Comparison of Film and Digital Fundus Photographs in Eyes of Individuals with Diabetes Mellitus

Sapna Gangaputra,¹ Talat Almukhtar,² Adam R. Glassman,² Lloyd Paul Aiello,³ Neil Bressler,⁴ Susan B. Bressler,⁴ Ronald P. Danis,¹ and Matthew D. Davis,¹ for the Diabetic Retinopathy Clinical Research Network⁵

PURPOSE. To compare grading of diabetic retinopathy (DR) and diabetic macular edema (DME) from stereoscopic film versus stereoscopic digital photographs obtained from a subset of Diabetic Retinopathy Clinical Research Network (DRCR.net) participants.

METHODS. In this photographic media comparison study, digital and film images were obtained at a single study visit from some of the subjects enrolled in active DRCR.net clinical study protocols. Digital camera systems and digital and film photographers were certified to obtain images according to standard procedures. Images were graded for DR severity and DME in a masked fashion by Fundus Photograph Reading Center (Madison, WI) graders. Agreement between gradings was assessed by calculating the percentage of agreement and κ statistics.

RESULTS. Images obtained with both film and digital media were submitted for 155 eyes of 96 study participants. On a nine-step Early Treatment Diabetic Retinopathy study DR severity scale, grading agreed exactly in 74%, and was within one step of agreement in 93%, with a weighted κ statistic of 0.82 (95% confidence interval [CI], 0.71–0.92). On a nine-step DME severity scale and three-step clinically significant macular edema (CSME) scale, grading agreed exactly in 39% and 88%, respectively, and within one step in 70% and 92% (weighted κ statistic, 0.44 [95% CI, 0.34–0.54] and 0.72 [95% CI, 0.55–0.90], respectively).

CONCLUSIONS. Among clinical sites participating in the DRCR.net, agreement between film and digital images was substantial to almost perfect for DR severity level and moderate to substantial for DME and CSME severity levels, respectively. Replacement of film fundus images with digital images for DR

severity level should not adversely affect clinical trial quality. (ClinicalTrials.gov numbers, NCT00367133, NCT00369486, NCT00444600, NCT00445003, NCT00709319.) (*Invest Ophthalmol Vis Sci.* 2011;52:6168–6173) DOI:10.1167/iovs.11-7321

Digital cameras for retinal photography have largely replaced film fundus cameras, in part because of the immediate availability of images and the convenience of their storage, reproduction, and transmission. In addition, commercial film production continues to decrease, and film is likely to be unavailable for retinal photography in the future.¹ Digital color fundus photography differs in some potentially important aspects from film photography of the fundus. Digital camera systems for retinal photography vary a great deal with respect to how they handle tonal balance and illumination,² which may result in differences in evaluation of lesions between film and digital photographs.² This variation presents challenges with respect to evaluating the images of subtle lesions in eyes with diabetic retinopathy (DR) in multicenter clinical trials. If grading of DR severity between film and digital images differs, there may be consequences for study interpretation.

For digital ophthalmic images to be suitable for use in clinical trials evaluating treatment of diabetic eye disease, the sensitivity of diagnosing vascular lesions has to be on a par with the gold-standard modified seven-field (7-field) stereoscopic images, as described in the Early Treatment Diabetic Retinopathy Study (ETDRS) report number 10.³ Some currently available digital camera systems include the option of wide-angle image capture at 45° to 60° (4-field wide protocol), versus the 30° to 35° fields used in conventional 7-field imaging protocols for clinical trials of DR.⁴ Wide-angle photography offers the possibility of imaging an equivalent area of retina with fewer frames, thereby increasing efficiency and exposing the subject to fewer light flashes. Decreasing the number of images also reduces the total file storage requirement, which may be an important consideration for clinical centers and centralized reading centers. Each uncompressed color image, one side of a stereo pair and independent of angle, is approximately 14 megabytes when obtained with a 6-megapixel digital camera system (14 MB × 14 images for modified 7-field and 14 MB × 8 images for 4-field wide).

We compared film and digital images from a subset of Diabetic Retinopathy Clinical Research Network (DRCR.net) clinical centers, to determine the extent of agreement between the grading of these images with respect to diabetic retinal disease assessment, including severity of DR, presence and extent of diabetic macular edema (DME), and clinically significant macular edema (CSME). In addition, we evaluated potential differences in the grading of digital 4-field wide-angle photographs versus digital modified 7-field standard photographs relative to traditional 7-field film grading.

From the ¹Fundus Photograph Reading Center, Department of Ophthalmology and Visual Sciences, University of Wisconsin-Madison, Madison, Wisconsin; the ²Jaeb Center for Health Research, Inc., Tampa, Florida; the ³Beetham Eye Institute and Eye Research Section, Joslin Diabetes Center, Department of Ophthalmology, Harvard Medical School, Boston, Massachusetts; and the ⁴Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland.

⁵A list of the Diabetic Retinopathy Clinical Research Network, as of October 19, 2010, is available at www.drcr.net.

Supported by a cooperative agreement from the National Eye Institute and the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Department of Health and Human Services Grants EY14231, EY018817, and EY14229.

Submitted for publication February 3, 2011; revised April 18, 2011; accepted April 18, 2011.

Disclosure: S. Gangaputra, None; T. Almukhtar, None; A.R. Glassman, None; L.P. Aiello, None; N. Bressler, None; S.B. Bressler, None; R.P. Danis, None; M.D. Davis, None

No reprints will be available.

Corresponding author: Talat Almukhtar, Jaeb Center for Health Research, 15310 Amberly Drive, Suite 350, Tampa, FL 33647; drcrstat6@jaeb.org.

METHODS

Study Participants

This study was conducted by the DRCR.net at 29 clinical sites in the United States. The protocol and Health Insurance Portability and Accountability Act informed-consent forms were approved by multiple institutional review boards. The study complied with the Declaration of Helsinki. Before participation in this study, all clinical centers completed certification of their digital systems through the Fundus Photograph Reading Center (FPRC), Department of Ophthalmology and Visual Sciences, University of Wisconsin-Madison. As part of this process, the photographers were also certified by sending in sample photographs for both film and digital media to confirm optimal illumination, color balance, and field definition, which helped to maintain standardization of equipment and technique.

All study participants were enrolled in another DRCR.net protocol and were recruited for this study at the visits specified in their primary protocols, when they were otherwise scheduled to undergo modified 7-field fundus photography with color film. Photographs were obtained from both eyes or only the study eye, depending on the requirement of the DRCR.net protocol that the subject was primarily enrolled in. The sample size was selected to be at least 50 participants, without respect to statistical principles.

Photography

All study participants underwent pharmacologic pupil dilation followed by 7-field modified stereoscopic film photography.³ Depending on system availability at the clinical center, either modified 7-field stereoscopic digital photography or 4-field wide-angle stereoscopic digital photography was performed (Fig. 1). The location and field size (30° or 35°) for the 7-field digital photograph protocol were the same as for the 7-field color film protocol. The 4-field wide-angle digital technique used 45° to 60° fields, obtained according to standard procedures.⁵ Capture and export settings were provided for each digital system. Uncompressed images were saved on a CD or DVD and submitted to the FPRC according to standard procedures.

Grading

Trained and certified graders at the FPRC evaluated each eye using the ETDRS classifications of retinopathy abnormalities for overall DR severity and DME severity.^{3,6,7} Grading was performed independently by two graders for the digital and film images in a standardized manner, using a multistep grading process with adjudication of key variables including DR severity level, presence and area of DME, and degree of center retina involvement.^{3,6-8} When there was a discrepancy be-

tween graders, the images were reviewed by an adjudicator, and the adjudicator's grade was accepted as the final grade of record. Grading of film and digital images of each eye was separated by a minimum period of 2 weeks, to minimize any memory effect on the part of the graders, with film graded first and digital images graded subsequently. Because of the random distribution of images among the trained evaluators, some graders were involved in both film and digital assessments. Film sets were viewed on a standard light box (6500°K color temperature) with a Donaldson stereo viewer (5×). Digital images were displayed on calibrated 20.5-in. LCD monitors and were viewed with hand-held stereo viewers (Screen-Vu Stereoscope; PS Manufacturing Co., Portland, OR). Optimum image illumination, contrast, and color balance for digital images were achieved by a standardized procedure,² in which the luminance histograms for each of the red/green/blue color channels were analyzed and manually adjusted to enhance color contrast and standardize illumination. Digital images were reviewed in color only, not monochromatically.

The quality of both film and digital images was rated by the graders based on the ability of the grader to view and grade the pertinent retinopathy features in the eye. The photographs were graded as good to fair if they were of good enough quality to grade the features confidently. They were graded borderline when they were reasonably good, but had some features that could not be graded because of impaired image quality, such as poor stereo, poor focus, or inadequate field definition. They were deemed ungradable if there was either a very poor view or no view of the fundus.

Retinopathy severity was graded in all images according to a nine-step ETDRS DR severity scale³ and categorized as follows: DR absent (levels 10 and 12), minimal DR (levels 14, 15 and 20), mild nonproliferative DR (NPDR) (level 35), moderate NPDR (level 43), moderately severe NPDR (level 47), severe NPDR (level 53), prior scars of panretinal photocoagulation or mild proliferative DR (PDR) (levels 60 and 61), moderate PDR (level 65), and high-risk PDR (levels 71, 75, 81, and 85). Images were graded with a summary method, in which the evaluator reviewed all fields and then assigned the grade based on the most severe lesion(s) seen in the eye.

The modified ETDRS DME severity scale was used to grade macular edema.⁷ The scale cross-classifies the area of retinal thickening (RT) within 1 disc diameter (DD) of the center of the macula (RTwi1DD), which has a total area equivalent to 4 disc areas (DAs), by retinal thickness at the center (compared with the maximum thickness of normal retina about 0.5 DD from the center, designated reference thickness). The 24 resulting cells of the cross-classification table were combined on the basis of concurrent visual acuity to create a nine-step scale extending from RTwi1DD absent to RTwi1DD more than 3 DA with central retinal thickness at least twice the reference thickness.⁷

Region of posterior pole covered by each imaging protocol

FIGURE 1. *Left:* modified 7-field stereoscopic digital photography; *right:* 4-field, wide-angle stereoscopic digital photography.

The resulting scale has more steps and tends to have more even distribution of information across its steps than do scales using either area of RT or severity of central thickening alone. CSME was defined as in the ETDRS: retinal thickening 1 DA or larger, any part of which is within 1 DD of the macular center or RT or adjacent hard exudates within 500 μm of the macular center.⁹

Statistical Analysis

Cross tabulations were performed of the distributions of DR severity, DME severity, and CSME determined by film versus digital photographs. Primary analyses pooled photographs from modified 7-field and 4-field wide-angle digital images. The assessed variables included DR severity level, DME severity level, and CSME, with and without center involvement. For the various graded features, the calculated measures of agreement included the percentage of agreement (exact agreement and agreement within one step), κ statistic, and weighted κ statistic. Weights were specified as 1 for exact agreement, 0.75 for one step of discordance, and 0 for all others.³ Landis and Koch¹⁰ defined benchmarks for interpreting simple and weighted κ statistics in which κ statistics in the ranges of <0.00, 0.00 to 0.20, 0.21 to 0.40, 0.41 to 0.60, 0.61 to 0.80, and 0.81 to 1.00 indicated poor, weak, fair, moderate, substantial, and almost perfect agreement, respectively. Only gradable images were included in the calculation of measures of agreement; however, simple κ statistics between gradable versus nongradable images were calculated when possible. Sensitivity and specificity were calculated after treating DR severity level as a binary variable with the cutoff point, excluding nongradable photographs, as severe NPDR or better versus mild PDR or worse, in which film photographs were considered the gold standard. Sensitivity and specificity were adjusted for the correlation between right and left eyes of a study participant when both eyes of the participant were included in the analysis; κ and weighted κ statistics were not adjusted, because the correlation in agreement between the right and left eyes of the same participant was low (<0.20). Exploratory subgroup analyses included stratifying agreement measures by the following: (1) type of digital procedure (4-field wide or modified 7-field); (2) grader (same or different); (3) volume of images obtained by the photographer according to the DRCR.net database since its inception in 2003 (high, moderate, or low when the number of images obtained by the photographer was >200, 51–200, and 1–50, respectively); and (4) volume of images obtained by the clinical site over the duration of study enrollment (19 months), using the same categorization as the volume by photographer in (3). The Mantel-Hansel approach of Kuritz et al.¹¹ was applied to test marginal homogeneity. Statistical analyses were conducted with several programs of commercial or freely available software (SAS system ver. 9.1, SAS Institute, Cary, NC; StatXact 6, Cytel Software Corp., Cambridge, MA; and R statistical software, ver. 2.10, R Foundation for Statistical Computing, Vienna, Austria).¹¹

RESULTS

One hundred fifty-five eyes of 96 study participants (mean age, 62 years) had images submitted using both film and digital modes from 29 DRCR.net clinical sites between March 2007 and October 2008; 48 (31%) digital images were obtained with wide-angle settings and 109 (69%) with the conventional 7-field setting. Both eyes of one subject had digital images obtained with 4-field wide and 7-field cameras, which resulted in 157 total scanned pairs (sets) included in the cohort. One hundred forty-seven (94%) sets of DR images and 119 (76%) sets of DME and CSME images were gradable on both film and digital media. Of the 157 pairs included in this study, overall photograph quality assessment agreed exactly in 67% (95% CI, 59%–74%) and 90% (95% CI, 85%–95%) were within one step of agreement. The overall distribution of photograph quality appeared to be similar between film and digital grading (P for marginal homogeneity = 0.76). Four (3%) film images compared with

eight (5%) digital images were ungradable for DR (P = 0.16). Study participant's characteristics at enrollment are described in Table 1.

In analysis of 147 pairs, grading of DR severity in film and digital images agreed exactly in 74% of eyes and was within one step in 93% of eyes with a weighted κ statistic of 0.82 (95% CI, 0.71–0.92; Table 2). In 22 (58%) of the 38 discordant pairs, retinopathy was graded higher in the digital image (P for marginal homogeneity = 0.45). The sensitivity and specificity of digital images for detecting severe NPDR or better, as identified on film images, were 94% and 96%, respectively.

DME severity level graded on a nine-step scale for 119 gradable image pairs agreed exactly in 39% of the eyes and within one step in 70% (weighted κ statistic, 0.44; 95% CI, 0.34–0.54; Table 3). A difference in grading DME between film and digital photographs could not be identified (P = 0.10). Grades were two or more steps higher in the film images of 24 eyes and in the digital images of 12 eyes. Analysis of CSME showed exact agreement between film and digital images in 88% and within one step of agreement in 92% of the eyes (weighted κ statistic, 0.72; 95% CI, 0.55–0.90; Table 4). Among the 119 eyes with images gradable for both image types, center involvement was graded as present in 83 eyes in both image types, in 10 eyes in film only, and 2 eyes in digital only. Images were ungradable for CSME in both image types in 17 eyes, in film only in 11 eyes, and in digital only in 10 eyes.

Agreement in grading of DR and DME with film photographs appeared similar regardless of the type of digital image (4-field wide versus modified 7-field; Table 5). However, the difference between the weighted κ statistics between the 4-field wide (0.55; 95% CI, 0.25–0.85) and modified 7-field (0.78; 95% CI, 0.56–0.99) images for grading CSME are less clear.

TABLE 1. Subject and Study Characteristics

Study Participants' Characteristics (<i>n</i> = 96)	
Age, mean (min, max) y	62 (37–86)
Women, <i>n</i> (%)	45 (47)
Race, <i>n</i> (%)	
White	82 (85)
African American	10 (10)
Asian	1 (1)
More than one race	1 (1)
Unknown/not reported	2 (2)
Type of diabetes, <i>n</i> (%)	
Type I	11 (11)
Type II	82 (85)
Uncertain	3 (3)
Duration of diabetes, y (mean $\pm$ SD)	19 $\pm$ 11
Eyes, <i>n</i> (%) (<i>n</i> = 155)	
One study eye	37 (39)
Two study eyes	59 (61)
Visual acuity letter score median (25th, 75th percentiles)	71 (61,77)
Approximate Snellen equivalent	20/40 (20/63, 20/32)
Characteristics of Matched Film and Digital Scans (<i>n</i> = 157)	
Digital procedure	
Four-field wide	48 (31)
Modified 7-field	109 (69)
Digital cameras used*	
Topcon 35	48 (31)
Zeiss FF 30	56 (36)
Topcon 50	19 (12)
Zeiss FF 450	33 (21)

* Carl Zeiss Meditec, Jena, Germany; Topcon, Tokyo, Japan.

TABLE 2. Cross Tabulation of Level of Diabetic Retinopathy in Graggable Film and Digital Photographs

DR in Film	DR in Digital										Total
	DR Absent	Minimal DR	Mild NPDR	Moderate NPDR	Moderately Severe NPDR	Severe NPDR	Mild PDR	Moderate PDR	High-Risk PDR	Cannot Grade	
DR Absent (Levels 10 and 12)	2	0	0	0	0	0	0	0	0	0	2
Minimal DR (Levels 14, 15, and 20)	0	0	0	0	0	0	0	0	0	0	0
Mild NPDR (Level 35)	0	1	8	7	0	0	2	0	0	2	20
Moderate NPDR (Level 43)	0	0	1	7	7	0	0	0	0	1	16
Moderately severe NPDR (Level 47)	0	0	2	5	34	1	1	1	0	1	45
Severe NPDR (Level 53)	0	0	0	0	2	3	1	0	0	0	6
Prior scars of panretinal photocoagulation or mild PDR (Levels 60 and 61)	0	0	0	2	1	0	39	1	1	2	46
Moderate PDR (Level 65)	0	0	0	0	0	0	2	11	0	0	13
High-risk PDR (Levels 71, 75, 81, and 85)	0	0	0	0	0	0	0	0	5	0	5
Cannot grade	0	0	0	0	1	0	1	0	0	2	4
Total	2	1	11	21	45	4	46	13	6	8	157

$n = 147$. The following measures do not include photographs that could not be graded by both film and digital. Exact agreement, 74% (95% CI, 66%–81%); within one step of agreement, 93% (95% CI, 88%–97%); weighted κ statistic, 0.82 (95% CI, 0.71–0.92). DR, diabetic retinopathy; NPDR, nonproliferative diabetic retinopathy; PDR, proliferative diabetic retinopathy.

Agreement in grading of DR severity did not appear to differ in subgroups when analysis was stratified by the same grader, the volume of scans obtained by photographer, or the volume by clinical site (Table 6). However, statistical power for these subgroup comparisons was relatively low, as evidenced by the wide CIs of the agreement measures.

DISCUSSION

We assessed intergrader agreement between film and digital images in classifying the severity levels of DR, DME, and CSME in a subset of study participants enrolled in DRCR.net studies. In general, agreement between digital and film grading of DR

severity was substantial regardless of whether the digital images were obtained using 7 standard ETDRS fields or 4-field, wide-angle images (Tables 2, 5). Agreement of macular edema grading ranged from moderate to substantial.

Recent reports on comparison between film and digital media for DR graded by the FPRC, following the same procedures, support these study results. Li et al.¹³ compared pairs of stereoscopic images from 152 eyes of 85 patients across the entire ETDRS scale. They showed 67.8% exact agreement and 96.1% agreement within one step in assessment of DR severity level (weighted κ statistic, 0.86; 95% CI, 0.82–0.90). These images were from a single academic center obtained by a single photographer. The similarity to the results found in our present

TABLE 3. Cross Tabulation of Level of Diabetic Macular Edema (DME) in Film and Digital Photographs

DME in Film	DME in Digital										Total
	Level 1A, 1B	Level 1C	Level 2	Level 3A	Level 3B	Level 4	Level 5A	Level 5B	Level 5C	Cannot Grade	
Levels 1A, 1B	12	0	2	0	1	1	0	0	0	2	18
Level 1C	2	4	0	1	0	0	0	0	0	1	8
Level 2	5	1	0	1	2	1	0	0	0	1	11
Level 3A	1	1	2	0	0	0	1	0	0	0	5
Level 3B	1	0	2	0	0	2	1	0	0	0	6
Level 4	0	0	2	0	1	5	2	2	0	0	12
Level 5A	0	0	2	0	2	1	3	12	0	3	23
Level 5B	2	0	0	1	0	5	7	19	4	3	41
Level 5C	0	0	0	0	0	0	0	2	3	0	5
Cannot grade	6	0	1	0	0	0	2	2	0	17	28
Total	29	6	11	3	6	15	16	37	7	27	157

The following measures do not include photographs that could not be graded by both film and digital or film or digital alone. $n = 119$. Exact agreement, 39% (95% CI, 30%–48%); within one step of agreement, 70% (95% CI, 61–78%); weighted κ statistic, 0.44 (95% CI, 0.34–0.54). Approximate severity levels are as follows: 1A and 1B, retinal thickening (RT) within 1 DD of the macular center (MC) (RTwi1DD) absent; 1C, RTwi1DD present, MC uninvolved; 2, RTwi1DD <1 DA with MC questionably involved or RTwi1DD <0.5 DA with MC definitely involved; 3A and 3B, RTwi1DD >1 DA with MC questionably involved or RTwi1DD 0.5–2 DA with MC mildly involved; 4 and 5A–C, combinations of RTwi1DD ranging from 1 to 4 DA with MC involvement ranging from mild to severe.¹²

TABLE 4. Cross Tabulation of the Level of Clinically Significant Macular Edema (CSME) in Film and Digital Photographs

CSME in Film	CSME in Digital				Total
	None	Noncenter CSME	Center CSME	Cannot Grade	
None	19	2	2	4	27
Noncenter CSME	0	3	0	0	3
Center CSME	8	2	83	6	99
Cannot grade	6	0	5	17	28
Total	33	7	90	27	157

$n = 119$. The following measures did not include photographs classified as cannot grade by both film and digital or by film or digital alone. Exact agreement, 88% (95% CI, 81%–93%); within one step of agreement, 92% (95% CI, 85%–96%); weighted κ statistic, 0.72 (95% CI, 0.55–0.90).

study suggests that rigorous certification of photographers minimizes potential variability between photographers and camera systems. Results were not compared to those in earlier publications of film digital analyses, which showed poor sensitivity

TABLE 5. Agreement between Film and Digital Photographs Stratified by the Type of Digital Procedure

	Four Fields Wide ($n = 48$)	Modified 7-Field ($n = 109$)
DR		
Nongradable photographs, n (%)		
Digital	2 (4)	6 (6)
Film	0 (0)	4 (4)
Agreement with film, % (range)		
Exact agreement	83 (69–92)	70 (60–79)
Within one step of agreement	93 (82–99)	93 (86–97)
Weighted κ statistic*	0.85 (0.66–1.00)	0.80 (0.67–0.93)
DME		
Nongradable photographs, n (%)		
Digital	5 (10)	22 (20)
Film	8 (17)	20 (18)
Agreement with film, % (range)		
Exact agreement	37 (22–54)	40 (29–51)
Within one step of agreement	71 (54–85)	69 (58–79)
Weighted κ statistic†	0.43 (0.26–0.60)	0.45 (0.33–0.57)
CSME		
Nongradable photographs, n (%)		
Digital	5 (10)	22 (20)
Film	8 (17)	20 (18)
Agreement with film, % (range)		
Exact agreement	84 (69–94)	90 (81–96)
Within one step of agreement	89 (75–97)	92 (85–97)
Weighted κ statistic‡	0.55 (0.25–0.85)	0.78 (0.56–0.99)

DR, diabetic retinopathy; DME, diabetic macular edema; CSME, clinically significant macular edema. Analysis did not include images that could not be graded. P value for difference in weighted κ statistic:

* $P = 0.67$.

† $P = 0.87$.

‡ $P = 0.22$.

TABLE 6. Agreement between Film and Digital Photographs in Grading of Diabetic Retinopathy Stratified by Covariates

Covariates	<i>n</i>	Weighted κ (95% CI)	<i>P</i> (Difference in Weighted κ)
Same grader for both media			
Yes	51	0.78 (0.60–0.96)	0.61
No	96	0.84 (0.70–0.97)	
Volume of photographs obtained, by photographer			
Low	24	0.79 (0.53–1.00)	0.98
Moderate	94	0.81 (0.54–1.00)	
High	29	0.77 (0.47–1.00)	
Volume of photographs obtained, by site			
Low-moderate*	68	0.79 (0.61–0.94)	0.58
High	79	0.84 (0.70–0.98)	

$n = 147$.

* The two sites deemed low volume were grouped with the moderate volume sites.

of digital images.^{12,14} As digital cameras have improved substantially in recent years, the process of post hoc manipulation of color balance and illumination have enhanced the quality of digital images. Li et al.¹⁵ compared a film image evaluation with three different formats of digital images acquired from different camera types. They noted substantial agreement between all digital formats and film in the evaluation of diabetic retinopathy severity level. Additionally, they noted that severity assessment remained the same, even without stereoscopic images. In the present study, disagreements on DR severity did not appear to be the result of more severe retinopathy being assessed by one specific medium, as the discrepancies were fairly equally distributed on both sides and were without any specific orientation. Distribution of DR severity grading on film and digital images appeared similar (P for marginal homogeneity = 0.45). There were five images in which the photograph pairs were more than two steps apart. Side-by-side comparisons of these sets showed that four of the five images had different field definition leading to a lesion being seen on one medium and not the other. The lesion was visible on two film images and on two digital images.

The agreement rate between film and digital images was similar to that expected with film–film reproducibility for DR severity level. Grader reproducibility for DR severity on film photographs, as published in ETDRS report number 12⁶ (a grading process similar to that used in the present study), was 53% exact agreement, and 88% were within one step (weighted κ statistic, 0.65). Contemporaneous intergrader reproducibility of DR severity level at the FPRC is assessed by a biannual regrading of images across multiple ongoing studies. Results from May 2010 showed 70% exact agreement and 91% agreement within one step (weighted κ statistic, 0.81). These data included a combination of film and digital images graded over the previous 6 months (unpublished results analyzed by the FPRC, 2010).

In the present study, the agreement rates were lower for the ETDRS DME severity scale than for CSME (weighted κ , 0.44 and 0.72, respectively). This difference may have resulted in part from the differing length of the scales (nine and three steps, respectively) and in part from the presence, in most eyes, of either center-involved DME or no DME at all. Li et al.¹⁶ compared an evaluation of presence or absence of CSME in a set of 152 paired stereoscopic film and digital images. The κ statistic for identification of CSME was 0.86 (95% CI: 0.77–0.95), some-

what higher than the weighted κ statistic for CSME in the present study. In a previous DRCR.net report using the ETDRS DME severity scale in eyes with a broader range of DME severity, the weighted κ statistic for a film–film comparison was 0.58 (SE 0.05).⁸

Photographic gradability and quality are important aspects of the grading process, as subtle DR lesions can easily be obscured without optimal illumination, contrast, and color balance.² In the present study of eyes with DR, with post hoc optimization of digital images, no systematic differences were noted between film and digital photograph quality. The frequency of ungradable images in assessing DR on digital media was slightly greater than the number on film—however, not to the extent that future clinical trial results would be significantly affected. Technical components that can affect data quality of a digital image are the make, model, and magnification of the cameras. Spatial resolution in most modern cameras is not an issue, as most modern digital cameras have resolution of 6 megapixels or higher, which is close to the resolution of color film. The FPRC procedure of certifying cameras reduces variability and also allows for “scaling” of images, to attain measurement data that can be compared uniformly across various camera types.

Strengths of the study are the inclusion of a large number of clinical sites, digital systems, and photographers in a clinical trial setting, using measures to maximize the quality of digital images. Limitations include fewer subjects with severe NPDR and mild degrees of DME and a small sample size for comparison of agreement between 4-field, wide and modified 7-field fundus cameras. Grading was performed by summary assessment and therefore the exact ability of each modality to assess individual retinopathy lesions with respect to one another other could not be evaluated.

In conclusion, grading results from digital fundus images obtained on 4-field wide or modified 7-field fundus cameras, when optimized in a standardized manner, have substantial to almost perfect agreement with grading obtained from film photographs for assessing DR level. Differences between gradings of DR severity obtained on film and digital images may be similar to differences in replicate gradings of film images. Digital fundus photographs are suitable for clinical trials where DR level is an outcome variable.

References

1. Kodachrome film. <http://www.kodak.com/global/en/professional/products/films/catalog/kodachrome64ProfessionalFilmPKR.jhtml>. Accessed October, 19, 2010.
2. 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(8):3269–3282.
3. Early Treatment Diabetic Retinopathy Study Research Group. Grading diabetic retinopathy from stereoscopic color fundus photographs: an extension of the modified Airle House classification. ETDRS report number 10. *Ophthalmology*. 1991;98:786–806.
4. Vujosevic S, Benetti E, Massignan F, et al. Screening for diabetic retinopathy: 1 and 3 nonmydriatic 45-degree digital fundus photographs vs 7 standard early treatment diabetic retinopathy study fields. *Am J Ophthalmol*. 2009;148(1):111–118.
5. Diabetic Retinopathy Clinical Research Network (DRCR.net). Information about the Network. Information for investigators. Procedures. Available at <http://www.drcr.net>. Accessed on October 19, 2010.
6. Early Treatment Diabetic Retinopathy Study Research Group. Fundus photographic risk factors for progression of diabetic retinopathy. ETDRS report number 12. *Ophthalmology*. 1991;98:823–833.
7. Gangnon R, Hubbard LD, Aiello LM, et al. A severity scale for diabetic macular edema (DME) developed from ETDRS data. *Invest Ophthalmol Vis Sci*. 2008;49:5041–5047.
8. 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(5):1745–1752.
9. Early Treatment Diabetic Retinopathy Study Research Group. Photocoagulation for diabetic macular edema. Early Treatment Diabetic Retinopathy Study report number 1. *Arch Ophthalmol*. 1985;103(12):1796–1806.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
11. Kuritz SJ, Landis JR, Koch GG. A general overview of Mantel-Haenszel methods: applications and recent developments. *Annu Rev Public Health*. 1988;9:123–160.
12. Lin DY, Blumenkranz MS, Brothers RJ, Grosvenor DM. The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. *Am J Ophthalmol*. 2002;134(2):204–213.
13. 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(11):5846–5852.
14. 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(3):572–585.
15. Li HK, Danis RP, Hubbard LD, Florez-Arango JF, Esquivel A, Krupinski EA. Comparability of digital photography with the ETDRS film protocol for evaluation of diabetic retinopathy severity. *Invest Ophthalmol Vis Sci*. 2011;52(7):4717–4725.
16. 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(12):6753–6761.