Comparison of Digital and Film Grading of Diabetic Retinopathy Severity in the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study

Larry D. Hubbard, MAT; Wanjie Sun, MS; Patricia A. Cleary, MS; Ronald P. Danis, MD; Dean P. Hainsworth, MD; Qian Peng, MS; Ruth A. Susman, BS; Lloyd Paul Aiello, MD, PhD; Matthew D. Davis, MD; for the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study Research Group

Objective: To compare diabetic retinopathy (DR) severity as evaluated by digital and film images in a long-term multicenter study, as the obsolescence of film forced the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study (DCCT/EDIC) to transition to digital after 25 years.

Methods: At 20 clinics from 2007 through 2009, 310 participants with type 1 diabetes with a broad range of DR were imaged, per the Early Treatment Diabetic Retinopathy Study (ETDRS) protocol, with both film and digital cameras. Severity of DR was assessed centrally from film and tonally standardized digital cameras. For retinopathy outcomes with greater than 10% prevalence, we had 85% or greater power to detect an agreement κ of 0.7 or lower from our target of 0.9.

Results: Comparing DR severity, digital vs film yielded a weighted κ of 0.74 for eye level and 0.73 for patient level ("substantial"). Overall, digital grading did not systematically underestimate or overestimate severity (McNemar bias test, $P=.14$). For major DR outcomes (≥3-step progression on the ETDRS scale and disease presence at ascending thresholds), digital vs film κ values ranged from 0.69 to 0.96 ("substantial" to "nearly perfect"). Agreement was 86% to 99%; sensitivity, 75% to 98%; and specificity, 72% to 99%. Major conclusions were similar with digital vs film gradings (odds reductions with intensive diabetes therapy for proliferative DR at EDIC years 14 to 16: 65.5% digital vs 64.3% film).

Conclusion: Digital and film evaluations of DR were comparable for ETDRS severity levels, DCCT/EDIC design outcomes, and major study conclusions, indicating that switching media should not adversely affect ongoing studies.

Arch Ophthalmol. 2011;129(6):718-726

Author Affiliations: Department of Ophthalmology and Visual Sciences, University of Wisconsin, Madison (Drs Danis, and Davis, Mr Hubbard, and Mss Peng and Susman); Biostatistics Center, George Washington University, Washington, DC (Mss Sun and Cleary); Department of Ophthalmology, University of Missouri, Columbia (Dr Hainsworth);and Joslin Diabetes Center, Department of Ophthalmology, Harvard Medical School, Boston, Massachusetts (Dr Aiello).
Group Information: See page 726 for group member information.

LONG-TERM MULTICENTER studies such as the Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications (EDIC) require consistent measurements of key outcome parameters over time and across clinics, especially when technology evolves during the study. The DCCT (1983-1993) demonstrated that intensive therapy aimed at maintaining blood glucose levels as close to normal as possible substantially reduced the risk of development and/or progression of diabetic retinopathy (DR) and other microvascular complications compared with conventional therapy.[1-3] The EDIC (1994-2016 [ongoing]), an observational follow-up study of the DCCT cohort,[4] demonstrated that the differences in DR and other microvascular (and macrovascular) outcomes between the former intensive and conventional treatment groups persisted for at least 10 years after the DCCT despite the loss of glycemic separation after the clinical trial ended.[5-9] Since the inception of the DCCT in 1983, recording of retinal images, from which DR status and progression are evaluated, has inexorably moved from film to digital. Commercial digital fundus camera systems have markedly improved in quality, have been widely adopted by clinics, and offer substantial convenience and economy compared with film cameras.

Changing retinal imaging methods in the DCCT/EDIC, while perhaps unavoidable, might alter study analysis results and conclusions. Although several cross-sectional studies have reported that digital systems provide results that are similar to the film "gold standard," most represent single-center experience and some lack a wide range of retinopathy severity. Therefore, the DCCT/EDIC Research Group undertook a formal due-diligence ancillary study to gauge the effect

Table 1. Clinical Characteristics of the 310 DCCT/EDIC Subjects With Gradable Digital and Film Photographs in the Digital-Film Ancillary Study

Characteristics	Percentage		
	DCCT Baseline (1983-1989)	DCCT Closeout or EDIC Baseline (1992-1993)	Digital-Film Ancillary Study (2007-2009)
Sample, No.		310	
Primary cohort		57	
Intensive therapy		50	
Female sex		53	
Age, mean (SD), y	26 (7.3)	33 (7.1)	48 (7.1)
Diabetes duration, mean (SD), y	5.4 (3.9)	12.0 (5.1)	27.2 (5.0)
BMI, mean (SD)	23.3 (2.9)	26.0 (4.0)	28.7 (5.5)
BMI ≥ 30	1.6	12.5	35.9
Current smoker	20.0	21.4	21.4
Blood pressure, mean (SD), mm Hg[a]	86.4 (9.2)	88.1 (9.2)	88.5 (9.1)
Hypertension[b]	3.5	5.8	46.4
AER ≥ 30 mg/d in DCCT/EDIC	11.6	12.3	60.4
AER ≥ 300 mg/d in DCCT/EDIC	0	1.9	9.6
Hyperlipidemia[c]	0	25.5	59.7
Retinopathy levels based on film photo			
No retinopathy (10/10)	56.5	25.8	5.8
Microaneurysms (MA) only (20/<20)	28.1	37.1	35.5
Mild NPDR (35/<35)	12.9	25.5	21.6
Moderate NPDR (43/<43)	1.9	6.1	17.7
Moderately severe NPDR (47/<47)	0.6	2.9	4.8
Severe NPDR (53/<53)	0	0	0.3
PDR (61/<61)	0	2.6	14.2
CSME based on film photograph	0	3.2	7.3
Hemoglobin A_{1c}, mean (SD)	9.0 (1.5)	8.1 (1.6)	7.9 (1.1)
Mean hemoglobin A_{1c} during DCCT or EDIC, mean (SD)		8.0 (1.3)	8.0 (1.0)

Abbreviations: AER, albumin excretion rate; BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); CSME, clinically significant macular edema; DCCT, Diabetes Control and Complications Trial; EDIC, Epidemiology of Diabetes Interventions and Complications Study; NPDR, nonproliferative diabetic retinopathy; PDR, proliferative diabetic retinopathy.

SI conversion factor: To convert hemoglobin A_{1c} to proportion of total hemoglobin, multiply by 0.01.

[a] Mean blood pressure defined as two-thirds of the diastolic blood pressure plus one-third of the systolic blood pressure.

[b] Hypertension is defined as systolic blood pressure of 140 mm Hg or greater or diastolic blood pressure of 90 mm Hg or greater, documented hypertension, or use of antihypertensive agents.

[c] Hyperlipidemia is defined as low-density lipoprotein cholesterol level of 130 mg/dL (to convert to micromoles per liter, multiply by 0.0357) or greater or use of lipid-lowering agents.

on retinal outcomes of switching from film to digital photography. In addition to examining conventional measures of agreement between digital and film grading results, we were also able to evaluate retrospectively the degree to which DCCT/EDIC primary study outcomes and conclusions might be altered by transitioning between the different imaging media.

METHODS

STUDY DESIGN

This was a masked, cross-sectional comparison study for determining results of film and digital imaging in assessing DR. Sample size calculations[10,11] indicated that, for outcomes with 10% or greater prevalence, 300 subjects would provide 85% or greater power to detect a κ of 0.7 or lower compared with our target κ of 0.9. The target and alternative κ were based on the test/retest κ on film photographs in the DCCT/EDIC.[2,6]

SUBJECTS

Twenty DCCT/EDIC centers certified for both film and digital imaging (of the 28 clinical centers) studied 319 subjects with type 1 diabetes at their regular visits; 9 subjects (2.8%) were excluded because they had ungradable digital (n=6) and/or film (n=5) photography sets in one or both eyes.

Inclusion and exclusion criteria for the DCCT have been published previously.[1] Clinical characteristics in the 310 subjects included in the study are given in **Table 1** at DCCT baseline (1983-1989), EDIC baseline, and at the time of the digital-to-film transition study (EDIC years 14-16). Comparison of the 310 participants with the remaining 1131 persons enrolled in the DCCT showed no important differences except that more nonparticipants were male, from the secondary cohort, and had higher mean hemoglobin A_{1c} levels during DCCT (eTable; www.archophthalmol.com), largely because 6.9% who had died and 6.2% who were inactive were included as substudy nonparticipants. Because the primary focus of this article is not on treatment effect, this imbalance does not introduce bias to most digital-film comparisons.

DCCT/EDIC DATA COLLECTION

Retinopathy was assessed by standard film fundus photography in the whole cohort every 6 months during DCCT, in approximately one-quarter of the cohort each year during EDIC, and in the entire cohort at EDIC years 4 and 10.[6] Reproducibility of the film grading procedure and its stability over time

were evaluated in each study by annual masked regrading of a sample of images (both eyes of each subject) that included a broad spectrum of DR severity. During DCCT, there were 7 annual replicate gradings of 42 and, later, 60 subjects; during EDIC, there were 10 annual replicate gradings of 50 subjects.[4]

FUNDUS PHOTOGRAPHY PROCEDURE

Both film and digital photography used the standard 7-field, nonsimultaneous stereoscopic, 30° color procedure established by the Diabetic Retinopathy Study,[12] as modified by the Early Treatment Diabetic Retinopathy Study (ETDRS).[13] Sets of fundus photographs of both eyes included central views of disc and macula, adjacent views of each of the 4 major vascular arcades, and an adjacent view just temporal to the macula. Although recent studies of macular edema have shifted the disc and temporal-to-macula fields slightly to include the center of the macula, DCCT/EDIC has retained the original ETDRS definitions of fields 1 and 3.

Film photographs were taken on Zeiss FF2-4 fundus cameras (Carl Zeiss Meditech, Inc, Oberkochen, Germany) (or approved alternatives) by certified photographers. Digital images were obtained using camera systems with a minimum of 3 megapixels; 19 of 20 clinics had 5-megapixel or higher systems. Clinics were required to submit images taken of nonstudy volunteers to obtain reading center certification of photographers and digital camera systems.

FUNDUS IMAGE HANDLING AND DISPLAY

At the clinic, film photographs were mounted in plastic sheets in approximate anatomic position and digital photographs were indexed as "proof sheets," with personal identifying information removed except for study identification number. At the reading center, all digital images were loaded for unified handling into the Topcon IMAGEnet system (Topcon Medical Imaging Inc, Paramus, New Jersey) and were JPEG-compressed at the IMAGEnet "maximum" quality setting, with an average compression ratio of approximately 20:1.

Film sets were retroilluminated on a standard light box (6500° K color temperature) and viewed with the Donaldson stereo viewer (George Davco, Holbrook, Massachusetts). Digital images were displayed on calibrated 20.5-in liquid crystal display monitors ($\gamma = 2.2$; color temperature, 6500° K; luminance, 110-170 candelas per m^2) and viewed with handheld stereo viewers (Screen-Vu Stereoscope; PS Manufacturing, Portland, Oregon).

Imposition on images of the ETDRS macular grid and measurements of distances/areas were done in film by superimposing grids and measuring circles printed on transparent acetate stock and in digital by superimposing a digital version of the grid and by using the standard distance and planimetry tools of the digital system. For stereo viewing, gridding, and measurement, graders invoked the IMAGEnet stereo analyzer function. For digital images, grids and measuring tools were scaled for each camera, according to the spatial calibration factor established by the reading center at the time of system certification.

Image illumination, contrast, and color balance were controlled in film by specifying acceptable film emulsions (Kodak Ektachrome Professional ASA [Kodak Inc, Rochester, New York] or equivalent) and development processes (E-6 process by a Kodak Q-certified laboratory). Digital image tonal characteristics were optimized via the standardized enhancement model published by the Age-Related Eye Disease Study 2.[14] An automated processor-computed luminance histogram for each of the red/green/blue color channels and the curves for each channel were adjusted via algorithm to conform to a model image derived from exemplars.

Quality of both film and digital images was rated by the graders, based on proper field definition, crisp focus, and stereo effect. Graders assigned an image confidence score of high, adequate, or inadequate for answers to the main DR questions as affected by image quality.

DIABETIC RETINOPATHY GRADING PROCEDURE

Certified graders evaluated each eye using the ETDRS classifications of DR abnormalities, diabetic macular edema,[12,13,15] and overall DR severity.[16] Data were entered into computerized forms, with checks for internal consistency and completeness. The grading program included independent assessments of each eye by 2 graders (from a pool of 6), with adjudication of substantial differences by a senior grader (from a pool of 3). Grading of film and digital images of each eye was separated by a minimum of 2 weeks (in most cases, several months) to minimize any memory effect. Another senior grader not involved in the original grading compared film and digital images side-by-side, with knowledge of the grades from both, to explore possible reasons for differences in grading between the two media.

GRADING AND OUTCOMES

Diabetic retinopathy severity at the eye level was assigned one of the following ETDRS levels: 10 (including levels 14 and15—eyes without microaneurysms but with cotton-wool spots or retinal hemorrhages, respectively), 20, 35, 43, 47, 53, 61 (including level 60—panretinal photocoagulation scars without extant proliferative DR), 65, 71, 75, 81, and 85.[15] The ETDRS person-level combines eye results (worse eye emphasized method) as previously done in the DCCT/EDIC.[3]

To estimate the effect of digital/film grading differences on DCCT/EDIC design outcomes, we collapsed grading scales into dichotomous categories of particular interest to the study: any retinopathy (including microaneurysms only, ie, level 20 or worse in either eye), mild nonproliferative DR (NPDR) or worse (≥35 in either eye), moderate NPDR or worse (≥43 in either eye), moderately severe NPDR or worse (≥47 in either eye), severe NPDR or worse (≥53 in either eye), proliferative DR (PDR) (≥60/61 in either eye), and Diabetic Retinopathy Study high-risk characteristics or worse (≥71 in either eye). Proliferative DR is the primary EDIC retinopathy outcome after EDIC year 10. Retinopathy progression in the DCCT was defined as an increase of 3 or more steps on the ETDRS person scale from DCCT baseline. Further retinopathy progression in EDIC was defined as 3 or more steps progression from DCCT closeout. Progression of DR at the dual imaging visit was used to compare the outcomes from digital vs film images.

Diabetic macular edema was analyzed as the presence or absence of ETDRS clinically significant macular edema (CSME). Center-involved diabetic macular edema was insufficiently prevalent in our population for reliable comparison between media.

PRELIMINARY TEST OF GRADING PERFORMANCE ON DIGITAL IMAGES PRIOR TO STANDARDIZED ENHANCEMENT

After grading the digital images of 98 eyes (49 subjects) without standardized enhancement for tonal characteristics, the reading center performed a preliminary comparison of ETDRS retinopathy severity levels between digital and film gradings. There appeared to be a systematic difference between results from the

two media, with higher DR severity levels in some eyes on film compared with digital images (data not shown). Standardized enhancement (optimization) was then applied to these digital images, and they were independently regraded. The reduction in systematic differences between the two media achieved by optimization was substantial. Therefore, all digital images were optimized prior to being graded.

STATISTICAL ANALYSIS

Agreement between film and digital gradings on ordinal DR categories was analyzed by cross-tabulation and by rates of exact and near agreement. Cohen κ statistics, both unweighted[17] and weighted,[18] were calculated for multistep ordinal scales. A weight of 1 was assigned for exact agreement, 0.75 for 1-step difference on eye and patient scales, and 0.5 for 2-step differences on the patient scale. For 2-step or greater differences on the eye scale or 3-step or greater differences on the patient scale, the weight 0 was applied. We used guidelines for interpretation of κ proposed by Landis and Koch: 0.0-0.20 indicates slight; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; and 0.81-1.00, almost perfect.[19] The Bhapkar test of marginal homogeneity[20] was used to assess the agreement between film and digital in marginal distribution of the ordinal ETDRS scale. The McNemar overall bias test[21] was used to test for systematic overestimation or underestimation between film and digital gradings.

Film/digital agreement on dichotomous DCCT/EDIC DR categories was evaluated by prevalence, agreement rate, sensitivity, specificity, false-positive and false-negative rates, and Cohen unweighted κ, using film as the reference standard. For prevalence rates close to 0 or 1, Cohen κ was not reported because of its unreliability owing to substantial imbalance in the distribution of marginal totals.[22]

To assess the effect of switching from film to digital images, separate multivariate logistic regression models were constructed within each image type comparing the glycemic treatment effect (odds reduction of the former intensive therapy compared with conventional therapy) on several DR outcomes, especially risk of further 3-step DR progression during EDIC (our primary retinopathy outcome through EDIC year 10) and risk of onset of PDR during EDIC (our primary retinopathy outcome after year 10). These models adjusted for the same covariates as our published Weibull proportional hazard model, including primary or secondary cohort (no retinopathy or retinopathy at DCCT baseline), diabetes duration at DCCT baseline, hemoglobin A_{1c} levels at DCCT eligibility, and retinopathy levels at DCCT closeout.[6]

To evaluate historical reproducibility of film photography during DCCT/EDIC, Fleiss κ among multiple raters[17] was used to calculate κ for DR dichotomous categories, using data from annual replicate gradings on the quality control image samples. Reliability of the digital film grading across clinics was analyzed via the Cochran test of homogeneity.[23]

RESULTS

COMPARISON OF DIGITAL VS FILM GRADINGS OF DR SEVERITY

Figure 1 compares film and digital gradings on the ETDRS person-level scale. There were at least 12 persons in each of the lower retinopathy severity categories (from no retinopathy, level 10=10, through moderately severe NPDR in the worse eye, level 47<47) and in the 3 mildest PDR categories (levels 60<60, 60=60, and 65<65) but only 0 to 3 in the more severe NPDR (levels 47=47 through 53=53) and PDR categories (levels 65<65 through 71=71). There was exact agreement in 51% of subjects, agreement within 1 level in 82%, and agreement within 2 levels in 95% (DR progression is worsening of ≥3 levels). Weighted κ was 0.73 (95% confidence interval, 0.68-0.77), representing substantial agreement between digital and film gradings. The McNemar test of overall bias did not show significant systematic difference between gradings (film higher in 27% and lower in 22%; $P=.14$). The Bhapkar test of marginal homogeneity indicated a borderline significant imbalance between the marginal distributions of film vs digital gradings ($P=.08$; eFigure 1).

The corresponding analysis using ETDRS eye-level scale is shown in **Figure 2**. To gain power, we used all eyes with gradable film and digital photographs (N=628, including those with gradable photographs in only 1 eye). Agreement rates were 63% for exact agreement and 94% for agreement within 1 step. Weighted κ for agreement was 0.74 (95% confidence interval, 0.71-0.78). Gradings showed more severe DR with film than with digital (film higher in 141 eyes and digital higher in 92, $P=.001$ by McNemar test), and there was significant marginal heterogeneity ($P=.002$ by Bhapkar test; eFigure 2). The most noteworthy differences were in the 106 eyes placed in level 10 by 1 or both image types (film higher in 36 and digital higher in 14; $P=.002$) and in the 122 eyes in level 43 by 1 or both image types (film higher in 56 and digital higher in 31; $P=.004$).

Side-by-side review of a sample of these cases post hoc by a senior grader confirmed that small, subtle microaneurysms, intraretinal microvascular abnormalities, and retinal new vessels were sometimes more difficult to detect in digital color images than in film, even after tonal enhancement.

COMPARISON OF DIGITAL VS FILM GRADINGS OF DIABETIC MACULAR EDEMA

In this study, clinically significant diabetic macular edema occurred in only 6% to 7% of subjects and 6% to 7% of eyes, providing insufficient power for reliable analyses. However, agreement rates on presence or absence were 94% for subjects and 96.8% for eyes; digital was higher in 5.3% and film higher in 4.3% (McNemar bias test, $P=.56$); and marginal totals were not significantly different (Bhapkar test of marginal homogeneity, $P=.59$).

AGREEMENT ON DCCT/EDIC DR OUTCOMES BASED ON DIGITAL VS FILM IMAGES

Table 2 presents the agreement on dichotomous DCCT/EDIC DR categories determined from digital vs film images. In these categories, digital vs film κ ranged from 0.69 to 0.96, agreement proportion was 86% to 99%, sensitivity was 75% to 98%, and specificity was 72% to 99%. Agreement on the presence of any degree of PDR (including scars of prior photocoagulation treatment of it, with or without residual new vessels), the primary EDIC retinopathy outcome, was very good, leading to high sensitivity (96%-98%), specificity (99%), and κ (0.95-0.96)

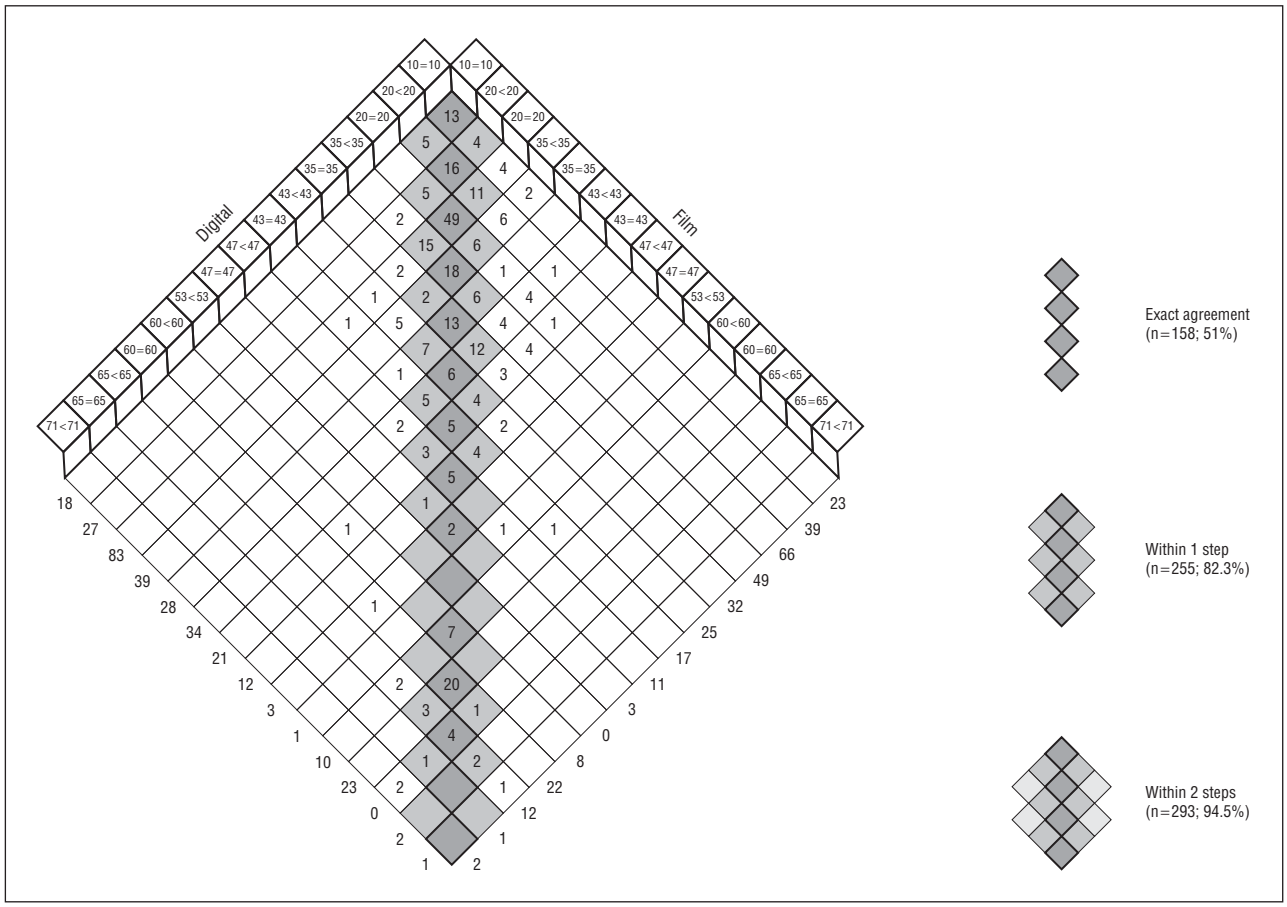

Figure 1. Cross-tabulation of film and digital gradings of final Early Treatment Diabetic Retinopathy Study scale based on person-level of 310 subjects with gradable dual image types. κ=0.44, SE=0.03, 95% confidence interval=0.38-0.5; weighted κ=0.7, SE=0.02, 95% confidence interval=0.65-0.74; weights are 1 for complete agreement, 0.75 for 1-step, 0.5 for 2-step, and 0 for all other disagreement.

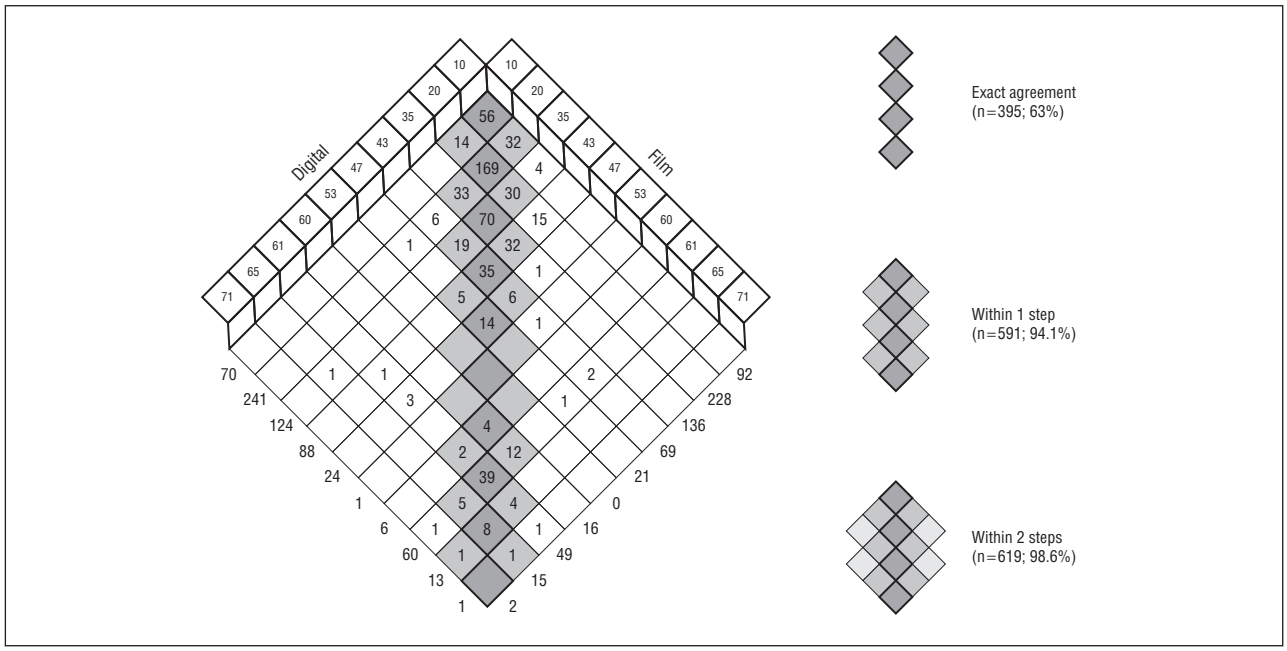

Figure 2. Cross-tabulation of film and digital gradings of final Early Treatment Diabetic Retinopathy Study scale based on eye level of 310 subjects with gradable dual-image types (n=628). Level 60 (scars of photocoagulation for proliferative diabetic retinopathy [DR] or severe nonproliferative DR without residual new vessels) and level 61 (mild retinal new vessels, with or without photocoagulation scars) are shown separately here rather than being pooled (into mild proliferative DR) as they are when change on the scale is calculated. κ=0.52, SE=0.02, 95% confidence interval=0.47-0.57; weighted κ=0.74, SE=0.02, 95% confidence interval=0.71-0.78; weights are 1 for complete agreement, 0.75 for 1-step, and 0 for all other disagreement.

Table 2. Reliability of Digital-Film Photography Grading in EDIC (N = 310)

Retinopathy Outcome	Percentage							κ (95% CI)[a]
	Prevalence Rate		Agreement Rate	Sensitivity	Specificity	False-Positive Rate	False-Negative Rate	
	Film	Digital						
3-Step progression from DCCT baseline	47.1	47.7	88	88	88	12	12	0.75 (0.68-0.83)
Further 3-step progression from DCCT closeout	32.9	31.3	90	82	94	6	18	0.77 (0.69-0.85)
Any retinopathy >10/10	94.2	92.6	95	97	72	28	3	
Mild NPDR or worse >20/20	58.7	58.7	86	88	84	16	12	0.72 (0.64-0.80)
Moderate NPDR or worse >35/35	37	33	86	75	92	8	25	0.69 (0.60-0.77)
Severe NPDR or worse >47/47	14.5	14.5	99	96	99	1	4	0.95 (0.90-1.00)
PDR or worse >53/53	14.2	14.5	99	98	99	1	2	0.96 (0.92-1.00)
CSME[b]	7.3	6.0	94	50	98	3	50	

Abbreviations: CI, confidence interval; CSME, clinically significant macular edema; DCCT, Diabetes Control and Complications Trial; EDIC, Epidemiology of Diabetes Interventions and Complications Study; NPDR, nonproliferative diabetic retinopathy; PDR, proliferative diabetic retinopathy.
[a] Cohen κ.[18] Cohen κ is not reliable when the prevalence of an outcome is close to 1 or 0.[22]
[b] N = 302 for CSME.

Table 3. Logistic Regression of DCCT Treatment Effect on Risk of Any Degree of PDR Based on Film vs Digital Photography at EDIC Years 14 Through 16 Among the Participants Free of PDR at DCCT Closeout After Adjustment for the Other Risk Factors (N = 302)

Covariate	Film-Based PDR		Digital-Based PDR	
	OR (95% CI)	*P* Value	OR (95% CI)	*P* Value
At DCCT entry				
HbA_{1c} level at DCCT eligibility, %	1.2 (0.9 to 1.5)	.28	1.1 (0.9 to 1.5)	.38
Cohort primary (vs secondary)	0.9 (0.3 to 3.1)	.86	0.9 (0.3 to 2.8)	.82
Type 1 diabetes mellitus duration, y	0.9 (0.8 to 1.0)	.12	0.9 (0.8 to 1.0)	.10
At DCCT closeout				
Retinopathy level				
Microaneurysms (vs no retinopathy)	3.0 (0.3 to 28.2)	.34	3.9 (0.4 to 35.2)	.23
Mild NPDR (vs no retinopathy)	24.9 (2.8 to 220)	.004	27.3 (3.1 to 238)	.003
Moderate or severe (vs no retinopathy)	129.8 (11.5 to >999)	<.001	116.1 (10.5 to >999)	<.001
DCCT treatment group conventional (vs intensive)	1.7 (0.7 to 4.1)	.27	1.7 (0.7 to 4.1)	.22

Abbreviations: CI, confidence interval; DCCT, Diabetes Control and Complications Trial; HbA_{1c}, hemoglobin A_{1c}; NPDR, nonproliferative diabetic retinopathy; OR, odds ratio; PDR, proliferative diabetic retinopathy.

for the PDR category. This result may be explained in part by panretinal photocoagulation scars, easily detected in images of either type in 25 of the 35 patients with mild proliferative DR. Proliferation consisting solely of early new vessels is sometimes more difficult to detect in digital than film images, although there was agreement on presence in 8 of 10 such eyes. Results for the severe NPDR (or worse) category could not be accurately determined because only 1 of the 310 participants was classified as having severe NPDR, and only using film (Figure 1). Similarly, the low sensitivity observed for CSME (50%) is of uncertain significance owing to low prevalence. There were very few subjects with no retinopathy in either eye (10 by film only, 5 by digital only, and 13 by both; Figure 1). Thus, the low specificity observed for the "any retinopathy" threshold (72%) is not statistically reliable.

Table 3 presents the agreement between digital and film grading regarding the effect of former DCCT treatment assignment (standard vs intensive glycemic control) on the risk of any degree of PDR, at the dual-imaging visit, among the 302 participants free of PDR at DCCT close out. Multivariate logistic regression revealed an almost identical treatment effect from film and digital gradings. Adjusted odds ratios (ORs) for risk of PDR, conventional vs intensive, were 1.7 for film (95% confidence interval, 0.7-4.1; *P* = .27) and 1.7 for digital (95% confidence interval, 0.7-4.1; *P* = .22). Models were adjusted for primary or secondary cohort (no retinopathy or retinopathy at DCCT baseline), diabetes duration at DCCT baseline, hemoglobin A_{1c} levels at DCCT eligibility, and retinopathy levels at DCCT closeout.

Additional multivariate logistic regression models on other retinopathy categories (**Table 4**) showed similar results. Adjusted ORs of conventional vs intensive treatment are comparable between film and digital at various levels: for further 3-step or greater progression, film OR was 1.6 (*P* = .07) vs digital, 1.5 (*P* = .10); for mild NPDR

Table 4. Logistic Regression of DCCT Treatment Effect on Risk of Various Retinopathy Categories Based on Film vs Digital Photography at EDIC Years 14 Through 16 Among the Participants Free of Respective Complications at DCCT Closeout After Adjustment for the Other Risk Factors[a]

Retinopathy Category	Participants, No.[b]	Prevalence		Adjusted OR of Conventional vs Intensive (95% CI)	*P* Value
		Intensive, %[c]	Conventional, %[c]		
3-Step progression from DCCT baseline					
Film	195	32.7	43.9	1.9 (1.0-3.5)	.05
Digital		34.5	41.5	1.5 (0.8-2.8)	.18
Further 3-step progression from DCCT closeout					
Film	304	28.4	37.6	1.6 (0.9-2.7)	.07
Digital		27.1	35.6	1.5 (0.9-2.6)	.10
Mild NPDR or worse					
Film	195	42.5	51.2	1.5 (0.8-2.6)	.22
Digital		41.6	50.0	1.5 (0.8-2.6)	.22
Moderate NPDR or worse					
Film	274	23.8	36.6	1.7 (0.9-3.0)	.09
Digital		18.9	31.3	1.8 (1.0-3.3)	.06
PDR or worse					
Film	302	7.8	16.2	1.7 (0.7-4.1)	.27
Digital		7.8	16.9	1.7 (0.7-4.1)	.22

Abbreviations: DCCT, Diabetes Control and Complications Trial; NPDR, nonproliferative diabetic retinopathy; PDR, proliferative diabetic retinopathy.
[a] The same logistic models as in Table 3 were used with the respective retinopathy category as the outcome, and the same covariates adjusted.
[b] Patients free of respective complications at DCCT closeout were included. For further 3-step progression, those with scatter photocoagulation in DCCT were excluded.
[c] Prevalences of respective complications within each treatment group of those free of the corresponding complications at DCCT closeout were reported.

Table 5. Reliability of Film Photography Grading in DCCT and EDIC

Retinopathy Outcome	DCCT			EDIC		
	Patients/ Regrading, No.	Prevalence Rate, %	κ (95% CI)[a]	Patients/ Regrading, No.[b]	Prevalence Rate, %	κ (95% CI)[b]
3-Step progression from DCCT baseline	NA	NA	NA	49/10	61	0.91 (0.87-0.95)
Any retinopathy >10/10	60/7	78	0.74 (0.68-0.79)	49/10	88	0.87 (0.83-0.91)
Mild NPDR or worse >20/20	60/7	46	0.80 (0.75-0.85)	49/10	82	0.93 (0.89-0.97)
Moderate NPDR or worse >35/35	60/7	23	0.83 (0.78-0.88)	49/10	70	0.91 (0.87-0.95)
Severe NPDR or worse >47/47	42/4	29	0.66 (0.54-0.78)	49/10	45	0.71 (0.67-0.75)
PDR or worse >53/53	42/4	13	0.72 (0.60-0.84)	49/10	30	0.82 (0.78-0.86)
High-risk characteristics or worse >65/65	42/4	7	0.90 (0.78-1.02)	49/10	11	0.85 (0.81-0.89)
CSME	42/4	14	0.91 (0.79-1.02)	49/10	29	0.65 (0.62-0.69)

Abbreviations: CI, confidence interval; CSME, clinically significant macular edema; DCCT, Diabetes Control and Complications Trial; EDIC, Epidemiology of Diabetes Interventions and Complications Study; NA, not applicable; NPDR, nonproliferative diabetic retinopathy; PDR, proliferative retinopathy.
[a] Fleiss κ for multiple raters.[25]
[b] One of the 50 subjects had ungradable photographs and was not included in the analysis.

or worse, film OR was 1.5 (*P*=.22) vs digital, 1.5 (*P*=.02); and for moderate NPDR or worse, film OR was 1.7 (*P*=.09) vs digital, 1.8 (*P*=.06). The greater-than-3-step progression from DCCT at baseline shows the largest discrepancy between image media, with adjusted ORs of 1.9 for film (*P*=.05) and 1.5 for digital (*P*=.18).

RELIABILITY OF κ ACROSS CLINICS

Comparison of κ for the dichotomous DCCT/EDIC DR outcomes across clinics via Cochran test of homogeneity[24] showed no significant difference among the 20 clinics from the United States and Canada (eFigure 3).

HISTORICAL REPRODUCIBILITY OF GRADING DR FROM FILM IN DCCT/EDIC

Weighted κ statistics for reproducibility on the ordinal ETDRS scale derived from film gradings in annual quality control exercises ranged from 0.72 to 0.84 in the DCCT[2] and from 0.69 to 0.80 in the EDIC—values somewhat greater than the κ of 0.70 from the film vs digital

comparison (Figure 1) using the same weighting scheme. For most dichotomous outcomes there were similar differences; for 3-step or greater progression, presence of mild NPDR or worse, and presence of moderate NPDR or worse, κ values ranged from 0.80 to 0.93 in DCCT and EDIC (**Table 5**), while corresponding values for film vs digital comparisons ranged from 0.69 to 0.77 (Table 2). In contrast, the film vs film quality control exercises produced lower κ values than the film vs digital comparison study for presence of PDR and presence of severe NPDR or worse, as might be expected in quality control sets selected to include eyes in level 53 and to minimize eyes with photocoagulation scars.

COMMENT

From the DCCT/EDIC perspective, the most important finding of this substudy is that, in the subset of subjects with dual images, the effects of DCCT intensive (relative to conventional) treatment on most measures of retinopathy progression were reasonably similar when assessed from digital compared with film images (Tables 3 and 4). For assessment of retinopathy severity level along the multistep ETDRS scale, agreement between gradings from film and digital images was also substantial ($\kappa = 0.70$) but appeared to be slightly lower than corresponding film vs film comparisons in the DCCT ($\kappa = 0.72$-0.84) and the more contemporaneous EDIC ($\kappa = 0.69$-0.80).

The comparability of grading digital vs film images for classification of DR severity has been described previously by others.[24,26-29] While some previous studies used the full ETDRS 7SF (7 standard field) imaging procedure,[27,29] others modified it by reducing the number of 30° fields or substituting wide-angle fields, switching to monochrome rather than color, dispensing with stereoscopic effect (in peripheral fields, or entirely), and/or using nonmydriatic (via dark adaptation) rather than pharmacologic pupillary dilation.[24,26,28] Many of these studies were primarily oriented toward screening programs for the purpose of referring persons with clinically important retinopathy to ophthalmologic care rather than conducting clinical trials or epidemiological studies. Most of these articles concluded that the comparability between film and digital grading was adequate to justify adoption of the digital medium for various clinical purposes. Thanks to these precedent studies, we were made aware of the limitations in emerging digital practice and were able to address some of these difficulties.

The DCCT/EDIC digital vs film ancillary study is the first formal comparison to be reported by an ongoing, multicenter clinical trial or epidemiological study. Several of our study design and implementation features may have enhanced the comparability between film and digital imaging for DR assessment: modern digital fundus cameras with higher spatial resolution, photographers and camera systems certified for digital performance, full ETDRS 7SF stereo imaging, standardized tonal enhancement of digital images to filmlike standard, and certified graders at a central reading center experienced in evaluating DR for many years with film and for the past few years with digital images.

A weakness of our study was the small number of cases with severe NPDR, severe PDR, and mild PDR in the absence of photocoagulation scars, resulting in lower power to examine differences between digital and film in these categories. We recruited all subjects within a specified time period rather than recruiting a stratified sample, and these levels are infrequent in our subjects. In most populations, severe NPDR is rare, being an acute stage through which eyes pass relatively quickly on their way to developing PDR.[15]

For retinopathy studies requiring discrimination between all of the individual levels on the ETDRS severity scale, we emphasize that we found worse performance currently with digital images at 2 points on the DR scale. For the presence of any retinopathy (driven at the lower end by microaneurysms only), digital sensitivity was 72% and its false-positive rate was 28%. For moderate NPDR (levels 43 and 47, driven mostly by intraretinal microvascular abnormalities), digital sensitivity was 75% and its false-negative rate was 25%. Our more recent work suggests that supplementing the view of the full-color image with the monochromatic green channel (the latter extracted from the former) improves performance of digital photography.[30] The green channel view maximizes the contrast of DR abnormalities against the retinal pigment epithelial background compared with the full-color view.

For studies that require evaluation of macular edema from fundus photography rather than ocular coherence tomography, we must also caution that sensitivity for detecting CSME with digital images appeared to be lower than with film, although this condition was too infrequent in our sample to draw robust conclusions. Our digital vs film results for CSME suggest high specificity (98%) but low sensitivity (50%) and a high false-negative rate (50%). Of note, most present-day clinical trials in ophthalmology now study diabetic macular edema primarily with ocular coherence tomography, which measures retinal thickening objectively rather than with grading of stereo color photographs (as done historically). However, the DCCT/EDIC has not yet elected to add ocular coherence tomographic examination, given the low incidence of CSME in our cohort. Work is ongoing at the reading center to improve grading of macular edema from digital photographs.

Given our ancillary study's finding of overall comparability of digital vs film gradings for evaluation of DR severity, the DCCT/EDIC Research Group and its external advisory committee voted in 2009 to approve the switch from film to digital imaging. At present, all 28 clinics have changed to digital photography.

In the context of a multicenter, long-term study, we found that ETDRS severity levels (the major DCCT/EDIC retinopathy outcomes) and our study conclusions drawn from them are comparable when DR is graded from digital rather than film images. Overall, these results support transition from the film to the digital imaging medium for research documentation of diabetic retinopathy.

Submitted for Publication: August 30, 2010; final revision received August 30, 2010; accepted October 5, 2010.
Correspondence: Larry D. Hubbard, MAT, Department

of Ophthalmology and Visual Sciences, University of Wisconsin, Madison, 8010 Excelsor Dr, Ste 100, Madison, WI 53717-0568 (hubbard@rc.opth.wisc.edu).

Group Information: A complete list of participants in the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study research group was published in *Arch Ophthalmol*. 2008; 126(12):1713.

Financial Disclosure: The authors report contributions from Abbott, Animas, Aventis, BD Bioscience, Bayer, (donated one time in 2008) Can-AM, Eli Lilly, Lifescan, Medtronic Minimed, Omron, Roche, and OmniPod to the trial, not attributed to any individual author.

Funding/Support: This study is supported by contracts with the Division of Diabetes, Endocrinology, and Metabolic Diseases of the National Institute of Diabetes and Digestive and Kidney Diseases (DK 034818), the National Eye Institute, the National Institute of Neurological Disorders and Stroke, the General Clinical Research Centers Program, the Clinical and Translational Science Awards Program, the National Center for Research Resources, and by Genentech through a Cooperative Research and Development Agreement with the National Institute of Diabetes and Digestive and Kidney Diseases.

Online-Only Material: The eTable and eFigures are available at http://www.archophthalmol.com.

REFERENCES

1. The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med*. 1993; 329(14):977-986.
2. The effect of intensive diabetes treatment on the progression of diabetic retinopathy in insulin-dependent diabetes mellitus: the Diabetes Control and Complications Trial. *Arch Ophthalmol*. 1995;113(1):36-51.
3. Diabetes Control and Complications Trial Research Group. Progression of retinopathy with intensive versus conventional treatment in the Diabetes Control and Complications Trial. *Ophthalmology*. 1995;102(4):647-661.
4. Epidemiology of Diabetes Interventions and Complications (EDIC) Research Group. Epidemiology of Diabetes Interventions and Complications (EDIC). Design, implementation, and preliminary results of a long-term follow-up of the Diabetes Control and Complications Trial cohort. *Diabetes Care*. 1999;22(1):99-111.
5. The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Retinopathy and nephropathy in patients with type 1 diabetes four years after a trial of intensive therapy. *N Engl J Med*. 2000;342(6):381-389.
6. White NH, Sun W, Cleary PA, et al. Prolonged effect of intensive therapy on the risk of retinopathy complications in patients with type 1 diabetes mellitus: 10 years after the Diabetes Control and Complications Trial. *Arch Ophthalmol*. 2008; 126(12):1707-1715.
7. Writing Team for the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Sustained effect of intensive treatment of type 1 diabetes mellitus on development and progression of diabetic nephropathy: the Epidemiology of Diabetes Interventions and Complications (EDIC) study. *JAMA*. 2003;290(16):2159-2167.
8. Martin CL, Albers J, Herman WH, et al; DCCT/EDIC Research Group. Neuropathy among the diabetes control and complications trial cohort 8 years after trial completion. *Diabetes Care*. 2006;29(2):340-344.
9. Nathan DM, Cleary PA, Backlund JY, et al; Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Study Research Group. Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *N Engl J Med*. 2005;353(25):2643-2653.
10. Donner A, Eliasziw M. A goodness-of-fit approach to inference procedures for the kappa statistic: confidence interval construction, significance-testing and sample size estimation. *Stat Med*. 1992;11(11):1511-1519.
11. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85(3):257-268.
12. Diabetic retinopathy study. Report number 6: design, methods, and baseline results: report number 7: a modification of the Airlie House classification of diabetic retinopathy: prepared by the Diabetic Retinopathy. *Invest Ophthalmol Vis Sci*. 1981;21(1, pt 2):1-226.
13. 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(5)(suppl):786-806.
14. Hubbard LD, Danis RP, Neider MW, et al; Age-Related Eye Disease 2 Research Group. 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.
15. Gardner TW, Sander B, Larsen ML, et al. An extension of the Early Treatment Diabetic Retinopathy Study (ETDRS) system for grading of diabetic macular edema in the Astemizole Retinopathy Trial. *Curr Eye Res*. 2006;31(6):535-547.
16. Early Treatment Diabetic Retinopathy Study Research Group. Fundus photographic risk factors for progression of diabetic retinopathy. ETDRS report number 12. *Ophthalmology*. 1991;98(5)(suppl):823-833.
17. Shoukri M. *Measures of Interobserver Agreement*. Boca Raton, FL: Chapman & Hall/CRC; 2004.
18. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70(4):213-220.
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
20. Bhapkar VP. A note on the equivalence of two test criteria for hypotheses in categorical data. *J Am Stat Assoc*. 1966;61:228-235. doi:10.2307/2283057.
21. McNEMAR Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*. 1947;12(2):153-157.
22. Feinstein AR, Cicchetti DV. High agreement but low kappa I: the problems of two paradoxes. *J Clin Epidemiol*. 1990;43(6):543-549.
23. Cochran WG. The combination of estimates from different experiments. *Biometrics*. 1954;10:101-120. doi:10.2307/3001666.
24. 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.
25. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76:378-382. doi:10.1037/h0031619.
26. Bursell SE, Cavallerano JD, Cavallerano AA, et al; Joslin Vision Network Research Team. 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.
27. Fransen SR, Leonard-Martin TC, Feuer WJ, Hildebrand PL; Inoveon Health Research Group. Clinical evaluation of patients with diabetic retinopathy: accuracy of the Inoveon diabetic retinopathy-3DT system. *Ophthalmology*. 2002;109 (3):595-601.
28. Rudnisky CJ, Tennant MT, Weis E, Ting A, Hinz BJ, Greve MD. Web-based grading of compressed stereoscopic digital photography versus standard slide film photography for the diagnosis of diabetic retinopathy. *Ophthalmology*. 2007; 114(9):1748-1754.
29. 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.
30. Reimers JL, Gangaputra S, Esser B, et al. Green channel vs color retinal images for grading diabetic retinopathy in DCCT/EDIC [ARVO abstract 2285]. *Invest Ophthalmol Vis Sci*. 2010;51:e-Abstract 2285. doi:10.1167/iovs.10-6303.