

# Comparability of Digital Photography with the ETDRS Film Protocol for Evaluation of Diabetic Retinopathy Severity

Helen K. Li,<sup>1,2,3,4</sup> Ronald P. Danis,<sup>5</sup> Larry D. Hubbard,<sup>5</sup> Jose F. Florez-Arango,<sup>1,6</sup> Adol Esquivel,<sup>1,7</sup> and Elizabeth A. Krupinski<sup>8</sup>

**PURPOSE.** To evaluate digital photography parameters affecting comparability with the Early Treatment Diabetic Retinopathy Study (ETDRS) film protocol for diabetic retinopathy (DR) severity grading.

**METHODS.** ETDRS protocol photographs and four variations of digital images (uncompressed stereoscopic, compressed stereoscopic, uncompressed monoscopic, and uncompressed monoscopic wide-angle mosaic) of 152 eyes were independently evaluated by using ETDRS classifications. Digital formats were compared to film and each other for agreement on severity level, DR presence at ascending threshold, presence of the DR index lesion, and repeatability of grading. Study parameters included image resolution sufficient to distinguish small lesions, color balancing of digital images to film, documenting essential ETDRS classification retinal regions, similar magnification, and supplementary green-channel viewing.

**RESULTS.** The  $\kappa$  statistic was substantial or near substantial between all digital formats and film for classifying severity levels ( $\kappa = 0.59-0.62$ ;  $\kappa_w$  [linear weighted] =  $0.83-0.87$ ). The distribution of DR levels in all digital formats was not significantly different from that of the film (Bhappkar test,  $P = 0.09-0.44$ ). The  $\kappa$  among digital formats for severity level was also substantial or near substantial ( $\kappa = 0.58-0.76$ ,  $\kappa_w = 0.82-0.92$ ). Differences between digital formats and film for grading severity level, severity threshold, or index lesions were not significant. The repeatability of grading between readers using film and all digital formats was also similar.

**CONCLUSIONS.** Digital format variations compared favorably with film for DR classification. Translating film characteristics (resolution, color/contrast) and protocol (magnification, retinal regions) to digital equivalents and augmentation of full color with green-channel viewing most likely contributed to the results. (*Invest Ophthalmol Vis Sci.* 2011;52:4717-4725) DOI: 10.1167/iovs.10-6303

The Early Treatment Diabetic Retinopathy Study (ETDRS) photography and grading protocols have been benchmarks for photographic evaluation of diabetic retinopathy (DR) since 1985.<sup>1</sup> Protocols were derived from the Airlie House Symposium's classification of DR,<sup>2</sup> modified for the Diabetic Retinopathy Study (DRS),<sup>3</sup> and extended for the ETDRS.<sup>4</sup> The classification has been used in defining DRS high-risk characteristics<sup>5,6</sup> and ETDRS clinically significant macular edema<sup>7,8</sup> as indications for panretinal and focal laser photocoagulation. Progression of retinopathy on the ETDRS DR severity scale has become a primary outcome, a surrogate for vision loss from DR in clinical research and epidemiology studies.<sup>9,10</sup> Progression along the severity scale was used by the Diabetes Control and Complications Trial (DCCT) to establish the efficacy of intensive blood glucose control in type 1 diabetes for DR<sup>11,12</sup> and by the Sorbinil Retinopathy Trial to show that an aldose reductase inhibitor lacks efficacy.<sup>13</sup> Progression has also been adopted as a secondary outcome in various Diabetes Retinopathy Clinical Research Network macular edema treatment trials.<sup>14,15</sup>

These protocols were limited by available technologies. The classic fundus camera's 2.5 $\times$  magnification and 30° field (Zeiss FF; Carl Zeiss Meditec, Dublin, CA) required seven 35-mm photographs to capture retinal regions of interest (optic disc, macula, four vascular arcades, and region temporal to the macula).<sup>16</sup> Color slide films (Kodachrome and Ektachrome; Eastman, Kodak) were adopted with standardized processing. Handheld 5 $\times$  Donaldson stereo viewers were used to examine slide pairs in approximate anatomical position inside plastic sheets on a light box.

Two forces now drive the replacement of film with digital fundus photography in DR research studies and clinical practice. Color slide film and development are less available (after 74 years, Kodachrome was discontinued in 2009). Perhaps more important, digital fundus cameras offer advantages such as "stat" viewing for disease diagnosis and patient consultation, more convenient image storage, indexing, retrieval, and transmission. However, the impact of parameters unique to digital photography is not fully understood. The ongoing DCCT/Epidemiology of Diabetes Interventions and Complications trial emphasizes the importance of digital image protocol comparability and historical continuity with the ETDRS film protocol.<sup>17</sup> The American Telemedicine Association recommends that ocular telehealth programs validate systems to the ETDRS film protocol.<sup>18</sup>

From the <sup>1</sup>School of Biomedical Informatics, University of Texas Health Science Center, Houston, Texas; the <sup>2</sup>Department of Ophthalmology, Thomas Jefferson University, Philadelphia, Pennsylvania; the <sup>3</sup>Department of Ophthalmology, Methodist Hospital, Houston, Texas; the <sup>4</sup>Weil Cornell Medical College, New York, New York; the <sup>5</sup>Department of Ophthalmology and Visual Sciences, University of Wisconsin-Madison, Madison, Wisconsin; <sup>6</sup>Universidad De Antioquia, Medellin, Columbia; the <sup>7</sup>St. Luke's Episcopal Health System, Houston, Texas; and the <sup>8</sup>Department of Radiology, University of Arizona, Tucson, Arizona.

Supported by a grant from Juvenile Diabetes Foundation Research International, New York, NY (HKL) and by unrestricted grants from Research to Prevent Blindness (Department of Ophthalmology and Visual Sciences, The University of Texas Medical Branch and the Department of Ophthalmology and Visual Sciences, University of Wisconsin School of Medicine and Public Health).

Submitted for publication July 29, 2010; revised December 12, 2010; accepted February 2, 2011.

Disclosure: **H.K. Li**, None; **R.P. Danis**, None; **L.D. Hubbard**, None; **J.F. Florez-Arango**, None; **A. Esquivel**, None; **E.A. Krupinski**, None

Corresponding author: Helen K. Li, 1400 Hermann Drive, Houston, TX; hlmed@mac.com.

In a multipart study, we compared the performance of digital color fundus photography to that of ETDRS film for grading DR.

Evaluation of diabetic macular edema was also studied and reported in another paper.<sup>19</sup> Parameters kept constant throughout the studies included resolution sufficient to distinguish subtle DR abnormalities (i.e., microaneurysm and intraretinal microvascular abnormalities), color balancing digital images to film, documenting essential retinal regions for ETDRS classification, viewing magnification, and supplementary green-channel viewing. In the first study, only the ETDRS seven-standard-field stereoscopic protocol recording medium was changed, from film (F) to digital (D).<sup>20</sup> Formats were systematically varied in additional studies that compared the ETDRS film protocol to:

*Digital image file compression (Dc).*<sup>21</sup> The volume of data from 32 ETDRS protocol color photographs per patient is substantial. Compression is often used to reduce storage requirements and speed transmission. Diagnostic image quality is, however, an important concern for clinical research and telemedicine. How well compressed digital photographs match the historical performance of ETDRS film in distinguishing full range DR has not been assessed.

*Elimination of stereo effect (Dm).*<sup>22</sup> There is limited information on how stereoscopic photography impacts DR diagnostic accuracy or whether stereo's added burden and cost are justified.

*Substitution of a monoscopic wide-angle mosaic image for the ETDRS 30° stereo photo array (DmMos).*<sup>23</sup> Wide-angle mosaics offer the possibility of viewing an entire retinal region of interest in a single photograph, increasing file storage and transmission efficiency. Evidence that full ETDRS severity scale DR evaluation is clinically achievable with a mosaic image is needed to reliably replace ETDRS protocol photography.

This report gives the results of the comparison of the four digital formats to film and each other for distinguishing a spectrum of DR severity levels in the same eye. It also weighs the importance of other parameters in achieving comparability with the ETDRS film protocol.

## MATERIAL AND METHODS

### Fundus Photographs

Patients from The University of Texas Medical Branch Department of Ophthalmology and Visual Sciences Eye Clinic gave written consent for the photography. Institutional review board approval was obtained, and the tenets of the Declaration of Helsinki were observed. Patients with a broad range of DR severity were informally screened in the clinic. Those needed to populate a complete range of ETDRS severity levels, from no retinopathy to severe proliferative retinopathy, were invited for photography. Patients having media opacities or limited pupil dilation preventing an adequate retina image, having other retinal vascular disease, or having had retinal laser photocoagulation were excluded. Each study eye had slides and digital photographs taken per the ETDRS protocol: seven nonsimultaneous color stereo pairs of the fundus. Film and digital photographs were taken using the same mydriatic camera (TRC-50EX/IX; Topcon Medical Systems, Paramus, NJ) at its 35° setting. Fundus camera optics were coupled with a 35-mm camera for one set of images and a digital camera (MegaVision, Santa Barbara, CA) for the other set. Patients also underwent dilated photography with a nonmydriatic camera (TRC-NW6X; Topcon Medical System) and with a digital camera (model D100; Nikon, Melville, NY) at its 45° setting for the mosaic format. Whereas ETDRS utilizes seven fields, the nonmydriatic camera's automosaic software (IMAGEnet 2000, ver.

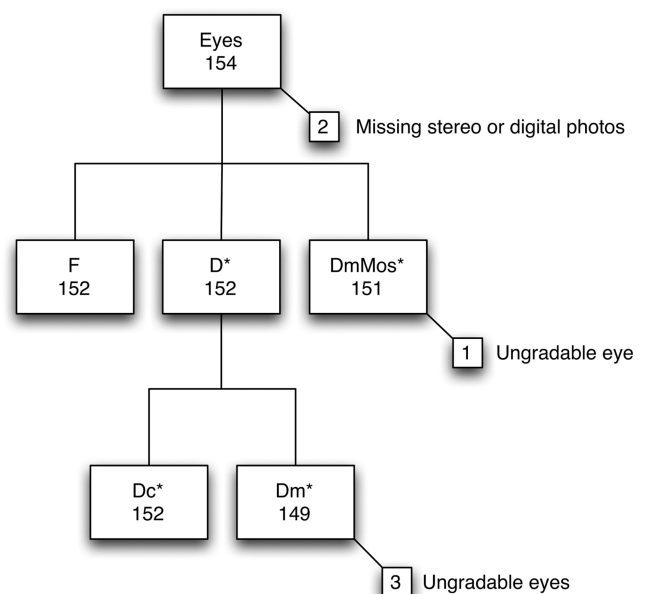
2.55; Topcon Medical System) uses nine: eight peripheral fields surrounding one central field of the posterior pole. The mydriatic camera's digital sensor resolution was  $2400 \times 2000$  pixels, and the nonmydriatic's was  $3000 \times 2000$  pixels. Resolving power governs the size of the smallest object that can be discerned and is determined by pixel spacing and camera optics.<sup>24,25</sup> Ektachrome 100 film is the equivalent of 2290 pixels per inch,<sup>26</sup> making the film's "pixel" spacing approximately  $4.6 \mu\text{m}/\text{pixel}$ , which corresponds to a resolving power of  $13 \mu\text{m}/\text{pixel}$ . The mydriatic system had  $4.6\text{-}\mu\text{m}/\text{pixel}$  spacing and  $13 \mu\text{m}/\text{pixel}$  resolving power. The nonmydriatic system had a  $6.2\text{-}\mu\text{m}/\text{pixel}$  spacing and  $17\text{-}\mu\text{m}/\text{pixel}$  resolving power. Both digital systems were sufficient to detect abnormalities 25 to  $30 \mu\text{m}$  in diameter.

Patients rested a minimum of 30 minutes between sessions. The same photographer, certified by the University of Wisconsin (UW) Fundus Photograph Reading Center for ETDRS protocol photography, took all photographs using the mydriatic camera followed by the nonmydriatic one in the sequence: D, F, and DmMos. Ektachrome slide film was processed at a Kodak-certified Q-Laboratory facility.<sup>27</sup> Stereoscopic photographs were taken with the mydriatic camera only.

Two photographed eyes were excluded due to missing stereo pair or digital photos (Fig. 1). A total of 152 eyes from an 85-patient cohort were selected to represent a stratified sample across the full range of DR severity levels. Stereo photographs were taken with the mydriatic camera only. Photographs were coded to remove identifying patient information.

A pilot study by the Wisconsin Fundus Photograph Reading Center found grading DR from digital color photos less sensitive than from film. Digital sensors have a narrower dynamic range, which lowers their brightness and contrast of red features against the reddish retinal pigment epithelium.<sup>24,25</sup> To overcome this limitation, we balanced the color and contrast of digital images to conform to a formal color model based on standard ETDRS slides.<sup>28</sup> Custom software generated red/green/blue (RGB) luminance histograms and adjusted each color channel curve to fit the model.

Digital stereo image sets were processed to create other digital formats (Fig. 1). One set was compressed using JPEG 2000 37:1 (lossy; ver. 3.98; IrfanView, Wiener Neustadt, Austria, with the Lura Wave



**FIGURE 1.** Cohort of eyes by format. Demographics: 32 (37.6%) males, 53 (62.4%) females; 37 (43.5%), Caucasians, 24 (28.2%), Hispanics, 24 (28.2%), African-Americans; age (median 60.5; mean 59.4 years, range, 33–83). Seventy-five right- and 77 left-eye images were used: both eyes from 67 (78.8%) patients, only right eye from 8 (9.4%), and only left from 10 (11.8%). \*Green-channel views available.

TABLE 1. ETDRS Diabetic Retinopathy Severity Level Assigned from Grading Digital versus Film Images

D	F										Total
	10	15/20	35	43	47	53	61	65	71/75	90	
10	25	1	1								27
15/20	1	5	1								7
35		2	26	3							31
43			4	10	7						21
47			2	11	8	2	1				24
53					8	3					11
61				1	1		7				9
65							2	9			11
71/75								1	10		11
90											0
<b>Total</b>	26	8	34	25	24	5	10	10	10	0	152
<b>%</b>	17.1	5.3	22.4	16.4	15.8	3.3	6.6	6.6	6.6		

Nine-step ETDRS scale: 10, no retinopathy; 15/20, microaneurysms or retinal hemorrhages only; 35, mild nonproliferative diabetic retinopathy (NPDR); 43, moderate NPDR; 47, moderately severe NPDR; 53, severe NPDR; 61, mild proliferative retinopathy (PDR); 65, moderate PDR; 71/75, severe PDR; and 90, cannot grade. F, stereoscopic film; D, digital stereoscopic. Dark gray shading indicates perfect agreement; light gray shading indicates agreement within one step.

JP2 plug-in; LuraTech, Inc., San Jose, CA).<sup>21</sup> Monochromatic green-channel images generally provide better contrast of small lesions than do full-color photographs.<sup>29</sup> To further compensate for the lower contrast of DR lesions in digital images, color images were supplemented with green-channel viewing to confirm subtle DR lesions when grading digital formats. Monochromatic photographs were created by extracting the green channel from seven-field and mosaic digital RGB color images. The left- or right-eye image from each stereo pair exhibiting the best focus and contrast was selected to create color and green-channel monoscopic formats.

**Grading Protocol**

Three University of Wisconsin-Madison certified readers independently evaluated all images. Batches of different formats were counterbalanced in the presentation order to minimize bias. To reduce recall, evaluation schedules were regulated by software that separated grading the same eye by at least 2 weeks. The software displayed digital images and provided screens for entering grades. Grading duration was recorded as the time between opening and closing the application for each eye.

Readers graded severity according to the abnormalities present in the relevant photographic fields, using the ETDRS classification system.<sup>4,9</sup> Findings were entered into the software, which calculated retinopathy severity on a nine-step ETDRS scale (Table 1).<sup>9</sup> Confidence levels were assigned to each study eye in each format on the basis of reader confidence in assessing retinopathy severity: high (good image quality and typical lesions), moderate (less satisfactory image quality and/or atypical lesions), or low (leading to the selection of “cannot grade”). Readers determined presence and severity of index abnormalities and diabetic macular edema (DME) according to ETDRS definitions.<sup>4,9</sup> Stereo sets of slides were viewed on daylight fluorescent light boxes using a Donaldson 5× stereo viewer (George Davco, Holcombe, MA) and graded as described in ETDRS report 10.<sup>4</sup> Magnification was approximately 12.5×, accounting for the combined magnification of the Topcon camera and the Donaldson viewer.

The software displayed seven-field digital images monoscopically with full-screen resolution at 13× magnification, which approximates slide magnification. Digital images were displayed on 21-in. CRT monitors (1600 × 1200) viewed approximately 26 in. from the screen. Monitors were set at a color temperature of 6500° K and 2.2 gamma and checked monthly (GretagMachbeth color checker; X-Rite Inc., Grand Rapids, MI). Digital images were displayed at 6.5× magnification when shown as stereo pairs and viewed with a handheld stereo viewer (Screen-Vu; Eye Supply USA, Tampa, FL). During grading, read-

ers first reviewed an online proof sheet of digital thumbnails before examining each 35° field monoscopically (13×) in color. For stereoscopic grading, each field was reviewed monoscopically (13×) and then stereoscopically (6.5×) in color.

Readers graded wide-angle mosaics without referring to their constituent images. The mosaic was first examined while fit to screen, then zoomed to 13× and panned to view each region at full magnification. A custom template indicating where individual fields would have been in a 30° standard field set was superimposed on the mosaic to guide readers in recording observations in a standard field-grading form (Fig. 2).



FIGURE 2. A custom template demarcates retinal regions in the mosaic corresponding to the standard fields: optic disc, macula, temporal to the macula, and each of the four vascular arcades. The template guided readers to appropriate sections of the seven-standard-field grading form for recording presence and severity of diabetic retinal abnormalities per ETDRS severity definitions.

## Statistical Analysis

DR severity level in each eye was calculated as the central tendency (median) among the three readers. Instead of duplicate grading with adjudication of differences, this method allowed pair-wise comparisons of all readers within each imaging format.

Stereo film grading results were considered the reference standard. The presence or absence of index abnormalities, severity level, and retinopathy severity at different thresholds were compared. We defined "threshold" as the cutoff for the presence of retinopathy at a particular ETDRS severity level or worse. Thresholds ranged from any retinopathy (level 15/20) to high-risk PDR (level 71/75).

Severity level agreement was cross-tabulated, and  $\kappa$  values (unweighted and weighted [linear scheme]) were calculated. Eyes with photographs classified as ungradable (level 90) were excluded. Based on the Landis and Koch ranges used in ETDRS report 10, the guidelines for interpretation were 0.0 to 0.2, slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and 0.81 to 1.00, almost perfect agreement.<sup>30</sup> McNemar's test for dichotomized scales and Bhapkar's test for multistep scales were used (cross-tab marginal homogeneity). Marginal distributions were assessed with McNemar's test of overall bias.

Grading comparisons for DR severity thresholds and index lesions were analyzed using  $\kappa$ , sensitivity and specificity percentages, positive and negative predictive values, and disease prevalence percentages. Agreements between digital and film, between digital formats, and within each format were assessed by percentages of exact and near agreement, and with  $\kappa$ .

Ungradable eyes were excluded from pair-wise interreader performance analyses within each format. Statistics comparing one image format to another were calculated using the common subset of gradable eyes. Analyses were performed with commercial software (Analyze-It; Analyze-It, Ltd.; Leeds, UK; MedCalc; MedCalc Software Bvba; Mariakerke, Belgium) and the MH (marginal homogeneity) program (ver. 1.2).<sup>31</sup>

## RESULTS

### Number of Eyes

Distributions of severity levels graded from F and from D are cross-classified in Table 1. Three Dm eyes were ungradable due to suboptimal focus and/or contrast. One DmMos eye was ungradable secondary to reduced clarity from corneal dryness (Fig. 1).

### Overall Results

Comparisons of determination of severity levels between the image formats are summarized in Table 2, ascending severity threshold in Figure 3, and index lesions in Figure 4. Severity level interreader reproducibility is summarized in Table 3. Severity level, severity threshold, and index abnormality pair-wise agreement comparisons between all digital formats and film were not significantly different. Confidence intervals of each pair overlap with those of other pairs.

Severity level pair-wise agreement between all digital formats and film was substantial except DmMos, which was moderate (Table 2A). Distributions of ETDRS severity levels from grading digital formats were not significantly different from film ( $P = 0.09$  to  $0.53$ , Bhapkar test; Table 2A). Agreement between digital formats was higher than between digital formats and film (Table 2).

Within each threshold, the  $\kappa$  values were similar for each film-digital comparison, and all were 0.80 or above, except for the  $\geq 47$  and  $\geq 53$  levels (Fig. 3A). Sensitivity between digital formats and film was above 90% for all DR thresholds except  $\geq 47$ , where sensitivity was 88% across most digital formats (Fig. 3B). Only 10 eyes had severity level 71/75, which contributed to wide confidence intervals in sensitivity across all digital formats. The small number of eyes with venous beading (VB), neovascularization of the disc (NVD), fibrous proliferation of the disc (FPD), and vitreous hemorrhage accounted for these lesions' wide sensitivity confidence intervals (Fig. 4B).

TABLE 2. Diabetic Retinopathy Severity Level Distribution

#### A. Digital Formats Compared to Stereo Film

	D vs. F (n = 152 Eyes)	Dc vs. F (n = 152 Eyes)	Dm vs. F (n = 149 Eyes)	DmMos vs. F (n = 151 Eyes)
Kappa (SE) (95% CI)	0.62 (0.04) (0.54-0.71)	0.60 (0.04) (0.51-0.68)	0.65 (0.04) (0.56-0.73)	0.59 (0.04) (0.51-0.68)
Weighted Kappa (SE) (95% CI)	0.86 (0.02) (0.82-0.90)	0.86 (0.02) (0.82-0.90)	0.87 (0.02) (0.83-0.91)	0.83 (0.03) (0.78-0.88)
Exact agreement, %	67.8	65.1	69.8	64.9
Severity higher in digital, %	21.7	21.7	19.7	23.0
Severity higher in film, %	10.5	13.2	11.8	12.5
Agreement $\pm$ 1 step	96.1	96.1	96.0	94.7
Agreement $\pm$ 2 steps	99.3	99.3	98.7	98.0
Bhapkar	0.21	0.09	0.53	0.44
McNemar bias*	0.02†	0.07†	0.18	0.04†

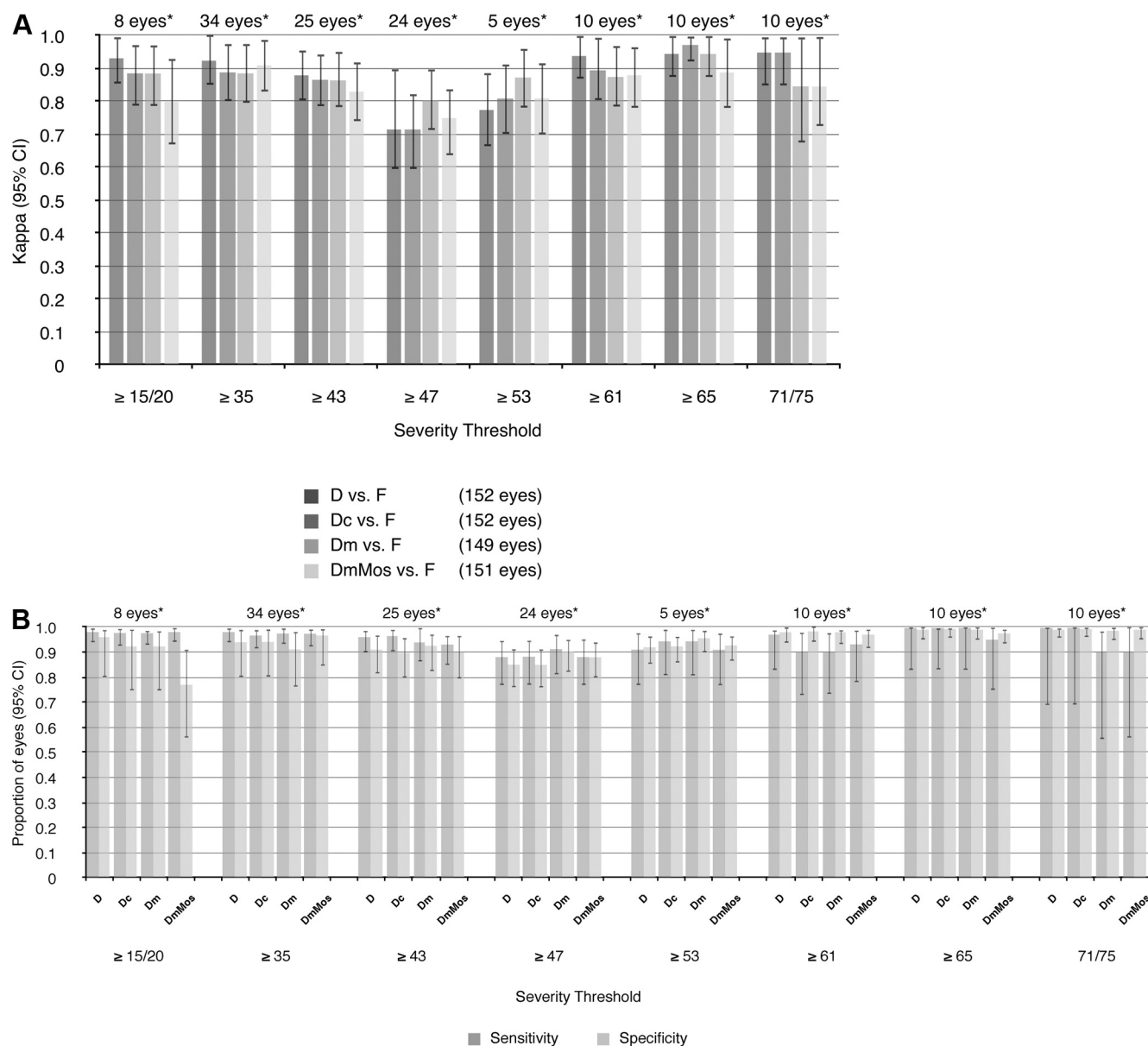
#### B. Uncompressed Digital Formats Compared to Other Digital Formats

	D vs. Dc (n = 152 Eyes)	D vs. Dm (n = 149 Eyes)	D vs. DmMos (n = 151 Eyes)	Dm vs. DmMos (n = 148 Eyes)
Kappa (95% CI)	0.76 (0.68-0.83)	0.66 (0.57-0.74)	0.62 (0.53-0.70)	0.58 (0.49-0.67)
Weighted Kappa (95% CI)	0.92 (0.89-0.95)	0.87 (0.84-0.91)	0.86 (0.82-0.90)	0.82 (0.77-0.87)
Exact agreement, %	78.9	70.7	66.9	63.5
Agreement $\pm$ 1 step	98.7	96.0	97.4	91.9
Agreement $\pm$ 2 steps	100	99.3	98.7	98.0

F, stereoscopic film; D, digital stereoscopic; Dc, digital compressed stereoscopic; Dm, digital monoscopic; DmMos, digital monoscopic mosaic.

\* McNemar test for overall bias shows that readers often evaluated DR severity level as higher in digital formats than in film.

† P-value statistically significant or near significant.



**FIGURE 3.** (A) Diabetic retinopathy severity threshold compared to film by digital format ( $\kappa$ ). (B) Diabetic retinopathy severity threshold compared to film by digital format (sensitivity and specificity). D versus F (152 eyes); Dc versus F (152 eyes); Dm versus F (149 eyes); DmMos versus F (151 eyes). \*Number of eyes with severity threshold in film.

### Uncompressed Stereoscopic Seven Standard Fields (D)

Format D was the most similar to the ETDRS protocol. The only difference between D and F was the recording medium, a digital sensor rather than film. Agreement with film for severity level was substantial. Readers assigned a higher level of severity with D images more often than with F. The bias between D and F was the most significant among all digital formats (D versus F,  $P = 0.02$ , McNemar bias; Table 2A).

### Compressed Stereoscopic Seven Standard Fields (Dc)

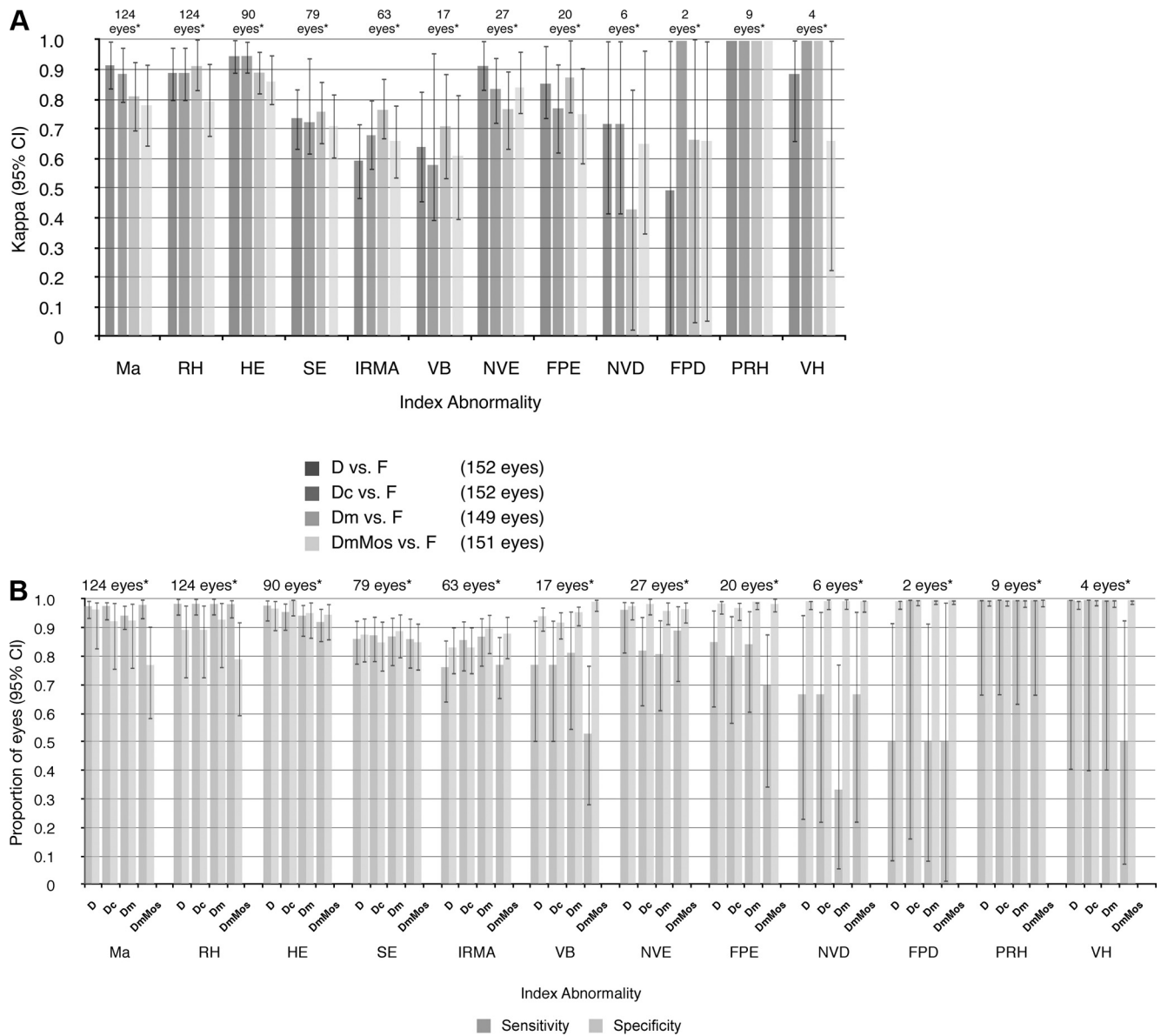
Comparisons of *uncompressed* and *compressed* formats to film for severity level grading yielded similar results (Table 2A). Agreement of D versus Dc was somewhat higher than D versus Dm or DmMos, although no comparison was significantly different (Table 2B).

### Monoscopic Seven Standard Fields (Dm)

Monoscopic agreement compared well to stereoscopic formats versus film (Fig. 2A). Although readers assigned a higher level of severity in Dm than F, the difference was not significant ( $P = 0.18$ , McNemar bias). NVD was present in only six eyes. NVD grading sensitivity in Dm photos was lower than in the stereoscopic and DmMos formats (Fig. 4B), probably because of an oversaturated red channel or suboptimal monoscopic quality.<sup>22</sup>

### Monoscopic Wide-Angle Mosaics (DmMos)

Wide-angle DmMos images were created by combining individual photos into an integrated mosaic. DmMos differed from the film protocol more than any other format. Severity level agreement between DmMos and F was the lowest among the digital formats (Table 2A). DmMos also documented a slightly different retinal region and larger retinal area than F or other digital



**FIGURE 4.** (A) DR index abnormalities compared to film by digital format ( $\kappa$ ). PRH kappa 95% CI, 1.00–1.00 for all format comparisons. (B) DR index abnormality compared to film by digital format (sensitivity and specificity). D versus F (152 eyes); Dc versus F (152 eyes); Dm versus F (149 eyes); DmMos versus F (151 eyes). Ma, microaneurysms; RH, retinal hemorrhages; HE, hard exudates; SE, soft exudates; IRMA, intraretinal microvascular abnormalities; VB, venous beading; NVE, neovascularization elsewhere; FPE, fibrous proliferation elsewhere; NVD, new vessels disc; FPD, fibrous proliferation disc; PRH, preretinal hemorrhage; VH, vitreous hemorrhage. \*Number of eyes with index abnormalities in film.

formats. Table 2B reveals that Dm versus DmMos had lower agreement than stereoscopic formats versus DmMos, suggesting that the stereo effect made little difference (Dm versus DmMos,  $\kappa = 0.58$ ; F versus DmMos,  $\kappa = 0.59$ ; and D versus DmMos,  $\kappa = 0.62$ ). Threshold level  $\geq 15/20$  had lower agreement and specificity (Fig. 3), Ma and RH had lower specificity, and VB and VH had lower sensitivity in DmMos than other formats (Fig. 4B).

### Reproducibility of Interreader Severity Level Grading

Pair-wise interreader agreement within each format was similar (Table 3). Reproducibility of grading within digital formats or film compared favorably to historical ETDRS results.<sup>9</sup> Interreader variability within formats was similar to agreement variability between digital formats versus film, suggesting that human variability accounted for most grading differences.

### Similarities between Digital Formats

Agreement between digital formats and film was lower for extraretinal than intraretinal abnormalities (extraretinal,  $\kappa$  as low as 0.43; intraretinal,  $\kappa = 0.58$ –0.91; Fig. 4A). Sensitivity was  $\geq 86\%$  for intraretinal lesions across all digital formats other than intraretinal microvascular abnormalities (IRMA) and VB (Fig. 4B). Absence or presence of IRMA separates level 35 (mild NPDR) from level 43 (moderately NPDR) and higher. Similarly, VB separates level 43 from level 47 (moderately severe NPDR) and higher. Thus, lower sensitivity of IRMA and VB resulted in lower  $\kappa$  for thresholds  $\geq 47$  and  $\geq 53$  (Fig. 3A). Specificity was also high across formats,  $\geq 85\%$  for almost all intraretinal lesions. Sensitivity for extraretinal abnormalities was  $\geq 70\%$  except NVD and FPD, which had lower agreement and sensitivity across most digital formats (Fig. 4B).

TABLE 3. Interreader Agreement on Diabetic Retinopathy Severity Level by Format

	F (n = 152 Eyes)	D (n = 152 Eyes)	Dc (n = 152 Eyes)	Dm (n = 149 Eyes)	DmMos (n = 151 Eyes)	ETDRS Report 12
Exact Agreement, %						
Reader A vs. B	50.3	56.9	52.0	63.1	46.4	53
Reader A vs. C	53.6	53.6	55.3	61.7	54.3	
Reader B vs. C	63.4	63.4	63.8	74.5	66.9	
Agreement ± 1 step, %						
Reader A vs. B	76.5	83.0	82.2	89.9	88.1	88
Reader A vs. C	83.7	86.3	90.8	88.6	88.7	
Reader B vs. C	93.5	94.8	94.1	95.3	94.7	
Agreement ± 2 steps, %						
Reader A vs. B	92.2	90.8	96.1	97.3	98.0	
Reader A vs. C	91.5	94.1	98.7	97.3	98.7	—
Reader B vs. C	98.0	98.7	98.0	98.0	98.0	
Kappa, 95% CI						
Reader A vs. B	0.43 (0.34–0.52)	0.51 (0.42–0.60)	0.46 (0.37–0.54)	0.58 (0.49–0.66)	0.39 (0.30–0.48)	
Reader A vs. C	0.47 (0.38–0.56)	0.47 (0.39–0.56)	0.49 (0.40–0.58)	0.56 (0.47–0.65)	0.48 (0.39–0.57)	0.42
Reader B vs. C	0.57 (0.48–0.66)	0.57 (0.49–0.66)	0.58 (0.49–0.67)	0.70 (0.62–0.78)	0.61 (0.53–0.70)	
Weighted Kappa, 95% CI						
Reader A vs. B	0.67 (0.59–0.75)	0.73 (0.66–0.80)	0.75 (0.69–0.81)	0.81 (0.76–0.87)	0.76 (0.71–0.81)	
Reader A vs. C	0.71 (0.64–0.78)	0.74 (0.67–0.81)	0.80 (0.75–0.85)	0.80 (0.75–0.86)	0.78 (0.73–0.83)	(0.65)*
Reader B vs. C	0.83 (0.78–0.87)	0.82 (0.77–0.88)	0.83 (0.79–0.88)	0.88 (0.84–0.92)	0.84 (0.80–0.89)	

Abbreviations are as in Table 2.

\* A weighting scheme different from linear was used.

Notably, readers more often assigned a higher level of severity to digital images than to film ( $P = 0.02-0.07$ , McNemar bias, Table 2A).

### Grading Duration

Average time grading an eye using digital formats was longer than using film (Table 4). The average duration of formats D and Dc were 40% longer and Dm, 10% longer. DmMos average grading time was the same as that of Dm.

### DISCUSSION

To our knowledge, this is the first study to compare multiple digital image formats to the ETDRS film protocol for DR evaluation in the same eyes. Results showed that the accuracy in classifying full-scale ETDRS severity level using studied digital formats was comparable to that of 35-mm film. The absence of stereoscopic viewing, the use of 37:1 JPEG 2000 compression, or substituting a wide-angle mosaic for the ETDRS seven standard fields did not compromise assessment of the severity level or threshold. Interreader reliability for all digital formats was similar to that of film. Although agreement suggests that 37:1 compression had less effect on DR severity classification than removing the stereo effect, differences between digital formats were not statistically significant.

Readers more often assigned a higher level of severity using digital formats than film. Other investigators assigned higher levels of severity using film than digital.<sup>32-34</sup> Differences between our results and those in other reports include algorithmic

color balancing and supplementary green-channel viewing, which may have contributed to higher digital severity level grading.

There was lower severity level agreement using DmMos compared with film and other digital formats, perhaps because of the larger DmMos retinal area and slightly different retinal region. These differences may also have contributed to lower Ma and RH specificity and lower  $\geq 15/20$  threshold specificity, compared with film. Digital format differences did not otherwise affect grading agreement. There was wider variability in agreement between digital and film when grading extraretinal than when grading intraretinal lesions, although there was no pattern of variability between formats. Consistent with ETDRS findings, IRMA, VB, NVD, and FPD were demanding lesions regardless of format. Our extraretinal abnormality results may be confounded by the smaller number of eyes and the challenge of any media used to photograph abnormalities in more than one plane.

In population studies or telemedicine programs, severity thresholds and pooled severity categories may be more relevant than discrete severity levels. An epidemiologic study may involve populations with vision-threatening DR (e.g.,  $\geq$  level 53). Threshold information is necessary for planning DR evaluation programs using fundus photographs. A genotype-phenotype linkage study may use three thresholds to analyze phenotypic effect: clearly unaffected (e.g., level  $\leq 20$ ), indeterminate (levels 35–43), and clearly affected (e.g., level  $\geq 47$ ).<sup>35</sup> Table 5 shows a substantial  $\kappa$  for a three-part and a clinical five-part<sup>36</sup> threshold.

Grading DR from digital images took longer than film, because readers could more quickly move a Donaldson viewer among stereo slide pairs than loading digital image files. Large files take time to load, even on fast computers and networks. With half as many files, reviewing monoscopic digital formats took less time than viewing stereoscopic digital. For the same reason, time differences between grading Dm or DmMos versus F were minor compared with D or Dc. No time was saved in reviewing one mosaic image compared with seven fields. There may be a minimum time necessary for readers to examine and classify DR regardless of format. Although the readers had many years' experience in grading seven standard fields

TABLE 4. Grading Duration by Format Compared to Film

	Mean ± SD (median) [range]
F	1.0 ± 0.7 (0.8) [0.2–3.2]
D	1.4 ± 0.9 (1.0) [0.3–3.5]
Dc	1.4 ± 0.8 (0.9) [0.3–3.5]
Dm	1.1 ± 0.6 (0.8) [0.3–3.3]
DmMos	1.1 ± 0.7 (0.8) [0.2–3.0]

Abbreviations are as in Table 2. Film grading = 1 unit of time.

TABLE 5. Agreement on Diabetic Retinopathy Severity Level and Distribution by Various Scales

Scale	D vs. F	Dc vs. F	Dm vs. F	DmMos vs. F
<b>Exact Agreement, % [<math>\kappa</math> (95% CI)]</b>				
3-Part phenotype	83.6 [0.75 (0.65-0.84)]	82.2 [0.73 (0.63-0.82)]	86.6 [0.79 (0.71-0.88)]	84.8 [0.77 (0.68-0.86)]
5-Part clinical	87.5 [0.81 (0.73-0.89)]	84.9 [0.77 (0.69-0.85)]	87.9 [0.81 (0.73-0.89)]	84.1 [0.76 (0.67-0.84)]
9-Step ETDRS	67.8 [0.62 (0.54-0.71)]	65.1 [0.60 (0.51-0.68)]	69.8 [0.65 (0.56-0.73)]	64.9 [0.59 (0.51-0.68)]
<b>Bhapkar (McNemar Bias)</b>				
3-Part phenotype	0.31 (0.16)	0.22 (0.34)	0.56 (0.37)	0.25 (0.83)
5-Part clinical	0.35 (0.11)	0.11 (0.53)	0.48 (0.64)	0.16 (0.22)
9-Step ETDRS	0.21 (0.02)*	0.09 (0.07)*	0.53 (0.18)	0.44 (0.04)*

3-Part phenotype, unaffected ( $\leq 20$ ); indeterminate (level 35-43); clearly affected ( $\geq 47$ ). 5-Part clinical, no diabetic retinopathy (level 10-14); mild nonproliferative diabetic retinopathy (level 15-20); moderate non-proliferative diabetic retinopathy (level 35-47); severe nonproliferative diabetic retinopathy (level 53); proliferative diabetic retinopathy ( $\geq 61$ ). ETDRS 9-Step scale agreement from Table 2 is included for comparison. McNemar test for overall bias shows readers often evaluated DR severity level as present in digital formats higher than film.

\* Statistically significant or nearly significant.

and 35-mm film, they may have been less fluent in grading mosaicked images, particularly with a customized grid simulating a seven-field retinal division.

Because there is a low prevalence of advanced retinopathy in the general population, this study is limited by having a small sample of eyes with level 53 (severe NPDR), NVD, and FPD.

Film has been the basis for diabetic retinal evaluation for many years. Criteria for color film slides in DR studies are well established. There are no widely accepted digital photography standards for acquiring and reviewing DR. Our results suggest that under controlled conditions, compression, absence of stereo effect, or deviation from ETDRS standard fields do not have a negative effect on DR assessment according to the ETDRS scale. Parameters maintained across all digital formats replicated properties of the ETDRS film protocol: resolution high enough to distinguish the smallest DR lesion, color balance similar to film, documentation of retinal regions essential to the ETDRS classification, and sufficient viewing magnification. We also augmented grading color digital images with green-channel views. All studied digital formats were comparable to 35-mm film. These results may be primarily due to the translation of important film protocol characteristics into digital equivalents.

### Acknowledgments

The authors are grateful to staff from the Department of Ophthalmology and Visual Sciences, University of Wisconsin-Madison. From the Ocular Epidemiology Group of Barbara E.K. Klein, MD and Ronald Klein, MD: Andrew F. Ewen, Anne E. Mosher, and Maria K. Swift for grading diabetic retinopathy images, Stacy M. Meuer for grading supervision, and Daniel P. Murach for computer support. From the Fundus Photograph Reading Center directed by Ronald P. Danis, MD: Trina M. Harding for grading orientation, Qian Peng for statistical advice, Jeff T. Klaves for statistical analyses, and Matthew D. Davis, MD, for suggestions regarding this manuscript.

### References

1. Early Treatment Diabetic Retinopathy Study (ETDRS) Manual of Operations. National Technical Information Service; Springfield, VA: 1985. Accession No. PB85223006.
2. Davis MD, Norton WD, Myers FL. Airlie classification of diabetic retinopathy. In: Goldberg MF, Fine SL, eds. *Symposium on the Treatment of Diabetic Retinopathy*. Publication 1890. Arlington, VA: U.S. Department of Health, Education, and Welfare. 1968: 7-22.
3. Diabetic Retinopathy Research Study Group. A modification of the Airlie House classification of diabetic retinopathy. DRS Report Number 7. *Invest Ophthalmol Vis Sci*. 1981;21:210-216.
4. Grading diabetic retinopathy from stereoscopic color fundus photographs: an extension of the modified Airlie House classification. ETDRS report number 10. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology*. 1991;98:786-806.
5. Photocoagulation treatment of proliferative diabetic retinopathy: the second report of diabetic retinopathy study findings. *Ophthalmology*. 1978;85:82-106.
6. Four risk factors for severe visual loss in diabetic retinopathy. The third report from the Diabetic Retinopathy Study. The Diabetic Retinopathy Study Research Group. *Arch Ophthalmol*. 1979;97: 654-655.
7. Photocoagulation for diabetic macular edema. Early Treatment Diabetic Retinopathy Study report number 1. Early Treatment Diabetic Retinopathy Study research group. *Arch Ophthalmol*. 1985;103:1796-1806.
8. Treatment techniques and clinical guidelines for photocoagulation of diabetic macular edema. Early Treatment Diabetic Retinopathy Study Report Number 2. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology*. 1987;94:761-774.
9. Fundus photographic risk factors for progression of diabetic retinopathy. ETDRS report number 12. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology*. 1991;98:823-833.
10. Klein R, Klein BE, Moss SE. How many steps of progression of diabetic retinopathy are meaningful?—the Wisconsin epidemiologic study of diabetic retinopathy. *Arch Ophthalmol*. 2001;119: 547-553.
11. White NH, Sun W, Cleary PA, et al. Prolonged effect of intensive therapy on the risk of retinopathy complications in patients with type 1 diabetes mellitus: 10 years after the Diabetes Control and Complications Trial. *Arch Ophthalmol*. 2008;126:1707-1715.
12. The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med*. 1993;329:977-986.
13. Sorbinil Retinopathy Trial Research Group. A randomized trial of sorbinil, an aldose reductase inhibitor, in diabetic retinopathy. *Arch Ophthalmol* 1990;108:1234-1244.
14. Diabetic Retinopathy Clinical Research Network Web site. Study protocols. Available at <http://drcrnet.jaeb.org/studies.aspx>.
15. Bressler NM, Edwards AR, Beck RW, et al. Exploratory analysis of diabetic retinopathy progression through 3 years in a randomized clinical trial that compares intravitreal triamcinolone acetonide with focal/grid photocoagulation. *Arch Ophthalmol* 2009;127: 1566-1571.
16. Modified 7-Standard Field Color Fundus Photography (7M-F) and Film Fluorescein Angiography (FA-F). University of Wisconsin-Madison Fundus Photograph Reading Center. 2009. Available at: <http://eyephoto.ophth.wisc.edu/Photography/Protocols/Mod7&FA-ver1.pdf>.
17. Hubbard LD, Sun W, Cleary PA, et al. Comparison of digital and film grading of diabetic retinopathy severity in DCCT/EDIC. *Arch Ophthalmol*. 2011;129:718-726.



18. Cavallerano J, Lawrence MG, Zimmer-Galler I, et al. Telehealth practice recommendations for diabetic retinopathy. *Telemed J E Health* 2004;10:469-482.
19. Li HK, Hubbard LD, Danis RP, Florez-Arango JF, Esquivel A, Krupinski EA. Comparison of multiple stereoscopic and monoscopic digital image formats to film for diabetic macular edema evaluation. *Invest Ophthalmol Vis Sci*. 2010;51:6753-6761.
20. Li HK, Hubbard LD, Danis RP, et al. Digital versus film fundus photography for research grading of diabetic retinopathy severity. *Invest Ophthalmol Vis Sci*. 2010;51:5846-5852.
21. Li HK, Florez-Arango JF, Hubbard LD, Esquivel A, Danis RP, Krupinski EA. Grading diabetic retinopathy severity from compressed digital retinal images compared with uncompressed images and film. *Retina*. 2010;30:1651-1661.
22. Li HK, Hubbard LD, Danis RP, Esquivel A, Florez-Arango JF, Krupinski EA. Monoscopic versus stereoscopic retinal photography for grading diabetic retinopathy severity. *Invest Ophthalmol Vis Sci*. 2010;51:3184-3192.
23. Li HK, Esquivel A, Hubbard LA, Florez-Arango JF, Danis RP, Krupinski EA. Mosaics versus ETDRS seven standard fields for evaluation of diabetic retinopathy severity. *Retina*. 2011 Apr 23. [Epub ahead of print].
24. Tyler ME, Hubbard LD, Boydston K, et al. Characteristics of digital fundus camera systems affecting tonal resolution of color retinal images. *J Ophthalmic Photogr*. 2009;31:9-14.
25. Hubbard LD. Digital color fundus image quality: the impact of tonal resolution. *J Ophthalmic Photogr*. 2009;31:15-20.
26. Vitale, TJ. Projecting Digital 'Slide' Images. Available at: <http://cool.conservation-us.org/byauth/vitale/digital-projection/index.html>. 2003. Accessed December 26, 2009.
27. Diabetic Retinopathy Clinical Research Network. DRCRnet Manual of Operations. Version 3.0. Tampa, FL: Jaeb Center for Health Research Foundation; April 8, 2005.
28. Hubbard LD, Danis RP, Neider MW, et al. Brightness, contrast, and color balance of digital versus film retinal images in the age-related eye disease study 2. *Invest Ophthalmol Vis Sci*. 2008;49:3269-3282.
29. Yassur Y, Siegel R, Karp M, Laron Z, Topper E. The detection of early microangiopathy in juvenile diabetics by monochromatic light photography. *Pediatr Adol Endocrinol*. 1980;9:181-184.
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
31. Uebersax JS. User guide for the MH program (vers. 1.2). Statistical Methods for Rater Agreement. 2006. Available at: <http://johnuebersax.com/stat/mh.htm>. Accessed May 1, 2009.
32. Bursell SE, Cavallerano JD, Cavallerano AA, et al. Stereo nonmydriatic digital-video color retinal imaging compared with Early Treatment Diabetic Retinopathy Study seven standard field 35-mm stereo color photos for determining level of diabetic retinopathy. *Ophthalmology*. 2001;108:572-585.
33. Fransen SR, Leonard-Martin TC, Feuer WJ, Hildebrand PL. Clinical evaluation of patients with diabetic retinopathy: accuracy of the Inoveon diabetic retinopathy-3DT system. *Ophthalmology*. 2002;109:595-601.
34. Rudnisky CJ, Tennant MT, Weis E, Ting A, Hinz BJ, Greve MD. Web-based grading of compressed stereoscopic digital photography versus standard slide film photography for the diagnosis of diabetic retinopathy. *Ophthalmology*. 2007;114:1748-1754.
35. Arar NH, Freedman BI, Adler SG, et al. Heritability of the severity of diabetic retinopathy: the FIND-Eye study. *Invest Ophthalmol Vis Sci*. 2008;49:3839-45.
36. American Academy of Ophthalmology Preferred Practice Pattern®. Diabetic Retinopathy. September 2008. <http://one.aao.org/CE/PracticeGuidelines/PPP.aspx?sid=ab789157-5312-4bbe-86ed-8d164ffa9567>. Accessed May 1, 2010.