Published in final edited form as:

Retina. 2012 September ; 32(8): 1629–1635. doi:10.1097/IAE.0b013e3182483361.

Validation of tablet-based evaluation of color fundus images

Mark Christopher^{1,*}, Daniela C. Moga^{2,3}, Stephen R. Russell⁴, James C. Folk⁴, Todd Scheetz^{1,4}, and Michael D. Abràmoff^{1,4,5,6}

¹Department of Biomedical Engineering, University of Iowa, Iowa City, IA 52242, USA

²Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, IA 52242, USA

³Biostatistics/Epidemiology/Research Design, Institute for Clinical and Translational Science, University of Iowa, Iowa City, IA 52242, USA

⁴Department of Ophthalmology and Visual Sciences, University of Iowa Hospitals and Clinics, 200 Hawkins Drive, Iowa City, IA 52242, USA

⁵Department of Veterans Affairs, Iowa City VA Medical Center, 601 Highway 6 West, Iowa City, IA 55242, USA

⁶Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242, USA

Abstract

Purpose—To compare diabetic retinopathy (DR) referral recommendations made by viewing fundus images using a tablet computer to recommendations made using a standard desktop display.

Methods—A tablet computer (iPad) and a desktop PC with a high-definition color display were compared. For each platform, two retinal specialists independently rated 1200 color fundus images from patients at risk for DR using an annotation program, Truthseeker. The specialists determined whether each image had referable DR, and also how urgently each patient should be referred for medical examination. Graders viewed and rated the randomly presented images independently and were masked to their ratings on the alternative platform. Tablet- and desktop display-based referral ratings were compared using cross-platform, intra-observer kappa as the primary outcome measure. Additionally, inter-observer kappa, sensitivity, specificity, and area under ROC (AUC) were determined.

Results—A high level of cross-platform, intra-observer agreement was found for the DR referral ratings between the platforms ($\kappa=0.778$), and for the two graders, ($\kappa=0.812$). Inter-observer agreement was similar for the two platforms ($\kappa=0.544$ and $\kappa=0.625$ for tablet and desktop, respectively). The tablet-based ratings achieved a sensitivity of 0.848, a specificity of 0.987, and an AUC of 0.950 compared to desktop display-based ratings.

Conclusions—In this pilot study, tablet-based rating of color fundus images for subjects at risk for DR was consistent with desktop display-based rating. These results indicate that tablet computers can be reliably used for clinical evaluation of fundus images for DR.

*corresponding author: mark-christopher@uiowa.edu.

Financial disclosures: None (all authors)

Contributions of Authors: Design of study (M.C., J.F., T.S., M.A.); Conduct of study (M.C., S.R., J.F., M.A.); Data collection and management (M.C., S.R., J.F., M.A.); Data analysis and interpretation (M.C., D.M., S.R., T.S., M.A.); Preparation of manuscript (M.C., M.A.); Review and approval of manuscript (M.C., D.M., S.R., J.F., T.S., M.A.).

Keywords

Diabetic retinopathy; Fundus images; iPad; Tablet computers

Introduction

Recently, tablet computers, such as the Apple iPad (Apple Corporation, Cupertino, CA) and tablet PCs (Samsung¹, HP², Motorola³, and Fujitsu⁴), have become widely available. These devices provide a substitute or supplementary viewing device for patient records and diagnostic images. The tablet computers are compact, portable, and high-capacity. In addition to a tablet's portability they can be ergonomically handled as a paper medical record, and viewed while conversing with or examining a patient. Because of their attributes that enhance face-to-face time, they offer a less obtrusive interface to electronic health records (EHR) and image archives than a standard desktop display. Several studies have shown that physicians accessing a patients' EHR using a stationary desktop display and keyboard, negatively impacts patients' perceptions of their care, especially in high-throughput outpatient settings.⁵⁻⁷ Several vendors have released tablet-based software allowing physicians to view clinical data and images^{8,9} and federal incentive programs encouraging healthcare providers to use EHR systems¹⁰ may encourage physicians to adopt tablet computers for clinical use.

There is no reason to suspect that textual or numeric information may be interpreted differently on a tablet compared to a desktop display. The critical evaluation of clinical images, such as retinal images, might differ however and any disparities could affect care. Perhaps differences in pixel resolution, contrast, or illumination of tablet displays substantially alter the visibility or appreciation of important diagnostic details. Of special concern in retinal images are small lesions such as microaneurysms, exudates, drusen, and small hemorrhages, which can serve as important diagnostic or prognostic markers.^{11,12} For example, small abnormalities may be imperceptible to physicians if resolution or time limitations preclude use of tablet zoom features to examine questionable retinal areas. We have extensive experience comparing the proficiency and detection thresholds of clinicians using desktop displays for review and analysis of retinal color fundus images for diabetic retinopathy (DR) and macular degenerations.¹³⁻¹⁹

In this study, we evaluated whether retinal specialists differed in the detection of referable DR when using a tablet compared to a high-definition desktop display. We examined the agreement and accuracy of two physicians' ratings for the presence of referable DR and severity in using a tablet to view the images and then independently, a high-definition desktop display.

Methods

Review Protocol

Two retinal specialists (JCF and SRR) rated each image for referable DR in masked fashion on each platform, an iPad with a resolution of 1024 by 768 using iOS version 3 (Apple Inc, Cupertino, CA) and a desktop PC using a high-definition stationary desktop 27" LCD display (2707WFP, Dell, Austin, Texas) with a resolution of 1920 by 1200 and standard saturation / brightness settings, keyboard, and mouse. To minimize bias, one retinal specialist rated the images on the tablet first, and then on the desktop display, whereas the other rated images using the desktop display first and then the tablet. To enhance masking, the order in which the images were displayed was re-randomized for each platform and grader, and graders were masked to any previous referral ratings. To rate each image for

referable DR, the retinal specialists were asked to determine if and when the patient should be referred from a primary care setting to an ophthalmologist or retinal expert based on the severity of the DR. The specialists chose one of the following referral ratings for each image:

- 0 - Noreferral, return in 12 months or more
- 1 - Yes referral, 6 months
- 2 - Yesreferral, 3 months
- 3 - Yes referral, immediately

Based on the response, each image was assigned a rating between 0 (no referable DR) and 3 (immediately referable DR). The retinal specialists were instructed to perform the rating as follows:

Imagine that you have only this image for this patient, that the image of the other eye has the same level of diabetic retinopathy, that the patient is in primary care, and that this is all you know about the patient. Based on the retinopathy, select whether the patient with this fundus image should return for imaging in 12 months, or be referred to an ophthalmologist/retinal specialist in 6 months, 3 months or immediately. The 12 month selection is for those with no/minimal retinopathy or anyone who doesn't require a referral to a specialist for at least one year. The 'immediately' selection means that the patient should be referred to a specialist for evaluation and possible treatment immediately. The 3 or 6 month selections may be chosen if you feel the patient doesn't need to be seen immediately but should be followed up in less than one year by a ophthalmologist/retinal specialist, and that it is not safe to simply image the patient again 12 months from now. You should just use your best judgment since there will often be no absolutely right or wrong answer.

To rate the images on a desktop computer, a standard mouse-based Java program, Truthseeker, was used.^{19,20} In order to perform tablet-based rating, Truthseeker was re-implemented for the Apple iPad platform (software will be made available in the Apple App store under the name Truthmarker). Truthseeker allows rapid setup for large image annotation projects, and has been used for cross-platform annotation of stereoscopic disc images, fundus images, and OCT images for a variety of lesions and normal structures including arteries and veins, soft and hard drusen, microaneurysms, retinal pigmented epithelial pigment, choroidal pigmentation, vessel branching and crossings.¹⁵ Implicitly incorporated into this rating scheme, is the recognition that severe DR corresponds to observation or treatment thresholds resulting in a higher numerical referral rating.

To characterize intra-grader variability, each grader was asked to rate the images one additional time using a desktop display. The images were once again randomized, graders were masked to any previous ratings, and the same review protocol was used.

Image Database

As the target population of individuals at risk of DR, we chose a set of 1200 publicly available high quality fundus images, the Messidor dataset.²¹ The Messidor images were derived from patients with diabetes seen in tertiary care clinics in France. These images are available in TIFF format using 8 bits per color channel and at resolutions ranging from 1440x960 to 2304x1526 pixels obtained with a color video 3CCD camera on a Topcon TRC NW6 non-mydiatic fundus camera with a 45 degree field of view.²¹

Data Analysis

Cross-platform, intra-observer agreement was measured by calculating the κ statistic between DR referral ratings assigned by a grader using a desktop display to those assigned by the same grader using the tablet.^{22,23} For comparison, intra-observer agreement was also measured by calculating the κ statistics between the two sets of desktop display-based ratings assigned by each of the graders. Inter-observer comparisons were performed by calculating κ statistics between DR referral ratings annotated by different graders using the same platform (for both the desktop display and tablet). In each comparison (intra- and inter-observer comparisons) both an unweighted κ statistic for binary classification and a linearly weighted κ statistic for ordinal classes was calculated. To compute the unweighted statistic, the DR referral ratings were converted to a binary classification by setting (“dichotomizing”) the rating of “no referable DR” to 0, while “immediate, 3, and 6 month referrals” were set to 1. Cross-platform agreement was also measured using a Bhapkar test for marginal homogeneity.²⁴ This test was performed by pooling the graders’ ratings for each platform and calculating a χ^2 test statistic for both the dichotomized and full scale of DR referral ratings.

The accuracy of tablet-based DR referral ratings was assessed by determining the sensitivity, specificity, and area under receiver operating characteristic curves (AUC) for each DR referral rating.²⁵ A reference standard was created by averaging the DR referral ratings assigned by the graders using a desktop display. Using this reference standard, the accuracy of the tablet-based ratings was calculated for each grader by class as DR referral rating 1 or higher, 2 or higher, and 3.

Analyses were conducted using SAS software version 9.2 of the SAS System and SPSS software version 19 for Windows.

Results

Cross-Platform and Inter-Observer Agreement

A distribution and categorization of the DR referral ratings assigned by each grader is shown in Tables 1 and 2. These tables compare the ratings assigned by each grader using the tablet with the set of referral ratings assigned by the same grader using a desktop display. The data summarized in these tables were used to perform a Bhapkar test of marginal homogeneity for both the dichotomized and full set of ordinal DR referral ratings. In both cases, the Bhapkar test showed no statistically significant difference ($p > 0.05$) between ratings assigned using a desktop display to those assigned using a tablet.

The resulting κ statistics measuring intra-observer agreement of DR referral ratings are shown in Table 3. The resulting κ statistics comparing ratings annotated using a tablet to those using a desktop display were 0.778 and 0.812 for the two graders. The weighted cross-platform κ statistics were 0.776 and 0.795 for the two graders. The intra-observer κ statistics comparing the two sets of ratings assigned using a desktop display were 0.800 and 0.784 for the two graders. The weighted κ statistics were 0.796 and 0.768.

The inter-observer κ statistics are shown in Table 4. The resulting κ statistic comparing the graders was 0.544 for tablet-based ratings and 0.625 for desktop display-based ratings. The weighted κ statistics for these DR referral ratings were 0.648 and 0.675 for the tablet and desktop display, respectively.

Tablet Grading Accuracy

Comparing the tablet DR referral ratings to the reference standard resulted in a sensitivity of 0.848, a specificity of 0.0987, and an AUC of 0.950 for the two graders in detecting DR referral rating 1 or higher. For detecting DR referral rating 2 or higher, the sensitivity was 0.948, specificity was 0.911 and AUC was 0.942. Finally, for detecting DR referral rating 3, the sensitivity, specificity, and AUC were 0.987, 0.611, and 0.801, respectively. A fuller illustration of the sensitivity, specificity, and AUC analysis are shown in Table 5.

Discussion

This pilot study shows that whether an expert views color fundus images on a tablet or a desktop display, their recommendations regarding DR referral are similar. This is relevant because evaluating color fundus images for the presence and severity of DR is a key task in preventing blindness and visual loss in this group of patients.²⁶

The comparison of DR referral ratings determined using a desktop display to ratings determined using a tablet showed high agreement between the two platforms. In addition, the κ (0.778 and 0.812) and weighted κ statistics (0.776 and 0.795) measuring cross-platform, intra-observer agreement did not differ significantly ($p > 0.05$) from the κ (0.800 and 0.784) and weighted κ (0.796 and 0.768) statistics measuring intra-observer agreement of ratings assigned using a desktop display. The similarity of ratings produced using these platforms is further supported by the results of the Bhapkar test of marginal homogeneity. This test showed no significant difference ($p > 0.05$) between ratings assigned using each of the platforms. Diagnostic accuracy for each grader, as measured by sensitivity and specificity, was also comparable to previously published results for DR diagnosis by ophthalmologists using digital fundus images.^{27,28}

As represented by the tablet-based sensitivities and specificities (Table 5), the observed sensitivity for each grader increased for higher DR referral ratings while the specificity decreased with higher DR referral ratings. This trend could be due in part to the rigorous reference standard used. The chosen, conservative reference standard used was the average of DR referral ratings assigned by the two graders using a desktop display to view the images. Because the reference standard was an average of DR referral ratings, it required that both graders independently assigned the highest DR referral rating to an image for that image to receive the highest rating in the standard. The use of this rigorous standard may have resulted in a reduction of the number of images assigned high DR referral ratings. The algebraic effect would be to decrease the observed number of false-negatives and increase the observed number of false-positives.

Tablet computers allow clinicians to review medical images and other clinical data while interacting with patients face-to-face, or for an ophthalmologist, eye-to-eye. Tablets also allow for annotations that are made during patient interactions to be easily and rapidly integrated into EHR systems. Currently, tablet-based evaluation of text and images are being introduced into clinical care with little or no scrutiny concerning whether tablet-based evaluation may affect the interpretation of these data and images.^{8,9} It is reassuring that the high agreement found in this study of tablet-based with desktop display-based evaluation supports the clinicians' confidence in the reliability of these tablet-based clinical applications.

One aspect of tablet-based review that was not examined in detail was the time taken by the graders to assess each image. Additionally, no characterization of how the graders used zooming and scrolling features during image assessment was performed for either platform. These data could give insight into the usability and expected efficiency of tablet-based

image review in clinical settings. The tablet and data management methods used for this study did not increase the time required to load image data for viewing or to save the assigned referral ratings, though. Image loading on the tablet took less than a second, and the waiting time for saving the annotations for each subject was negligible compared to the several seconds spent waiting for the images to load on the PC. Anecdotally, the turn-around for tablet-based grading was several days shorter than desktop-based grading. This work, though, focused on an evaluation of the tablet platform with respect to the quality of the resulting referral ratings rather than efficiency.

In addition to the direct patient care uses, researchers needing expert interpretation and annotation of medical images can benefit from the use of tablet-based image viewing. The portability, ergonomics, and ease-of-use of handheld tablets increase the likelihood that experts will be willing to undertake annotation projects. Tablet computers may allow researchers to efficiently generate annotation data sets and perform larger-scale experiments using them. It should be noted, though, that a strict grading scale for DR, such as Early Treatment of Diabetic Retinopathy (ETDRS) or international clinical diabetic retinopathy and diabetic macular edema disease severity scales was not used in this study.²⁹ Rather, the scale used was based on the experts' assessment regarding the need for clinical evaluation or intervention.

There are some limitations to this study. First, the number of retinal specialists evaluated was small (n=2). Larger scale studies that include additional numbers of clinicians comparing tablet-and desktop display-based DR referral ratings are needed to fully characterize cross-platform use. Second, only one disease (DR) and one imaging modality (color fundus images) were evaluated here. While the results are encouraging, we are not aware that an evaluation of tablets for use in grading any other image type, such as OCT, or for other disease such as AMD or glaucoma, has yet been performed.

In summary, the observed agreement and accuracy of tablet-based evaluation of retinal images to evaluation using a desktop display are encouraging for clinicians that want to use tablets clinically because they facilitate face-to-face interaction with patients. Though additional studies are needed to confirm and extend these findings for other types of images and ocular disorders, we are currently increasing our clinical use of color fundus, tablet-based image evaluation.

Acknowledgments

Financial support:

Abràmoff: NEI-EY017066; R01 EY018853; R01 EY019112

Dr. Russell is the Dina J. Schrage Professor for Macular Degeneration Research

Dr. Folk is the Judith Gardner and Donald H. Beisner, M.D., Professor of Vitreoretinal Diseases and Surgery

This publication was made possible by Grant Number UL1RR024979 from the National Center for Research Resources (NCRR), a part of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the CTSA or NIH.

The authors M.C., T.S., and M.A. plan to make software (Truthmarker) used in this study available for purchase through the Apple App Store.

The study was prospectively approved by and conducted in compliance with the Institutional Review Board of the University of Iowa.

The authors would like to thank Kyle R. Taylor for his help in implementing Truthmarker for the Apple iPad platform.

References

1. [Accessed Jan 24, 2011] Samsung GALAXY Tab Opens a New Chapter in Mobile Industry. Samsung GALAXY Tab. 2010. <http://galaxytab.samsungmobile.com/press/pressrelease.html>
2. [Accessed Jan 24, 2011] HP Slate 500 Tablet PC -Overview and Features. HP Mini. 2011. http://h10010.www1.hp.com/wwpc/us/en/sm/WF05a/321957-321957-64295-3841267-3955550-4332585.html?jumpid=re_r602_slate_body_psg_oct10_product
3. Motorola Xoom. [Accessed Jan 24, 2011] 2011. <http://www.motorola.com/Consumers/US-EN/Consumer-Product-and-Services/Tablets/ci.MOTOROLA-XOOM-US-EN.overview>
4. STYLISTIC ST Series. [Accessed Jan 24, 2011] Notebooks & Tablet PCs. 2011. http://ts.fujitsu.com/products/mobile/tablet_pcs/stylistic_st.html
5. Margalit RS, Roter D, Dunevant MA, Larson S, Reis S. Electronic medical record use and physician-patient communication: An observational study of Israeli primary care encounters. *Patient Educ Couns*. Apr; 2006 61(1):134–141. [PubMed: 16533682]
6. Shachak A, Hadas-Dayagi M, Ziv A, Reis S. Primary Care Physicians' Use of an Electronic Medical Record System: A Cognitive Task Analysis. *J Gen Intern Med*. Jan; 2009 24(3):341–348. [PubMed: 19130148]
7. Nissman SA. Electronic Health Records. *Ophthalmology*. May; 2009 116(5):1018–1019. [PubMed: 19410969]
8. [Accessed November 30, 2010] ClearPractice Announces General Release of Nimble™ – A Comprehensive EMR for the iPad™. Press Room. 2010. <http://www.clearpractice.com/ehr/press-9-28-2010.cfm>
9. [Accessed November 30, 2010] if a presents the iPad version; Top News. 2010. p. 3 Available at: <http://www.ifa4emr.com/index.php?view=article&catid=9:home&id=75:home&format=pdf>
10. [Accessed November 30, 2010] EHR Incentive Programs. 2010. <https://www.cms.gov/EHRIncentivePrograms/>
11. 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. [PubMed: 12140027]
12. Olson JA, Strachan FM, Hipwell JH, et al. A comparative evaluation of digital imaging, retinal photography and optometrist examination in screening for diabetic retinopathy. *Diabet Med*. 2003; 20(7):528–534. [PubMed: 12823232]
13. van Dijk HW, Verbraak FD, Stehouwer M, et al. Association of visual function and ganglion cell layer thickness in patients with diabetes mellitus type 1 and no or minimal diabetic retinopathy. *Vision Res*. Aug 27.2010
14. Niemeijer M, Loog M, Abramoff MD, Viergever MA, Prokop M, van Ginneken B. On Combining Computer-Aided Detection Systems. *IEEE Trans Med Imaging*. Sep 2.2010
15. Abramoff MD, Reinhardt JM, Russell SR, et al. Automated early detection of diabetic retinopathy. *Ophthalmology*. Jun; 2010 117(6):1147–1154. [PubMed: 20399502]
16. Niemeijer M, van Ginneken B, Cree MJ, et al. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. *IEEE Trans Med Imaging*. Jan; 2010 29(1):185–195. [PubMed: 19822469]
17. Niemeijer M, Abramoff M, van Ginneken B. Information Fusion for Diabetic Retinopathy CAD in Digital Color Fundus Photographs. *IEEE Trans Med Imaging*. Jan 13.2009 13:13.
18. Niemeijer M, van Ginneken B, Russell SR, Suttorp-Schulten MS, Abramoff MD. Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. *Invest Ophthalmol Vis Sci*. May; 2007 48(5): 2260–2267. [PubMed: 17460289]
19. Abramoff MD, Niemeijer M, Suttorp-Schulten MS, Viergever MA, Russell SR, van Ginneken B. Evaluation of a system for automatic detection of diabetic retinopathy from color fundus

- photographs in a large population of patients with diabetes. *Diabetes Care*. Feb; 2008 31(2):193–198. [PubMed: 18024852]
20. Abramoff MD, Alward WL, Greenlee EC, et al. Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features. *Invest Ophthalmol Vis Sci*. Apr; 2007 48(4):1665–1673. [PubMed: 17389498]
 21. [Accessed Nov 15, 2010] MESSIDOR: Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology. TECHNO-VISION Project. 2005.
<http://messidor.crihan.fr/download-en.php>
 22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. Mar; 1977 33(1):159–174. [PubMed: 843571]
 23. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. Oct; 1968 70(4):213–220. [PubMed: 19673146]
 24. Bhapkar VP. A Note on Equivalence of 2 Test Criteria for Hypotheses in Categorical Data. *J Am Stat Assoc*. 1966; 61(313):228.
 25. Chiang MF, Starren J, Du YE, et al. Remote image based retinopathy of prematurity diagnosis: a receiver operating characteristic analysis of accuracy. *Br J Ophthalmol*. Oct; 2006 90(10):1292–1296. [PubMed: 16613919]
 26. Klonoff DC, Schwartz DM. An economic analysis of interventions for diabetes. *Diabetes Care*. 2000; 23(3):390–404. [PubMed: 10868871]
 27. Suansilpong A, Rawdaree P. Accuracy of single-field nonmydriatic digital fundus image in screening for diabetic retinopathy. *J Med Assoc Thai*. Sep; 2008 91(9):1397–1403. [PubMed: 18843870]
 28. Scanlon P, Malhotra R, Thomas G, et al. The effectiveness of screening for diabetic retinopathy by digital imaging photography and technician ophthalmoscopy. *Diabet Med*. Jun; 2003 20(6):467–474. [PubMed: 12786681]
 29. Wilkinson CP, Ferris FL III, Klein RE, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*. 2003; 110(9):1677–1682. [PubMed: 13129861]

Figure 1. Example color fundus images with DR referral ratings of 0 (no referable DR) through 3 (immediately referable DR) in the reference standard used here.

Figure 2.
The graphical interfaces used by graders during rating for referable DR using a desktop display (left) and tablet (right) computer.

Table 1

Cross-Platform Comparison of Dichotomized DR Referral Ratings

<i>Grader 1</i>	Tablet DR Rating		
	Desktop DR Rating	No Referable DR	Referable DR
No Referable DR		915 (76.51)	25 (2.09)
Referable DR		60 (5.02)	196 (16.39)
Total		975 (81.52)	221 (18.48)
			1196* (100.00)

<i>Grader 2</i>	Tablet DR Rating		
	Desktop DR Rating	No Referable DR	Referable DR
No Referable DR		727 (60.58)	60 (5.00)
Referable DR		43 (3.58)	370 (30.83)
Total		770 (64.17)	430 (35.83)
			1200 (100.00)

Comparison of the dichotomized DR referral ratings assigned using a tablet to those assigned using a desktop display for each grader. Values indicate numeric freq. and (%).

* : Grader 1 missed grading of 4 images so all analyses were performed using the 1196 rated images

Table 2

Cross-Platform Comparison of DR Referral Ratings

Grader 1	Tablet DR Rating					
	Desktop DR Rating	0	1	2	3	Total
	0	915 (76.51)	20 (1.67)	3 (0.25)	2 (0.17)	940 (78.60)
	1	29 (2.42)	23 (1.92)	0 (0.00)	15 (1.25)	67 (5.60)
	2	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	3	31 (2.59)	17 (1.42)	1 (0.08)	140 (11.71)	189 (15.80)
	Total	975 (81.52)	60 (5.02)	4 (0.33)	157 (13.13)	1196* (100.00)
Grader 2	Tablet DR Rating					
	Desktop DR Rating	0	1	2	3	Total
	0	727 (60.58)	55 (4.58)	4 (0.33)	1 (0.08)	787 (65.58)
	1	42 (3.50)	196 (16.33)	11 (0.92)	17 (1.42)	266 (22.17)
	2	1 (0.08)	10 (0.83)	7 (0.58)	9 (0.75)	27 (2.25)
	3	0 (0.00)	15 (1.25)	7 (0.58)	98 (8.17)	120 (10.00)
	Total	770 (64.17)	276 (23.00)	29 (2.42)	125 (10.42)	1200 (100.00)

Comparison of the DR referral ratings assigned using a tablet to those assigned using a desktop display for each grader. Values indicate numeric freq. and (%).

* : Grader 1 missed grading of 4 images so all analyses were performed using the 1196 rated images

Table 3

Intra-Observer Agreement

Grader		κ (95% CI)	weighted κ (95% CI)
1	Tablet to Desktop	0.778 (0.733–0.823)	0.776 (0.734–0.818)
	Desktop to Desktop	0.800 (0.758–0.842)	0.796 (0.758–0.834)
2	Tablet to Desktop	0.812 (0.777–0.846)	0.795 (0.765–0.826)
	Desktop to Desktop	0.784 (0.746–0.821)	0.768 (0.736–0.800)

The agreement of DR referral ratings assigned by the same grader. Both cross-platform (tablet to desktop) and single-platform (desktop to desktop) κ statistics are tabulated.

Table 4

Inter-Observer Agreement

Platform	κ (95% CI)	weighted κ (95% CI)
Tablet	0.544 (0.496–0.593)	0.648 (0.605–0.691)
Desktop	0.625 (0.578–0.672)	0.675 (0.637–0.714)

The agreement between DR referral ratings assigned by the graders for each of the platforms (iPad tablet and desktop display).

Table 5

Accuracy of Tablet-Based Referable DR Ratings

<i>DR Grade 1 or Higher</i>			
	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
Grader 1	0.872 (0.853–0.891)	0.987 (0.980–0.994)	0.950 (0.938–0.962)
Grader 2	0.848 (0.828–0.869)	1.0 (0.998–1.0)	0.967 (0.957–0.977)
<i>DR Grade 2 or Higher</i>			
	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
Grader 1	0.966 (0.956–0.976)	0.911 (0.895–0.927)	0.942 (0.928–0.955)
Grader 2	0.948 (0.936–0.961)	0.976 (0.967–0.985)	0.974 (0.965–0.983)
<i>DR Grade 3</i>			
	Sensitivity (95% CI)	Specificity(95% CI)	AUC(95% CI)
Grader 1	0.991 (0.9985–0.997)	0.611 (0.585–0.640)	0.801 (0.779–0.824)
Grader 2	0.987 (0.980–0.994)	0.728 (0.704–0.754)	0.859 (0.840–0.879)

The accuracy of the referable DR ratings assigned by each grader using the iPad tablet assessed using sensitivity, specificity, and area under the ROC curve (AUC). An average of the ratings assigned by the two graders using a desktop display was used as a standard to compute the values.