Journal of Diabetes Science and Technology
Volume 1, Issue 4, July 2007

ORIGINAL ARTICLES

Accuracy of Digital Images for Assessing Diabetic Retinopathy

Hyungjin Myra Kim, Sc.D.,[1] Julie C. Lowery, Ph.D.,[2] and Ronald Kurtz, M.D.[3]

Abstract

Background:
To determine the accuracy of diabetic retinopathy status assessments with and without pupil dilation using digital fundus photographs acquired by a clinic staff person and interpreted remotely by ophthalmologists.

Method:
Using early treatment diabetic retinopathy study (EDTRS) grading criteria, diabetic retinopathy status assessments were made by an experienced (nonphysician) retinal grader (NPG) based on seven standard field 35-mm stereoscopic slides acquired by an experienced ophthalmic photographer. These assessments were compared with those of the same eyes made by two ophthalmologists and the NPG using digital photographs acquired by a clinic staff person using a high-resolution (800 × 600) digital color camera system.

Results:
Based on 35-mm slides, 38% of 244 diabetic patients had ETDRS ≥35 in at least one eye and 5% had vision-threatening diabetic retinopathy (ETDRS ≥53 or macular edema). The proportion of ungradable images was significantly greater for nonmydriatic than mydriatic assessments (30% versus 10% ungradable as determined by the NPG). For ETDRS level ≥35, specificity ranged from moderate to high (0.70 to 0.96) for the three graders, while sensitivity was poor to moderate (0.38 to 0.71), and the area under the receiver-operating characteristic curves was less than satisfactory (0.67 to 0.71).

Conclusions:
The low sensitivity of the digital assessments indicates a significant proportion of patients in need of referral would not have been referred. These findings suggest that implementation of a simplified screening system using nonphotographer clinic staff acquiring nonmydriatic images, with interpretation by an ophthalmologist, should take place with an understanding of potential limitations.

J Diabetes Sci Technol 2007;1(4):531-539

Author Affiliations: [1]Health Services Research & Development Center of Excellence, VA Ann Arbor Healthcare System, Center for Statistical Consultation and Research, University of Michigan, Ann Arbor, Michigan; [2]Health Services Research & Development Center of Excellence, VA Ann Arbor Healthcare System, Ann Arbor, Michigan; and [3]Department of Ophthalmology, University of California, Irvine, California, and Department of Ophthalmology and Visual Science, Kellogg Eye Center, University of Michigan, Ann Arbor, Michigan

Abbreviations: (AUROC) area under the receiver-operating characteristic, (DR) diabetic retinopathy, (ETDRS) early treatment diabetic retinopathy study, (NPDR) nonproliferative diabetic retinopathy, (NPG) nonphysician grader, (VTDR) vision-threatening diabetic retinopathy

Keywords: diabetic retinopathy, screening, sensitivity and specificity, telemedicine

Corresponding Author: Julie C. Lowery, Ph.D., Health Services Research & Development Center of Excellence (11H) VA Ann Arbor Healthcare System, 2215 Fuller Road, Ann Arbor, MI 48105; email address julie.lowery@va.gov

Introduction

There is sound rationale for screening for the presence of diabetic retinopathy (DR). Diabetic patients are 25 times more likely than the general population to become blind, with an estimated 40,000 Americans developing severe vision loss and irreversible blindness annually.[1,2] DR will affect most diabetic patients at some stage during the course of their lifetime, leading to significant health care and disability costs.[3] If DR is detected and treated early enough, the risk of vision loss and blindness, as well as the complexity and cost of treatment, can be reduced significantly, utilizing well-established and widely available treatments.[4–10]

The desirability of DR screening has led to well-established clinical recommendations by the American Diabetes Association, American Academy of Ophthalmology, and other professional societies and health care organizations. Despite these recommendations, only about half of diabetic patients in Medicaid (48.6%) or commercial health insurance plans (54.8%) received recommended DR screening examinations in 2006, and about two-thirds of Medicare patients (66.5%) received screening.[10]

To increase rates of DR screening, a number of alternatives to direct examination by an ophthalmologist or optometrist have been proposed and tested utilizing various retinal-imaging technologies. Standardized, stereoscopic 35-mm retinal photography through a dilated pupil, with image grading by a specially trained retinal grader, is a gold standard for retinal imaging and examination that has been established in several prospective clinical trials.[7–9,11,12] While highly useful for clinical trials, the high acquisition, interpretation, and storage costs of this technique have limited its application as a screening method.

Alternative screening techniques that use digital imaging by ophthalmic photographers and remote interpretation by trained graders (telemedicine) are being used as an alternative to in-person examinations or 35-mm slides. These successful programs have led some to consider even more streamlined protocols using digital image acquisition by trained, but nonspecialized, medical office staff with review by local ophthalmic or optometric staff. The purpose of the current study was to determine if a screening technique that incorporated these methods met expectations for DR disease detection accuracy in a likely target population.

Research Design and Methods

Study Design

The study had a prospective cohort design. Remote interpretations of digital fundus images acquired by clinic staff through both dilated and undilated pupils were compared with those from gold standard, photographer-acquired, seven standard field stereo images using a modified early treatment diabetic retinopathy study (ETDRS) grading scale.[8,11]

Study Population

Participants were outpatients from general medicine clinics at a Department of Veterans Affairs Medical Center diagnosed with type 2 diabetes mellitus for at least 5 years. Exclusion criteria were (1) any previous treatment for retinopathy (e.g., laser surgery) and (2) blind or nearly blind in either eye. A research assistant telephoned eligible patients and explained the study to them. Patients who agreed to participate were scheduled for an appointment to have their eyes photographed. Participants were mailed a copy of the informed consent document. Written consent was obtained at the time of the visit. Patients were paid $50 for their participation. Enrollment occurred between June 2000 and October 2001.

Retinal Imaging

A digital camera vendor trained a clinic administrative assistant (less than half a day) on use of the equipment. All digital photos were taken using a nonmydriatic digital fundus camera (Canon CR6-45NM) plus a Medical Video Concepts Image Capture high-resolution (800 × 600) digital color camera system. Two photos of each eye without mydriasis were taken first: one 45° digital image nasal to disk and a second 45° digital image temporal to disk. The patients' eyes were then dilated with 1% Midriacyl (tropicamide). After a period of 20 minutes, two identical photos of each eye were again taken with the digital fundus camera. An experienced ophthalmic photographer then completed the seven standard field photographs using a 35-mm fundus camera.

Grading

The standard ETDRS grading system was used for the 35-mm images. A slightly modified grading system was used for the digital images, as these did not have standard photographs available and did not provide stereoscopic images. An experienced nonphysician grader (NPG), trained at the Fundus Photograph Reading Center, University of Wisconsin, examined both 35-mm and digital images. Two ophthalmologists, one of whom was a retinal subspecialist and the other a retinal fellow, also graded two random subsets of digital images. The digital assessments were made without results from the assessments using the 35-mm images, and vice versa. The nonmydriatic and mydriatic digital images for a given patient were sent to the graders at different time periods, with a minimum of 1 month lag in between, to mitigate possible confounding effects of reviewing sets of photographs of the same patient. The graders were blinded to each other's assessments, and no patient-identifying information was included with the photographs.

Statistical Analysis

Ungradable 35-mm images (as determined by the nonphysician grader) were excluded from the analysis. Using assessments based on 35-mm images as the gold standard, the diagnostic accuracy of the three graders' assessments was evaluated for both mydriatic and nonmydriatic digital images.

The primary outcome was per-patient diagnostic accuracy of the digital system in making dichotomous referral decisions. To allow comparison of our results with other published results, patient referral was defined in four ways based on the presence of (1) ETDRS level ≥35, (2) ETDRS level ≥53, (3) macular edema (including "questionable"), and (4) vision-threatening diabetic retinopathy (VTDR, defined as ETDRS level ≥53 and/or macular edema) in at least one eye. Secondary outcome was per-eye diagnostic accuracy of clinical retinopathy defined similarly. Each ungradable digital image was classified as positive (i.e., warranting a patient referral) in order to more closely approximate an actual screening system. As sensitivity analyses, analyses were repeated after excluding ungradable images.

For each of the per-patient and per-eye assessments, agreement between 35-mm and digital image assessments was measured using κ statistics.[13] In addition, using the 35-mm image assessment as the gold standard, sensitivity, specificity, and area under the receiver-operating characteristic (AUROC) curve were calculated along with the 95% confidence intervals.

Results

Demographics

Two hundred and forty-nine patients were enrolled into the study. After excluding those with an ungradable 35-mm image and those without at least one gradable digital image, data from 244 different patients and 485 eyes were included in the analyses. The primary grader, NPG, graded all three sets of 485 35-mm images and mydriatic and nonmydriatic digital images, while the two ophthalmologists graded different random subsets of mydriatic and nonmydriatic digital images. Mean age of the sample was 65.0 years (SD = 9.8, *N* = 228); 93% of the sample were 50 years or older. Ninety-eight percent (of 232) were male. Eighty-eight percent of the sample categorized themselves as "white," 8% as "black," and 4% as "other" (*N* = 229). The mean number of years since they were diagnosed with diabetes was 12.1 years (SD = 7.4, *N* = 229).

Patient Referrals Based on 35-mm Slides

The prevalence of patients requiring referral based on the NPG's assessment of 35-mm slides was 38.1% for ETDRS level ≥35 in at least one eye and 3.8% for ETDRS level ≥53 (**Table 1**). The prevalence of patients with VTDR was 5.0%.

Agreement between 35-mm and Digital Assessments: Patient Referrals

The percentage of patients assigned a rating of "cannot grade" to at least one of the digital images varied significantly among graders (with the NPG consistently assigning a rating of "cannot grade" to a higher proportion of patients) and varied somewhat across the diagnoses (i.e., retinopathy, macular edema, or VTDR) (**Table 1**). For all three graders, a much larger proportion of nonmydriatic images than mydriatic images were ungradable. For example, for digital image assessments made by the NPG, 10% of the mydriatic images were ungradable for retinopathy, whereas 30% of the nonmydriatic digital images were ungradable.

Both mydriatic and nonmydriatic digital assessments generally underestimated the prevalence of patients requiring referral for DR level ≥35 (**Table 1**). Referral rates for DR ≥53 and macular edema were overestimated, especially for nonmydriatic assessments. The large degree of overestimation for nonmydriatic assessments was primarily because of the substantial proportion of ungradable images, all of which were included as positive referrals. When the referral rate was analyzed after removing patients with ungradable digital images of one or both eyes from the analysis, the magnitude of overestimation decreased for both mydriatic and nonmydriatic assessments and resulted in a greater underestimation of the prevalence of patients requiring referral defined as DR ≥35 (results not shown).

As measured by the κ statistic, agreement between digital and 35-mm assessments tended to be no better than

Table 1.
Per-Patient Analyses—Percentage of Ungradable Digital Images, Prevalence of Patients Requiring Referral, and Agreement between 35-mm and Digital Image Assessments (κ)

	Mydriatic			Nonmydriatic		
Graders[a]	A	B	C	A	B	C
Retinopathy (at least one eye)						
N	239	136	146	239	137	216
% ungradable digital images[b]	10.4	6.6	3.4	30.5	23.4	4.6
Mild NPDR or worse (ETDRS level ≥35)						
% prevalence in 35-mm[c]	38.1	38.2	36.3	36.4	38.7	37.0
% prevalence in digital	28.9	20.6	16.4	44.8	32.1	16.7
κ	0.55	0.45	0.51	0.40	0.41	0.37
Severe NPDR or worse (ETDRS level ≥53)						
% prevalence in 35-mm[c]	3.8	2.2	4.1	2.9	1.5	3.2
% prevalence in digital	12.1	7.4	6.2	33.5	24.8	8.3
κ	0.05	0.12	0.37	0.02	0.03	0.37
Macular edema (at least one eye)						
N	241	135	145	241	137	216
% ungradable digital images[b]	8.7	6.7	4.8	34.4	23.4	7.9
% prevalence in 35-mm[c]	4.1	2.2	4.8	4.1	2.9	4.2
% prevalence in digital	12.5	11.1	11.0	36.5	24.8	13.0
κ	0.20	0.19	0.39	0.10	0.00	0.34
Vision-threatening diabetic retinopathy[d]						
N	238	135	145	239	137	216
% ungradable digital images[b]	12.2	6.7	4.8	34.3	23.4	8.3
% prevalence in 35-mm[c]	5.0	3.0	5.5	4.6	3.7	5.1
% prevalence in digital	16.4	11.1	11.7	36.8	25.6	14.4
κ	0.17	0.17	0.48	0.09	-0.01	0.33

[a] A, nonphysician grader; B, retinal subspecialist; C, retinal fellow.
[b] Digital images for a patient are considered ungradable if one eye is negative for the diagnosis and the other eye has an ungradable digital image or digital images for both eyes are ungradable. Ungradable digital images for a patient were considered positive for the diagnosis.
[c] Only grader A graded 35-mm images, but prevalence of a positive diagnosis for 35-mm images varies across graders and mydriasis according to the sample of patients each graded.
[d] Defined as ETDRS ≥53 and/or macular edema in at least one eye.

moderate, but it was worse for nonmydriatic images and when referral was defined at ≥53 than at ≥35 (**Table 1**). The low κ values for referral defined at ≥53, macular edema, and VTDR are partly because of a known paradox of κ as a measure of agreement when the statistic being compared (i.e., referral rate) is small in one of the two groups (35-mm slides in this case).[13]

The diagnostic accuracy estimates of patient referral assessments were generally better than a coin toss (**Table 2**), but no grader had an AUROC curve greater than 0.85. Specificity was generally higher than sensitivity regardless of the use of mydriatics. The contrast between specificity and sensitivity was larger for mydriatic than nonmydriatic image assessments. The analysis after deleting patients with ungradable digital images did not improve sensitivity, but increased specificity to close to 100% (results not shown).

Agreement between 35-mm and Digital Assessments: Per-Eye Analysis

For digital image assessments made by the NPG, 7% of the mydriatic images were ungradable for retinopathy, whereas 25% of the nonmydriatic digital images were ungradable (**Table 3**). For the two ophthalmologists who assessed subsets of the digital images, the percentage of ungradable images was similarly higher for nonmydriatic than for mydriatic images.

The prevalence of eyes with an ETDRS level ≥35 was about 30% (based on NPG's assessment of 35-mm

Table 2.
Per-Patient Analyses—Diagnostic Accuracy of Patient Referral Based on Digital Images (with 35-mm Assessments as Reference Standard)

	Mydriatic			Nonmydriatic		
Graders[a]	A	B	C	A	B	C
At least one eye with ETDRS ≥35						
Sensitivity/specificity	0.62/0.91	0.46/0.95	0.45/1.00	0.71/0.70	0.57/0.83	0.38, 0.96
Area (95% CI for area)[b]	0.76 (0.71, 0.82)*	0.71 (0.63, 0.78)*	0.73 (0.66, 0.79)*	0.71 (0.65, 0.77)*	0.70 (0.62, 0.78)*	0.67 (0.61, 0.72)*
At least one eye with ETDRS ≥53						
Sensitivity/specificity	0.22/0.88	0.33/0.93	0.50/0.96	0.43/0.67	0.50/0.76	0.71/0.94
Area (95% CI for area)[b]	0.55 (0.41, 0.70)	0.63 (0.31, 0.96)	0.73 (0.51, 0.95)*	0.55 (0.35, 0.75)	0.63 (0.14, 1.00)	0.83 (0.64, 1.00)*
At least one eye with macular edema						
Sensitivity/specificity	0.50/0.89	0.67/0.90	0.71/0.92	0.80/0.65	0.25/0.75	0.78/0.90
Area (95% CI for area)[b]	0.70 (0.53, 0.86)*	0.78 (0.46, 1.00)	0.82 (0.64, 1.00)*	0.73 (0.59, 0.86)*	0.50 (0.25, 0.75)	0.84 (0.69, 0.98)*
At least one eye with vision-threatening diabetic retionpathy[c]						
Sensitivity/specificity	0.54/.85	0.40/0.90	0.78/0.92	0.73/0.65	0.20/0.40	0.73/0.89
Area (95% CI for area)[b]	0.70 (0.55, 0.84)*	0.65 (0.41, 0.89)	0.85 (0.70, 0.99)*	0.69 (0.55, 0.83)*	0.47 (0.27, 0.67)	0.81 (0.67, 0.95)*

[a] A, nonphysician grader; B, retinal subspecialist; C, retinal fellow.
[b] Area under the receiver-operating characteristic curve with 95% confidence interval; an asterisk indicates an area significantly greater than 0.5.
[c] Defined as ETDRS ≥53 and/or macular edema in at least one eye.

slides), which mydriatic digital assessments by the NPG underestimated at 22% (**Table 3**). The prevalence of eyes with an ETDRS level ≥53 was 3%, which mydriatic digital assessments by the NPG overestimated at 8%. In addition, mydriatic digital assessments overestimated the prevalence for macular edema and VTDR. The trend was similar for nonmydriatic assessments and assessments by the two ophthalmologists. As with the patient analysis, the reason for the overestimation of prevalence with ETDRS level ≥53, macular edema, and VTDR, especially for nonmydriatic images, is primarily because of the substantial number of ungradable digital images, which are included as positive for the diagnosis.

The agreement between 35-mm and digital image assessments was generally poor, but tended to be higher for mydriatic assessments than nonmydriatic assessments, regardless of the ETDRS cutoff (**Table 3**). However, the overall diagnostic accuracy as measured by AUROC curves was similar between mydriatic and nonmydriatic assessments, but the areas were generally below 0.75 (**Table 4**). Similar to per-patient analyses, specificity was higher than sensitivity for all three graders. Sensitivity was generally higher and specificity lower for nonmydriatic image assessments compared with mydriatic image assessments because a large proportion of the nonmydriatic images were ungradable and therefore considered positive. The overall agreement (κ) and accuracy (AUROC curve) results were similar when ungradable digital images were excluded from the analysis, but with decreased sensitivity and increased specificity (results not shown) for both mydriatic and nonmydriatic assessments.

Ungradable Digital Images

Of the ungradable digital images graded by NPG, the most common reasons given for ungradable images were haziness (67%) for 36 mydriatic digital images and shadow (74%) or haziness (19%) for 122 nonmydriatic digital images. Of patients with at least one ungradable eye, both eyes of a patient were ungradable in 58% by grader A (NPG), 47% by grader B, and 50% by grader C, showing high correlation between two eyes of a patient with respect to ungradable images across graders.

Intergrader Agreement

Intergrader agreement among the three graders measured with κ ranged between 0.35 and 0.62 for per-eye analysis and between 0.33 and 0.67 for per-patient analysis and was better for mydriatic assessments compared with nonmydriatic assessments (**Table 5**). Although

Table 3.
Per-Eye Analyses—Percentage of Ungradable Digital Images, Prevalence of Eyes Requiring Referral, and Agreement between 35-mm and Digital Image Assessments (κ)

	Mydriatic			Nonmydriatic		
Graders [a]	A	B	C	A	B	C
Retinopathy						
N	485	277	293	485	279	436
% ungradable digital images [b]	7.4	5.4	2.1	25.2	18.6	3.4
Mild NPDR or worse (ETDRS level ≥35)						
% prevalence in 35-mm [c]	30.3	31.4	27.7	28.9	31.9	29.4
% prevalence in digital	21.9	16.3	12.6	34.9	25.8	13.1
κ	0.53	0.42	0.47	0.42	0.38	0.40
Severe NPDR or worse (ETDRS level ≥53)						
% prevalence in 35-mm [c]	3.1	1.4	3.4	2.9	1.4	3.2
% prevalence in digital	7.6	6.1	4.1	25.4	19.7	6.7
κ	0.03	0.17	0.43	0.01	0.04	0.44
Macular edema						
N	484	276	292	485	279	436
% ungradable digital images [b]	4.6	5.4	2.4	25.2	18.6	5.5
% prevalence in 35-mm [c]	3.1	1.5	3.4	3.1	1.8	3.4
% prevalence in digital	7.2	7.6	6.9	26.8	19.7	9.4
κ	0.33	0.14	0.44	0.09	0.04	0.40
Vision-threatening diabetic retinopathy [d]						
N	484	276	292	485	279	436
% ungradable digital images [b]	7.6	5.4	2.4	26.4	18.6	6.0
% prevalence in 35-mm [c]	4.3	2.2	5.1	4.1	2.5	4.6
% prevalence in digital	11.5	8.0	8.2	28.3	20.1	11.0
κ	0.26	0.19	0.53	0.11	0.02	0.43

[a] A, nonphysician grader; B, retinal subspecialist; C, retinal fellow.
[b] Ungradable digital images for a patient were considered positive for the diagnosis.
[c] Only Grader A graded 35-mm images, but prevalence of a positive diagnosis for the 35-mm images varies across graders and mydriasis according to the sample of patients each graded.
[d] Defined as ETDRS ≥53 and/or macular edema.

agreement among the three graders is only moderate at best, intergrader agreement for digital images was better than the agreement between assessments by 35-mm and digital images, both for patient referral (**Tables 1** and **2**) and for image grading (**Tables 3** and **4**).

Conclusions

Incorporating digital retinal photography into a comprehensive primary care visit has the potential to improve patient compliance with preventative screening initiatives for DR. A number of reports have documented high levels of accuracy for such screening utilizing relatively sophisticated technical and grading methods, as well as a dedicated staff for image acquisition and interpretation. This success has sparked interest in less sophisticated screening programs. The primary purpose of this study was to determine the accuracy of retinopathy status assessments made under conditions that might commonly exist in such ad hoc programs, including (1) the acquisition of digital images by an inexperienced retinal photographer using commonly available equipment with and without mydriasis and (2) image grading by a practicing ophthalmologist.

The British Diabetic Association has proposed a sensitivity of >80%, a specificity of >95%, and a technical failure rate of <5% as standards for effective screening criteria.[14] Most published studies of the accuracy of digital images, which use seven standard field 35-mm stereoscopic slides as reference criteria, meet (or come close to meeting) these standards,[15–20] but our findings did not. One other study obtained similarly low sensitivity rates for macular edema.[21] These studies have also shown moderate

Table 4.
Per-Eye Analyses—Diagnostic Accuracy of Eyes Based on Digital Images (with 35-mm Assessments as Reference Standard)

	Mydriatic			Nonmydriatic		
Graders[a]	A	B	C	A	B	C
ETDRS ≥35						
Sensitivity/specificity	0.56/0.93	0.41/0.95	0.41/0.98	0.66/0.78	0.51/0.86	0.37/0.97
Area (95% CI for area)[b]	0.74 (0.70, 0.79)*	0.68 (0.63, 0.74)*	0.69 (0.64, 0.75)*	0.72 (0.68, 0.74)*	0.68 (0.62, 0.74)*	0.67 (0.62, 0.71)*
ETDRS ≥53						
Sensitivity/specificity	0.13/0.93	0.50/0.95	0.50/0.98	0.29/0.75	0.50/0.81	0.71/0.96
Area (95% CI for area)[b]	0.53 (0.44, 0.62)	0.72 (0.44, 1.00)	0.74 (0.57, 0.90)*	0.52 (0.39, 0.64)	0.65 (.37, 0.94)	0.83 (0.71, 0.96)*
Macular edema						
Sensitivity/specificity	0.60/0.94	0.50/0.93	0.70/0.95	0.67/0.74	0.40/0.81	0.80/0.93
Area (95% CI for area)[b]	0.77 (0.64, 0.90)*	0.72 (0.43, 1.00)*	0.83 (0.68, 0.98)*	0.71 (0.58, 0.83)*	0.60 (0.36, 0.84)*	0.87 (0.76, 0.97)*
Vision-threatening diabetic retionpathy[c]						
Sensitivity/specificity	0.52/0.91	0.50/0.93	0.73/0.95	0.70/0.74	0.29/0.80	0.80/0.92
Area (95% CI for area)[b]	0.72 (0.61, 0.83)*	0.71 (0.49, 0.93)	0.84 (0.73, 0.96)*	0.72 (0.61, 0.82)*	0.54 (0.36, 0.73)	0.86 (0.77, 0.95)*

[a] A, nonphysician grader; B, retinal subspecialist; C, retinal fellow.
[b] Area under the receiver-operating characteristic curve with 95% confidence interval; an asterisk indicates an area significantly greater than 0.5.
[c] Defined as ETDRS ≥53 and/or macular edema.

Table 5.
Intergrader Agreement for Image Assessments (Clinical Retinopathy Diagnosis and Macular Edema) and for Patient Referrals Based on Digital Images[a,b]

	Image grading				Patient referrals			
	ETDRS ≥35	ETDRS ≥53	Macular edema	VTDR[c]	ETDRS ≥35	ETDRS ≥53	Macular edema	VTDR[c]
Dilated	0.62	0.42	0.57	0.54	0.67	0.40	0.63	0.54
Nondilated	0.47	0.35	0.46	0.49	0.45	0.33	0.51	0.48

[a] Cell entries are κ.
[b] These analyses only include data where at least two graders made assessments for the same image. All κ values had *p* values <0.0001, testing for κ > 0; ungradable digital images were considered positive for image grading, and ungradable in at least one eye was considered positive for patient referral.
[c] Vision-threatening diabetic retinopathy is defined as ETDRS ≥53 and/or macular edema.

(κ between 0.4 and 0.6) to substantial (κ between 0.6 and 0.8) agreement, whereas our study showed low agreement for both image grading and patient referral.

Multiple factors could account for our poorer results. One possible explanation is the absence of an adjudication system, which was used in at least two previous studies.[15,16] The resolution of our camera (800 × 600) was similar to most other studies (except for Fransen *et al.*,[15] who used 1152 × 1152 resolution), but others took more images (three to seven) per eye.[15,16,19,21]

In addition, as noted earlier, by design the photographer in our study (the clinic administrative assistant) was not an experienced retinal photographer. This is evident by the higher percentage of ungradable images observed compared with those in other studies,[16,18,20] and we did not find the proportion of ungradable images to decrease as the imager gained more experience. Some studies used experienced ophthalmic photographers[15,18] or did not specify the type of training received by their photographers.[16,17,21] Training provided to our ophthalmologist graders was also limited. The digital graders in most studies are specially trained retinal graders.[15–17,20] Interestingly, Newsom and colleagues,[21] who used an approach similar to ours with ophthalmologist graders, obtained low accuracy readings for macular edema that were comparable to our findings.[21] Although ETDRS criteria were used for both sets of images, the grading for digital images was slightly different to account for definitions inherent in the standard ETDRS 35-mm system that could not be applied directly to the digital images. The specific grading criteria used by other studies are, in general, unpublished.

Other studies also tended to differ from ours in the proportion of type 1 versus type 2 diabetic patients, patient age, duration of diabetes, and the prevalence of retinopathy. Our patients were all type 2 diabetic patients, with a 30% prevalence of mild or worse nonproliferative diabetic retinopathy (NPDR) (ETDRS ≥35) in eyes and about a 5% prevalence of patients with VTDR. The prevalence of our sample is similar to the pooled prevalence published using eight population-based eye surveys in the United States.[4] However, other published studies had a substantial proportion of type 1 diabetes[16] or a higher prevalence of patients with VTDR.[15,16]

The mean age of patients in our study was 65, with 93% greater than or equal to 50 years of age. This contrasts sharply with other studies, which used younger patients.[15–17] Studies have shown that the screening failure rate increases with increasing patient age,[22] especially for nonmydriatic screening, which is consistent with the high rate of ungradable images in this study. Consistent with other studies, our study showed lower sensitivity than specificity across graders, suggesting a need to focus training on the reduction of false negatives. In addition, nonmydriatic assessments had a substantially greater percentage of ungradable images.

One review showed teleophthalmology to be an accurate and reliable test for detecting diabetic retinopathy and macular edema.[23] The review concluded that both seven and three field images yield acceptably high sensitivity and specificity values, and while a single field digital image performed as well as ophthalmoscopy in some cases, the sensitivity of single field digital imaging would likely be considered too low to function as an adoptable substitute for an ophthalmoscopic examination. Our study suggests that using two field imaging is insufficient particularly for nonmydriatic images.

In summary, a telemedicine system for detecting DR, employed in primary care physicians' offices where diabetic patients receive the majority of their care, has the potential to play a major role in improving the outcomes of diabetic patients with retinopathy. However, we found agreement between a trained grader's evaluation of retinopathy status using 35-mm slides and two ophthalmologists' evaluations using digital photos taken by an office assistant trained by the camera vendor to be only fair, and those between the same grader's evaluation using 35-mm slides and digital photos were similar. Nevertheless, the system's implementation in this study is likely to resemble actual implementation in primary care clinic settings. Findings from this study and other rigorous evaluations of the diagnostic accuracy of telemedicine technology need to guide the implementation and use of such systems, specifically for establishing policies and procedures governing such aspects as number of fields, training, grading criteria, and referral criteria.

Acknowledgement:

This study was funded by the Department of Veterans Affairs, Health Services Research & Development Service.

References:

1. National Society to Prevent Blindness: Visual problems in the US. Data, analysis, definitions, data sources, detailed tables, analysis, interpretations. New York: National Society to Prevent Blindness; 1980.
2. King H, Aubert RE, Herman WH. Global burden of diabetes, 1995-2025: prevalence, numerical estimates, and projections. Diabetes Care. 1998 Sep;21(9):1414-31.
3. Centers for Disease Control (CDC). Public health focus: prevention of blindness associated with diabetic retinopathy. MMWR Morb Mortal Wkly Rep. 1993 Mar 19;42(10):191-5.
4. Kempen JH, O'Colmain BJ, Leske MC, Haffner SM, Klein R, Moss SE, Taylor HR, Hamman RF; Eye Disease Prevalence Research Group. The prevalence of diabetic retinopathy among adults in the United States. Arch Opthalmol. 2004 Apr;122(4):552-63.
5. British Multicentre Study Group. Photocoagulation for proliferative diabetic retinopathy: a randomised controlled clinical trial using the xenon-arc. Diabetologia. 1984 Feb;26(2):109-15.
6. Diabetic Retinopathy Study Research Group. Photocoagulation treatment of proliferative diabetic retinopathy: the second report of diabetic retinopathy/vitrectomy study findings. Am J Ophthalmol. 1978;85:82-106.
7. The Diabetic Retinopathy Study Research Group. Four risk factors for severe visual loss in diabetic retinopathy. The third report from the Diabetic Retinopathy Study. Arch Ophthalmol. 1979 Apr;97(4):654-5.
8. Diabetic retinopathy study. Report Number 6. Design, methods, and baseline results. Invest Ophthalmol Vis Sci. 1981 Jul;21:149-209.
9. Anonymous. Proliferative diabetic retinopathy: treatment with xenon-arc photocoagulation. Interim report of multicentre randomised controlled trial. Br Med J. 1977 Mar 19;1(6063):739-41.
10. National Committee for Quality Assurance. The State of Health Care Quality 2006 [report on the Internet] [accessed 2006 Oct 23]. Available from: http://www.ncqa.org.

11. 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 10. Ophthalmology 1991 May;98(5 Suppl):786-806.

12. Early Treatment Diabetic Retinopathy Study Research Group. Fundus photographic risk factors for progression of diabetic retinopathy: ETDRS Report 12. Ophthalmology. 1991 May;98(5 Suppl):823-33.

13. Fleiss JL. Statistical methods for rates and proportions. New York: John Wiley; 1981.

14. British Diabetic Association. Retinal photographic screening for diabetic eye disease. London: British Diabetic Association; 1997.

15. 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 Mar;109(3):595-601.

16. Bursell SE, Cavallerano JD, Cavallerano AA, Clermont AC, Birkmire-Peters D, Aiello LP, Aiello LM; 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 Mar;108(3):572-85.

17. Taylor DJ, Fisher J, Jacob J, Tooke JE. The use of digital camera in a mobile retinal screening environment. Diabet Med. 1999 Aug;16(8):680-6.

18. Scanlon PH, Malhotra R, Greenwood RH, Aldington SJ, Foy C, Flatman M, Downes S. Comparison of two reference standards in validating two field mydriatic digital photography as a method of screening for diabetic retinopathy. Br J Ophthalmol. 2003 Oct;87(10):1258-63.

19. Massin P, Erginay A, Mehidi AB, Vicaut E, Quentel G, Victor Z, Marret M, Guillausseau PJ, Gaudric A. Evaluation of a new non-mydratic digital camera for detection of diabetic retinopathy. Diabet Med. 2003 Aug;20(8):635-41.

20. 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 Aug;134(2):204-13.

21. Newsom R, Moate B, Casswell T. Screening for diabetic retinopathy using digital colour photography and oral fluorescein angiography. Eye. 2000 Aug;14(Pt 4):579-82.

22. Scanlon PH, Foy C, Malhotra R, Aldington SJ. The influence of age, duration of diabetes, cataract, and pupil size on image quality in digital photographic retinal screening. Diabetes Care. 2005 Oct;28(10):2448-53.

23. Whited JD. Accuracy and reliability of teleophthalmology for diagnosing diabetic retinopathy and macular edema: a review of the literature. Diabetes Technol Ther. 2006 Feb;8(1):102-11.