

A novel device for accurate and efficient testing for vision-threatening diabetic retinopathy



April Y. Maa^{a,b}, William J. Feuer^{c,d}, C. Quentin Davis^e, Ensa K. Pillow^{f,g,h}, Tara D. Brown^f, Rachel M. Caywood^f, Joel E. Chasan^b, Stephen R. Fransen^{g,h,i,*}

^a Atlanta VA Medical Center, Ophthalmology, 1670 Clairmont Road MC 112E, Decatur, GA, 30033, USA

^b Emory University School of Medicine, Emory Eye Center, Comprehensive Ophthalmology, 1365B Clifton Road NE, Atlanta, GA, 30322, USA

^c University of Miami Miller School of Medicine, Biostatistics, Dominion Tower, Box C210, 1400 NW 10th Avenue, Suite 506, Miami, FL, 33136, USA

^d Bascom Palmer Eye Institute, 900 NW 17th Street, Miami, FL, 33136, USA

^e LKC Technologies, Inc., 2 Professional Drive, Suite 222, Gaithersburg, MD, 20879, USA

^f Oklahoma City VA Medical Center, Ophthalmology, 921 NE 13th Street, Oklahoma City, OK 73104, USA

^g University of Oklahoma College of Medicine, Ophthalmology, 608 Stanton L. Young Blvd, Oklahoma City, OK, 73104, USA

^h Dean McGee Eye Institute, 608 Stanton L. Young Blvd, Oklahoma City, OK, 73104, USA

ⁱ Inoveon Corporation, 800 Research Parkway, Suite 370, Oklahoma City, OK, 73104, USA

ARTICLE INFO

Article history:

Received 24 September 2015

Received in revised form 30 November 2015

Accepted 1 December 2015

Available online 9 December 2015

Keywords:

Diabetic retinopathy
Diabetic macular edema
Retinal diagnostic test
ERG
Pupillography
Diabetic eye exam

ABSTRACT

Aims: To evaluate the performance of the RETeval device, a handheld instrument using flicker electroretinography (ERG) and pupillography on undilated subjects with diabetes, to detect vision-threatening diabetic retinopathy (VTDR).

Methods: Performance was measured using a cross-sectional, single armed, non-interventional, multi-site study with Early Treatment Diabetic Retinopathy Study 7-standard field, stereo, color fundus photography as the gold standard. The 468 subjects were randomized to a calibration phase (80%), whose ERG and pupillary waveforms were used to formulate an equation correlating with the presence of VTDR, and a validation phase (20%), used to independently validate that equation. The primary outcome was the prevalence-corrected area under the receiver operating characteristic (ROC) curve for the detection of VTDR.

Results: The area under the ROC curve was 0.86 for VTDR. With a sensitivity of 83%, the specificity was 78% and the negative predictive value was 99%. The average testing time was 2.3 min.

Conclusions: With a VTDR prevalence similar to that in the US, the RETeval device will identify about 75% of the population as not having VTDR with 99% accuracy. The device is simple to use, does not require pupil dilation, and has a short testing time.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

This study was funded by a grant from the US National Eye Institute (R44EY021121) and by LKC Technologies, Inc. This study is a result of work supported with resources and the use of facilities at the Atlanta and Oklahoma City VA medical centers. The content is solely the responsibility of the authors and does not necessarily represent the official views of the US National Eye Institute, the National Institutes of Health, the US Department of Veterans Affairs, or the US government.

Registry: NCT01950663, ClinicalTrials.gov

Conflict of interest: These conflicts of interest are relevant to this research: grant funding (A.Y. Maa, W.J. Feuer, E.K. Pillow, S.R. Fransen); employment (C.Q. Davis); stock ownership (C.Q. Davis, S.R. Fransen). The remaining authors report no conflict of interest (T.D. Brown, R.M. Caywood, J.E. Chasan).

* Corresponding author at: University of Oklahoma College of Medicine, Ophthalmology, Dean McGee Eye Institute, 608 Stanton L. Young Blvd, Oklahoma City, OK, 73104, USA. Tel.: +1 405 209 7531; fax: +1 405 271 7821.

E-mail addresses: april.maa2@va.gov (A.Y. Maa), wfeuer@med.miami.edu (W.J. Feuer), qdavis@lkc.com (C.Q. Davis), ensa.pillow@va.gov (E.K. Pillow), tara.brown2@va.gov (T.D. Brown), rachel.caywood@va.gov (R.M. Caywood), jchasan@emory.edu (J.E. Chasan), stephen-fransen@ouhsc.edu (S.R. Fransen).

1. Introduction

Diabetic retinopathy (DR) remains the leading cause of blindness among working age adults in the US ([American Academy of Ophthalmology Retina/Vitreous Panel, 2014](#); [Centers for Disease Control and Prevention, 2011](#)) and is a major cause of blindness worldwide ([Yau et al., 2012](#)). In the US, less than 50% of the patients with diabetes receive an annual DR examination ([Lee, Feldman, Ostermann, Brown, & Sloan, 2003](#); [Lee et al., 2014](#); [Paz et al., 2006](#); [Rosenberg, Friedman, & Gurland, 2011](#); [Saadine, Fong, & Yao, 2008](#)). Many of those with vision-threatening diabetic retinopathy (VTDR) ([Zhang et al., 2010](#)) are not diagnosed in time to benefit from the remarkable efficacy of DR therapy established by the Diabetic Retinopathy Study (DRS) ([Ferris, 1993](#)), Early Treatment Diabetic Retinopathy Study (ETDRS) ([ETDRS Research Group, 1991a](#); [Chew et al., 2003](#)) and clinical trials of intravitreal anti-VEGF therapy ([Brown](#)

et al., 2015; Nguyen et al., 2012; Wang, Sun, Liu, & Xu, 2012). VTDR is defined as severe non-proliferative or proliferative DR, with or without clinically significant macular edema (CSME) (American Academy of Ophthalmology Retina/Vitreous Panel, 2014; Zhang et al., 2010). Patients with VTDR are at risk for blindness or moderate vision loss and should be referred to an eye care provider.

This study reports the performance characteristics of the RETeval device to identify subjects with VTDR. ETDRS 7-standard field, stereo, color fundus photography, read by a qualified reading center was the gold standard (Ku, Landers, Henderson, & Craig, 2013; Lawrence, 2004; Williams et al., 2004) for comparison. Unlike ophthalmoscopy and retinal photography, the RETeval device combines flicker electroretinogram (ERG) and pupillary responses to generate a single numerical output that can be compared to a cutoff value in order to minimize the subjectivity of testing for VTDR.

The ERG measures the electrical activity of the retina in response to an intermittent flash stimulus (McCulloch et al., 2015). The flicker ERG waveform is typically characterized by a time delay between the light stimulus and the peak electrical response (implicit time), and the peak-to-peak amplitude of the electrical response. The flicker ERG is a response from the cone system, and is therefore representative of the whole retina. While anatomical studies (Curcio, Sloan, Kalina, & Hendrickson, 1990) show a 40-fold increase in cone density in the fovea, the area of the fovea is quite small (3.6 mm² out of 1000 mm²) and the rest of the retina is relatively uniform in cone density. Changes in these ERG parameters correlate strongly with DR severity (Bresnick & Palta, 1987; Han & Ohn, 2000; Holopigian, Seiple, Lorenzo, & Carr, 1992; Kim, Lee, Bae, Cho, & Kang, 1997; Messias et al., 2012; Park, Sun, Lee, Park, & Ohn, 2010; Satoh, Iijima, Imai, Abe, & Shibuya, 1994; Tahara, Matsuura, & Otori, 1993; Tyrberg et al., 2011). DR progression leads to increasing retinal ischemia (American Academy of Ophthalmology Retina/Vitreous Panel, 2014) which prolongs the implicit time and decreases the amplitude (Park et al., 2010; Satoh et al., 1994; Severns & Johnson, 1993).

The pupillary response to a light stimulus has also been shown to correlate with DR severity (Nakayama et al., 2001; Ortube et al., 2013; Smith & Smith, 1983; Straub, Jeron, & Kerp, 1992; Straub, Thies, Jeron, Palitzsch, & Scholmerich, 1994). Both the rate and extent of pupil constriction in response to a light stimulus decrease with increasing DR severity (Ortube et al., 2013).

The objectives of this study were first to determine the best method of combining ERG and pupillary response measurements for the detection of VTDR, and second to evaluate the performance using an independent cohort of subjects.

2. Materials and methods

2.1. RETeval device

The handheld RETeval device (LKC Technologies Inc., Gaithersburg, MD; Welch Allyn, Inc., Skaneateles Falls, NY), simultaneously measures the full-field flicker ERG and pupillary response to light (Kato, Kondo, Sugimoto, Ikesugi, & Matsubara, 2015). A 28.3 Hz flickering white-light stimulus (CIE 1931 chromaticity of 0.33, 0.33) is produced by brief (<5 ms) flashes from red, green, and blue LEDs in a ganzfeld (i.e., an integrating sphere). The ganzfeld also has a red fixation LED to direct the subject's gaze during the test. The device has an IR-sensitive camera and an IR LED to video the eye in the infrared at the flicker frequency. The pupil size is measured in real-time from the video. The timing among the stimulus, data acquisition, and video recording are hardware-synchronized through the use of a single clock crystal. The pupillary measurements are used to dynamically adjust the white light stimulus according to the formula flash retinal illuminance (Troland-second or Td·s) = flash luminance (cd·s/m²) × pupil area (mm²), thereby providing an ERG stimulus that is largely independent of pupil size (Kato et al., 2015). A sensor strip skin electrode, which has separate

patches for a positive, negative, and an active ground (right leg drive) contact, was placed below each lower eyelid (Kato et al., 2015). While skin electrodes have traditionally been avoided in ERG testing due to lower signal levels (McCulloch et al., 2015), improvements in hardware data acquisition and Fourier-based analysis methods have enabled the reproducible results described in this study while avoiding the need to touch the eye with corneal electrodes. The pupils are not artificially dilated, so that the device can use the pupillary response as an independent indicator of DR severity.

2.2. Study design

To ensure representation of all levels of disease, subjects with diabetes were selectively recruited, via chart review, to tentatively target one of five DR severity strata (ETDRS Research Group, 1991b; Wilkinson et al., 2003). The severity strata were (1) no DR, (2) mild non-proliferative DR (NPDR) without clinically significant macular edema (CSME), (3) moderate NPDR without CSME, (4) mild or moderate NPDR with CSME, and (5) severe NPDR or proliferative DR with or without CSME.

This multi-site, non-interventional study was cross-sectional and had a single arm. In a single session, each subject was first tested with the RETeval device followed by an Amsler grid psychophysical test. The Amsler grid is used to detect metamorphopsia, a change in the perceived shape of objects that implies macular edema or other macular pathology. A randomly selected subset was retested with the RETeval device to assess test-retest variability. The subjects were then dilated with tropicamide and phenylephrine drops. After dilation, subjects underwent ETDRS 7-standard field, stereo, color fundus photography, which completed their participation in the study.

The RETeval device tested each eye with 4, 8, 16, and 32 Td·s flicker stimuli (28.3 Hz) that each lasted between 5 and 15 s, depending on the standard error of mean for the implicit time measurement. There was no background light, as previous studies have shown improved detection of VTDR without a background light (Bresnick & Palta, 1987; Tyrberg et al., 2011). The stimuli were presented in randomized order, with about 1 s of darkness between each brightness tested. The brightness of the stimuli was selected to maintain subject comfort while providing a large enough response to have good signal to noise ratio.

The subject's ETDRS 7-standard field images were double graded by readers masked to the other readers' results and to the RETeval results, in a dedicated reading center (Inoveon Corp, Oklahoma City, OK). Results differing by more than one ETDRS level (ETDRS Research Group, 1991b) or with respect to the VTDR referral criterion, were adjudicated by the two readers overseen by a retinal specialist. The adjudicated results for the subject's worst eye (because in clinical practice subjects, not individual eyes, are referred for evaluation by an eye care provider) served as the gold standard to which the RETeval device results were compared.

Technical failures from ungradeable ETDRS 7-standard field photographs created two more DR severity strata: CSME with ungradable DR severity, and ungradable CSME and DR severity. For a subject to be considered ungradable, either both eyes were ungradable or one eye was ungradable while the other did not have VTDR. Subjects with ungradable ETDRS photographs were excluded by necessity from further analysis. Subjects with ungradable RETeval results were considered to have tested positive for VTDR, as is done in screening programs in order to reduce the likelihood of false negatives (Castell, 2012).

At the end of the trial, subjects were randomized to a calibration portion and validation portion that were separately analyzed, as described below.

The research followed the tenets of the Declaration of Helsinki; informed consent was obtained from the subjects after explanation of the nature and possible consequences of the study; and the institutional review boards of the participating institutions approved the research. The study was overseen by an independent study monitor and is registered at ClinicalTrials.gov (NCT01950663).

2.3. Subjects

468 subjects were enrolled at two centers in the United States (Atlanta VA Medical Center and Oklahoma City VA Medical Center). The enrollment target was 80 subjects in each of the five DR severity strata. Inclusion criteria were that subjects be diagnosed with diabetes and treated with at least one oral hypoglycemic medication or insulin. Exclusion criteria were (1) a history of photosensitive epilepsy, (2) previous laser or drug treatment for DR or CSME, (3) eye diseases other than diabetic retinopathy or macular edema that, in the opinion of the recruiting ophthalmologist, might affect the ERG or result in ungradable ETDRS photographs, (4) or an inability or unwillingness of the subject or legal representative to provide written informed consent. Previous laser or drug treatment for DR or CSME was an exclusion criteria because we wanted to target subjects who did not know if they had VTDR under the assumption that those people who knew they had VTDR were already receiving adequate care.

2.4. Outcomes

The primary outcome was the area under the receiver operating characteristic (ROC) curve for the detection of VTDR in either eye. VTDR is severe non-proliferative (ETDRS level 53) or proliferative (ETDRS levels 61–85) DR or the presence of CSME with any ETDRS level (American Academy of Ophthalmology Retina/Vitreous Panel, 2014; Zhang et al., 2010). In the calibration and validation portions of the analysis, the area under the ROC curve (AUROC) was not adjusted for the intentional oversampling of DR severity strata. However, a final ROC analysis did correct for the intentional oversampling. These calculations used prevalence data from a commercial ETDRS 7-standard field diabetic retinopathy evaluation service on 55,000 right eyes from subjects' first visits over the past decade (Inoveon Corp, Oklahoma City, OK). These data, shown in Table 1, were used for the prevalence correction because, to our knowledge, prevalence data for the six DR severity strata used in this study have not been published.

2.5. Sample size

The sample size for this study was based on comparison of paired AUROC curves and confidence interval widths with 90% power assuming a 5% two-sided alpha error. We assumed the correlation between two measurements of the same person is 0.5. While during the calibration phase of the trial many such comparisons would be made, there was no adjustment for alpha error inflation because any such error would be caught during the validation phase. Validation was based on an independent cohort of subjects that were assessed only once.

2.6. Reproducibility

A 30% random sample of subjects was selected for retesting to assess RETeval device test–retest variability. Study personnel were not informed if a subject was selected for retesting until after the first RETeval test and the Amsler grid test had been performed.

2.7. Statistical methods

Following dataset closure, measurements from 80% of the subjects in each disease severity group were randomly selected to calibrate the RETeval's detection algorithm, while 20% were reserved for validation.

2.7.1. Calibration phase

The purpose of the calibration phase was to determine the best way to combine the information obtained by the RETeval device to predict the presence of VTDR as determined by the gold standard.

The measurements used in the analysis included the amplitude and implicit time for each ERG brightness (both as using the whole waveform and just the fundamental of the waveform as suggested in the literature as being more robust (McAnany & Nolan, 2014; Severns, Johnson, & Merritt, 1991)). The pupillary response used in the analysis was the ratio of the pupil area from the 32 Td·s and the 4 Td·s stimuli after the initial 2.5 s of the stimulus. The brightness of the iris in the infrared images was also used as a measurement in the analysis, as it had been reported that blue irides are relatively transparent and don't attenuate pupillary responses (Kardon, Hong, & Kawasaki, 2013).

The influence of age on RETeval measurements was assessed by using the 74 subjects in the calibration phase with no DR in either eye (ETDRS level 10), none of whom had CSME in either eye. This group was used for age correction so as to not confound changes in DR disease state with age while still using a relevant population (subjects with diabetes) to determine the age dependencies. Ages in this group ranged from 23 to 77 years (mean = 59.7, SD = 10.8). Measurements for the two eyes of each subject were averaged and fit using linear regression models of each parameter onto age. Residual plots were examined to assess for non-linear effects and none were noted.

Age-corrected measurements from the two eyes of each subject were characterized as either best eye (BE) or worst eye (WE). These best eye and worst eye deviations were included into a forward stepwise logistic regression model with referral (yes/no) as the dependent variable. The criterion for inclusion of a parameter into the forward stepwise logistic regression model was a p-value of 0.01 by the likelihood ratio method; this criterion, more stringent than the usual 0.05 p-value, was selected to minimize over-fitting of the prediction model to the calibration dataset. Model coefficients were used to create a prediction equation and a prediction probability was generated for each subject in the calibration stage. The prediction probability was in turn used to create the numerical output of the RETeval device.

By varying a cutoff value above which subjects are considered to have tested positive for VTDR (and therefore referred), a Receiver Operating Characteristic (ROC) curve was constructed and the area under the curve and its asymptotic 95% confidence interval were determined.

2.7.2. Validation phase

The prediction equation was then applied to the 20% validation sample. An ROC curve was generated, and the area under the curve and its asymptotic 95% confidence interval were similarly determined.

2.7.3. Application to primary care prevalence

These analyses include all subjects measured with the RETeval device, excluding those that could not be assessed with ETDRS 7-standard field photography. Because the area under the ROC curve (AUROC) in the

Table 1
Severity groups.

International Clinical Classification (Wilkinson et al., 2003)	ETDRS Level	CSME	Calibration phase (N = 371)	Validation phase (N = 96)	Application to primary care prevalence (N = 467)	Primary care prevalence
No NPDR	10–12	–	74	19	93 (20%)	74.2%
Mild NPDR	14–35	–	101	26	127 (27%)	20.1%
Moderate NPDR	43–47	–	72	18	90 (19%)	2.5%
CSME with No, Mild, or Moderate NPDR	10–47	+	31	8	39 (8%)	0.8%
Severe NPDR or Proliferative DR	53–85	+/-	41	11	52 (11%)	2.3%
Ungradable ETDRS Level	?	+	12	3	15 (3%)	0.5%
Ungradable fundus photographs	?	?	N/A	N/A	51 (11%)	

validation sample exceeded the AUROC of the calibration sample, the inclusion of all subjects should not lead to an overly optimistic estimate of RETeval performance in a primary care setting. Thus, all subject data were used to generate an ROC curve after prevalence-correcting the results using the data shown in Table 1. The area under the ROC curve was generated asymptotically, and a bootstrap method was used (1000 replicates) for the confidence interval.

The prevalence-corrected referral statistics for a variety of prediction probabilities were computed. The positive and negative predictive values were based on the prevalence of vision-threatening diabetic retinopathy in the US (4.4%) (Zhang et al., 2010). If the worldwide prevalence was used (10.2%) (Yau et al., 2012), the NPVs would be lower and the PPVs would be higher. The lower confidence limit (LCL) and upper confidence limit (UCL) of the 95% confidence interval were computed from the Clopper–Pearson interval for binomial distributions.

Reproducibility was determined using the intraclass correlation (ICC). A widely used guide to interpreting ICCs (Fleiss, 1981) characterizes ICC < 0.4 to be poor reproducibility, ICC > 0.75 excellent reproducibility, and ICC from 0.4 to 0.75 as fair to good.

3. Results

3.1. Subject flow

A total of 468 subjects were enrolled between September 2013 and April 2014; 467 completed testing (99.8%). The subject who did not complete the study left after being tested with the RETeval device, the Amsler grid, and after being dilated but before ETDRS 7-standard field photography. Recruitment ended after the study sites exhausted their pool of potential subjects in the low prevalence categories that had fewer than 80 subjects. For reproducibility, 137 subjects were randomly assigned to duplicate the RETeval test; data were missing for 9 of those subjects (6 were missing due to procedural issues unrelated to the RETeval device and 3 were missing due to RETeval device technical failures).

3.2. Subject characteristics

The characteristics of the subjects are shown in Table 2. The percentage of female subjects in the study (12.4%) was substantially greater than the percentage of female users of the VHA health care system (5%).

Table 2
Subject characteristics.

Characteristic	Calibration phase (N = 371)	Validation phase (N = 96)	Application to primary care prevalence (N = 467)	P-value between calibration and validation phases
Age (yr)				0.39 ^a
Mean (SD)	60 (10)	61 (12)	61 (11)	
Range	23–88	24–82	23–88	
Race, no. (%)				0.92 ^b
Caucasian	193 (52%)	50 (52%)	243 (52%)	
African American	147 (40%)	37 (39%)	184 (39%)	
Native American	19 (5%)	5 (5%)	24 (5%)	
Hispanic	6 (2%)	3 (3%)	9 (2%)	
Other	4 (1%)	1 (%)	5 (1%)	
Asian	2 (0.5%)	0 (%)	2 (0.4%)	
Gender, no. (%)				0.60 ^c
Female	48 (13%)	10 (10%)	58 (12%)	
Male	323 (87%)	86 (90%)	409 (88%)	
Diabetes type, no. (%)				0.79 ^c
1	17 (5%)	5 (5%)	22 (5%)	
2	354 (95%)	91 (95%)	445 (95%)	
Medication type, no. (%)				0.78 ^b
Oral	121 (33%)	28 (29%)	149 (32%)	
Insulin	119 (32%)	31 (32%)	150 (32%)	
Both	131 (35%)	37 (39%)	168 (36%)	

^a 2-sample t-test.
^b chi-squared test.
^c Fisher exact test.

3.3. Outcomes

Fig. 1 shows representative ERG and pupillary data for a subject with and without VTDR, while Fig. 2 shows summary statistics for all subjects. In the presence of VTDR, the best eye's 32 Td·s ERG timing is delayed, the best eye's 16 Td·s ERG amplitude is reduced, and the worst eye's pupillary response (4 Td·s compared to 32 Td·s) is reduced. These three parameters (age-corrected) were all highly statistically significantly associated with referral status ($p \leq 0.002$), and formed the prediction equation used to generate the RETeval device's numerical output (called DR Score in Fig. 2).

Fig. 3 shows the receiver operating characteristic curves from the calibration phase, validation phase, and the overall result utilizing the prevalence found in primary care settings. The area under the curve for the validation (0.81; 95% confidence interval of 0.71–0.92) is larger than that of the calibration (0.78; 95% confidence interval of 0.72–0.84), giving assurance that the prediction equation generated with the calibration data was not over-fitted to peculiarities of those data. The stratified recruitment strategy employed in this trial created a population that was heavily concentrated with cases near the VTDR threshold which was useful to calibrate the device to best distinguish VTDR by oversampling the cases most difficult to categorize. After correcting to the prevalence seen in a primary care setting, the area under the curve increased (0.86; 95% confidence interval of 0.77–0.93).

Table 3 shows the performance of the RETeval device for detecting VTDR at 5 points along the prevalence-corrected ROC curve of Fig. 3. For example, with a cutoff value of ≥ 20 , the device has a sensitivity of 83%, a specificity of 78%, and a negative predictive value of 99% for VTDR. In a primary care setting, for every 1000 subjects tested, about 44 will have VTDR. With a cutoff of ≥ 20 , it is expected that about 753 subjects will be below the referral threshold, 7 of whom will have VTDR. Of the 247 referred, 37 will have VTDR. Thus the device should eliminate over three-fourths of the tested population while missing very few cases. By lowering the referral threshold, the negative predictive value can be improved at the expense of more subjects referred for further testing. For example, with a cutoff of ≥ 17.6 the negative predictive value is 99.5%. For every 1000 subjects tested it is expected that about 500 (half) will be below the referral threshold, only 3 of which would have VTDR.

Among the 93 subjects with no retinopathy or CSME, there was no statistically significant difference in the mean RETeval result between either genders ($p = 0.44$, 2 sample t-test) or between Caucasians and African Americans ($p = 0.28$, one way ANOVA).

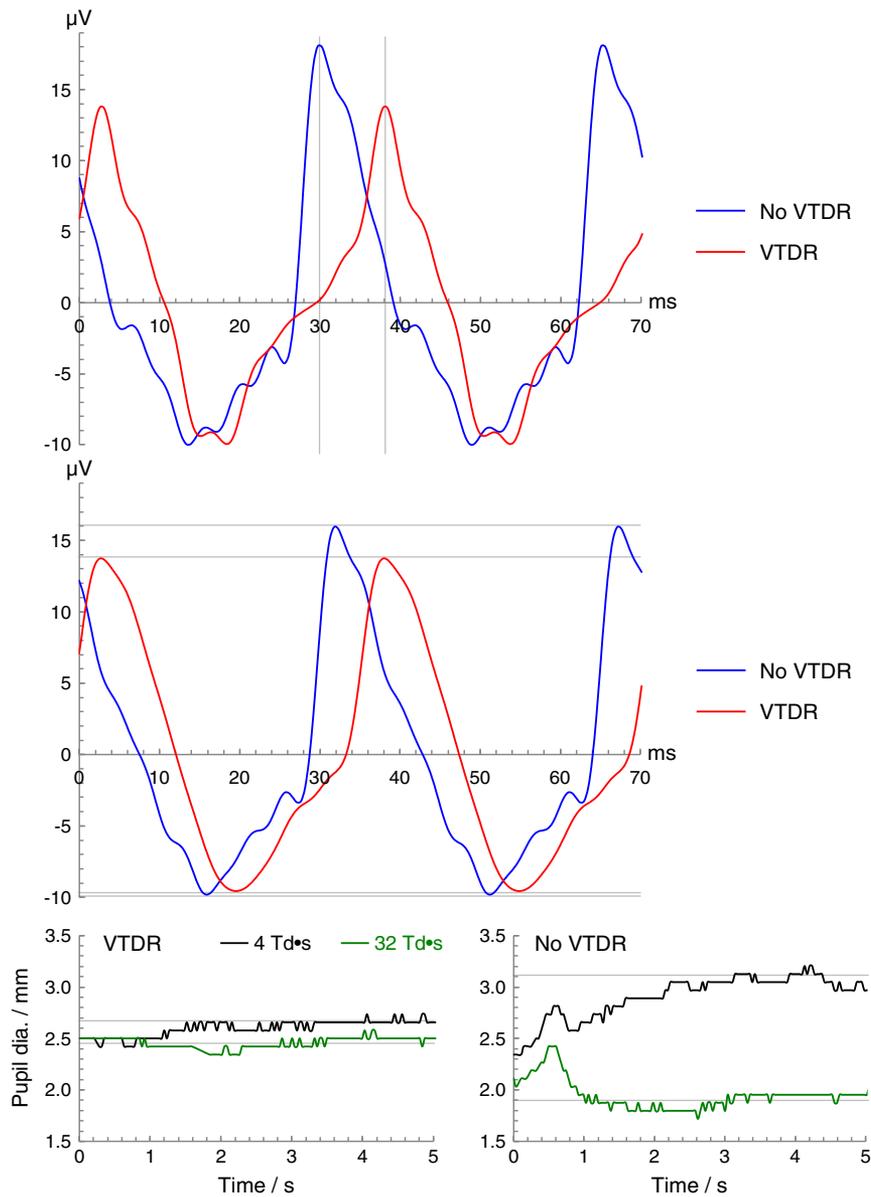


Fig. 1. Representative raw data from the RETeval device from two subjects, one with and one without VTDR. The top plot shows the timing difference between the subject with VTDR (red line) and the subject without VTDR (blue line) for the 32 Td·s flicker ERG. In this steady-state response, the flashes occurred at time = 0, ± 35.33 ms, ± 70.66 ms, etc. The vertical grid lines show the implicit time between the flash and the eye's peak electrical response to be about 30 ms for the subject without VTDR, faster than the 38 ms for the subject with VTDR. The middle plot shows the amplitude difference for the 16 Td·s flicker ERG, where the horizontal grid lines show the peak-to-peak amplitude for the subject without VTDR was 26 μ V, larger than the 23 μ V for the subject with VTDR. The bottom plots show the difference in pupillary response between a 4 Td·s and 32 Td·s flickering stimulus for the subject with VTDR (bottom left) and without VTDR (bottom right). The time scale for the pupillary response is from the onset of the flickering light (0 s) to 5 s. The grid lines show that after initial transients, the pupil size for the subject with VTDR was 2.45 mm with the brighter 32 Td·s flicker stimulus, and was a slightly larger 2.66 mm with the dimmer 4 Td·s flicker stimulus. The subject without VTDR had a larger change in pupil diameter between the two stimuli: 3.10 mm vs 1.89 mm. These changes in pupil size are summarized by the ratio of the pupil areas. The subject without VTDR had a pupil area ratio of 2.7 ($= 3.10^2/1.89^2$), larger than the 1.2 for the subject with VTDR.

Using the results of the Amsler grid test did not improve the RETeval device's performance of detecting VTDR.

3.4. Ancillary analyses

3.4.1. Testing time

RETeval testing time averaged 2.3 min (standard deviation, 0.8 min) to test both eyes.

3.4.2. Reproducibility

Using the 128 subjects with duplicate RETeval measurements, the test–retest standard deviation was 1.25 and the intraclass correlation (ICC) was 90.2%. Thus, the RETeval device measurements have

excellent reproducibility (Fleiss, 1981). The RETeval measurements in this group ranged from 11.1 to 31.8, with a mean of 19.7.

3.4.3. Performance excluding CSME

If CSME is ignored in the primary-care prevalence analysis, the RETeval device's performance is improved, having a sensitivity of 87%, a specificity of 78%, and a negative predictive value of 99.2% with a cutoff of ≥ 20 . For every 1000 subjects tested, 752 would not be referred, 6 of which have severe NPDR or PDR. In the 248 referred, 38 will have severe NPDR or PDR. By lowering the referral threshold, to a cutoff of ≥ 17.9 , the sensitivity is 94%, specificity is 54%, and the negative predictive value is 99.5%.

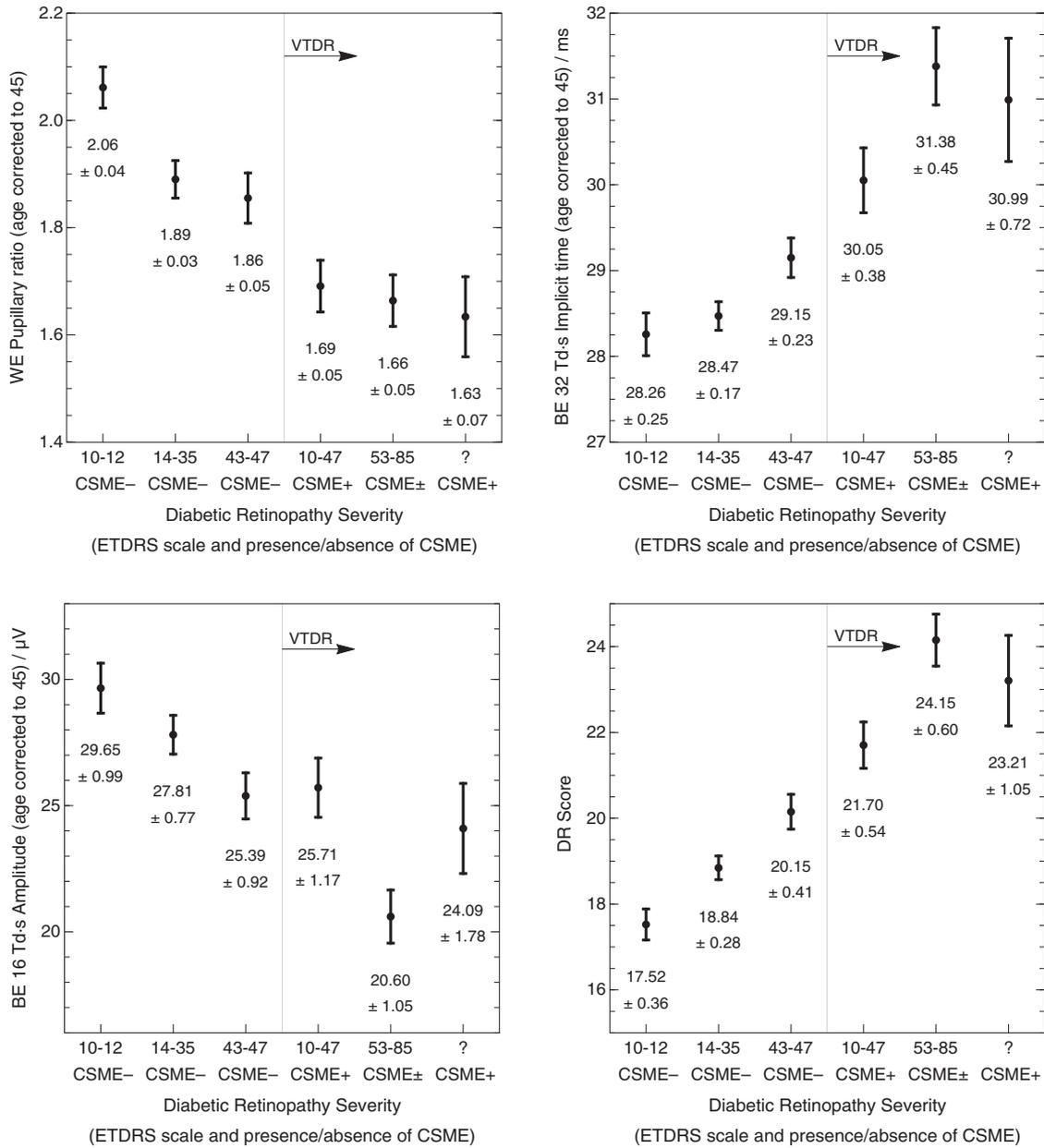


Fig. 2. Dependence of RETeval measurements on diabetic retinopathy severity level. Plots show the mean and standard error of the mean for three measurements and the overall RETeval measurement (DR Score) for each severity group. Severity group definitions and the number of subjects in each group can be found in Table 1. BE and WE stand for “best eye” and “worst eye” respectively.

3.4.4. Technical failures

The RETeval device had a technical failure rate (no results generated) of 1% (5/467) whereas ETDRS 7-standard field photography (ungradable images) had a significantly higher ($P < 0.001$, exact McNemar test) technical failure rate of 11% (51/467). The RETeval device generated results on 98% (50/51) of the subjects who had ungradable ETDRS photographs.

3.5. Safety

No adverse events were reported.

4. Discussion

Using a cutoff score of 20.0, our analysis (Table 3) suggests that if 100 unselected subjects with diabetes are tested with the RETeval device, 76 will have a negative test result and of those 75 (99%) will

not have VTDR. Therefore, over three fourths of the subjects will be told they do not have VTDR with 99% accuracy. This allows providers to focus on the remaining 24 who may have VTDR or another ocular disease that requires attention. Test time averaged 2.3 min in this study. The device had a technical failure rate of 1%. This combination of accuracy and efficiency should improve the quality and cost-effectiveness of DR testing and compliance. In comparison to earlier ERG studies and pupillary response studies, this study shows improved performance in part due to the combination of these formally disparate measures into a combined score, which to the authors’ knowledge has not been done before.

The RETeval device’s performance compares favorably to point of care digital retinal photography when using ETDRS 7-standard field photography as the gold standard. One point of care digital retinal imaging system (Joslin Vision Network, Boston, MA) photographed 3 stereoscopic nonmydriatic fields and reported a sensitivity of 85% and specificity of 100% for the detection of severe NPDR or worse in the subject’s worst eye

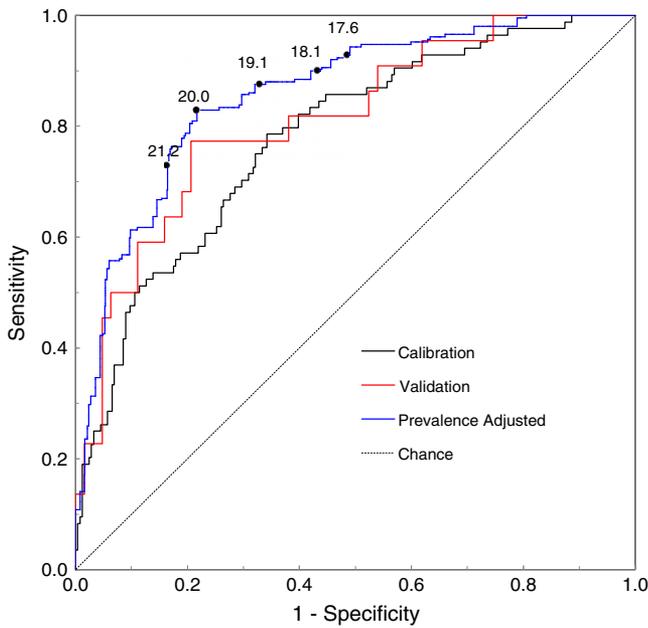


Fig. 3. Receiver operating characteristic (ROC) curves for the calibration phase, validation phase, and prevalence-corrected overall result. The cutoff values from Table 3 are labeled on the prevalence-adjusted curve.

(Bursell et al., 2001). The RETeval device had slightly better sensitivity (87%), and worse specificity (78%) when using the same criteria. The digital retinal imaging system did not report performance when CSME is included; nonetheless, subjects with CSME should be included in these analyses if the results are used to identify subjects at risk for vision loss (American Academy of Ophthalmology Retina/Vitreous Panel, 2014). While this photographic system takes stereo photographs and therefore can in principle measure CSME, many photography systems are not stereo and therefore cannot directly assess macular edema.

The RETeval device's technical failure rate (1%) was much better than that reported for the digital retinal imaging system (35%) (Ahmed et al., 2006). If subjects with ungradeable images are referred, the specificity of the digital retinal imaging system decreases from 100% to 66% when using a disease prevalence of 4.4% (Zhang et al., 2010). Thus, the RETeval device's sensitivity and specificity compare favorably to nonmydriatic digital retinal imaging when subjects with ungradable images are referred and included in the performance analysis.

Table 3
Performance metrics with the primary care prevalence.

RETeval cutoff value	Sensitivity	Specificity	Predictive Power of Test		Percent of Tested Sample			
			Positive	Negative	Positive Test		Negative Test	
					True	False	True	False
17.6	93%	52%	8%	99.5%	4.1%	46%	50%	0.2%
LCL	73%	47%	5%	97%	2.4%	41%	45%	0.0%
UCL	99.9%	57%	13%	100.0%	6.5%	51%	55%	1.3%
18.1	90%	57%	9%	99.2%	3.8%	41%	55%	0.5%
LCL	65%	52%	5.0%	97%	2.2%	36%	50%	0.1%
UCL	99%	62%	14%	99.9%	6.2%	46%	59%	1.7%
19.1	88%	67%	11%	99.1%	3.8%	31%	64%	0.5%
LCL	65%	62%	6%	97%	2.2%	27%	59%	0.1%
UCL	99%	72%	17%	99.9%	6.2%	36%	69%	1.7%
20.0	83%	78%	15%	99%	3.6%	21%	75%	0.7%
LCL	59%	74%	9%	97%	2.0%	17%	71%	0.1%
UCL	96%	82%	23%	99.9%	5.9%	25%	79%	2.1%
21.2	73%	84%	17%	99%	3.1%	15%	80%	1.2%
LCL	47%	80%	9%	97%	1.7%	12%	76%	0.4%
UCL	90%	87%	27%	99.5%	5.3%	19%	84%	2.8%

LCL and UCL represent the lower and upper 95% confidence limits, respectively.

The RETeval device's performance also compares favorably to ophthalmologists when using ETDRS 7-standard field photography as the gold standard. In one large study (Pugh et al., 1993) (n = 352), ophthalmologists performed indirect ophthalmoscopy followed by either direct ophthalmoscopy or slit lamp biomicroscopy to classify each subject's worst eye as positive (moderate to severe NPDR or PDR) or negative. Two retina specialists and eight general ophthalmologists performed the examinations. The study did not report performance when CSME is included. The ophthalmologists' sensitivity was 33% and specificity was 99%.

Table 4 summarizes the results of these studies, all of which used the ETDRS 7-standard field photography gold standard. The sensitivity and specificity are those described above. The remaining metrics assume a VTDR prevalence of 4.4% (Zhang et al., 2010). The RETeval device has the smallest number of false negatives, the most important factor from an initial detection point of view.

4.1. Limitations

Several demographic characteristics of the subjects enrolled in this study differed from the US population. Although 58 female subjects were enrolled, study subjects were predominantly male Caucasians and African Americans. Although this could affect generalizability, we saw no statistically significant difference in the mean RETeval results between genders or between those two races in our no-retinopathy group.

Subjects with concurrent eye disease that could affect the ERG (e.g., retinal vascular occlusive disease (Severns & Johnson, 1993; Yasuda, Kachi, Ueno, Piao, & Terasaki, 2015; Kjeka, Jansson, Bredrup, & Krohn, 2013; Larsson, Bauer, & Andreasson, 2000)) were excluded from this study to avoid confounding the results. These diseases have a similar effect on the ERG and therefore are likely to cause a false positive result. When testing unselected subjects, as would be done in practice, these "false positive" subjects would actually improve the negative predictive value of the test. The positive predictive value would decrease although these "false positive" subjects likely have an eye disease and would benefit from referral to an eye care provider.

5. Conclusions

The RETeval device offers a new approach for DR testing. Validated using gold standard ETDRS 7-standard field photography, this handheld device measures the eye's electrical and pupillary responses rather than photographing the retina. The benefits of this method include no dilation, short test time, minimal personnel training, immediate results, and low technical failure rates. The flicker ERG is largely unaffected by cataracts (Ratanapakorn et al., 2010) and the

Table 4
Detection of severe NPDR or PDR using ETDRS 7-standard field stereo photography as the gold standard in a subject's worst eye, as adjusted for ungradable subjects.

	RETeval device (this study)	Nonmydriatic 3-field stereoscopic photography (Ahmed et al., 2006; Bursell et al., 2001)	Ophthalmologist (Pugh et al., 1993)
Study sample size (n)	468	54	352
Sensitivity	87%	85%	33%
Specificity	78%	66%	99%
Positive Predictive Value	15%	10%	60%
Negative Predictive Value	99.2%	99.0%	97.0%
Per 1000 people tested, expected number of:			
False positives	210	325	10
False negatives	6	7	29
True negatives	746	631	946
True positives	38	37	15

RETeval device generates results even with small pupils (the smallest pupil measured in this study was 1.4 mm). We believe that the RETeval test can be performed easily in the primary care physician's office, or other locations where subjects with diabetes receive care or obtain medications and supplies.

Role of the funding source

The corresponding author (S.R. Fransen) and study biostatistician (W.J. Feuer) accept full responsibility for the conduct of this study and had control of the data at all times. LKC Technologies, Inc. participated in the design of the study but had no role in conducting the study, data collection, data management, data analysis, or interpretation of the data, other than to answer technical questions from the corresponding author and study biostatistician regarding parameters measured by the RETeval device.

Acknowledgments

The authors thank the study subjects for their contribution to the research, and Vanessa Bergman, Tim Booher, Carl Gibson, Dawn Gittemeier, Regina Hansen, Frank Hunleth, Dawn Lee, and Martin Milner for assistance in study implementation.

References

- Ahmed, J., Ward, T. P., Bursell, S. E., Aiello, L. M., Cavallerano, J. D., & Vigersky, R. A. (2006). The sensitivity and specificity of nonmydriatic digital stereoscopic retinal imaging in detecting diabetic retinopathy. *Diabetes Care*, *29*, 2205–2209.
- American Academy of Ophthalmology Retina/Vitreous Panel (2014). Preferred practice guidelines. Diabetic retinopathy. In American Academy of Ophthalmology (Ed.), (San Francisco, CA). <http://www.aaof.org/preferred-practice-pattern/diabetic-retinopathy-ppp-2014>.
- Early photocoagulation for diabetic retinopathy. ETDRS report number 9. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology*, *98*(1991), 766–785.
- Fundus photographic risk factors for progression of diabetic retinopathy. ETDRS report number 12. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology*, *98*(1991), 823–833.
- Bresnick, G. H., & Palta, M. (1987). Temporal aspects of the electroretinogram in diabetic retinopathy. *Archives of Ophthalmology*, *105*, 660–664.
- Brown, D. M., Schmidt-Erfurth, U., Do, D. V., Holz, F. G., Boyer, D. S., Midena, E., ... Korobelnik, J. F. (2015). Intravitreal aflibercept for diabetic macular edema: 100-Week results from the VISTA and VIVID studies. *Ophthalmology*, *122*, 2044–2052.
- Bursell, S. E., Cavallerano, J. D., Cavallerano, A. A., Clermont, A. C., Birkmire-Peters, D., Aiello, L. P., & Aiello, L. M. (2001). 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*, *108*, 572–585.
- Castell, P. (2012). Internal quality assurance guidance and best practice toolkit. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/416856/Internal_Quality_Assurance_Guidance_and_Best_Practice_Toolkit_vers_1_2_23_Jan_2012_1_1_.pdf (NHS Screening Programmes, Diabetic Eye)
- Centers for Disease Control and Prevention (2011). National diabetes fact sheet. In Centers for Disease Control and Prevention (Ed.), US Department of Health and Human Services. http://www.cdc.gov/diabetes/pubs/pdf/ndfs_2011.pdf.
- Chew, E. Y., Ferris, F. L., III, Csaky, K. G., Murphy, R. P., Agron, E., Thompson, D. J., ... Schachat, A. P. (2003). The long-term effects of laser photocoagulation treatment in patients with diabetic retinopathy: The early treatment diabetic retinopathy follow-up study. *Ophthalmology*, *110*, 1683–1689.
- Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *The Journal of Comparative Neurology*, *292*, 497–523.
- Ferris, F. L., III (1993). How effective are treatments for diabetic retinopathy? *JAMA*, *269*, 1290–1291.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). xvii. New York: Wiley (321 p. See page 218).
- Han, Y. -K., & Ohn, Y. -H. (2000). Changes of ERG parameters in diabetic retinopathy. *Journal of Korean Ophthalmological Society*, *41*, 149–155. http://210.101.116.107/kos/journalSearch_index.asp?year=2000&page=149&vol=41&iss=1.
- Holopigian, K., Seiple, W., Lorenzo, M., & Carr, R. (1992). A comparison of photopic and scotopic electroretinographic changes in early diabetic retinopathy. *Investigative Ophthalmology & Visual Science*, *33*, 2773–2780. <http://iovs.arvojournals.org/article.aspx?articleid=2160442>.
- Kardon, R. H., Hong, S., & Kawasaki, A. (2013). Entrance pupil size predicts retinal illumination in darkly pigmented eyes, but not lightly pigmented eyes. *Investigative Ophthalmology & Visual Science*, *54*, 5559–5567.
- Kato, K., Kondo, M., Sugimoto, M., Ikusugi, K., & Matsubara, H. (2015). Effect of pupil size on flicker ERGs recorded with RETeval system: New mydriasis-free full-field ERG system. *Investigative Ophthalmology & Visual Science*, *56*, 3684–3690.
- Kim, S. H., Lee, S. H., Bae, J. Y., Cho, J. H., & Kang, Y. S. (1997). Electroretinographic evaluation in adult diabetics. *Documenta Ophthalmologica. Advances in Ophthalmology*, *94*, 201–213.
- Kjeka, O., Jansson, R. W., Bredrup, C., & Krohn, J. (2013). Early panretinal photocoagulation for ERG-verified ischaemic central retinal vein occlusion. *Acta Ophthalmologica*, *91*, 37–41.
- Ku, J. J., Landers, J., Henderson, T., & Craig, J. E. (2013). The reliability of single-field fundus photography in screening for diabetic retinopathy: The Central Australian Ocular Health Study. *The Medical Journal of Australia*, *198*, 93–96.
- Larsson, J., Bauer, B., & Andreasson, S. (2000). The 30-Hz flicker cone ERG for monitoring the early course of central retinal vein occlusion. *Acta Ophthalmologica Scandinavica*, *78*, 187–190.
- Lawrence, M. G. (2004). The accuracy of digital-video retinal imaging to screen for diabetic retinopathy: An analysis of two digital-video retinal imaging systems using standard stereoscopic seven-field photography and dilated clinical examination as reference standards. *Transactions of the American Ophthalmological Society*, *102*, 321–340.
- Lee, P. P., Feldman, Z. W., Ostermann, J., Brown, D. S., & Sloan, F. A. (2003). Longitudinal rates of annual eye examinations of persons with diabetes and chronic eye diseases. *Ophthalmology*, *110*, 1952–1959.
- Lee, D. J., Kumar, N., Feuer, W. J., Chou, C. F., Rosa, P. R., Schiffman, J. C., ... Lam, B. L. (2014). Dilated eye examination screening guideline compliance among patients with diabetes without a diabetic retinopathy diagnosis: The role of geographic access. *BMJ Open Diabetes Research & Care*, *2*, e000031.
- McAnany, J. J., & Nolan, P. R. (2014). Changes in the harmonic components of the flicker electroretinogram during light adaptation. *Documenta Ophthalmologica. Advances in Ophthalmology*, *129*, 1–8.
- McCulloch, D. L., Marmor, M. F., Brigell, M. G., Hamilton, R., Holder, G. E., Tzekov, R., & Bach, M. (2015). ISCEV standard for full-field clinical electroretinography (2015 update). *Documenta Ophthalmologica. Advances in Ophthalmology*, *130*, 1–12.
- Messias, A., Ramos Filho, J. A., Messias, K., Almeida, F. P., Costa, R. A., Scott, I. U., ... Jorge, R. (2012). Electroretinographic findings associated with panretinal photocoagulation (PRP) versus PRP plus intravitreal ranibizumab treatment for high-risk proliferative diabetic retinopathy. *Documenta Ophthalmologica. Advances in Ophthalmology*, *124*, 225–236.
- Nakayama, M., Nakamura, J., Hamada, Y., Chaya, S., Mizubayashi, R., Yasuda, Y., ... Hotta, N. (2001). Aldose reductase inhibition ameliorates pupillary light reflex and F-wave latency in patients with mild diabetic neuropathy. *Diabetes Care*, *24*, 1093–1098.
- Nguyen, Q. D., Brown, D. M., Marcus, D. M., Boyer, D. S., Patel, S., Feiner, L., ... Ehrlich, J. S. (2012). Ranibizumab for diabetic macular edema: Results from 2 phase III randomized trials: RISE and RIDE. *Ophthalmology*, *119*, 789–801.
- Ortubé, M. C., Kiderman, A., Eydelman, Y., Yu, F., Aguilar, N., Nusinowitz, S., & Gorin, M. B. (2013). Comparative regional pupillography as a noninvasive biosensor screening method for diabetic retinopathy. *Investigative Ophthalmology & Visual Science*, *54*, 9–18.
- Park, S. E., Sun, H. J., Lee, H. J., Park, T. K., & Ohn, Y. -H. (2010). The role of electroretinography in assessing the progression of diabetic retinopathy. *Journal of Korean Ophthalmological Society*, *51*, 693–699. <http://synapse.koreamed.org/DOIx.php?id=10.3341/jkos.2010.51.5.693>.
- Paz, S. H., Varma, R., Klein, R., Wu, J., Azen, S. P., & Los Angeles Latino Eye Study G. (2006). Noncompliance with vision care guidelines in Latinos with type 2 diabetes mellitus: The Los Angeles Latino Eye Study. *Ophthalmology*, *113*, 1372–1377.
- Pugh, J. A., Jacobson, J. M., Van Heuven, W. A., Watters, J. A., Tuley, M. R., Lairson, D. R., ... Velez, R. (1993). Screening for diabetic retinopathy. The wide-angle retinal camera. *Diabetes Care*, *16*, 889–895.
- Ratanapakorn, T., Patarakittam, T., Sinawat, S., Sanguansak, T., Bhoombunchoo, C., Kaewpanna, S., & Yospaiboon, Y. (2010). Effect of cataract on electroretinographic response. *Journal of the Medical Association of Thailand = Chotmaihet Thangphaet*, *93*, 1196–1199. <http://www.scopus.com/record/display.uri?eid=2-s2.0-77958084751&origin=inward&txid=0>.
- Rosenberg, J. B., Friedman, I. B., & Gurland, J. E. (2011). Compliance with screening guidelines for diabetic retinopathy in a large academic children's hospital in the Bronx. *Journal of Diabetes and Its Complications*, *25*, 222–226.
- Saadine, J. B., Fong, D. S., & Yao, J. (2008). Factors associated with follow-up eye examinations among persons with diabetes. *Retina*, *28*, 195–200.
- Satoh, S., Iijima, H., Imai, M., Abe, K., & Shibuya, T. (1994). Photopic electroretinogram implicit time in diabetic retinopathy. *Japanese Journal of Ophthalmology*, *38*, 178–184.
- Severns, M. L., & Johnson, M. A. (1993). Predicting outcome in central retinal vein occlusion using the flicker electroretinogram. *Archives of Ophthalmology*, *111*, 1123–1130.
- Severns, M. L., Johnson, M. A., & Merritt, S. A. (1991). Automated estimation of implicit time and amplitude from the flicker electroretinogram. *Applied Optics*, *30*, 2106–2112.
- Smith, S. A., & Smith, S. E. (1983). Reduced pupillary light reflexes in diabetic autonomic neuropathy. *Diabetologia*, *24*, 330–332.
- Straub, R. H., Jeron, A., & Kerp, L. (1992). The pupillary light reflex. 2. Prevalence of pupillary autonomic neuropathy in diabetics using age-dependent and age-independent pupillary parameters. *Ophthalmologica. Journal international d'ophthalmologie. International journal of ophthalmology. Zeitschrift für Augenheilkunde*, *204*, 143–148.
- Straub, R. H., Thies, U., Jeron, A., Palitzsch, K. D., & Scholmerich, J. (1994). Valid parameters for investigation of the pupillary light reflex in normal and diabetic subjects shown by factor analysis and partial correlation. *Diabetologia*, *37*, 414–419.

- Tahara, K., Matsuura, T., & Otori, T. (1993). Diagnostic evaluation of diabetic retinopathy by 30-Hz flicker electroretinography. *Japanese Journal of Ophthalmology*, 37, 204–210.
- Tyrberg, M., Lindblad, U., Melander, A., Lovestam-Adrian, M., Ponjavic, V., & Andreasson, S. (2011). Electrophysiological studies in newly onset type 2 diabetes without visible vascular retinopathy. *Documenta Ophthalmologica. Advances in Ophthalmology*, 123, 193–198.
- Wang, H., Sun, X., Liu, K., & Xu, X. (2012). Intravitreal ranibizumab (Lucentis) for the treatment of diabetic macular edema: A systematic review and meta-analysis of randomized clinical control trials. *Current Eye Research*, 37, 661–670.
- Wilkinson, C. P., Ferris, F. L., III, Klein, R. E., Lee, P. P., Agardh, C. D., Davis, M., ... Verdaguer, J. T. (2003). Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*, 110, 1677–1682.
- Williams, G. A., Scott, I. U., Haller, J. A., Maguire, A. M., Marcus, D., & McDonald, H. R. (2004). Single-field fundus photography for diabetic retinopathy screening: A report by the American Academy of Ophthalmology. *Ophthalmology*, 111, 1055–1062.
- Yasuda, S., Kachi, S., Ueno, S., Piao, C. H., & Terasaki, H. (2015). Flicker electroretinograms before and after intravitreal ranibizumab injection in eyes with central retinal vein occlusion. *Acta Ophthalmologica*, 93, e465–e468.
- Yau, J. W., Rogers, S. L., Kawasaki, R., Lamoureux, E. L., Kowalski, J. W., Bek, T., ... Wong, T. Y. (2012). Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care*, 35, 556–564.
- Zhang, X., Saaddine, J. B., Chou, C. F., Cotch, M. F., Cheng, Y. J., Geiss, L. S., ... Klein, R. (2010). Prevalence of diabetic retinopathy in the United States, 2005–2008. *JAMA*, 304, 649–656.