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

April Maa, Emory University
William J. Feuer, University of Miami
C. Quentin Davis, LKC Technologies, Inc.
Ensa K. Pillow, Oklahoma City VA Medical Center
Tara D. Brown, Oklahoma City VA Medical Center
Rachel M. Caywood, Oklahoma City VA Medical Center
Joel E. Chasan, Emory University
Stephen R. Fransen, University of Oklahoma

Journal Title: Journal of Diabetes and its Complications
Volume: Volume 30, Number 3
Publisher: Elsevier | 2015-12-09, Pages 524-532
Type of Work: Article | Post-print: After Peer Review
Publisher DOI: 10.1016/j.jdiacomp.2015.12.005
Permanent URL: https://pid.emory.edu/ark:/25593/rzsxs

Final published version: http://dx.doi.org/10.1016/j.jdiacomp.2015.12.005

Copyright information:
© 2016 The Authors.
This is an Open Access work distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Accessed April 24, 2019 12:45 AM EDT
A novel device for accurate and efficient testing for vision-threatening diabetic retinopathy

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

Abstract

\textsuperscript{*}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.

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).

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.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
**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 sensitivity of 83%, specificity was 78% and negative predictive value was 99%. The average testing time was 2.3 minutes.

**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.

**Keywords**
Diabetic retinopathy; Diabetic macular edema; Electroretinography; Pupillography

---

1. Introduction

Diabetic retinopathy (DR) remains the leading cause of blindness among working age adults in the US\(^1\) and is a major cause of blindness worldwide\(^2\). In the US, less than 50% of the patients with diabetes receive an annual DR examination\(^4\)–\(^8\). Many of those with vision-threatening diabetic retinopathy (VTDR)\(^9\) are not diagnosed in time to benefit from the remarkable efficacy of DR therapy established by the Diabetic Retinopathy Study (DRS)\(^10\), Early Treatment Diabetic Retinopathy Study (ETDRS)\(^11\),\(^12\) and clinical trials of intravitreal anti-VEGF therapy\(^13\)–\(^15\). VTDR is defined as severe non-proliferative or proliferative DR, with or without clinically significant macular edema (CSME)\(^2\),\(^9\). 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\(^16\)–\(^18\) 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\(^19\). 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\(^20\) show a 40-fold increase in cone density in the fovea, the area of the fovea is quite small (3.6 mm\(^2\) out
of 1000 mm$^2$) and the rest of the retina is relatively uniform in cone density. Changes in these ERG parameters correlate strongly with DR severity$^{21–29}$. DR progression leads to increasing retinal ischemia$^2$ which prolongs the implicit time and decreases the amplitude$^{24, 29–31}$.

The pupillary response to a light stimulus has also been shown to correlate with DR severity$^{32–36}$. Both the rate and extent of pupil constriction in response to a light stimulus decrease with increasing DR severity$^{35}$.

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$^{37}$. 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/m$^2$) × pupil area (mm$^2$), thereby providing an ERG stimulus that is largely independent of pupil size$^{37}$. 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$^{37}$. While skin electrodes have traditionally been avoided in ERG testing due to lower signal levels$^{19}$, 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$^{38, 39}$. 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 seconds, 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\(^{21,26}\). The stimuli were presented in randomized order, with about 1 second 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\(^{38}\) 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\(^{40}\).

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, 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\(^2\)-\(^9\). 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\cite{41, 42}). 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 seconds 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\cite{43}.

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 was 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 was 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 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%). If the worldwide prevalence was used (10.2%), 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 characterizes ICC<0.4 to be poor reproducibility, ICC>0.75 excellent reproducibility, and ICC from 0.4–0.75 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 was 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%).

3.3. Outcomes

Figure 1 shows representative ERG and pupillary data for a subject with and without VTDR, while Figure 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 ≤0.002), and formed the prediction equation used to generate the RETeval device’s numerical output (called DR Score in Figure 2).

Figure 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 Figure 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).

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 minutes (standard deviation, 0.8 minutes) 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. 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%.

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 minutes 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. 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. 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%). If subjects with ungradable images are referred, the specificity of the digital retinal imaging system decreases from 100% to 66% when using a disease prevalence of 4.4%. 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.

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 (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%. 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\textsuperscript{30, 48–50}) 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.

4. 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\textsuperscript{51} and the RETeval device generates results even with small pupils (the smallest pupil measured in this study was 1.4 mm). We believe 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.

Acknowledgments

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

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.

References


218


Figure 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 μV, larger than the 23 μV for the subject without 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 seconds) to 5 seconds. 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.
Figure 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.
Figure 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.
### Table 1

<table>
<thead>
<tr>
<th>Severity groups</th>
<th>International Clinical Classification&lt;sup&gt;39&lt;/sup&gt;</th>
<th>ETDRS Level</th>
<th>CSME</th>
<th>Calibration phase (N=371)</th>
<th>Validation phase (N = 96)</th>
<th>Application to primary care prevalence (N=467)</th>
<th>Primary care prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>No NPDR</td>
<td>10 – 12</td>
<td>-</td>
<td>74</td>
<td>19</td>
<td>93 (20%)</td>
<td>74.2%</td>
<td></td>
</tr>
<tr>
<td>Mild NPDR</td>
<td>14 – 35</td>
<td>-</td>
<td>101</td>
<td>26</td>
<td>127 (27%)</td>
<td>20.1%</td>
<td></td>
</tr>
<tr>
<td>Moderate NPDR</td>
<td>43 – 47</td>
<td>-</td>
<td>72</td>
<td>18</td>
<td>90 (19%)</td>
<td>2.5%</td>
<td></td>
</tr>
<tr>
<td>CSME with No, Mild, or Moderate NPDR</td>
<td>10 – 47</td>
<td>+</td>
<td>31</td>
<td>8</td>
<td>39 (8%)</td>
<td>0.8%</td>
<td></td>
</tr>
<tr>
<td>Severe NPDR or Proliferative DR</td>
<td>53 – 85</td>
<td>+ / −</td>
<td>41</td>
<td>11</td>
<td>52 (11%)</td>
<td>2.3%</td>
<td></td>
</tr>
<tr>
<td>Ungradable ETDRS Level</td>
<td>?</td>
<td>+</td>
<td>12</td>
<td>3</td>
<td>15 (3%)</td>
<td>0.5%</td>
<td></td>
</tr>
<tr>
<td>Ungradable fundus photographs</td>
<td>?</td>
<td>?</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>51 (11%)</td>
<td></td>
</tr>
</tbody>
</table>
### Table 2

Subject characteristics

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

* 2-sample t-test; † chi-squared test; ‡ Fisher exact test
Table 3

Performance metrics with the primary care prevalence. LCL and UCL represent the lower and upper 95% confidence limits, respectively.

<table>
<thead>
<tr>
<th>RETeval cutoff value</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Predictive Power of Test</th>
<th>Percent of Tested Sample</th>
<th>Positive Test</th>
<th>Negative Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Positive</td>
<td>Negative</td>
<td>True</td>
<td>False</td>
</tr>
<tr>
<td>17.6</td>
<td>93%</td>
<td>52%</td>
<td>8%</td>
<td>99.5%</td>
<td>4.1%</td>
<td>46%</td>
</tr>
<tr>
<td>LCL</td>
<td>73%</td>
<td>47%</td>
<td>5%</td>
<td>97%</td>
<td>2.4%</td>
<td>41%</td>
</tr>
<tr>
<td>UCL</td>
<td>99.9%</td>
<td>57%</td>
<td>13%</td>
<td>100.0%</td>
<td>6.5%</td>
<td>51%</td>
</tr>
<tr>
<td>18.1</td>
<td>90%</td>
<td>57%</td>
<td>9%</td>
<td>99.2%</td>
<td>3.8%</td>
<td>41%</td>
</tr>
<tr>
<td>LCL</td>
<td>65%</td>
<td>52%</td>
<td>5.0%</td>
<td>97%</td>
<td>2.2%</td>
<td>36%</td>
</tr>
<tr>
<td>UCL</td>
<td>99%</td>
<td>62%</td>
<td>14%</td>
<td>99.9%</td>
<td>6.2%</td>
<td>46%</td>
</tr>
<tr>
<td>19.1</td>
<td>88%</td>
<td>67%</td>
<td>11%</td>
<td>99.1%</td>
<td>3.8%</td>
<td>31%</td>
</tr>
<tr>
<td>LCL</td>
<td>65%</td>
<td>62%</td>
<td>6%</td>
<td>97%</td>
<td>2.2%</td>
<td>27%</td>
</tr>
<tr>
<td>UCL</td>
<td>99%</td>
<td>72%</td>
<td>17%</td>
<td>99.9%</td>
<td>6.2%</td>
<td>36%</td>
</tr>
<tr>
<td>20.0</td>
<td>83%</td>
<td>78%</td>
<td>15%</td>
<td>99%</td>
<td>3.6%</td>
<td>21%</td>
</tr>
<tr>
<td>LCL</td>
<td>59%</td>
<td>74%</td>
<td>9%</td>
<td>97%</td>
<td>2.0%</td>
<td>17%</td>
</tr>
<tr>
<td>UCL</td>
<td>96%</td>
<td>82%</td>
<td>23%</td>
<td>99.9%</td>
<td>5.9%</td>
<td>25%</td>
</tr>
<tr>
<td>21.2</td>
<td>73%</td>
<td>84%</td>
<td>17%</td>
<td>99%</td>
<td>3.1%</td>
<td>15%</td>
</tr>
<tr>
<td>LCL</td>
<td>47%</td>
<td>80%</td>
<td>9%</td>
<td>97%</td>
<td>1.7%</td>
<td>12%</td>
</tr>
<tr>
<td>UCL</td>
<td>90%</td>
<td>87%</td>
<td>27%</td>
<td>99.5%</td>
<td>5.3%</td>
<td>19%</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th></th>
<th>RETeval device (this study)</th>
<th>Nonmydriatic 3-field stereoscopic photography(^{45, 46})</th>
<th>Ophthalmologist (^{47})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study sample size (n)</td>
<td>468</td>
<td>54</td>
<td>352</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>87%</td>
<td>85%</td>
<td>33%</td>
</tr>
<tr>
<td>Specificity</td>
<td>78%</td>
<td>66%</td>
<td>99%</td>
</tr>
<tr>
<td>Positive Predictive Value</td>
<td>15%</td>
<td>10%</td>
<td>60%</td>
</tr>
<tr>
<td>Negative Predictive Value</td>
<td>99.2%</td>
<td>99.0%</td>
<td>97.0%</td>
</tr>
</tbody>
</table>

Per 1000 people tested, expected number of:

<table>
<thead>
<tr>
<th></th>
<th>RETeval device (this study)</th>
<th>Nonmydriatic 3-field stereoscopic photography(^{45, 46})</th>
<th>Ophthalmologist (^{47})</th>
</tr>
</thead>
<tbody>
<tr>
<td>False positives</td>
<td>210</td>
<td>325</td>
<td>10</td>
</tr>
<tr>
<td>False negatives</td>
<td>6</td>
<td>7</td>
<td>29</td>
</tr>
<tr>
<td>True negatives</td>
<td>746</td>
<td>631</td>
<td>946</td>
</tr>
<tr>
<td>True positives</td>
<td>38</td>
<td>37</td>
<td>15</td>
</tr>
</tbody>
</table>