Comparison of Multiple Stereoscopic and Monoscopic Digital Image Formats to Film for Diabetic Macular Edema Evaluation

Helen K. Li,[1,2] *Larry D. Hubbard,*[3] *Ronald P. Danis,*[3] *Jose F. Florez-Arango,*[1,4] *Adol Esquivel,*[5] *and Elizabeth A. Krupinski*[6]

Purpose. To assess agreement between evaluations of monoscopic and stereoscopic digital images versus stereo film photographs in diabetic macular edema (DME).

Methods. A 152-eye group of digital monoscopic macular images (seven-field sets and wide-angle mosaics) were compared with digital stereoscopic images (uncompressed and compressed seven-field sets) and stereo 35-mm film photos (Early Treatment Diabetic Retinopathy Study protocol) for the presence of hard exudates (HE), retinal thickening (RT), clinically significant macular edema (CSME), and RT at the center of the macular (RTCM).

Results. Agreement, according to the κ statistic, was almost perfect in identifying HE and RT between all digital formats and stereo film (HE, $\kappa = 0.81–0.87$; RT, $\kappa = 0.87–0.92$). Distribution in all digital formats was not significantly different from that in film (Bhapkar test: HE, $P = 0.20–0.40$; RT, $P = 0.06–1.0$). CSME and RTCM grading differences were either significant or trended toward significance. The readers detected CSME and RTCM in film images more often than in digital formats. In identifying DME features, agreement between evaluations of monoscopic digital formats and film was similar to that between stereo digital formats and film, and the performance of uncompressed images versus film was similar to that of compressed images versus film. Repeatability between readers was similar in evaluations of film and all digital formats. Repeatability in identifying RTCM was lower than that of other DME components in film and all digital formats.

Conclusions. Stereoscopic digital formats are equivalent to monoscopic for DME evaluation, but digital photography is not as sensitive as film in detecting CSME and RTCM. (*Invest Ophthalmol Vis Sci.* 2010;51:6753–6761) DOI:10.1167/iovs.10-5504

Diabetic macular edema (DME) is a common cause of reduced vision in diabetic patients.[1] The Wisconsin Epidemiologic Study reported that 50% of diabetic patients had vision <20/40 at 14 years' follow-up, once DME was present, and that 20% had acuity <20/200.[2] Characterizing DME is important for vision prognosis and determining laser treatment. The Early Treatment Diabetic Retinopathy Study (ETDRS) defined edema involving or threatening the macular center as clinically significant macular edema (CSME).[1] These eyes have retinal thickening (RT) in one disc area or more, part of which is within 1 disc diameter of the macular center, or RT or hard exudates (HE) adjacent to RT within 500 μm of the macular center.[1] There is a 33% vision loss risk of three or more ETDRS chart lines at 3 years when the macular center is involved.[1] The prognosis is somewhat better when the macular center is spared, ranging from 17% to 22%, depending on whether the edema qualifies as CSME.[1] Risk is also associated with DME severity, determined by the area of RT and the degree of thickening at the macular center.[3] The rate of visual acuity loss ranges from 1 to 17 ETDRS letters per year, depending on whether the DME is mild or very severe.[4] Focal laser photocoagulation increases the chances of visual improvement and lowers the risk of moderate visual loss[1] by reducing the severity and duration of the edema.[4] CSME eyes with large areas of edema within 1 disc diameter of the macular center and/or greater degrees of edema at the macular center benefit most from treatment.[3]

DME is assessed by identifying HE, presence and area of RT, proximity of RT or HE to the macular center, and presence and degree of RT at the macular center. Although slit lamp biomicroscopy is standard clinical practice, the ETDRS stereo film photography protocol has been used in research studies for almost 30 years.[5] Good-quality photographs provide information on the location and the severity of HE and RT comparable to contact lens slit lamp biomicroscopy.[6] Optical coherence tomography (OCT) is now supplanting photography in clinical research, with more accurate and reproducible retinal thickness measurements. Stereophotography is still used for documenting baseline status in new studies and for follow-up studies begun before OCT. Instead of film, however, digital photography is now used, allowing stereographic diabetic retinopathy (DR) evaluation in telemedicine.[7–9]

The performance of digital images versus ETDRS slide film in assessing DME has not been thoroughly studied. Most stereo digital-to-stereo film comparisons are from studies of customized telemedicine programs. Limitations of existing reports include: a small number of DME eyes,[10] agreement per patient

From the [1]School of Biomedical Informatics, University of Texas Health Science Center, Houston, Texas; the [2]Department of Ophthalmology, Thomas Jefferson University, Philadelphia, Pennsylvania; the [3]Department of Ophthalmology and Visual Sciences, University of Wisconsin-Madison, Madison, Wisconsin; [4]Universidad De Antioquia, Medellin, Columbia; the [5]Houston Veterans Affairs Medical Center and Baylor College of Medicine, Houston, Texas; and the [6]Department of Radiology, University of Arizona, Tucson, Arizona.

Presented at the annual meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, Florida, May, 2009.

Supported by a grant from Juvenile Diabetes Foundation Research International, New York, NY (HKL) and by unrestricted grants from the 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 March 9, 2010; accepted June 14, 2010.

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

Corresponding author: Helen K. Li, 1400 Hermann Drive, Houston, TX 77004; hlimed@mac.com.

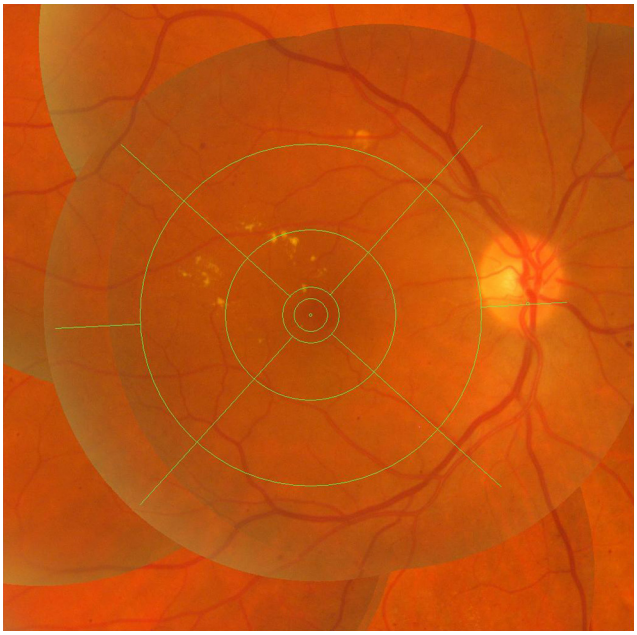

FIGURE 1. Scaled ETDRS grid superimposed on the macula.

instead of per eye,[8] compressed but not uncompressed digital images compared with film,[9] and no grading retest to measure reliability.[9–11]

Some have questioned the need for stereophotography, noting better patient convenience, time efficiency, and cost savings in digital monoscopic photography. New imaging protocols should be examined to affirm that the digital photography assessment of DME, whether stereo- or monoscopic, offers the same reliability as the ETDRS stereoscopic film standard. We assessed the efficacy of digital images versus film in evaluating DME by comparing the grading of DME features in uncompressed stereo, compressed stereo, monoscopic, and monoscopic wide-angle mosaic images with that in film images. We also compared the reproducibility of grading within each medium.

MATERIAL AND METHODS

Study Design

We compared color digital photographs with ETDRS 35-mm slide film for the evaluation of DME features. This study is one part of a comprehensive evaluation of multiple digital photography formats compared with ETDRS film. Other reports describe each digital format's performance compared with that of film in evaluating the severity of diabetic retinopathy.

Fundus Photographs

Patients from The University of Texas Medical Branch Department of Ophthalmology and Visual Sciences Eye Clinic gave written consent for eye photography. Institutional review board approval was obtained, and the tenets of the Declaration of Helsinki were observed. Patients with a broad range of severity of DR were informally screened in the clinic. Those with the severity levels needed for the study were invited for photography. Patients with media opacities or limited pupil dilation preventing an adequate retina image, with other retinal vascular disease, or with previous retinal laser photocoagulation were excluded, to avoid possible grader bias in assessing macular edema or proliferative retinopathy. Each study eye had 16 slides and 16 digital photographs taken per the ETDRS protocol: seven nonsimultaneous color stereo field pairs of the fundus and one pair of the anterior segment. Film and digital photographs were taken with the same mydriatic camera (TRC-50EX/IX; Topcon Medical Systems, Paramus, NJ) at its 35° setting. Fundus camera optics was 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) with a digital sensor (model D100; Nikon, Melville, NY) at its 45° setting for the mosaic format. ETDRS utilizes seven fields (IMAGEnet 2000, ver. 2.55; Topcon Medical System), and nonmydriatic camera mosaic software uses nine fields: eight peripheral fields surrounding one central field of the posterior pole. Mydriatic camera resolution was 2400 × 2000 pixels and nonmydriatic was 3000 × 2000 pixels.

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 camera in the same sequence. The slide film (Ektachrome) was processed at a certified facility (Kodak certified Q-Laboratory; Eastman Kodak, Rochester, NY).[12]

A total of 152 eyes from an 85-patient cohort was selected to represent a stratified sample across the full range of DR severity levels. The patients included 32 (37.6%) men, 53 (62.4%) women, 37 (43.5%) Caucasians, 24 (28.2%) Hispanics, and 24 (28.2%) African Americans. Seventy-five right-eye and 77 left-eye images were used. Photographs included both eyes of 67 (78.8%) patients, only the right of 8 (9.4%), and only the left of 10 (11.8%). The patients ranged in age from 33 to 83 years, with a median of 60.5 and a mean of 59.4 years. The photographs were coded to remove the identifying patient information.

The color and contrast of the digital images were adjusted to conform to a formal color model based on standard ETDRS slides. Custom software generated red/green/blue (RGB) luminance histograms and adjusted each color channel curve to fit model parameters. This process maximized the contrast of DR abnormalities against retinal pigment epithelial (RPE) backgrounds without creating artifacts. The algorithm was modified from one used in the Age Related Eye Disease Study 2 (AREDS2).[13]

The 152 digital stereo pair image sets were processed to create other digital formats. One set was JPEG2000 compressed to 37:1

TABLE 1. Graded DME Features and Scales

HE within Grid*	RT within Grid*	CSME	RTMC
(0) No	(0) No	(0) No	(0) No
(1) Questionable	(1) Questionable	(1) Questionable	(1) Questionable
(2) Definite	(2) Outside the grid	(2) RT ≥1 DA, part ≤1 DD to center	(2) <1× reference thickness
(8) Cannot grade	(3) Inside the grid	(3) RT or adjacent HE <500 μm to center	(3) ≥1×, <2× reference thickness
	(8) Cannot grade	(8) Cannot grade	(4) ≥2× reference thickness
			(8) Cannot grade

DA, disc area; DD, disc diameter.
* Within two disc diameters of the macular center.

(ver. 3.98; IrfanView, Wierner Neustadt, Austria, using the LuraWave JP2 plug-in; LuraTech, Inc., San Jose, CA). The compression algorithm and ratio are discussed in another report.[14] Monochromatic green channel images generally provide better contrast of small lesions than does full color photography.[15] These were created by extracting the green channel from digital RGB color images and were routinely used to confirm suspected subtle DR lesions.

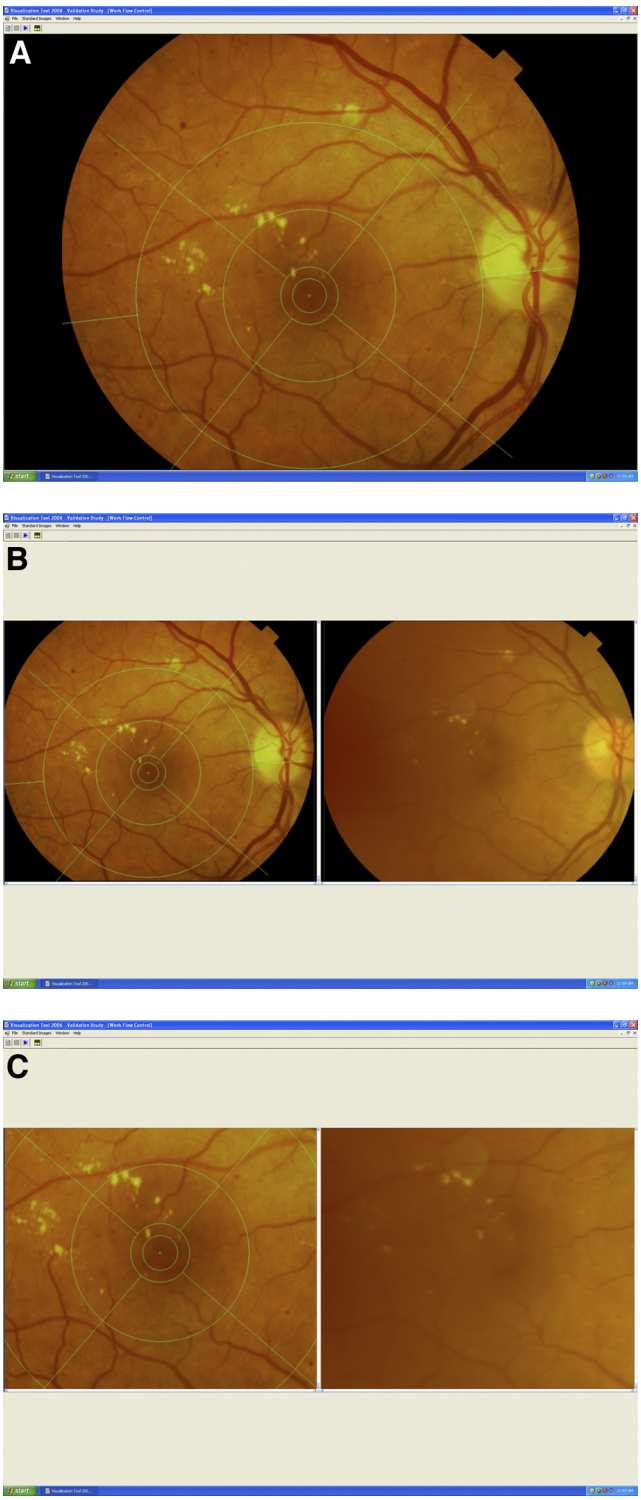

FIGURE 2. A 35° digital image displayed monoscopically at 13× magnification (**A**), stereoscopically at 6.5× (**B**), and stereoscopically zoomed to approximately 13× (**C**).

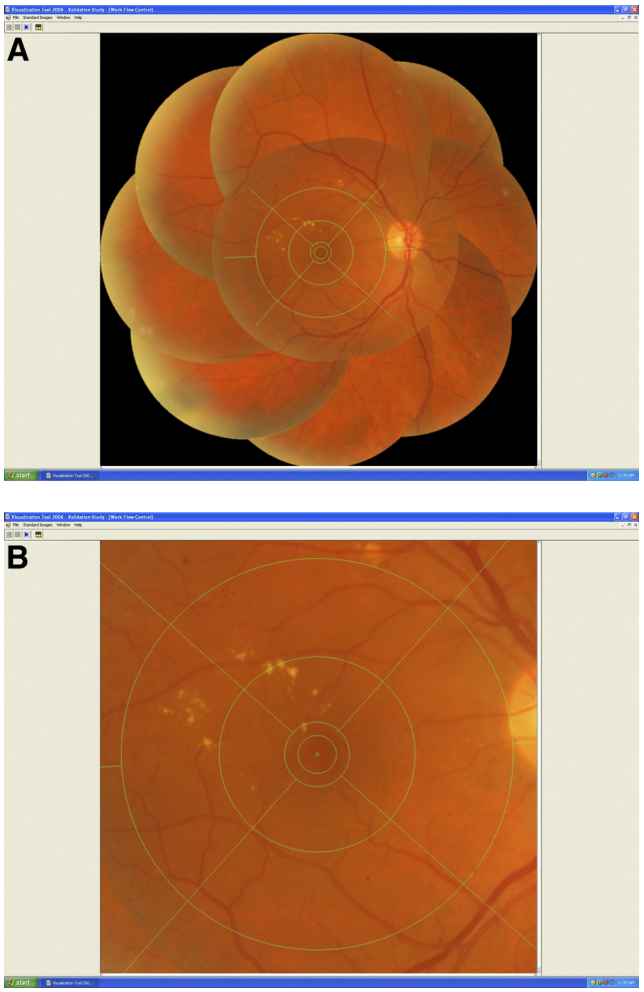

FIGURE 3. Mosaicked digital image displayed at 5.5× (**A**) and zoomed to approximately 13× (**B**).

The left or right eye image from each stereo pair exhibiting the best focus and contrast was selected to create the color and green-channel monoscopic formats.

Grading Photographs

Three readers with a minimum of 10 years' experience from the University of Wisconsin, Madison, independently evaluated all images in each format. To minimize recall, evaluation was regulated by custom scheduling software that separated grading of the same eye by at least 2 weeks. Batches of different formats were counterbalanced in the order of presentation.

Readers graded DME abnormalities per ETDRS field 2, supplemented by oblique views from fields 1 and 3, from the 35° photo sets.[16] For each format, including mosaic, appropriately scaled ETDRS macular grids were superimposed on the macular view to demarcate the macular subfields (Fig. 1). Four DME features were graded according to a modified ETDRS classification: HE within the macular grid (HE), retinal thickening within the grid (RT), ETDRS CSME, and retinal thickening at the center of the macula (RTCM; Table 1).[16] The presence and severity of the abnormality were determined according to ETDRS definitions.[16] Each abnormality included in the ETDRS retinopathy severity classification was graded in relevant photographic fields and entered into a computerized grading form.

Stereo sets of slides were graded per ETDRS report 10.[16] Film sets were viewed on daylight fluorescent light boxes with a Donaldson 5×

Table 2. DME Features Assigned from Grading Digital Formats Compared with Stereo Film (# eyes)

		HE within Grid			RT within Grid			CSME			RTMC		
		Absent	**Present**	**Total**	**Absent**	**Present**	**Total**	**Absent**	**Present**	**Total**	**Absent**	**Present**	**Total**
						Digital Stereoscopic							
	Absent	67	5	72	87	3	90	103	1	104	103	8	111
	Present	8	72	80	4	58	62	8	40	48	15	26	41
	Total	75	77	152	91	61	152	111	41	152	118	34	152
						Digital Compressed Stereoscopic							
	Absent	68	4	72	87	3	90	102	2	104	105	6	111
	Present	7	73	80	3	59	62	10	38	48	14	27	41
	Total	75	77152	90	62	152	112	40	152	119	33	152	
Film													
						Digital Monoscopic							
	Absent	67	5	72	84	5	89	101	2	103	104	4	108
	Present	9	68	77	4	56	60	8	38	46	15	26	41
	Total	76	73	149	88	61	149	109	40	149	119	30	149
						Digital Monoscopic Mosaic							
	Absent	68	3	71	88	1	89	102	1	103	107	4	111
	Present	7	73	80	6	56	62	10	38	48	15	25	40
	Total	75	76	151	94	57	151	112	39	151	122	29	151

stereo viewer (George Davco, Holcombe, MA). Magnification was approximately 12.5×, accounting for the combined magnification of the Topcon camera and the Donaldson viewer.

Readers first reviewed a proof sheet of digital thumbnails. Fields were then reviewed monoscopically at full screen with 13× magnification, approximating slide magnification review (Fig. 2A). For stereoscopic grading, the images were viewed side by side in full color, first in fit-to-screen mode at 6.5× magnification (Fig. 2B), then zoomed to approximately 13× (Fig. 2C). When examining the mosaicked set of 45° images, the readers first viewed fit-to-screen images at 5.5× magnification (Fig. 3A), then zoomed to approximately a 13× magnification of the 35° individual-field images (Fig. 3B).

Customized software facilitated viewing of the digital images at approximately 26 in. from a 21-in. CRT display (model G225F; ViewSonic Corp., Walnut, CA) at 1600 × 1200-pixel resolution. The monitors were set at a color temperature of 6500°K and 2.2 gamma and checked monthly (Greytag Macbeth; X-Rite Inc.; Grand Rapids, MI). Stereo digital images were viewed with a handheld viewer (Screen-Vu; Eye Supply, Tampa, FL).

Statistical Analysis

DME features in each eye were summarized as the central tendency (median) among the three graders. Instead of duplicate grading with adjudication of differences, this multigrader method allowed pair-wise comparisons of all readers within each format.

The presence or absence of DME features in the images was compared, with stereo film grading results serving as the reference standard. For HE, RT, and CSME, the results for questionable grades were combined with the absent results. For RTCM, questionable grades were combined with grades signifying definite presence, in accordance with ETDRS.

Eyes with photographs classified as ungradable were excluded from the analyses. Format comparisons were made by using the common subset of eyes gradable in both formats. Three eyes were ungradable in the monoscopic format, and one from the nonmydriatic camera had incomplete photo sets. Analyses included uncompressed and compressed stereoscopic digital images (152 eyes, each set), monoscopic digital images (149 eyes), and monoscopic wide-angle mosaic digital images (151 eyes).

Agreement was cross-tabulated and the κ statistic calculated. Based on Landis and Koch[17] ranges used in ETDRS Report 10, 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 high to perfect agreement. The McNemar test for dichotomized scales and the Bhapkar test for multistep scales were used

Table 3. DME Feature Distribution: Digital Formats Compared with Stereo Film

	D vs. F	**Dc vs. F**	**Dm vs. F**	**Dmm vs. F**
HE within grid				
Bhapkar *P*-value	0.4043	0.3644	0.2832	0.2035
McNemar bias *P*-value	0.4054	0.3657	0.2850	0.2059
RT within grid				
Bhapkar *P*-value	0.7053	1.0000	0.7388	0.0588*
McNemar bias *P*-value	1.0000	1.0000	1.0000	0.1250
CSME				
Bhapkar *P*-value	0.0175*	0.0187*	0.0548*	0.0054*
McNemar bias *P*-value	0.0391*	0.0209*	0.0578*	0.0067*
RTMC				
Bhapkar *P*-value	0.1416	0.0706*	0.0099*	0.0099*
McNemar bias *P*-value	0.1444	0.0736*	0.0116*	0.0116*

McNemar test for overall bias shows that readers evaluated DME features as present in film more often than in digital formats. F, stereoscopic film; D, digital stereoscopic (n = 152 eyes); Dc, digital compressed stereoscopic (n = 152 eyes); Dm, digital monoscopic (n = 149 eyes); Dmm, digital monoscopic mosaic (n = 151 eyes).

* *P*-value statistically significant or near statistically significant.

TABLE 4. DME Features: Digital Formats Compared with Stereo Film

Features	Sensitivity	Sensitivity 95% CI	Specificity	Specificity 95% CI	*n* Eyes with Features	Positive Predictive Value	Negative Predictive Value	Rate of Agreement	κ	κ 95% CI
					HE within Grid					
D vs. F	0.90	0.81–0.96	0.93	0.85–0.98	80	0.94	0.89	0.91	0.83	0.74–0.92
Dc vs. F	0.91	0.83–0.96	0.94	0.86–0.98	80	0.95	0.91	0.93	0.86	0.77–0.94
Dm vs. F	0.88	0.79–0.95	0.93	0.85–0.98	77	0.93	0.88	0.91	0.81	0.72–0.91
Dmm vs. F	0.91	0.83–0.96	0.96	0.88–0.99	80	0.96	0.91	0.93	0.87	0.79–0.95
					RT within Grid					
D vs. F	0.94	0.84–0.98	0.97	0.91–0.99	62	0.95	0.96	0.95	0.90	0.84–0.97
Dc vs. F	0.95	0.86–0.99	0.97	0.91–0.99	62	0.95	0.97	0.96	0.92	0.85–0.98
Dm vs. F	0.93	0.84–0.98	0.94	0.87–0.98	60	0.92	0.95	0.94	0.87	0.80–0.95
Dmm vs. F	0.90	0.80–0.96	0.99	0.94–1.00	62	0.98	0.94	0.95	0.90	0.83–0.97
					CSME					
D vs. F	0.83	0.70–0.93	0.99	0.95–1.00	48	0.98	0.93	0.94	0.86	0.77–0.95
Dc vs. F	0.79	0.65–0.90	0.98	0.93–1.00	48	0.95	0.91	0.92	0.81	0.71–0.91
Dm vs. F	0.83	0.69–0.92	0.98	0.93–1.00	46	0.95	0.93	0.93	0.84	0.74–0.93
Dmm vs. F	0.79	0.65–0.90	0.99	0.95–1.00	48	0.97	0.91	0.93	0.82	0.72–0.92
					RTMC					
D vs. F	0.63	0.47–0.78	0.93	0.86–0.97	41	0.77	0.87	0.85	0.59	0.45–0.74
Dc vs. F	0.66	0.49–0.80	0.95	0.89–0.98	41	0.82	0.88	0.87	0.64	0.50–0.79
Dm vs. F	0.63	0.47–0.78	0.96	0.94–1.00	41	0.87	0.87	0.87	0.65	0.51–0.79
Dmm vs. F	0.63	0.46–0.77	0.96	0.91–0.99	40	0.86	0.86	0.87	0.65	0.50–0.79

Abbreviations and number of eyes are as in Table 3.

to test for significant differences (cross-tab marginal homogeneity). Marginal distributions were assessed with McNemar's test of overall bias.

DME grading comparisons were tested using the κ statistic, sensitivity and specificity percentages, positive and negative predictive values, and disease prevalence percentages. Agreements between monoscopic digital and stereoscopic digital grading and within each format between readers were assessed by percentage of agreement and the κ statistic.

Statistical analyses were performed with commercial software (Analyze-It, Ltd., Leeds, UK, and MedCalc, MedCalc Software, Bvba, Mari-

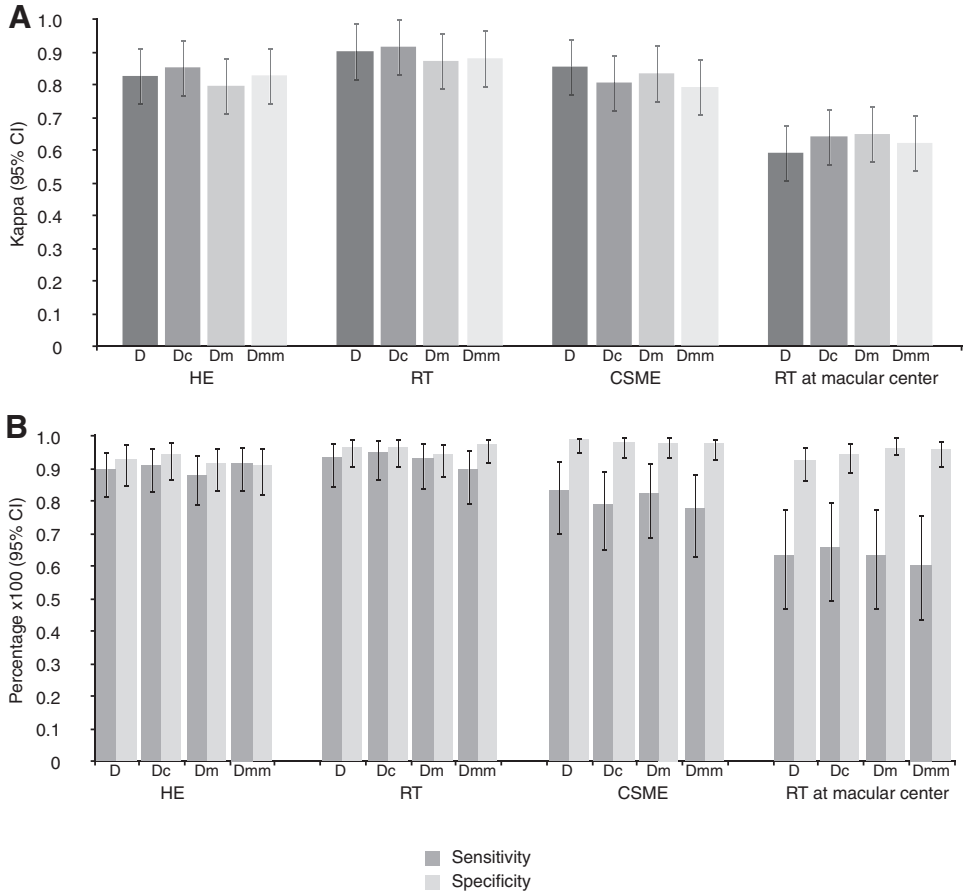

FIGURE 4. Diabetic macular edema features, digital formats compared to stereo film. Kappa (**A**); sensitivity and specificity (**B**). HE, hard exudate within grid; RT, RT within grid; CSME, clinically significant macular edema; D, digital stereoscopic (n = 152 eyes); Dc, digital compressed stereoscopic (n = 152 eyes); Dm, digital monoscopic (n = 149 eyes); and Dmm, digital monoscopic mosaic (n = 151 eyes).

TABLE 5. Interreader DME Feature Agreement by Format

	F	D	Dc	Dm	Dmm
			Agreement (%)		
			HE within Grid		
Reader A vs. B	79.6	83.6	89.5	89.3	86.8
Reader A vs. C	83.6	82.9	92.1	90.6	84.8
Reader B vs. C	92.1	95.4	89.5	91.9	90.1
			RT within Grid		
Reader A vs. B	87.5	88.0	87.5	86.6	90.7
Reader A vs. C	88.2	90.8	88.2	91.3	90.1
Reader B vs. C	91.4	91.4	91.4	94.0	94.0
			CSME		
Reader A vs. B	85.5	86.2	90.1	88.6	88.1
Reader A vs. C	86.8	89.5	87.5	90.6	90.1
Reader B vs. C	93.4	92.8	89.5	89.9	91.4
			RTMC		
Reader A vs. B	82.9	86.8	84.2	83.9	83.4
Reader A vs. C	85.5	88.2	82.9	85.9	80.1
Reader B vs. C	88.2	89.5	85.5	89.9	90.1
			κ (95% CI)		
	F	**D**	**Dc**	**Dm**	**Dmm**
			HE within Grid		
Reader A vs. B	0.60 (0.47–0.72)	0.67 (0.56–0.79)	0.79 (0.70–0.89)	0.79 (0.69–0.88)	0.74 (0.63–0.84)
Reader A vs. C	0.67 (0.55–0.79)	0.66 (0.54–0.78)	0.84 (0.76–0.93)	0.81 (0.72–0.91)	0.70 (0.59–0.81)
Reader B vs. C	0.84 (0.76–0.93)	0.91 (0.84–0.97)	0.79 (0.69–0.89)	0.84 (0.75–0.93)	0.80 (0.71–0.90)
			RT within Grid		
Reader A vs. B	0.74 (0.64–0.85)	0.76 (0.66–0.87)	0.74 (0.64–0.85)	0.71 (0.60–0.83)	0.80 (0.70–0.90)
Reader A vs. C	0.76 (0.65–0.86)	0.81 (0.71–0.90)	0.76 (0.65–0.86)	0.81 (0.72–0.91)	0.79 (0.68–0.89)
Reader B vs. C	0.82 (0.73–0.91)	0.82 (0.73–0.91)	0.82 (0.73–0.91)	0.87 (0.79–0.95)	0.87 (0.79–0.95)
			CSME		
Reader A vs. B	0.68 (0.56–0.80)	0.65 (0.51–0.78)	0.76 (0.65–0.88)	0.71 (0.59–0.84)	0.70 (0.57–0.82)
Reader A vs. C	0.70 (0.58–0.82)	0.73 (0.61–0.85)	0.69 (0.57–0.82)	0.74 (0.62–0.87)	0.72 (0.59–0.85)
Reader B vs. C	0.85 (0.76–0.94)	0.83 (0.73–0.92)	0.75 (0.63–0.86)	0.76 (0.64–0.87)	0.79 (0.68–0.90)
			RTMC		
Reader A vs. B	0.57 (0.43–0.72)	0.63 (0.49–0.78)	0.58 (0.44–0.72)	0.54 (0.38–0.70)	0.55 (0.41–0.70)
Reader A vs. C	0.64 (0.50–0.78)	0.68 (0.54–0.82)	0.57 (0.43–0.72)	0.62 (0.48–0.77)	0.47 (0.32–0.62)
Reader B vs. C	0.68 (0.55–0.82)	0.68 (0.53–0.82)	0.57 (0.41–0.73)	0.68 (0.52–0.83)	0.63 (0.47–0.80)

Abbreviations and number of eyes are as in Table 3.

akerke, Belgium) and a noncommercial program (the MH [marginal homogeneity] Program, ver. 1.2).[18]

RESULTS

Digital versus Film

Table 2 presents the cross-tabulation for the presence or absence of each DME feature. Film grading of 152 eyes found HE in 80 (52.6%) eyes; RT in 62 (40.8%) eyes; CSME in 48 (31.6%) eyes; and RTCM in 41 (27.0%) eyes.

TABLE 6. Unanimous DME Feature Presence/Absence Agreement among All Three Readers

	F	D	Dc	Dm	Dmm
HE within grid	77.6	86.2	85.5	85.9	80.8
RT within grid	83.6	85.5	85.5	85.9	87.4
CSME	82.9	82.9	83.6	84.6	74.8
RTMC	77.6	82.2	76.3	79.9	83.4

Data are percentage of eyes. Abbreviations and number of eyes are as in Table 3.

Bhapkar and McNemar bias tests determined whether the DME gradings of digital formats were significantly different from those with film (Table 3). The HE and RT results were not significantly different between the different media. The differences in recognizing CSME between all digital formats and film were significant or trended toward significance.

For RTCM, the differences were significant for all digital formats except uncompressed stereoscopic photos, which showed a nonsignificant trend ($P = 0.14$ Bhapkar, $P = 0.14$ McNemar). The McNemar test showed that readers identified CSME and RTCM with film more than with digital formats (Table 3).

The κ between digital formats and film for HE and RT was almost perfect (Table 4, Fig. 4A). Sensitivity and specificity were mostly 90% or higher in each digital format compared with those of film (Table 4, Fig. 4B).

For CSME, κ between digital and film remained high ($\kappa = 0.81–0.86$). However, systematic differences between digital and film resulted in lower sensitivity with all digital formats (79%–83%). For RTCM, κ was markedly lower for digital images ($\kappa = 0.59–0.65$), and the sensitivity of all digital formats was substantially lower (63%–66%; Table 4, Fig. 4B). The specificities for recognizing CSME and RTCM was high (≥90%)

TABLE 7. HE and RT: Digital Formats Compared with Stereo Film

	F	D	Dc	Dm	Dmm
HE within grid	80	77	77	73	76
No RT within grid	18 (22.5)	16 (20.8)	15 (19.5)	12 (16.4)	19 (25.0)
No CSME	32 (40.0)	36 (46.8)	37 (48.1)	33 (45.2)	37 (48.7)
No RTMC	39 (48.8)	43 (55.8)	44 (57.1)	43 (58.9)	48 (63.2)

Data are number of eyes (%). Abbreviations and total eyes in the groups are as in Table 3.

for all digital formats (Table 4). All digital formats were somewhat less sensitive than film for CSME and were least sensitive for RTCM.

Monoscopic and Stereoscopic Digital Formats versus Film

For all DME features, agreement between *monoscopic* digital formats (seven fields and mosaic) and film was similar to that between *stereoscopic* digital formats (uncompressed seven fields and compressed seven fields) and film (Table 4). Comparison between monoscopic and stereoscopic *digital* formats showed good agreement: HE, $\kappa = 0.83$ (95% CI, 0.73–0.92); RT, $\kappa = 0.92$ (95% CI, 0.85–0.98); CSME, $\kappa = 0.90$ (95% CI, 0.82–0.98); and RTCM, $\kappa = 0.78$ (95% CI, 0.65–0.90). Agreement of the assessments of stereo digital images was comparable to those of stereo film for HE and RT, but trended lower for CSME and especially for RTCM.

Intergrader Agreement

Pair-wise intergrader agreement on digital format assessments was similar to that with film (Table 5). Agreement on monoscopic formats was also similar to that on stereoscopic digital images and stereo film. Pair-wise intergrader repeatability on compressed or uncompressed digital images was similar. RTCM had the lowest agreement of all DME features in all digital formats and film ($\kappa = 0.47$–0.68; Table 5).

The proportion of eyes for which all three graders agreed on the presence or absence of each DME feature was similar (~80%) between all digital formats and film (Table 6). The rate of unanimity on assessments of monoscopic formats also compared well with that of stereo photograph evaluations. Compared with other DME features, RTCM showed an insignificant trend toward lower unanimity with all digital formats and film (Table 6). If we had not used the ETDRS method of combining eyes with questionable thickening in the macular center and eyes with definite thickening for analysis, intergrader agreement for RTCM would have been similar to that for other features (data not shown).

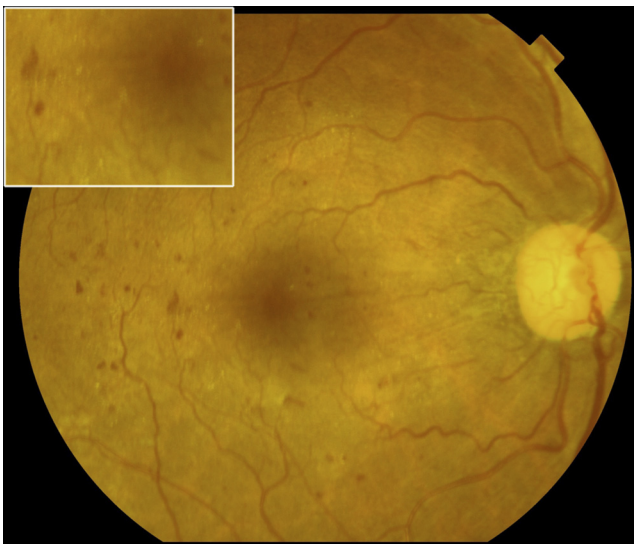

FIGURE 5. Small flecks of HE.

HE versus Other DME Features

Table 7 shows the number of eyes graded as having HE and the percentage of those as having no RT. The results confirm that the readers did not consider HE to be definitive evidence of RT.

As seen in Table 6, the graders were unanimous in the presence or absence of HE in >80% of all eyes. Despite this high level of intergrader agreement within the media, there were disagreements: eight eyes identified as having HE on film but not on uncompressed digital images and five eyes with HE in uncompressed digital but not on film. Because HE is a feature that arguably depends less on stereo imaging than does RT, an experienced reader uninvolved in the study grading reviewed these digital and film photographs side by side, to characterize the nature of the disagreements. In 3 eyes, the abnormalities were judged to be drusen, not HE, and in the other 10, there were only one or two small flecks of HE, smaller than a large druse (>125 μm diameter; Fig. 5).

Of the 13 eyes with HE disagreement, RT was identified in 2. In 1 eye, HE but not RT was recognized on film, whereas RT not HE was found on the uncompressed digital image. In the other eye, CSME with subtle central macular involvement and one small speck of HE were identified on film. These abnormalities were not detected in the uncompressed digital image.

Cannot Grade

RTCM had the highest number of eyes assigned cannot grade, and HE had the fewest (Table 8). No eyes were deemed cannot grade on film. Compressed images had the most cannot grades and the uncompressed ones had the fewest.

DISCUSSION

To the best of our knowledge, this is the only study in which uncompressed stereo, compressed stereo, monoscopic, and monoscopic wide-angle mosaicked digital images have been compared to ETDRS film for the evaluation of four DME fea-

TABLE 8. Eyes Graded Cannot Grade: Digital Formats Compared with Stereo Film

	F	D	Dc	Dm	Dmm
HE within grid	0	0	1	0	0
RT within grid	0	0	2	0	0
CSME	0	0	2	1	0
RTMC	0	2	3	3	3

Data are number of eyes. Abbreviations and number of eyes are as in Table 3.

tures. Uncompressed stereoscopic, monoscopic, monoscopic wide-angle mosaic, and compressed stereo digital images were comparable to ETDRS stereoscopic slides for detecting HE and RT. The studied formats were not, however, comparable to film for detecting CSME and RTCM. There was also no significant difference in detecting DME features using monoscopic images versus stereo. This result may be due to the readers' use of subtle retinal color changes as cues for edema in monoscopic images. Disagreements appeared attributable more to film versus digital differences than to the presence or absence of stereo effect.

The agreement of the compressed images with film was similar to that of uncompressed images with film. The reproducibility of detecting most DME features was similar between digital, film, monoscopic, stereo, uncompressed images, and compressed images. The reproducibility of grading RTCM, however, was less than that for other DME features in film and all digital formats.

Sharp focus and stereo quality are important for viewing RT borders. Readers look for changes in the plane of the perifoveal vessels. HE unusually elevated above the RPE plane in the outer retina is a cue. A normal RPE appears granular; thus, RPE that appears blurred when viewed through an edematous retina is a telltale RT sign. RTCM is more challenging than CSME, because it is not easy to appreciate the foveal avascular retinal surface without vessels for reference. DME's small area and amorphous shape contribute to ambiguous stereoscopic identification, because smooth objects are harder to perceive than structures with clearly defined angles.[19]

We found significant differences between all digital formats and film grading of CSME. A previous study compared compressed images (16:1 JPEG stereo 3040 × 2008) and film for CSME, reporting 87.2% sensitivity, 92.8% specificity, κ of 0.8, and McNemar $P = 0.65$.[9] Those investigators applied a lower threshold for CSME than we used in our study, and CSME was graded present if a reader was at least 50% sure. We used the ETDRS classification definitions, assigning a grade of questionable when a reader was 50% to 90% certain of CSME presence and a grade of present at 90% certainty or higher.

HE had the fewest eyes assigned cannot grade (Table 8). Grading HE depends least on stereo viewing or discriminating fine detail. Fine detail and stereo are, however, necessary for determining RTCM. Accounting for optics and camera sensors, our mydriatic system pixel spacing was 4.6 μm/pixel and the nonmydriatic 6.2 μm/pixel. Assuming that Kodak Ektachrome 100 film represents 2290 pixels per inch,[20] then the film's equivalent pixel spacing is approximately 4.6 μm/pixel. Grading CSME required viewing two images fit to the screen to locate RT in reference to the macular center. This limited the monitors' effective resolution to approximately 21.6 μm/pixel. Although our digital camera sensors were five (mydriatic) and six (nonmydriatic) megapixels, we believe using any conventional monitor could be a limiting factor, even for images acquired by a 10-megapixel camera.

Among eyes with disagreement between uncompressed images and film, tiny flecks of HE appeared slightly better on film. Subsequent comparison using a wide-screen monitor (FlexScan S2242W; Eizo Nanao Corp., Ishikawa, Japan) revealed better display of subtle flecks of HE. Its 1.6 aspect ratio (compared to our monitor's 1.3) resulted in 50% more pixels for stereo viewing. This wide-screen monitor came on the market after our study was completed and is now offered on several digital fundus camera systems. The color medical-grade monitors used in radiology may also be beneficial. Future studies of DME with high-resolution, wide-screen monitors may determine their ultimate value for detection of DME features.

Uncompressed digital images had the fewest instances of features assigned cannot grade (Table 8). Compressed digital images had the highest, with monoscopic formats somewhere in the middle. Although compressed digital images fared the worst in cannot grade eyes, the number of cannot grade eyes between formats was not statistically significant. In another study, a κ of 0.77 was found for CSME JPEG images compressed 55:1 compared with uncompressed 3040 × 2008 TIFF images and a κ of 0.89 for JPEG images compressed 113:1.[21] However, that study had only five eyes with CSME.

Monoscopic images in our study were not acquired as single photographs. Instead, we used the left or right image of a stereo pair that exhibited the best focus and contrast. To achieve an adequate stereo base, we took the first photograph of a stereo pair as far to one side of a pupil as possible without introducing shading or edge artifacts. This limitation accounted for three of the ungradable eyes in the monoscopic format excluded from analysis. It is possible that suboptimal monoscopic image quality also accounted for one CSME eye and three RTCM eyes rated as cannot grade (Table 8). Our monoscopic format image quality would likely have been better if directly acquired monoscopically. The photographer could have imaged through the middle of the pupil instead of the pupillary margin. Time and the patient's comfort for each photograph would have also increased, as only half the number of images would have been needed. A previous comparison of monoscopic film to stereo film for detecting CSME also showed high agreement using one photo from a stereo pair. Interreader repeatability was lower for stereo than monoscopic photographs.[22]

Mosaic images were created with a nonmydriatic camera, as it was the only system available with an automosaic feature at the time of the study. We adjusted mosaic viewing magnification to equal that of other formats. Although the added opportunity for examining macular centers in adjoining ETDRS fields 1 and 3 was not available in the mosaic format, agreement of DME features was not lower than the seven-field formats. Our study also did not include grading the posterior vitreous, even though an abnormal vitreoretinal interface can contribute to DME.

Not including OCT in our study comparison is a limitation. OCT is more sensitive than stereo film for detecting RT and, when available, is the method of choice for evaluation of DME.[23,24] Another caveat is that our readers' extensive experience grading DME may have enhanced their ability to perceive subtle edema cues in monoscopic images that are suggestive of RT.

Under our test conditions, the digital images were not significantly different from film for RT grading but were for evaluating CSME. Surprisingly, we also found that digital stereo evaluation of DME offered no advantage over digital monoscopic photographs.

Digital media's replacement of film in ophthalmic research and clinical practice has created the need for improved procedures to address digital photography's current poorer performance than film in assessing CSME and RTCM. Higher resolution digital camera sensors, better display systems, and improved viewing techniques may resolve the discrepancies. Until then, options for identifying CSME include using HE as a surrogate[25] or complementary imaging with OCT. Slide film for documenting DME retains benefits, but may become impractical as that technology is phased out.

Acknowledgments

The authors thank the staff from the Department of Ophthalmology and Visual Sciences, University of Wisconsin-Madison: Andrew F. Ewen, Anne E. Mosher, and Maria K. Swift for grading diabetic retinopathy images; Stacy M. Meuer for grading supervision; Daniel P. Murach for computer support (Ocular Epidemiology Group of Barbara E. K. Klein, MD, and Ronald Klein, MD); and 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 the manuscript (Fundus Photograph Reading Center directed by Ronald P. Danis, MD).

References

1. Photocoagulation for diabetic macular edema: Early Treatment Diabetic Retinopathy Study Report no. 4. The Early Treatment Diabetic Retinopathy Study Research Group. *Int Ophthalmol Clin.* 1987;27:265–272.
2. Moss SE, Klein R, Klein BE. The 14-year incidence of visual loss in a diabetic population. *Ophthalmology.* 1998;105:998–1003.
3. Early Treatment Diabetic Retinopathy Study Research Group. Focal photocoagulation treatment of diabetic macular edema. Relationship of treatment effect to fluorescein angiographic and other retinal characteristics at baseline: ETDRS report no. 19. *Arch Ophthalmol.* 1995;113:1144–1155.
4. Gangnon RE, Davis MD, Hubbard LD, et al. A severity scale for diabetic macular edema developed from ETDRS data. *Invest Ophthalmol Vis Sci.* 2008;49:5041–5047.
5. 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.
6. Kinyoun J, Barton F, Fisher M, et al. Detection of diabetic macular edema: Ophthalmoscopy versus photography—Early Treatment Diabetic Retinopathy Study Report Number 5. The ETDRS Research Group. *Ophthalmology*. 1989;96:746–750.
7. 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.
8. 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.
9. Rudnisky CJ, Tennant MT, Weis E, et al. Web-based grading of compressed stereoscopic digital photography versus standard slide film photography for the diagnosis of diabetic retinopathy. *Ophthalmology.* 2007;114:1748–1754.
10. Tennant MT, Greve MD, Rudnisky CJ, et al. Identification of diabetic retinopathy by stereoscopic digital imaging via teleophthalmology: a comparison to slide film. *Can J Ophthalmol.* 2001;36:187–196.
11. Rudnisky CJ, Hinz BJ, Tennant MT, et al. High-resolution stereoscopic digital fundus photography versus contact lens biomicroscopy for the detection of clinically significant macular edema. *Ophthalmology.* 2002;109:267–274.
12. Diabetic Retinopathy Clinical Research Network. *DRCRnet Manual of Operatons*. Tampa, FL: Jaeb Center for Health Research Foundation, Inc.; 2005.
13. 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.
14. Li H, Florez-Arango JF, Hubbard LD, et al. Grading diabetic retinopathy severity from compressed digital retinal images compared with uncompressed images and film. *Retina*. In press.
15. 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.
16. Early Treatment Diabetic Retinopathy Study Research Group. Grading diabetic retinopathy from stereoscopic color fundus photographs: an extension of the modified Airlie House classification. ETDRS report number 10. *Ophthalmology.* 1991;98:786–806.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–174.
18. Uebersax JS. *User Guide for the MH Program* (ver. 1.2). Statistical Methods for Rater Agreement website. 2006. Available at: http://john-uebersax.com/stat/mh.htm. Accessed May 1, 2009.
19. Marr D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* New York: WH Freeman & Co.; 1982:8–40.
20. Vitale TJ. Projecting digital 'slide' images. 2003. Available at: http://cool.conservation-us.org/byauth/vitale/digital-projection/index.html.. Accessed December 26, 2009.
21. Baker CF, Rudnisky CJ, Tennant MT, et al. JPEG compression of stereoscopic digital images for the diagnosis of diabetic retinopathy via teleophthalmology. *Can J Ophthalmol.* 2004;39:746–754.
22. Welty CJ, Agarwal A, Merin LM, Chomsky A. Monoscopic versus stereoscopic photography in screening for clinically significant macular edema. *Ophthalmic Surg Lasers Imaging.* 2006;37:524–526.
23. Strom C, Sander B, Larsen N, et al. Diabetic macular edema assessed with optical coherence tomography and stereo fundus photography. *Invest Ophthalmol Vis Sci.* 2002;43:241–245.
24. Davis MD, Bressler SB, Aiello LP, et al. Comparison of time-domain OCT and fundus photographic assessments of retinal thickening in eyes with diabetic macular edema. *Invest Ophthalmol Vis Sci.* 2008;49:1745–1752.
25. Bresnick GH, Mukamel DB, Dickinson JC, Cole DR. A screening approach to the surveillance of patients with diabetes for the presence of vision-threatening retinopathy. *Ophthalmology.* 2000; 107:19–24.