Digital versus Film Fundus Photography for Research Grading of Diabetic Retinopathy Severity

Helen K. Li,[1,2] *Larry D. Hubbard,*[3] *Ronald P. Danis,*[3] *Adol Esquivel,*[4] *Jose F. Florez-Arango,*[1,5] *Nicola J. Ferrier,*[6,7] *and Elizabeth A. Krupinski*[8]

Purpose. To assess agreement between digital and film photography for research classification of diabetic retinopathy severity.

Methods. Digital and film photographs from a 152-eye cohort with a full spectrum of Early Treatment Diabetic Retinopathy Study (ETDRS) severity levels were assessed for repeatability of grading within each image medium and for agreement on ETDRS discrete severity levels, ascending severity thresholds, and presence or absence of diabetic retinopathy index lesions, between digital and 35-mm slides (film). Digital photographs were color balanced to match film.

Results. There was substantial agreement ($\kappa = 0.61$, κ_w [linear weighted] $= 0.87$) in classification of ETDRS diabetic retinopathy severity levels between digital images and film. Marginal homogeneity analyses found no significant difference in frequency distributions on the severity scale ($P = 0.21$, Bhapkar test). The κ results ranged from 0.72 to 0.95 for presence or absence of eight ascending diabetic retinopathy severity thresholds. Repeatability of grading between readers viewing digital images was equal to or better than that obtained with film (pair-wise interreader κ for digital images ranged from 0.47 to 0.57 and for film from 0.43 to 0.57. The κ results for identifying diabetic retinopathy lesions ranged from moderate to almost perfect. Moderate agreement of intraretinal microvascular abnormalities and venous beading between digital images and film accounted for slightly lower concordance for severity thresholds $\geq$47 and for slightly lower interreader agreement within digital and film images at severity thresholds $\geq$43 and $\geq$47.

Conclusions. Under controlled circumstances, digital photography can equal the reliability of 35-mm slides for research classification of ETDRS severity level. (*Invest Ophthalmol Vis Sci.* 2010;51:5846–5852) DOI:10.1167/iovs.09-4803

Fundus photography has been the foundation of diabetic retinopathy (DR) clinical research for more than 40 years. Photography criteria were established by international researchers at the Airlie House Symposium in 1968,[1] followed in 1981 by the Diabetic Retinopathy Study's standards for detecting and grading DR severity by using stereoscopic 35-mm slides (film).[2–4] The evolving protocol and DR severity classification system were expanded in 1991 by the Early Treatment Diabetic Retinopathy Study (ETDRS).[5] Today, ETDRS is the gold standard against which other DR assessment approaches are measured.[6–8] ETDRS defines a 30°, seven standard-field, stereoscopic, color film protocol. Its severity guidelines specify the retinal vascular abnormalities graded in a set of photographs, the importance of each abnormality, the density or quantity of abnormalities, and for some lesions, a location (e.g., neovascularization on the disc).[5,9]

Today's technology is changing clinical and research photography practices. Digital photography offers significant convenience and other advantages over film. In some settings, digital photography is being integrated into health care as an alternative to face-to-face DR evaluation. ETDRS film practices have also been adapted for telemedicine.[10–12] Several important studies have compared ETDRS photography and retinopathy classification with protocols modified for digital photographs.[13–16] Camera resolutions ranged from 640 × 480[15–17] to 3040 × 2008[13] pixels, reflecting the technologies available at the time of each study. Modifications to ETDRS protocols continue as teleretinal imaging programs develop disease referral criteria based on program goals. Some telemedicine programs use digital photographs for DR screening.[18–20] Others follow the clinical course of retinopathy and/or identify disease that meets treatment criteria.[13,21,22]

Despite the transition from film to digital media in DR clinical research, there are critical unknowns. No published report has shown the equivalency of digital photographs in matching the historical performance of ETDRS film in distinguishing the entire spectrum of DR. Some studies comparing ETDRS film with digital photography enrolled patients consecutively, resulting in cohorts with a limited range of retinopathy severity levels. Other studies had a large proportion of no or mild retinopathy, reflecting the low prevalence of advanced DR in the general population.[23]

Protocol requirements for DR research are more demanding than those for clinical care. Fundus photos in epidemiologic studies and clinical trials are used to determine smaller differences in DR progression than clinical photography. Digital photography introduces variables unknown in film and thus falls outside the scope of the historical ETDRS protocol. Just as clinical research reading centers have developed time-proven

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 Departments of [3]Ophthalmology and Visual Sciences, [6]Mechanical Engineering, and [7]Biomedical Engineering, University of Wisconsin-Madison, Madison, Wisconsin; the [4]Houston Veterans Affairs Medical Center and Baylor College of Medicine, Houston, Texas; [5]Universidad De Antioquia, Medellin, Columbia; and the [8]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, 2008.

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

Submitted for publication October 23, 2009; revised December 29, 2009, and January 31 and March 11, 2010; accepted April 19, 2010.

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

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

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

	Film										
	10	**15,20**	**35**	**43**	**47**	**53**	**61**	**65**	**71,75**	**90**	**Total**
Digital											
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	0
Total	26	8	34	25	24	5	10	10	10	0	152

Dark gray shading indicates perfect agreement; light gray shading indicates agreement within one step.

best practices to guide the acquisition and review of ETDRS film photographs,[24] similar best practices are needed for digital photography. In this study, we assessed digital images for use in DR research grading across the full severity scale, using *uncompressed* color stereoscopic digital retinal images versus color stereoscopic slides.

MATERIAL AND METHODS

Study Design

This study is one arm of a comprehensive evaluation of digital photography formats compared with ETDRS film: uncompressed color stereoscopic, compressed color stereoscopic, monoscopic, and monoscopic wide-angle mosaic.

Fundus Photographs

Patients with a history of type 1 or 2 diabetes mellitus from a University of Texas Medical Branch (UTMB) 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 were informally screened in clinic by a retinal physician (HKL). Those with severity levels not yet adequately represented were invited for photography. Those with media opacity or limited pupil dilation preventing an adequate retina image, with other retinal vascular disease, or with previous retinal laser photocoagulation were excluded from photography, to avoid possible bias in grading macular edema or proliferative retinopathy. Each eye had 16 slides and 16 digital photographs taken according to ETDRS protocol: seven nonsimultaneous color stereoscopic field pairs of the fundus and one pair of the anterior segment using the same camera (TRC-50EX/IX; Topcon Medical Systems, Paramus, NJ) coupled with 35-mm and digital camera backs (MegaVision, Santa Barbara, CA). The same photographer, certified by the University of Wisconsin (UW) Fundus Photograph Reading Center for ETDRS protocol photography, took all photographs in the same sequence. Patients rested a minimum of 30 minutes between sessions. Digital images were saved as uncompressed 2392 × 2048 TIFF files. The film was processed at a Kodak-certified Q-Laboratory facility (Eastman Kodak, Rochester, NY).[24]

Two photographed eyes were excluded because of missing stereo pair or digital photos. A total of 152 eyes of 85 patients were selected to represent the full range of DR severity levels. Patients included 32 (37.6%) males, 53 (62.4%) females, 37 (43.5%) Caucasians, 24 (28.2%) Hispanics, and 24 (28.2%) African Americans. Seventy-five right and 77 left eye images were used. Photographs included both eyes of 67 (78.8%) patients, the right eye only of 8 (9.4%), and the left eye only of 10 (11.8%). Ages ranged from 33 to 83 years with a median of 60.5 and mean of 59.4 years. Photographs were de-identified by assigning coded ID numbers.

Color and contrast of 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 maximized the contrast of DR abnormalities against retinal pigment epithelial (RPE) backgrounds without creating artifacts. This algorithm was modified from one used in the Age Related Eye Disease Study 2 (AREDS2).[25] Monochromatic green channel images generally provide better contrast of small lesions than do full-color photographs.[26] These were created by extracting the green channel from digital RGB color images and were used routinely to confirm suspected subtle DR lesions.

Grading of Photographs

Three UW readers independently evaluated all images in each format. To minimize recall, evaluation was regulated by custom scheduling

TABLE 2. DR Severity Thresholds, Digital versus Film Images

Retinopathy Threshold	Sensitivity	Sensitivity 95% CI	Specificity	Specificity 95% CI	Eyes at or Above Threshold (*n*)	Positive Predictive Value	Negative Predictive Value	Rate of Agreement	κ	κ 95% CI
≥15,20	0.98	0.94–1.00	0.96	0.80–0.99	126	0.99	0.93	0.98	0.93	0.85–1.00
≥35	0.98	0.94–1.00	0.94	0.80–0.99	118	0.98	0.94	0.97	0.92	0.85–1.00
≥43	0.96	0.90–0.99	0.91	0.82–0.97	84	0.93	0.95	0.94	0.88	0.80–0.96
≥47	0.88	0.77–0.95	0.85	0.76–0.92	59	0.79	0.92	0.86	0.72	0.60–0.90
≥53	0.91	0.77–0.98	0.92	0.85–0.96	35	0.76	0.97	0.91	0.77	0.66–0.89
≥61	0.97	0.83–0.99	0.98	0.94–1.00	30	0.94	0.99	0.98	0.94	0.87–1.00
≥65	1.00	0.83–1.00	0.99	0.95–1.00	20	0.91	1.00	0.99	0.94	0.87–1.00
71,75	1.00	0.69–1.00	0.99	0.96–1.00	10	0.91	1.00	0.99	0.95	0.85–1.00

n = 152 eyes.

TABLE 3. DR Severity Index Lesions, Digital versus Film Images

Abnormality	Sensitivity	Sensitivity 95% CI	Specificity	Specificity 95% CI	Eyes with Abnormality (*n*)	Positive Predictive Value	Negative Predictive Value	Rate of Agreement	κ	κ 95% CI
Ma	0.98	0.93–1.00	0.96	0.82–0.99	124	0.99	0.90	0.97	0.91	0.83–1.00
RH	0.98	0.94–1.00	0.89	0.72–0.98	124	0.98	0.93	0.97	0.89	0.79–0.98
HE	0.98	0.92–1.00	0.97	0.89–1.00	90	0.98	0.97	0.97	0.95	0.89–1.00
SE	0.86	0.77–0.93	0.88	0.78–0.94	79	0.88	0.85	0.87	0.74	0.63–0.84
IRMA	0.76	0.64–0.86	0.83	0.74–0.90	63	0.76	0.83	0.80	0.59	0.46–0.72
VB	0.77	0.50–0.93	0.94	0.89–0.97	17	0.62	0.97	0.92	0.64	0.45–0.83
NVE	0.96	0.81–0.99	0.98	0.93–1.00	27	0.90	0.99	0.97	0.91	0.83–1.00
FPE	0.85	0.62–0.97	0.99	0.95–1.00	20	0.90	0.98	0.97	0.85	0.73–0.98
NVD	0.67	0.23–0.95	0.99	0.96–1.00	6	0.80	0.99	0.98	0.72	0.41–1.00
FPD	0.50	0.08–0.92	0.99	0.96–1.00	2	0.50	0.99	0.99	0.49	0.00–1.00
PRH	1.00	0.66–1.00	1.00	0.97–1.00	9	1.00	1.00	1.00	1.00	1.00–1.00
VH	1.00	0.40–1.00	0.99	0.96–1.00	4	0.80	1.00	0.99	0.89	0.66–1.00

$n = 152$ 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; NVE, new vessels elsewhere; FPE, fibrous proliferation elsewhere; NVD, new vessels disc; FPD, fibrous proliferation disc; PRH, preretinal hemorrhage; VH, vitreous hemorrhage.

software that separated grading of the same eye by at least 2 weeks. Batches of different formats were counterbalanced in order of presentation. Readers graded presence and severity of each abnormality included in the ETDRS retinopathy severity classification in relevant photographic fields.[5,9] A computer algorithm assigned retinopathy severity on a nine-step ETDRS scale: 10, no retinopathy; 15,20, microaneurysms (Ma) 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, high-risk PDR; and 90, cannot grade.[9] Image sets were graded in three steps according to a reader's confidence level for retinopathy severity: *high* (good image quality and typical lesions), *moderate* (adequate, because of less satisfactory image quality and/or atypical lesions), or *low* (inadequate, leading to the selection of cannot grade). "Index" lesions are abnormalities, when their presence or severity qualified an eye for a specific ETDRS level of retinopathy. Readers determined presence and severity of index lesions and diabetic macular edema according to ETDRS definitions.[5]

Stereo sets of slides were graded as described in ETDRS Report 10[5] and viewed on daylight fluorescent light boxes by using a Donaldson 5× stereo viewer (George Davco, Holcombe, MA). Magnification was approximately 12.5×, accounting for the combined magnification of the camera (Topcon) and the Donaldson viewer.

Customized software facilitated viewing of digital images monoscopically at full screen using 13× magnification, approximating slide magnification. Digital images were viewed at a distance of 26 inches on 21-inch CRT displays (1600 × 1200 pixels), calibrated to a color temperature of 6500°K and 2.2 gamma, and checked monthly (Greytag Macbeth; X-Rite, Inc., Grand Rapids, MI). Images were displayed at 6.5× magnification as stereo pairs, viewed with a handheld stereo viewer (Screen-Vu; PS Manufacturing, Portland, OR). Readers first reviewed a proof sheet of thumbnail photographs before examining each 35° field image in detail. Each field was then reviewed monoscopically (13×) and stereoscopically (6.5×) in full color.

Statistical Analysis

The DR severity level for each eye was calculated as the central tendency (median) among the three readers. Instead of duplicate grading with adjudication of differences,[9] this multireader method allowed pairwise comparisons of all readers within each format.

Film grading results were considered the reference standard. We defined *threshold* as the cutoff for the presence of retinopathy at a particular ETDRS severity level or worse, ranging from any retinopathy (level 15/20) to high-risk PDR (level 71/75). Presence or absence of index lesions, severity level, and retinopathy severity at several thresholds were compared.

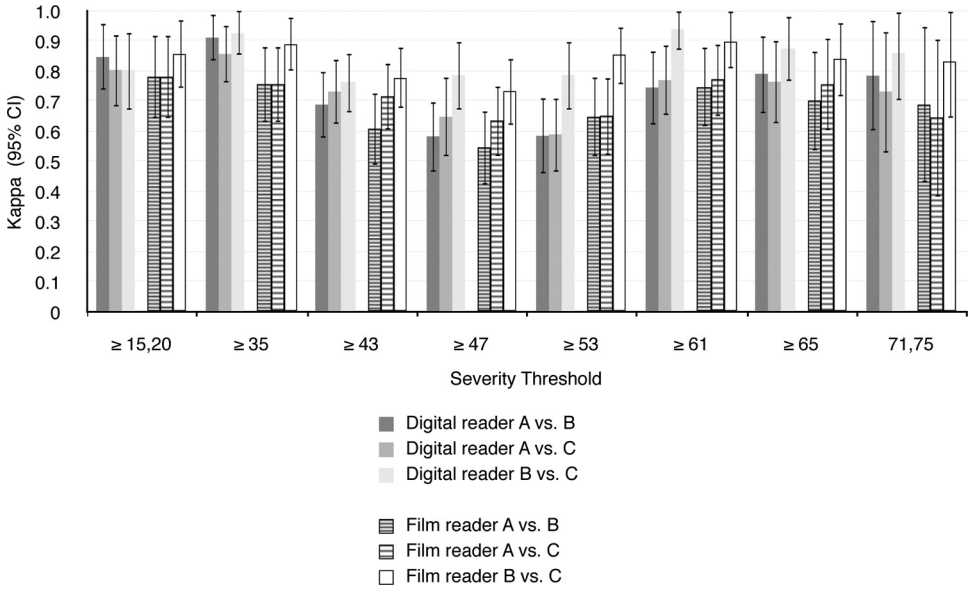


FIGURE 1. Interreader DR severity threshold agreement: digital compared with film images ($n = 152$ eyes).

Agreement on the level of severity was cross-tabulated and κ statistics (unweighted and weighted, linear scheme) were calculated. Eyes with film or digital photographs classified as ungradable (level 90) were excluded from the analysis. Interpretation guidelines were those used by Landis and Koch[27]: 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. The McNemar test for dichotomized scales and the Bhapkar test for multistep scales were used to test for significant differences (cross-tab marginal homogeneity). Marginal distributions were assessed with the McNemar test for overall bias.

Severity thresholds and index lesions were tested using κ, sensitivity and specificity percentages, positive and negative predictive values, and disease prevalence percentages. Agreements between digital and film and within each medium between readers were assessed by percentage of agreement and κ (Analyze-It, Ltd., Leeds, UK; MedCalc; MedCalc Software BVBA, Mariakerke, Belgium; and the MH [marginal homogeneity] program, ver. 1.2[28]).

Results

Digital and Film Agreement

Table 1 shows the distribution of severity levels among the studied media. There was substantial grading agreement between stereoscopic film and digital stereoscopic images over the entire ETDRS scale (κ = 0.62; 95%, 0.54–0.71; linear weighted κ = 0.86; 95%, 0.82–0.90). Agreement was exact in 67.8% of the eyes, within one step in 96.1%, and within two steps in 99.3%. ETDRS severity scale distribution differences between grading with digital images and film were not statistically significant (P = 0.21 by Bhapkar test).

Sensitivity and specificity between digital photographs and film was >90% in all divisions, except threshold ≥47, at which sensitivity was 88% and specificity 85% (Table 2). The κ statistic for all thresholds was almost perfect, except for threshold ≥47 and ≥53, for which it was substantial.

The κ for identifying DR lesions between film and digital ranged from moderate to perfect (Table 3). Intraretinal microvascular abnormalities (IRMAs), venous beading (VB), and fibrovascular proliferation on the optic disc (FPD) showed moderate agreement. The sensitivity of IRMA, VB, FPD, or neovascularization on the disc was lower than for other lesions. Digital images had specificity almost equal to that of film for all lesions except IRMA.

Interreader Agreement

Pair-wise (reader A versus B, A versus C, B versus C) reproducibility of digital image grading was similar to that achieved with film: digital image κ = 0.47–0.57, median = 0.51 (95% CI, 0.42–0.59); film κ = 0.43–0.57, median = 0.47 (95% CI, 0.38–0.56). Our intrareader agreement for digital images was also similar to that for film (data not shown). Reproducibility of grading eight severity thresholds using film was no better than grading using digital images (Fig. 1).

Digital and Film Disagreement

There were six eyes with differences of two or more severity levels between digital and film grading. Two eyes were graded at a higher level with film than digital. Four eyes were graded at a higher level with digital than film. Disagreement in four of the six eyes was due to uneven quality between digital and film images. Quality was better in film in two eyes but better in digital in another two. Grading differences in the remaining two eyes were due to reader misclassification of RPE abnormalities as laser photocoagulation in one eye and to a missed microaneurysm and retinal hemorrhage in the other.

Table 4. DR Severity Level in Various Studies, Digital versus Film Images

Study	Eyes (*n*)	Steps Graded (*n*)	Severity Levels	Exact Agreement (%)	Agreement ± 1 Step (%)	κ Unweighted (95% CI)	κ Weighted Linear (95% CI)	Bhapkar *P*	McNemar Bias *P*	Direction	Referral Severity Threshold Studied
Bursell et al.[15]	105	7	10 / 35 / 43 / 53 / 53E / 61<71 / ≥71	61.9	89.5	0.53 (0.41–0.64)	0.69 (0.59–0.79)	0.0277	0.0269	Film	≥53
Lin et al.[16]	181	8	10 / 14,15,20 / 35 / 43 / 47 / 53 / 60 / 70	59.1	93.4	0.44 (0.34–0.53)	0.64 (0.57–0.71)	0.0214	0.0201	Film	≥35
Fransen et al.[14]	513	9	10,14 / 15,20 / 35 / 43 / 47 / 53 / 61 / 65 / 71,81	85.8	97.9	0.78 (0.74–0.82)	0.90 (0.88–0.92)	0.0074	0.0140	Film	≥53
Rudnisky et al.[13]	198	9	10 / 15,20 / 35 / 43 / 47 / 53 / 60,61 / 65 / 71,75,80	71.7	96.0	0.65 (0.58–0.73)	0.87 (0.83–0.90)	0.0052	0.0001	Film	≥61
Present study	152	9	10 / 15,20 / 35 / 43 / 47 / 53 / 61 / 65 / 71,75	67.8	96.1	0.62 (0.54–0.71)	0.86 (0.82–0.90)	0.2137	0.0152	Digital	Not applicable

* Includes only eyes with readable digital and film images.

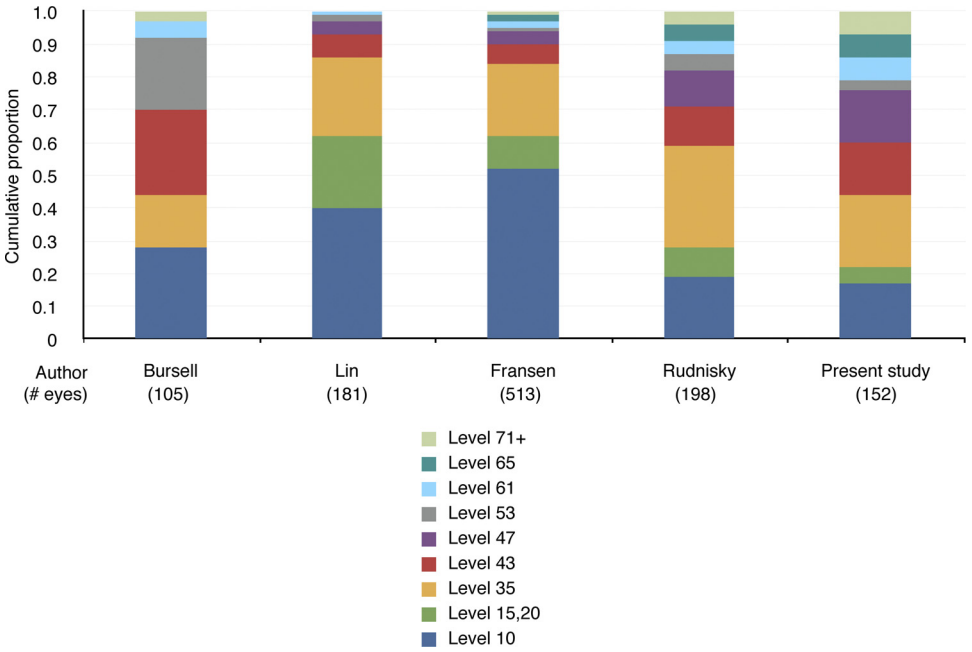


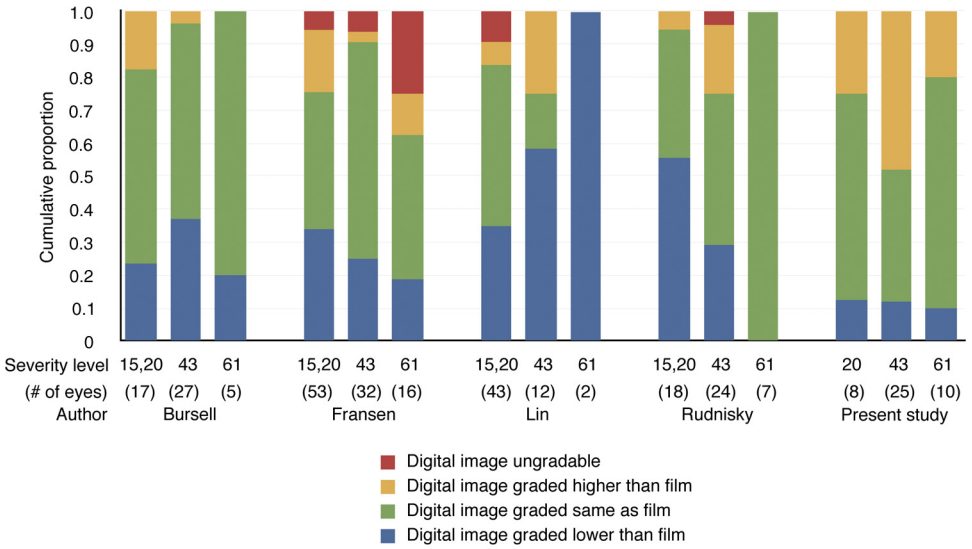
FIGURE 2. Distribution of DR severity level grading using film in various studies.

DISCUSSION

To our knowledge, this is the first study in which the full ETDRS photography protocol and classification system has been applied to digital photography in a stratified sample across all DR severity levels. Except for replacing a fundus camera's 35-mm film recording emulsion with a digital chip sensor, we made minimal changes to ETDRS protocol parameters.

Grading severity levels in digital images compared favorably to those in film and were reproducible between readers. Interreader digital image reproducibility (median, $\kappa = 0.51$) was comparable to ETDRS ($\kappa = 0.42$).[9] The accuracy of identifying DR lesions from digital photographs also compared favorably to film. Lower agreement between digital photographs and film in identifying IRMA and VB accounted for lower κ scores when grading severity thresholds $\geq$47 and $\geq$53 and slightly lower κ for threshold $\geq$43. Detecting and assigning grades to IRMA and VB, however, was equally challenging when using film. ETDRS found full lesion severity κ of IRMA = 0.15 and κ of VB = 0.19.[5] IRMA is critical to the definition of severity levels 43, 47, and 53. VB helps define levels 47 and 53. These important index lesions accounted for our study's lower interreader agreement for thresholds $\geq$43, $\geq$47, and $\geq$53 in both digital and film images. Lower agreement in identifying IRMA and VB was also found in other studies, whether comparing stereoscopic 640 $\times$ 480 compressed[15] or monoscopic 3040 $\times$ 2008 compressed images[13] to slides. Our camera system had a 4.6-μm/pixel spacing and 13-μm/pixel resolving power on the retina, sufficient to detect the smallest microaneurysm, IRMA, or neovascularization.

Analyses of previous studies revealed significant differences in DR severity level grading between film and digital ($P < 0.05$



Level 15,20 (microaneurysms or retinal hemorrhages).

Level 43 (intraretinal microvascular abnormalities or moderate microaneurysms/retinal hemorrhages).

Level 61 (fibrous proliferation on the disc or elsewhere).

FIGURE 3. DR severity level in various studies: digital compared with film images.

TABLE 5. Photography Details in Various Studies

Study	Eyes with Readable Images (*n*)	Consecutive Enrollment	Fundus Camera	Dilation	Image Resolution (Pixels, Bit Depth)	Sensor Size (Megapixels)	Fields, Field of view (*n*, deg)	Stereoscopic Viewing Device	Compression
Bursell et al.[15]	105	No	Nonmydriatic	No	640 × 480, 24-bit	0.3	3, 45	LCD glasses	10:1 JPEG
Lin et al.[16]	181	Yes	Nonmydriatic	No	640 × 480, 8-bit	0.3	1, 45	Monoscopic	No
Fransen et al.[14]	513	Yes	Mydriatic	Yes	1152 × 1152, 24-bit	1.3	7, 30	LCD glasses	No
Rudnisky et al.[13]	198	Yes	Mydriatic	Yes	3040 × 2008, 24-bit	6.1	7, 30*	LCD glasses*	16:1 JPEG
Present study	152	No	Mydriatic	Yes	2392 × 2048, 24-bit	4.9	7, 35	Handheld stereo prism viewer	No

* Only fields 1 and 2 are stereo.

Bhapkar test; Table 4). Consecutive enrollment in most of these studies resulted in unbalanced samples (Fig. 2) because there is a greater prevalence of no or mild DR severity in the general population. Analyses also showed that readers in previous studies tended to assign higher levels of severity grading to film images ($P < 0.05$ by McNemar bias test; Table 4). In contrast, our readers assigned higher severity levels when grading with digital images.

Further examination of severity levels defined by subtle index lesions in previous studies supports a trend toward reduced sensitivity with digital images for retinopathy detection. Figure 3 shows each study's digital photography versus film sensitivities at levels 15/20, 43, and 61. Compared with previous studies (Table 5), the proportion of digital images assigned a *higher* severity level than film in our study was equal to or greater than the proportion of digital images assigned a *lower* severity level. It is unsurprising that the study with the most abbreviated imaging protocol (smallest retinal area, low digital resolution, monoscopic rather than stereo, photographs taken without pharmacological dilation) had the most undergrading using digital images.[16] Other investigators whose protocols had digital photography retinal areas comparable to that covered by ETDRS' seven fields also had more undergrading than overgrading.[13–15]

We believe that this is the highest fidelity comparison of a digital imaging protocol to ETDRS. Optics were kept consistent between digital and film photography by using the same fundus camera. Digital photographs' viewing magnification, color, and contrast were adjusted to approximate slides. No digital file compression was used. Readers with a minimum of 10 years' experience in ETDRS protocol grading reviewed all images. Reading conditions were controlled by using calibrated monitors in a defined viewing environment via standardized display software. Reproducibility of grading was tested using several measures.

We are also not aware of another DR study in which algorithmic optimization was employed to more closely match digital images to film. Our relatively high agreement may be due to normalization of digital images to a formal color model to ensure color consistency. The ETDRS photography protocol required standardized slide film and processing to ensure color consistency.[29,30] An analysis of AREDS2 images found that slides have better brightness and contrast than do digital photographs, in the hands of most photographers.[25] The reddish nature of the fundus combined with the narrow dynamic range and the "hot" red response of silicon sensors makes digital fundus photography difficult.[31] Investigators in previous studies did not combine supplemental green channel viewing with color viewing for ETDRS classification. Readers' inspection of subtle lesions in green channel images may have contributed to our agreement outcome. In addition, we achieved high agreement in eyes without advanced lenticular opacity, other retinal vascular disease, or laser photocoagulation.

Because our primary concern was the performance of film versus digital images, it is possible that our results would not be achievable in other settings. For example, to achieve a relatively stratified sample, our study eye enrollment was not consecutive. Instead, we selected eyes with retinopathy levels needed to provide the full nine-step scale DR spectrum. As in population-based studies, however, it was difficult to find enough patients at level 53 (very severe NPDR). Enrollment in levels 15/20 (8 eyes) and 53 (5 eyes) fell short of our target of 10 eyes per category. Other limitations include those inherent in taking consistent-quality photographs of subtle index abnormalities using digital or film media: IRMA, NVE and perhaps VB. Human evaluation of retinal photographs for classification is intrinsically subjective, resulting in grading variability that cannot be fully eliminated, regardless of media.

Film has been the basis for research in DR evaluation for many years. Criteria for using color slides are well established. ETDRS required two emulsions: Kodak Professional Ekta-

chrome 100 or Kodachrome 64 (or their equivalents; Eastman Kodak) processed by approved laboratories.[29,30] Kodachrome was introduced in 1935 and Ektachrome in the early 1940s. There is a long history and a good understanding of 35-mm slide technology. In contrast, there are several digital sensor technologies on the market today, many variations in sensor parameters, and varying color responses. Far from being a mature platform, digital sensor technology is rapidly evolving. Nevertheless, we found that, under carefully controlled circumstances, grading a broad range of DR severity levels using uncompressed digital images produced results equivalent to those obtained with 35-mm slides.

Acknowledgments

The authors thank the staff of the Department of Ophthalmology and Visual Sciences, University of Wisconsin-Madison; 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 images for diabetic retinopathy; Stacy M. Meuer for grading supervision; and Daniel P. Murach for computer support; 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 the manuscript.

References

1. 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.
2. Diabetic Retinopathy Study Group. *Manual of Operations*. Baltimore: Diabetic Retinopathy Study Coordinating Center; 1972.
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. Diabetic Retinopathy Research Study Group: Design, methods, and baseline results. DRS report number 6. *Invest Ophthalmol Vis Sci.* 1981;21:149–209.
5. 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. 1991;98:786–806.
6. Cavallerano J, Lawrence MG, Zimmer-Galler I, et al. Telehealth practice recommendations for diabetic retinopathy. *Telemed J E Health.* 2004;10:469–482.
7. 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;148:111–118.
8. Neubauer AS, Rothschuh A, Ulbig MW, Blum M. Digital fundus image grading with the non-mydriatic Visucam(PRO NM) versus the FF450(plus) camera in diabetic retinopathy. *Acta Ophthalmol.* 2008;86:177–182.
9. 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.
10. Massin P, Chabouis A, Erginay A, et al. OPHDIAT: a telemedical network screening system for diabetic retinopathy in the Ile-de-France. *Diabetes Metab.* 2008;34:227–234.
11. Boucher MC, Desroches G, Garcia-Salinas R, et al. Teleophthalmology screening for diabetic retinopathy through mobile imaging units within Canada. *Can J Ophthalmol.* 2008;43:658–668.
12. Lemmetty R, Makela K. Mobile digital fundus screening of type 2 diabetes patients in the Finnish county of South-Ostrobothnia. *J Telemed Telecare.* 2009;15:68–72.
13. 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.
14. 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.
15. 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.
16. 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;134:204–213.
17. Lawrence MG. The accuracy of digital-video retinal imaging to screen for diabetic retinopathy: an analysis of two digital-video retinal imaging systems using standard stereoscopic seven-field photography and dilated clinical examination as reference standards. *Trans Am Ophthalmol Soc.* 2004;102:321–340.
18. Zeimer R, Zou S, Meeder T, et al. A fundus camera dedicated to the screening of diabetic retinopathy in the primary-care physician's office. *Invest Ophthalmol Vis Sci.* 2002;43:1581–1587.
19. Scanlon PH, Carter S, Foy C, et al. An evaluation of the change in activity and workload arising from diabetic ophthalmology referrals following the introduction of a community based digital retinal photographic screening programme. *Br J Ophthalmol.* 2005;89:971–975.
20. Taylor CR, Merin LM, Salunga AM, et al. Improving diabetic retinopathy screening ratios using telemedicine-based digital retinal imaging technology: the Vine Hill Study. *Diabetes Care.* 2007;30:574–578.
21. Cavallerano AA, Cavallerano JD, Katalinic P, et al. Use of Joslin Vision Network digital-video nonmydriatic retinal imaging to assess diabetic retinopathy in a clinical program. *Retina.* 2003;23:215–223.
22. Tennant MT, Rudnisky CJ, Hinz BJ, et al. Tele-ophthalmology via stereoscopic digital imaging: a pilot project. *Diabetes Technol Ther.* 2000;2:583–587.
23. Klein R, Klein BE, Moss SE, et al. The Wisconsin epidemiologic study of diabetic retinopathy. III. Prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years. *Arch Ophthalmol.* 1984;102:527–532.
24. Diabetic Retinopathy Clinical Research Network. *DRCRnet Manual of Operations*. Version 3. Tampa, FL: Jaeb Center for Health Research Foundation, Inc.; 2005.
25. 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.
26. 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.
27. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–174.
28. Uebersax JS. User Guide for the MH Program (ver. 1.2). *Statistical Methods for Rater Agreement.* 2006. Available at: http://john-uebersax.com/stat/mh.htm. Accessed May 1, 2009.
29. Early Treatment Diabetic Retinopathy Study (ETDRS) Manual of Operations. Springfield, VA: National Technical Information Service; 1985. Accession No. PB85223006.
30. 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. Accessed June 21, 2009.
31. Tyler ME, Hubbard LD, Boydston K, et al. Characteristics of digital fundus camera systems affecting tonal resolution of color retinal images. *J Ophthalmic Photography.* 2009;31:9–14.