Interobserver Agreement in Clinical Grading of Vitreous Haze Using Alternative Grading Scales

Dana M. Hornbeak, MD, MPH,[1] *Abhishek Payal, MBBS, MPH,*[1,4,5] *Maxwell Pistilli, MEd, MS,*[2,3] *Jyotirmay Biswas, MS,*[6] *Sudha K. Ganesh, MBBS, DO,*[7,8] *Vishali Gupta, MD,*[9] *Sivakumar R. Rathinam, MD, PhD,*[10,11] *Janet L. Davis, MD,*[12] *John H. Kempen, MD, PhD*[1,2,3]

Purpose: To evaluate the reliability of clinical grading of vitreous haze using a new 9-step ordinal scale versus the existing 6-step ordinal scale.

Design: Evaluation of diagnostic test (interobserver agreement study).

Participants: A total of 119 consecutive patients (204 uveitic eyes) presenting for uveitis subspecialty care on the study day at 1 of 3 large uveitis centers.

Methods: Five pairs of uveitis specialists clinically graded vitreous haze in the same eyes, one after the other using the same equipment, using the 6- and 9-step scales.

Main Outcome Measures: Agreement in vitreous haze grade between each pair of specialists was evaluated by the κ statistic (exact agreement and agreement within 1 or 2 grades).

Results: The scales correlated well (Spearman's $\rho = 0.84$). Exact agreement was modest using both the 6-step and 9-step scales: average $\kappa = 0.46$ (range, 0.28−0.81) and $\kappa = 0.40$ (range, 0.15−0.63), respectively. Within 1-grade agreement was slightly more favorable for the scale with fewer steps, but values were excellent for both scales: $\kappa = 0.75$ (range, 0.66−0.96) and $\kappa = 0.62$ (range, 0.38−0.87), respectively. Within 2-grade agreement for the 9-step scale also was excellent ($\kappa = 0.85$; range, 0.79−0.92). Two-fold more cases were potentially clinical trial eligible on the basis of the 9-step than the 6-step scale ($P<0.001$).

Conclusions: Both scales are sufficiently reproducible using clinical grading for clinical and research use with the appropriate threshold (≥2- and ≥3-step differences for the 6- and 9-step scales, respectively). The results suggest that more eyes are likely to meet eligibility criteria for trials using the 9-step scale. The 9-step scale appears to have higher reproducibility with Reading Center grading than clinical grading, suggesting that Reading Center grading may be preferable for clinical trials. *Ophthalmology 2014;■:1−6 © 2014 by the American Academy of Ophthalmology.*

Uveitis is an important cause of visual loss.[1] Treatment of uveitis with anti-inflammatory medication aims to improve or maintain vision by alleviating inflammation. Inflammatory cells and protein exudates in the vitreous make the view of the fundus hazy,[2] which usually has been graded by ophthalmoscopy with reference to standard photographs[3,4] and more recently by grading of color fundus photographs.[5] Improvement of vitreous haze is a goal of anti-inflammatory therapy and has been adopted as an appropriate primary or secondary outcome for several clinical trials studying the effect of new treatments on uveitis.[6−10] Thus, grading vitreous haze is useful for determining the course of patient management and as a quantifiable outcome for clinical trials.

The 6-step ordinal scale developed at the National Eye Institute in 1985 for grading vitreous haze was accepted by the Standardization of Uveitis Nomenclature (SUN) Working Group in 2005 as an appropriate scale for use in clinical research; the only change recommended in the original scale was recording the grade of "trace" as 0.5+.[3,4] Validation of the 6-step scale has been limited to a small initial study of 3 observers examining 6 eyes,[3] and a larger study evaluating the scale in a clinical setting, showing modest exact agreement but favorable within 1-grade agreement between 1 pair of observers (κ value for within 1-grade agreement was 0.75).[11] The scale has been implemented in published[6−10] and unpublished clinical trials. Many trials have required 2+ or higher vitreous haze for enrollment to permit detection of a change from 2+ to zero haze. In these studies, recruitment has been difficult, because ≥2+ vitreous haze is encountered infrequently in clinical practice.[6] In addition, the ordinal scale has been analyzed in some cases as if it were a numeric scale, taking the 0.5+ ordinal step as being 0.5,[8,10] which may introduce error given that the steps are not necessarily an equal distance apart, and the 0.5+ step is an ordinal step just like the other steps. Although alternative methods for analysis of ordered categoric outcomes exist, they are complicated to implement; thus, a scale based on an underlying quantitative foundation would have advantages.

Davis et al[5] recently proposed a 9-step scale to standardize the grading of vitreous haze using reading center gradings of color fundus photographs. The scale offers potential advantages in having more steps (with greater sensitivity to haze

ISSN 0161-6420/14/$ - see front matter
http://dx.doi.org/10.1016/j.ophtha.2014.02.018

differences at the lower end of the scale where most gradings fall) and in being based on a log-linear distribution of image haziness, such that the difference between any 2 steps has the same quantitative meaning. Also, the greater number of steps potentially would broaden patient eligibility for clinical trials in which the goal is to show a ≥2-step difference in vitreous haze. Davis et al[5] found high interobserver and intraobserver agreement (average κ value = 0.91 for within 1-grade agreement) using this scale in a Reading Center environment[5] and replicated these findings using baseline images from a major clinical trial, in which vitreous haze correlated well with visual acuity.[12] However, the scale has not been evaluated for use in a clinical grading environment, which would be less expensive to implement in clinical trials and clinical practice.

The purpose of this cross-sectional study was to evaluate intergrader agreement in clinical grading of vitreous haze, with the goal of more rigorously validating the reproducibility of the existing 6-step scale and evaluating the reproducibility of the proposed 9-step scale for use in clinical and research settings. The study also directly compares the use of the 2 scales in clinical settings and assesses correlations between vitreous haze grade and other clinical characteristics.

Methods

Five uveitis specialists participated from 3 large uveitis centers that had high numbers of cases with severe inflammation. Each clinician had more than 10 years in subspecialty practice and previously had participated in multicenter uveitis clinical trials using vitreous haze as an outcome. The specialists reviewed the grading criteria for the 6-step and 9-step scales and completed a brief run-in training session to make sure they had equivalent understanding of the grading process. Pairs consisted of 1 ophthalmologist paired with each of the 5 host ophthalmologists. The same patients were evaluated by each clinician in the pair, one clinician after the other. The centers were at the Aravind Eye Hospital, Madurai, India (center 1, pair 1); the Post-Graduate Institute, Chandigarh, India (center 2, pair 2); and Sankara Nethralaya, Chennai, India (center 3, pairs 3–5, grading the same group of patients). The study was conducted after obtaining the required approvals from the Institutional Review Boards of the University of Pennsylvania, Sankara Nethralaya/The Medical and Vision Research Foundation, the L. V. Prasad Eye Institute, and the Aravind Eye Hospital and Postgraduate Institute of Ophthalmology. The study was conducted in accordance with the precepts of the Declaration of Helsinki.

Patients presenting for uveitis subspecialty care were examined by each clinician pair as the patients presented, one after the other, without discussing the cases or gradings. Some patients had been asked to come in that day on the basis of an expectation of the host clinician that a high level of vitreous haze would be present. Each pair completed live gradings using the same slit-lamp biomicroscopes, indirect ophthalmoscopes, and lenses (equipment that had been used in clinical trials). Eyes judged to have a diagnosis of uveitis by at least 1 grader in each pair, on the basis of findings at the time of the gradings without reference to medical records or the clinician's memory of the case, were included in the analysis.

For the 6-step scale, vitreous haze was graded following the described method, by comparing indirect ophthalmoscopy findings with a printed poster previously used in a clinical trial displaying standard photographs, mentally subtracting the effect of media opacities other than vitreous haze.[3,4] For grading using the 9-step scale, indirect ophthalmoscopy findings with the best fundus view obtained were compared with photographs demonstrating the 9 levels shown on a laptop computer screen (the same screen at each center),[5] without mentally subtracting the effect of media opacities other than vitreous haze. Additional clinical characteristics that might have affected vitreous haze grade were noted by each ophthalmologist on the basis of slit-lamp biomicroscopy, again using the same equipment in succession. These included the presence of central corneal opacities, the degrees of posterior synechiae, and the lens status (clear lens, cataract, pseudophakia, posterior capsular opacification, or aphakia). Participants kept their response sheets separate to avoid biasing one another. Host clinicians who might have seen the patients previously were instructed to ignore any recollection of previous findings and to record findings solely on the basis of observations made on the day of examination.

Statistical Analysis

Frequencies of clinical characteristics and vitreous haze grade were compiled for all participant eyes, except that comments on uveitis type were ignored for eyes graded to have no uveitis activity (because there was no basis on which to determine what the site of uveitis was in this scenario). The probability of any particular 6-step grading given an observed 9-step grading was modeled using cumulative logistic regression, adjusting only for the grader, applying generalizations of general estimating equations[13] to account for correlation between the eyes of individual patients (SAS v9.3, SAS Inc, Cary, NC). Interobserver agreements within each scale and intraobserver agreements between the 2 scales were assessed using simple agreement and calculation of the κ statistic (Stata 11 Intercooled software, StataCorp, College Station, TX). A κ value of 0.00 to 0.20 was considered slight agreement, 0.21 to 0.40 was considered fair agreement, 0.41 to 0.60 was considered moderate agreement, 0.61 to 0.80 was considered substantial agreement, and 0.81 to 1.0 was considered almost perfect agreement.[14]

Similar to the approach we took in a previous study evaluating agreement in grading clinical characteristics related to uveitis,[11] we evaluated exact agreement and within 1- or 2-grade agreement (i.e., if 2 clinicians graded vitreous haze within 1 or 2 grades of each other, that grade was considered to reflect within 1- or 2-grade agreement, respectively). For the 6-step scale, we also evaluated agreement within 1 grade except requiring grades 0 and 4 to have exact agreement (modified 1-grade approach), in an effort to mimic the approach suggested by the SUN Working Group in which a change of 2 grades or a change of 1 grade to the floor or ceiling of the scale is considered a clinically important degree of change. Weighted κ values were calculated, and 95% confidence intervals were calculated via bootstrapping 1000 times. The proportion eligible for a hypothetical clinical trial were evaluated using McNemar's test based on random selection of one grader's grading for each eye (the average of 100 000 replications).

Results

The total number of uveitic eyes (patients) graded and included in the final analysis was 44 (30) at center 1 (pair 1), 79 (43) at center 2 (pair 2), and 81 (46) at center 3 (pairs 3–5).

In this group of cases, although several were asked to come in on the day of evaluation because they were expected to have high levels of vitreous haze, only 14% of uveitic eyes were given a grading of ≥2+ on the 6-step scale (Fig 1A). Comparatively, 29% and 20% of eyes were given a grading of ≥3 and ≥4, respectively, on the 9-step scale (Fig 1B).

Clinicians using the 6-step scale mentally subtract the effect of media opacities when grading vitreous haze. In contrast, when