinicians are trained to discount lens opacity in viewing the vitreous, the camera records any opacity. Comparisons of haze scores over time will still be valuable in clinical trials, providing that lens opacification is stationary during the interval needed to achieve a 2-step decrease in vitreous haze, a time-frame that seems feasible. Information about lens density grading should be available and considered when vitreous haze grades are interpreted. Other interfering factors include corneal scars, posterior synechiae, capsular opacification, and inhomogeneous vitreous opacities. The live observer can adapt the viewing angle to negate many of these obstructions, but the camera cannot. In addition, light scattering from photographic flash, focus, and patient cooperation may be significant confounding factors.

Despite these limitations, our scale introduces a standardized method for vitreous haze quantification. The grading levels correspond directly with Snellen visual acuity measurements with linearity on a log scale. This creates a more meaningful and objective clinical standard. More grades allow for discrimination of low levels of vitreous opacification, which are more representative of the distribution of vitreous haze in uveitis patients. Incorporation into reading centers for clinical trials will present significant advantage.

The usefulness of photographic grading of vitreous haze should be assessed further using fundus photographs acquired in clinical trials. Correlation of photographic grading with clinical haze measured by the Nussenblatt scale and with other clinical data will help to validate the use of the vitreous haze scale for clinical research in uveitis. Validation of the scale for live grading of vitreous haze by clinicians is potentially feasible but would require more study.

Acknowledgments

Publication of this article was supported by an unrestricted departmental grant from research to Prevent Blindness, Inc, New York, New York. Dr Davis has a consulting arrangement with Centocor and grant support from Centocor and Novartis. Dr Porciatti receives grant support from National Institutes of Health Center Grant P30-EY14801, NIH R01EY019077, NIH R01-EY014957. Mr Feuer receives support from National Institutes of Health Center Grant P30-EY14801. The study was supported by an unrestricted grant to the University of Miami from Research to Prevent Blindness, Inc. Involved in Design and conduct of the study (J.L.D., B.M., D.H., R.S., V.P.); Collection and management of data (J.L.D., B.M., J.C., D.H., R.S.); Analysis and interpretation of the data (J.L.D., B.M., W.J.F.); and Preparation and review (J.L.D., B.M., W.J.F.) and approval (J.L.D., B.M., J.C., R.S., D.H., V.P., W.J.F.) of the manuscript. The Human Subjects Research Committee at the University of Miami Miller School of Medicine approved conduct of the study and creation of a test set of photographs from patient files. Photography of the volunteer's eye was carried out with consent under a separate protocol. Grading of the test set in the simulated reading center environment was performed by Dr Brian Madow, Dr Ryan Rush, and Dr Namrata Kapoor.

REFERENCES

1. Hogan MJ, Kimura SJ, Thygeson P. Signs and symptoms of uveitis I. Anterior uveitis. *Am J Ophthalmol.* 1959; 47(5 part 2):155–170. [PubMed: 13649855]
2. Jabs DA, Nussenblatt RB, Rosenbaum JT. 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(3):509–516. [PubMed: 16196117]
3. Kimura SJ, Thygeson P, Hogan MJ. Signs and symptoms of uveitis. II. Classification of the posterior manifestations of uveitis. *Am J Ophthalmol.* 1959; 47(5 part 2):171–176. [PubMed: 13649856]
4. Nussenblatt RB, Palestine AG, Chan CC, Roberge F. Standardization of vitreal inflammatory activity in intermediate and posterior uveitis. *Ophthalmology.* 1985; 92(4):467–471. [PubMed: 4000641]

5. 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(6):813–818. [PubMed: 18687418]
6. Odell NV, Leske DA, Hatt SR, Adams WE, Holmes JM. The effect of Bangerter filters on optotype acuity, vernier acuity, and contrast sensitivity. *J AAPOS*. 2008; 12(6):555–559. [PubMed: 18706841]
7. Fleiss, JL. Statistical methods for rates and proportions. 2nd ed.. New York: John Wiley and Sons; 1981. p. 218-224.

Biography

Janet L. Davis, MD, Professor of Ophthalmology, is a member of the Retina Section, and the Chief of the Uveitis Section at the Bascom Palmer Eye Institute of the University of Miami Miller School of Medicine, Miami, Florida. She has participated in numerous multi-center clinical research trials in infectious and non-infectious uveitis. Her research interests include the assessment of intraocular fluids in uveitis and the metrics of monitoring uveitis.

FIGURE 1.

Vitreous haze scale grades 0 through 8. (Top row, from left to right) Grades 0, 1, and 2. (Second row, from left to right) Grades 3, 4, and 5. (Third row, from left to right) Grades 6, 7, and 8. The 9 images are displayed in increasing order of opacity each equivalent to approximately 0.3 log units of degradation in visual acuity based on the Bangerter calibration. Grade 8 is equivalent to or slightly more opaque than grade 3+ in the Nussenblatt vitreous haze scale. Diffraction rings from combining the Bangerter filters prevented creating increasing levels of haze beyond grade 8. A tenth grade equivalent to haze more than standard 8, and an eleventh grade in which no fundus details are visible are possible, but would not be linear.

FIGURE 2.

Vitreous haze scale grading of the test set of fundus images. The bar histogram displays the 97 test objects distributed according to the scale grading 0 through 8. The y-axis corresponds to the number of photographs in each of the 9 scale grades. The x-axis gives the mean category scores with their standard deviation. The most common grades of haze in the test set were in the 4 steps from 0 to 3, but all grades were represented.

TABLE 1

Vitreous Haze Scale Values with Corresponding Bangerter Filters, Estimated Snellen Acuities, and logMAR Equivalents

Haze Scale Value	Bangerter Filter	Estimated Snellen Acuity	LogMAR Equivalent
0	None	20/20	0
1	0.8	20/25	0.1
2	0.4	20/50	0.4
3	0.2	20/100	0.7
4	0.1	20/200	1
5	0.1*0.4	20/400	1.3
6	0.1*0.2	20/1000	1.7
7	0.1*0.1	20/2000	2
8	0.1*0.1*0.2	20/5000	2.7

logMAR = logarithm of the minimal angle of resolution.

TABLE 2

Source of Variability between Vitreous Haze Scores Generated by 3 Graders and 2 Grading Sessions in a Simulated Reading Center Environment Using the Vitreous Haze Scale

Source of Variability	Variance Component	% of Variance
Photograph (n = 97)	3.21	87.7
Grader (n = 3)	0.39	10.6
Grading (n = 2)	0.06	1.7

TABLE 3
Vitreous Haze Scale Score Differences Between 3 Graders Over 2 Gratings of 97 Test Items With Estimates of Excess Agreement Beyond Chance (κ)

	No. of Items With Difference in Score					κ within 1 Grade ^b
	0	± 1	± 2	± 3	κ^a	
Grader 1 vs 2, first grading	53	39	5	0	0.94	0.98
Grader 1 vs 3, first grading	45	43	6	3	0.91	0.95
Grader 2 vs 3, first grading	50	39	8	0	0.93	0.97
Grader 1 vs 2, second grading	38	41	16	2	0.87	0.91
Grader 1 vs 3, second grading ^c	40	40	12	4	0.86	0.89
Grader 2 vs 3, second grading	51	37	7	2	0.92	0.95
Grader 1, grading 1 vs 2	36	46	13	2	0.88	0.93
Grader 2, grading 1 vs 2	54	38	5	5	0.94	0.98
Grader 3, grading 1 vs 2	55	38	4	0	0.95	0.99

^a κ estimates the excess agreement beyond what is expected by chance. Values from 0.81 to 1.00 are considered near perfect.

^b κ estimates created by considering agreement within 1 grade to be equal to exact agreement.

^c There was one 4-grade difference between grader 1 and 3 in the second grading.