



Published in final edited form as:

*Ophthalmology*. 2015 February ; 122(2): 356–366. doi:10.1016/j.ophtha.2014.07.056.

## Subjective and Objective Screening Tests for Hydroxychloroquine Toxicity

Catherine Cukras, MD, PhD<sup>1</sup>, Nancy Huynh, MD<sup>2</sup>, Susan Vitale, PhD, MHS<sup>1</sup>, Wai T. Wong, MD, PhD<sup>3</sup>, Fredrick L. Ferris III, MD<sup>1</sup>, Paul A. Sieving, MD, PhD<sup>2,4</sup>

<sup>1</sup>Division of Epidemiology and Clinical Research, National Eye Institute, National Institutes of Health, Bethesda, Maryland. <sup>2</sup>Ophthalmic Genetics and Visual Function Branch, National Eye Institute, National Institutes of Health, Bethesda, Maryland. <sup>3</sup>Office of the Scientific Director, National Eye Institute, National Institutes of Health, Bethesda, Maryland. <sup>4</sup>Section for Translational Research in Retinal and Macular Degeneration, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, Maryland.

### Abstract

**Objective:** To compare subjective and objective clinical tests used in the screening for hydroxychloroquine retinal toxicity to multifocal electroretinography (mfERG) reference testing.

**Design:** Prospective, single-center, case control study.

**Participants:** Fifty-seven patients with a previous or current history of hydroxychloroquine treatment of more than 5 years' duration.

**Methods:** Participants were evaluated with a detailed medical history, dilated ophthalmologic examination, color fundus photography, fundus autofluorescence (FAF) imaging, spectral-domain (SD) optical coherence tomography (OCT), automated visual field testing (10–2 visual field mean deviation [VFMD]), and mfERG testing. We used mfERG test parameters as a gold standard to divide participants into 2 groups: those affected by hydroxychloroquine-induced retinal toxicity and those unaffected.

**Main Outcome Measures:** We assessed the association of various imaging and psychophysical variables in the affected versus the unaffected group.

**Results:** Fifty-seven study participants (91.2% female; mean age, 55.7±10.4 years; mean duration of hydroxychloroquine treatment, 15.0±7.5 years) were divided into affected (n = 19) and unaffected (n = 38) groups based on mfERG criteria. Mean age and duration of hydroxychloroquine treatment did not differ statistically between groups. Mean OCT retinal thickness measurements in all 9 macular subfields were significantly lower (<40 μm) in the affected group ( $P < 0.01$  for all comparisons) compared with those in the unaffected group.

---

Correspondence: Catherine Cukras, MD, PhD, Division of Epidemiology and Clinical Research, National Eye Institute, National Institutes of Health, Bethesda, MD 20892. cukrasc@nei.nih.gov.

Supplemental material is available at [www.aajournal.org](http://www.aajournal.org).

Financial Disclosure(s):

The author(s) have no proprietary or commercial interest in any materials discussed in this article.

Mean VFMD was 11 dB lower in the affected group ( $P < 0.0001$ ). Clinical features indicative of retinal toxicity were scored for the 2 groups and were detected in 68.4% versus 0.0% using color fundus photographs, 73.3% versus 9.1% using FAF images, and 84.2% versus 0.0% on the scoring for the perifoveal loss of the photoreceptor ellipsoid zone on SD-OCT for affected and unaffected participants, respectively. Using a polynomial modeling approach, OCT inner ring retinal thickness measurements and Humphrey 10–2 VFMD were identified as the variables associated most strongly with the presence of hydroxychloroquine as defined by mfERG testing.

**Conclusions:** Optical coherence tomography retinal thickness and 10–2 VFMD are objective measures demonstrating clinically useful sensitivity and specificity for the detection of hydroxychloroquine toxicity as identified by mfERG, and thus may be suitable surrogate tests.

Hydroxychloroquine is widely used in the treatment of various autoimmune diseases, but has the potential to cause severe retinal dysfunction and vision loss.<sup>1</sup> Current guidelines from the American Academy of Ophthalmology (AAO) recommend starting annual ophthalmic screening within 1 year of initiating hydroxychloroquine therapy. The guidelines further recommend that patients receiving hydroxychloroquine therapy for more than 5 years be evaluated using automated 10–2 visual field testing plus one or more of the following objective tests: spectral-domain (SD) optical coherence tomography (OCT), multifocal electroretinography (mfERG), or fundus autofluorescence (FAF) imaging.<sup>1</sup> The current recommendations exclude lower-yield tests such as color vision, and instead focus on subjective and objective tests believed to be associated with early toxicity.

Screening for hydroxychloroquine toxicity in the general ophthalmic community presents practical challenges. Although disparate testing methods can reveal changes consistent with hydroxychloroquine toxicity,<sup>2,3</sup> some methods, such as mfERG, are not widely available. Other imaging and psychophysical tests may not identify early changes associated with toxicity with high sensitivity and specificity and often rely on subjective expert interpretation where thresholds for determining toxicity are not well established. The optimal algorithm for hydroxychloroquine toxicity screening using different methods is still being debated.<sup>4,5</sup>

Considerations for screening recommendations include accessibility, reliability, ease of interpretation, and cost of testing. A recent article by Browning<sup>4</sup> reported that revisions in the AAO hydroxychloroquine screening guidelines from the 2002 version to its current revised 2011 version resulted in a 40% increase in total associated health expenditure costs, rising from an estimated \$29 million to \$40.7 million. Both Marmor<sup>5</sup> and Browning<sup>4</sup> point out in their exchange that the AAO guidelines do not explicitly discuss that a certain level of expertise is needed to interpret mfERG, visual field, and OCT data.<sup>5</sup> They recommended that further studies are needed to assess the relative usefulness of testing methods and to optimize guidelines to identify those affected by hydroxychloroquine toxicity.

In several recent studies, mfERG assessment has been considered to be the gold standard test for the detection of hydroxychloroquine toxicity because it has the dual characteristics of being both an objective test and a direct measure of retinal function (Invest Ophthalmol Vis Sci 2013;54:3597; Invest Ophthalmol Vis Sci 2013;54:5037; Invest Ophthalmol Vis Sci 2013;54:5105).<sup>6</sup> Hydroxychloroquine toxicity typically manifests on mfERG testing as a

characteristic ring of depressed responses in the perifoveal regions of the macula.<sup>7,8</sup> An increase in the ratio of central-to-paracentral response amplitudes (i.e., an increased R1-to-R2 ratio) is diagnostically useful, providing high sensitivity and specificity in predicting toxicity.<sup>6</sup> However, mfERG testing is not available in most ophthalmology practices and requires specialized training to perform and analyze the test results.

The purpose of this study was to evaluate the findings of subjective and objective screening tests recommended by the current AAO guidelines in a prospective study of participants receiving long-term hydroxychloroquine therapy. Study participants had at least 5 years of hydroxychloroquine therapy and were identified using mfERG as the reference gold standard as having or not having hydroxychloroquine toxicity. These 2 groups then were evaluated with various testing methods suggested by the AAO 2011 guidelines including (1) automated visual field testing, (2) SD-OCT imaging, (3) fundus photography, (4) FAF imaging, and (5) visual acuity measurements. The results of these tests were evaluated for association with the presence or absence of hydroxychloroquine toxicity as defined by mfERG testing to establish which tests could best serve as surrogates for mfERG testing results. These findings may help to enable screening ophthalmologists to have a more targeted approach, with more widely available tests, to identify patients with hydroxychloroquine toxicity.

## Methods

### Study Participants

This prospective case-control study was conducted at the eye clinic of the National Eye Institute, National Institutes of Health, Bethesda, Maryland. Inclusion criteria included a current or previous history of hydroxychloroquine treatment for a total duration exceeding 5 years and an absence of concomitant retinal disorders (e.g., diabetic retinopathy, retinal vein occlusion, age-related macular degeneration, or Stargardt's disease). Information on patient characteristics, including demographics, medical history, body weight and height, duration and cumulative dose of hydroxychloroquine therapy, and diagnostic indications for hydroxychloroquine treatment, were obtained by medical history evaluation.

The study protocol and informed consent forms were approved by a National Institutes of Health–based institutional review board and the study was registered at [www.clinicaltrials.gov](http://www.clinicaltrials.gov) (identifier, [NCT01145196](https://clinicaltrials.gov/ct2/show/study/NCT01145196)). The study protocol adhered to the tenets of the Declaration of Helsinki and complied with the Health Insurance Portability and Accountability Act.

### Study Procedures

All participants underwent a comprehensive ocular examination, including best-corrected visual acuity testing using the Early Treatment Diabetic Retinopathy Study (ETDRS) protocol, slit-lamp examination, and dilated fundus examination. In addition, all patients underwent mfERG testing, automated visual field testing, and retinal imaging, including SD-OCT, FAF imaging, and color fundus photography. Testing was performed in both eyes of all participants.

## Visual Field Testing and Analysis

Perimetric assessment was performed using a standard 10–2 Humphrey Visual Field Analyzer (Humphrey Instruments, Inc, San Leandro, CA) with a white test spot. The visual field mean deviation (VFMD) values, representing deviation from age-matched normal eyes, were obtained from the visual field output.

## Multifocal Electroretinography Testing and Analysis

Multifocal ERG testing was performed according to the International Society for Clinical Electrophysiology of Vision guidelines,<sup>9</sup> based on the 61-hexagon stimulus pattern of the VERIS Clinic system (Electro-Diagnostic Imaging, Inc, Redwood, CA). Each hexagon elicits a waveform consisting of a negative trough (N1), followed by a positive peak (P1), followed by another negative trough (N2). The 61 hexagon responses were grouped into 5 concentric rings (R1–R5), as shown in Figure 1. The average amplitude, measured as (P1–N1), was assessed for each ring outside the R1 hexagon. The average response densities (nanovolts per degrees squared) within concentric rings from the center (ring 1) to the periphery (ring 5) were generated by the mfERG VERIS software (Fig 1A). The ring ratios of the mfERG were defined as ratios of the central hexagon amplitude (R1) to each of the peripheral ring amplitudes (R2–R5). These ratios were calculated for all tested eyes.

## Spectral-Domain Optical Coherence Tomography Imaging and Analysis

We evaluated both the objective quantitative retina thickness in all ETDRS subfields as well as the subjective assessment of the OCT of all participants by 2 masked educated graders (C.C., N.H.). Foveal-centered SD-OCT volumes were obtained for both eyes from each participant on the Cirrus-HD system (Carl Zeiss Meditec, Inc, Dublin, CA) using the macular cube 512×128 scan pattern. The macular thickness map was divided into 3 concentric circles based on the ETDRS grading grid: a central circle (0.5 mm or 1.5° radius) centered on the fovea, a concentric inner ring (1.5 mm or 5° radius), and a concentric outer ring (3 mm or 10° radius). Radii at 45° and 135° angles were used to divide the circles into the 9 ETDRS subfields: the central subfield and 4 inner and 4 outer subfields (temporal, superior, nasal, and inferior subfields; Fig 1B). Mean retinal thicknesses in each of the 9 subfields were generated by the manufacturer's software version 6.5.0.772 (Carl Zeiss Meditec, Inc).

The OCT images also were acquired in parallel using the Heidelberg Spectralis HRA + OCT system (Spectralis; Heidelberg Engineering, Heidelberg, Germany). Horizontal 9.5-mm images through the fovea with 100 scans averaged were graded manually for the presence or absence of anatomic disruptions in the perifoveal ellipsoid zone (EZ; i.e., the mitochondrial rich layer near the inner segment-outer segment junction) located approximately 0.5 to 1 mm from the fovea. In cases where the quality of Spectralis images was insufficient to visualize this region clearly (n = 3 of 57 participants), corresponding Cirrus HD-OCT images were graded in their place. Two independent readers (C.C., N.H.) performed the grading in a masked fashion, with any discordant grades resolved by consensus after joint review.

## Fundus Photography and Fundus Autofluorescence Imaging

Digital fundus color images were obtained using the Topcon fundus camera (TRC-50EX; Topcon Medical Systems, Oakland, NJ). The FAF images (excitation, 488 nm; emission, >500 nm) were obtained with the Spectralis scanning laser ophthalmoscope (Heidelberg Engineering, Heidelberg, Germany). Color images were graded for both eyes of all study patients (n = 57). For 1 participant, FAF images were not obtained, and for another, FAF images for the left eye were of poor quality and could not be graded. The FAF and color fundus images were graded independently in a masked fashion by 2 ophthalmologists (C.C., N.H.). Grading was based on the scale developed by Marmor.<sup>3</sup> A score of 0 to 4 was given for each image based on the following criteria: 0 = normal, 1 = patchy damage, 2 = bull's-eye damage, 3 = bull's-eye damage involving fovea or retinal pigment epithelium, and 4 = diffuse posterior pole damage. For images that were discrepant between the 2 graders, a consensus grading was agreed on after joint review of the masked images. Masked grading of the fundus photographs was concordant for 82 (72%) of the 114 eyes. Of the 32 eyes (28%) with discordant grades, 17 (15%) were discordant on their normal (score, 0) versus abnormal (score, 1–4) status, whereas 15 (13%) were discordant on their severity score (score, 1–4). Masked grading of the FAF images was concordant for 110 (99.1%) of the 111 eyes on a 5-step severity scale similar to that used in the grading of color fundus photographs.

### Definition of Toxicity

Participants were divided into 2 groups according to the presence (the affected group) or absence (the unaffected group) of hydroxychloroquine-related toxicity using objective mfERG criteria as the gold standard. Participants were assigned to the affected group based on the presence of either of the following 2 conditions: (1) increased R1-to-R2 ratio (defined as exceeding the 99% confidence limits for the normal population), or (2) reduced R1 absolute amplitude (defined as less than the 99% confidence limits for the normal population).<sup>6</sup> All remaining participants were assigned to the unaffected group. Because R1 amplitudes vary with age, cutoffs for the lower and upper limits of normal were defined within each age group.<sup>6</sup> If one eye of a participant met criteria for toxicity but the other did not, the participant was assigned to the affected group.

### Determination of Study Eye for Statistical Analyses

Because analyses showed that, for all test parameters, the right eye and left eye of participants were correlated highly, only 1 eye was used for further statistical analyses. For affected participants, if R1 was normal, the eye with the higher R1-to-R2 ratio (i.e., the worse eye) was designated to be the study eye; if R1 was abnormal, the eye with the smaller R1 was designated the worse eye. In unaffected participants, the right eye was chosen arbitrarily as the study eye. Multivariate analyses were performed using study eyes only. Raw data are shown for both study and fellow eyes as well as right eye and left eye in Table 1 (available at [www.aaojournal.org](http://www.aaojournal.org)).

## Statistical Analysis

Analyses were performed using SAS software version 9.3 (SAS Inc, Cary, NC). Spearman correlations between right and left eyes were computed. We conducted preliminary univariate analyses to explore differences between affected and unaffected participants (Wilcoxon rank-sum test) using the study eye for each participant.

To identify the parameters most strongly associated with affected or unaffected status while controlling for type I error, we used a cross-validation approach.<sup>10</sup> A random number generator was used to divide the sample into 2 subsamples. For each parameter, we fit a logistic regression model on the first subsample to compute the area under the receiver operating characteristic (ROC) curve and to identify the cut-point associated with Youden's J statistic, an index that identifies as optimal the point on the ROC curve that is farthest from chance.<sup>11</sup> The optimal cut-point identified from the first subsample then was applied to the second subsample, and the sensitivity and specificity were determined in the second subsample.

Some researchers believe that the cross-validation approach may create highly variable cut-points depending on the sample used and biased estimates of sensitivity and specificity.<sup>12</sup> Therefore, we also used an analytic approach suggested by Royston and Altman<sup>13</sup> using polynomial models to identify the best-fitting curve to the data, followed by a series of hierarchical models to find the overall best-fitting model, while adjusting for age and other pertinent covariates. Based on comparisons of deviances,<sup>14</sup> we found that a linear model was the best fit for each of the individual parameters.

Because the parameters were highly correlated in many cases, we systematically analyzed subgroups of parameters in the full dataset, using a stepwise selection technique to identify parameters within each subgroup that were associated most strongly with affected or unaffected status. The subgroups of parameters were: (1) the 9 sector retinal thicknesses from the OCT output, (2) 2 summary parameters (inner ring and outer ring) from the OCT output, (3) VFMD, (4) visual acuity, (5) fundus color score, and (6) autofluorescence score. Within each subgroup of parameters, the most strongly associated parameters were identified and then combined with similarly selected parameters from the other subgroups. We then used stepwise logistic regression models to find the combination of variables that provided the best fit to the data.

The Spearman correlation coefficient was used to assess the linear relationship between mfERG ring ratios and OCT thicknesses, visual acuity, and VFMD. A *P* value of less than 0.05 was considered to be statistically significant. The accuracy of each measurement parameter in discriminating between affected and unaffected patients was evaluated by using ROC curves.

## Results

### Categorization of Study Participants into Hydroxychloroquine Toxicity Categories: Affected and Unaffected

For each of the 57 participants, mfERG testing was performed on both eyes and R1-to-R2 ratios were calculated and both R1 and R1-to-R2 ratios were compared with the published limits.<sup>6</sup> Eyes falling below these mfERG limits were defined as demonstrating evidence of toxicity. Sixteen participants had bilateral evidence of toxicity and were classified as affected (Fig 2, solid red circles). Another 3 participants had one eye meeting the criteria for toxicity and the fellow eye not meeting criteria (Fig 2, open red circles). These individuals also were included in the affected group. For the remaining 38 participants, neither eye met toxicity criteria and they were categorized as unaffected (Fig 2, black squares).

Mean values from mfERG parameters in the affected and unaffected groups are shown in Table 1 (available at [www.aaojournal.org](http://www.aaojournal.org)). Correlation coefficients between right and left eye mfERG R1 through R5 amplitudes were between 0.9 and 0.96 ( $P < 0.0001$  for all comparisons).

### Study Participant Characteristics According to Affected Status

Study participants had a mean age of  $55.7 \pm 10.7$  years (range, 31–72 years), with most being women (70%). Most participants (95%) received hydroxychloroquine therapy for either lupus or rheumatoid arthritis. The mean duration of hydroxychloroquine treatment was  $15.0 \pm 7.5$  years.

There were no statistically significant differences between the affected and unaffected groups with regard to mean age, indication for treatment, height, or weight (Table 2;  $P = 0.05$  for all comparisons). Some patient factors previously implicated in increasing the likelihood of hydroxychloroquine toxicity<sup>3,15</sup> did not differ significantly between affected and unaffected groups: mean dose of hydroxychloroquine of more than 6.5 mg/kg daily (76% vs. 69%;  $P = 0.54$ ), concomitant renal or liver disease (0% vs. 16%;  $P = 0.16$ ), and body mass index exceeding 30 kg/m<sup>2</sup> (21% vs. 37%;  $P = 0.36$ ). However, the proportion of participants older than 60 years was higher in the affected group than in the unaffected group (58% vs. 29%), a difference that reached borderline statistical significance ( $P = 0.05$ ).

### Visual Acuity and Automated Visual Field Testing

Mean visual acuity was 20/20 for right and left eyes in the unaffected group, although in the affected group, the mean was 20/32 and 20/25, respectively (Table 3). Although the mean differences were statistically significant, there was considerable overlap in the ranges. The VFMD was computed from Humphrey Visual Field 10–2 testing of both eyes of all participants and was correlated between fellow eyes ( $r = 0.97$ ;  $P < 0.0001$ ). Mean VFMD demonstrated a greater defect in the affected group ( $-12.3 \pm 8.8$  dB right eye;  $P < 0.0001$ ) relative to the unaffected group ( $-0.8 \pm 1.5$  dB right eye), with minimal overlap in their ranges.

## Optical Coherence Tomography Imaging

**Qualitative Assessment.**—The SD-OCT images obtained in both eyes of all participants were graded qualitatively by scoring masked images for the presence or absence of pericentral interruption of the photoreceptor EZ. For each participant, right eye and left eye grades were identical. Discontinuity or loss of the EZ was present in 84% (16/19) of affected participants and in 0% (0/38) of unaffected participants ( $P < 0.0001$ ). Representative SD-OCT images in affected participants are shown in Figure 3.

**Quantitative Assessment of Optical Coherence Tomography Thickness.**—The OCT images also were analyzed quantitatively by measuring the retinal thickness in the macula according to macular subfields. The mean thickness for each of the 9 macular subfields is summarized in Table 4. The correlation for all mean subfield thicknesses between right and left eyes ranged between 0.88 and 0.98 (all  $P < 0.0001$ ). For all subfields, the affected group had significantly thinner mean OCT thickness than the unaffected group ( $P < 0.001$ ; Table 5). The percentage reduction in thickness between affected and unaffected groups was quite similar for all regions, ranging between 16% and 21%.

## Fundus Color Images

Marmor score grading of right and left eye color fundus photographs were highly correlated ( $r = 0.95$ ;  $P < 0.0001$ ). All the eyes in the unaffected group were graded as normal by both graders. Among the affected group, 6 patients (31.6%) were graded as normal (grade 0), 2 patients (10.5%) had patchy damage (grade 1), 6 patients (31.6%) demonstrated bull's-eye damage (grade 2), 4 patients (21%) had bull's-eye damage involving the fovea or retinal pigment epithelium (grade 3), and 1 patient (5.3%) showed diffuse damage of the posterior pole (grade 4). Table 5 summarizes the fundus photograph grading for the unaffected and affected groups with examples of the reference scale used in Figure 4 (available at [www.aaojournal.org](http://www.aaojournal.org)).

## Fundus Autofluorescence Findings

Study and fellow eye FAF scores were correlated ( $r = 0.99$ ;  $P < 0.0001$ ). Most eyes (91.9%) were graded as normal in the unaffected group. For the affected group, the eyes were graded as follows: grade 0 (26.3%), grade 1 (5.3%), grade 2 (21%), grade 3 (15.8%), and grade 4 (31.6%). Table 6 summarizes the FAF grading with examples of the reference scale used in Figure 5 (available at [www.aaojournal.org](http://www.aaojournal.org)).

## Statistical Analysis and Optimal Cut-Points

**Cross-validation Analyses.**—In the cross-validation approach, the sample was divided randomly into 2 subsamples. The ROC curve, the area under the ROC curve (AUC), and Youden's J statistic were calculated for each parameter, and then the cut-point was applied to the second subsample to determine its performance (Table 7). For example, using Youden's J statistic for the OCT inner inferior subfield, the optimal cut-point for the first half of the sample was 278.5  $\mu\text{m}$  (AUC, 0.98; sensitivity, 100%; specificity, 88.2%). When this cut-point was applied to the second half-sample, the results were not as strong: of 8 affected subjects, 6 were identified correctly (sensitivity, 75%), and of 21 unaffected patients, 19



were classified correctly (specificity, 90.5%). When the second half-sample was analyzed independently, the optimal cut-point was 245.6  $\mu\text{m}$  and the AUC was 0.94. The variability of the optimal cut-points provides evidence that a simple classification rule based on findings from our dataset may be problematic.

**Stepwise Approach to Identify Parameters Most Closely Associated with Affected or Unaffected Status.**—We also analyzed the data by fitting separate stepwise logistic regression models combining the most strongly associated parameters. In each of the models, the OCT inner inferior subfield was the single most strongly associated variable with affected or unaffected status. The final best-fitting model included OCT inner inferior subfield thickness (odds ratio, 0.94; 95% CI, 0.89–1.00;  $P=0.045$ ) and VFMD (odds ratio, 0.55; 95% CI, 0.31–0.99;  $P=0.047$ ). Adjustment for age, total dose of hydroxychloroquine, and duration of hydroxychloroquine use (in a model separate from total hydroxychloroquine dose) showed that these covariates were not associated significantly with affected or unaffected status and did not alter the odds ratio estimates for the OCT inner inferior subfield and VFMD. To provide additional data for reference purposes, we computed percentiles for the affected and unaffected groups for inner inferior OCT thickness and VFMD (Fig 6).

## Discussion

This study evaluated the usefulness of various screening procedures currently recommended by the AAO relative to mfERG testing. In this dataset, the testing parameters most strongly associated with affected or unaffected status, as defined by mfERG, were OCT retinal thickness, especially of the inner inferior subfield, and VFMD. The results of the polynomial modeling and stepwise logistic regression approach were consistent with results based on computations of AUC and bolstered the validity of our conclusions. Further studies will be needed to provide more robust estimates of the associations of these parameters with affected or unaffected status.

Visual field testing has been the primary screening tool recommended by the AAO and widely available in the community. Humphrey Visual Field 10–2 testing is attractive in that it is a widely available test that measures visual function. However, it is a subjective test that is affected by the reliability of the tester and can reflect changes other than those of the retina. The test often is interpreted subjectively, with examiners assessing for paracentral visual loss with identification of a partial or full ring scotoma. The threshold for determining toxicity is not well established and is subject to the interpreter. Interpretation of the visual field test often is difficult, because too low of a threshold for identifying important field changes potentially subjects a patient to stopping a drug that is helping them systemically, whereas too high of a threshold will fail to identify signs of retinal damage. In this study, we used the VFMD as a quantitative output of this subjective test and, in our cohort, found that the results correlated well with the affected status as determined mfERG. Had our participants not been such reliable test-takers, the visual field data may not be as compelling.

Traditional interpretation of SD-OCT for determination of toxicity is qualitative, with hydroxychloroquine toxicity manifested as a loss or disruption of perifoveal photoreceptor

EZ, and relies on trained graders identifying what are often subtle findings.<sup>3,16</sup> In our study, we were able to identify EZ disruption in most (84%) of our affected patients, but some cases of toxicity were missed using this qualitative method of evaluation even under conditions of high suspicion by educated graders. The advantage of this evaluation is that it is specific for toxicity, but it requires training and, even with trained graders, lacks sensitivity.

As a quantitative measure, we found that there was a statistically significant difference in retinal thickness across all the OCT subfields between the unaffected and affected groups. Anatomically, the ring 2 on mfERG testing and the inner subfield of the OCT correspond to this area of paracentral retina (Fig 1A). Consistent with the findings of Kahn et al,<sup>17</sup> the greatest difference in retinal thickness was observed in the inner subfield, corresponding to the area 1 mm from the foveal center. In this study, the average inner OCT subfield thickness was correlated significantly with the mfERG R1-to-R2 ratio across all participants ( $r = -0.45$ ;  $P = 0.0007$ ). Furthermore, as Marmor<sup>3</sup> observed, thinning of the inner inferior subfield seems to be correlated especially with the presence of toxicity. The quantitative measure of OCT subfield thicknesses has the advantage of being an objective measurement of an objective test that has excellent correlation with mfERG testing. Depending on the cutoff used, it can be made to be very sensitive, and thus represents an attractive screening tool. It is difficult to recommend an optimal cutoff screening number given the relatively small dataset. We therefore provided percentile distributions for each group (Table 8) so that investigators can balance sensitivity and specificity with individual desires.

From Table 8, we can see that, in this cohort, if a cutoff for the OCT thickness of the inner inferior subfield of less than 305  $\mu\text{m}$  were used, for example, it would capture all of the affected participants in this cohort (100% sensitivity) and also would include slightly more than 25% of unaffected participants. Similarly, if a cutoff of worse than  $-1.2$  dB on the VFMD were used, it would include all of the affected participants (100% sensitivity) and slightly less than 25% of the unaffected participants.

Our results support that fundus examination and photography are not very sensitive methods and can miss cases of hydroxychloroquine toxicity, because 31.6% of patients with toxicity were not found to have abnormalities on the fundus photographs. Although examination of FAF images provided better results than examination of color photography, there was still a problem with sensitivity, and the method still relied on subjective interpretation by the grader. In the patients with toxicity, most patients (73.7%) showed some abnormality on FAF images, with 31.6% showing diffuse posterior pole damage. Here again, this test lacks adequate sensitivity and requires subjective interpretation.

Although most ophthalmology practices do not have the capacity for mfERG testing, many do have access to SD-OCT machines, and a quantitative OCT thickness measurement (especially thinning of the inner inferior subfield) can serve as a useful objective screening tool for possible of hydroxychloroquine toxicity. Also, OCT images can be examined to identify the presence of paracentral EZ disruption, and if present, this seems to be a quite specific finding of toxicity. Central visual field testing with the Humphrey Visual Field 10–2 is widely available and also can provide important visual function information.

Our data suggest that the combination of SD-OCT and Humphrey Visual Field 10–2 testing can be used as a screening tool to identify patients with possible hydroxychloroquine toxicity. Additional testing and the consistency of the evidence then can be used to decide whether to recommend that hydroxychloroquine be discontinued.

## Acknowledgment.

The authors thank Sam Dresner for his help with data collection.

Supported by the National Institutes of Health Intramural Research Programs of the National Eye Institute and the National Institute on Deafness and Other Communication Disorders. The sponsor or funding organization had no role in the design or conduct of this research.

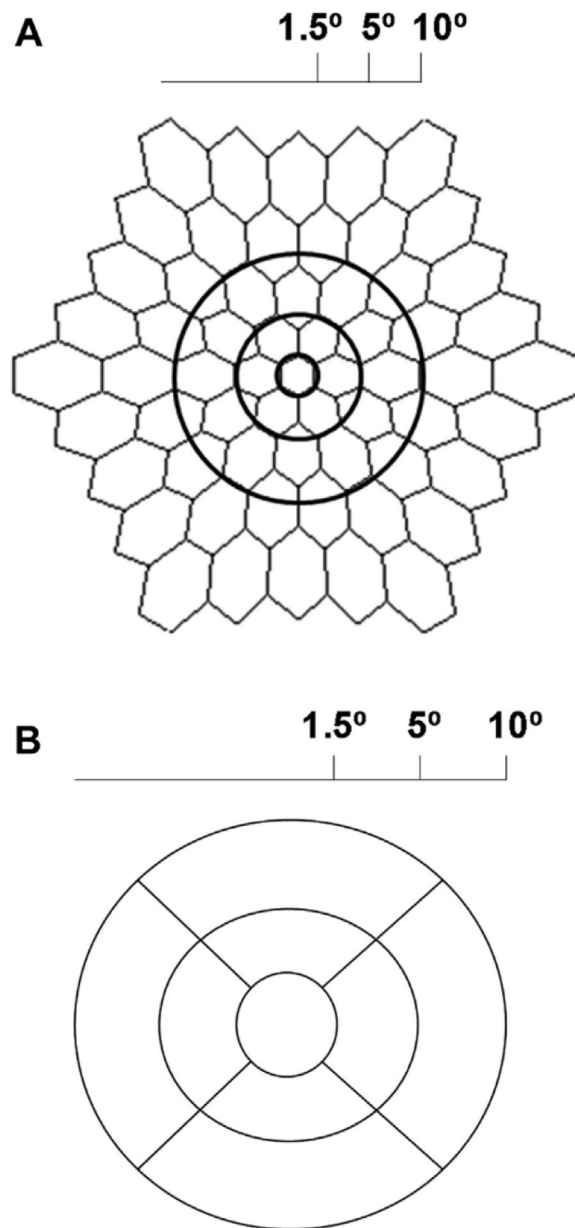
## Abbreviations and Acronyms:

<b>AUC</b>	area under the receiver operating characteristic curve
<b>ETDRS</b>	Early Treatment Diabetic Retinopathy Study
<b>EZ</b>	ellipsoid zone
<b>FAF</b>	fundus autofluorescence
<b>mfERG</b>	multifocal electroretinography
<b>OCT</b>	optical coherence tomography
<b>ROC</b>	receiver operating characteristic
<b>SD</b>	spectral domain
<b>VFMD</b>	visual field mean deviation

## References

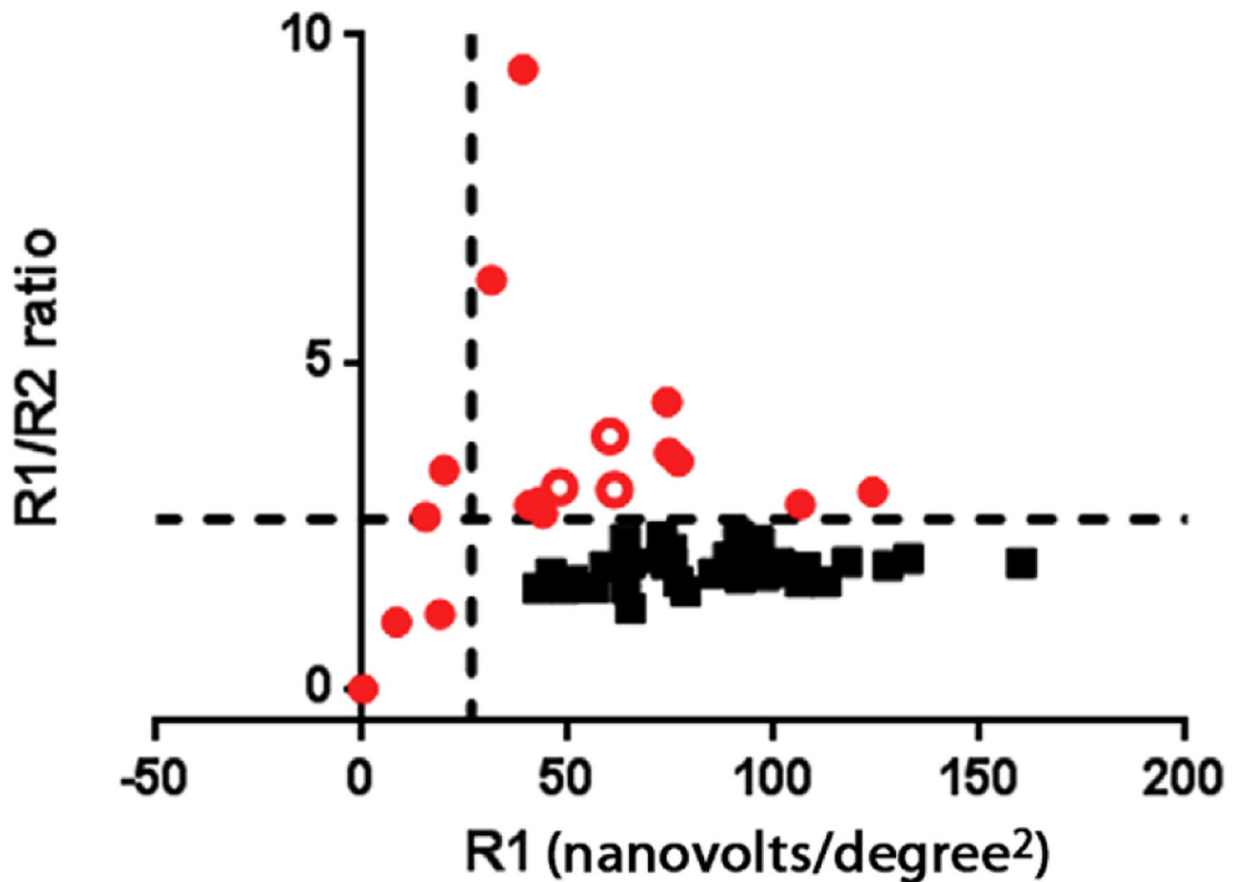
1. Marmor MF, Kellner U, Lai TY, et al.; American Academy of Ophthalmology. Revised recommendations on screening for chloroquine and hydroxychloroquine retinopathy. *Ophthalmology*2011;118:415–22. [PubMed: 21292109]
2. Kellner S, Kellner U, Weber BH, et al. Lipofuscin- and melanin-related fundus autofluorescence in patients with ABCA4-associated retinal dystrophies. *Am J Ophthalmol*2009;147:895–902. [PubMed: 19243736]
3. Marmor MF. Comparison of screening procedures in hydroxychloroquine toxicity. *Arch Ophthalmol*2012;130:461–9. [PubMed: 22159170]
4. Browning DJ. Impact of the revised American Academy of Ophthalmology guidelines regarding hydroxychloroquine screening on actual practice. *Am J Ophthalmol*2013; 155: 418–28. [PubMed: 23218706]
5. Marmor MF. Efficient and effective screening for hydroxychloroquine toxicity. *Am J Ophthalmol*2013; 155: 413–4. [PubMed: 23394729]
6. Lyons JS, Severns ML. Detection of early hydroxychloroquine retinal toxicity enhanced by ring ratio analysis of multifocal electroretinography. *Am J Ophthalmol*2007; 143: 801–9. [PubMed: 17336914]
7. Maturi RK, Yu M, Weleber RG. Multifocal electroretinographic evaluation of long-term hydroxychloroquine users. *Arch Ophthalmol*2004; 122:973–81. [PubMed: 15249360]

8. Lai TY, Chan WM, Li H, et al. Multifocal electroretinographic changes in patients receiving hydroxychloroquine therapy. *Am J Ophthalmol*2005; 140:794–807. [PubMed: 16310457]
9. Hood DC, Bach M, Brigell M, et al.; International Society for Clinical Electrophysiology of Vision. ISCEV standard for clinical multifocal electroretinography (mfERG) (2011 edition). *Doc Ophthalmol*2012;124:1–13.
10. Faraggi D, Simon R. A simulation study of cross-validation for selecting an optimal cutpoint in univariate survival analysis. *Stat Med*1996;15:2203–13. [PubMed: 8910964]
11. Perkins NJ, Schisterman EF. The inconsistency of “optimal” cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol*2006; 163: 670–5. [PubMed: 16410346]
12. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*2006;25:127–41. [PubMed: 16217841]
13. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Appl Stat*1994;43:429–67.
14. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol*1999;28:964–74. [PubMed: 10597998]
15. Marmor MF, Carr RE, Easterbrook M, et al.; American Academy of Ophthalmology. Recommendations on screening for chloroquine and hydroxychloroquine retinopathy: a report by the American Academy of Ophthalmology. *Ophthalmology*2002;109:1377–82. [PubMed: 12093666]
16. Chen E, Brown DM, Benz MS, et al. Spectral domain optical coherence tomography as an effective screening test for hydroxychloroquine retinopathy (the “flying saucer” sign). *Clin Ophthalmol*2010;4:1151–8. [PubMed: 21060664]
17. Kahn JB, Haberman ID, Reddy S. Spectral-domain optical coherence tomography as a screening technique for chloroquine and hydroxychloroquine retinal toxicity. *Ophthalmic Surg Lasers Imaging*2011;42:493–7. [PubMed: 21830746]



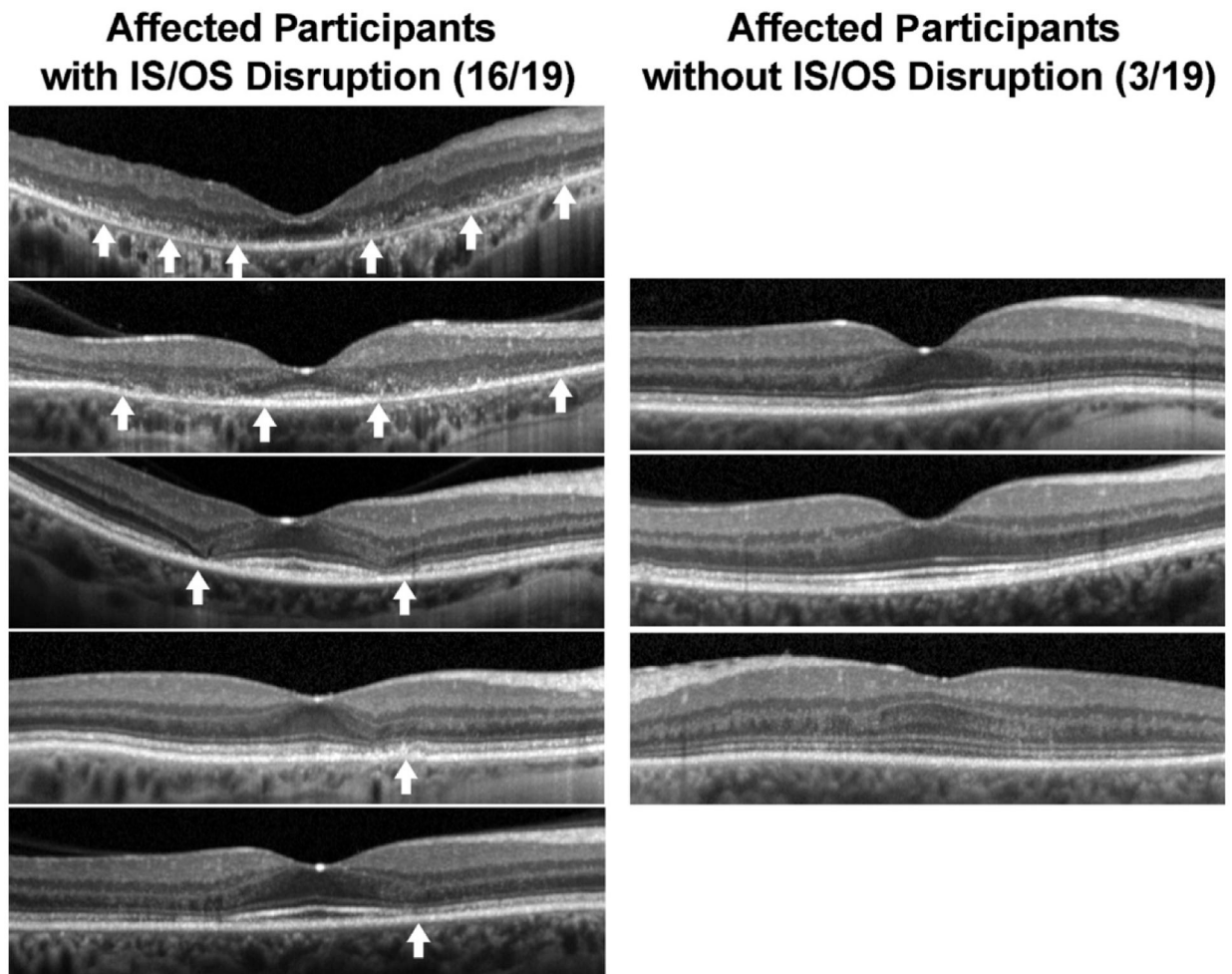
**Figure 1.**

**A**, Multifocal electroretinography image demonstrating anatomic location for generation of ring ratios. Ring 1 corresponds to the central 3°, ring 2 represents the area from 3° to 10°, and ring 3 covers the area from 10° to 20°. **B**, Diagram showing the anatomic location of the optical coherence tomography subfields. The central subfield represents the central 3°, the inner subfield represents the area from 5° to 10°, and the outer subfield corresponds to the area from 10° to 20°.

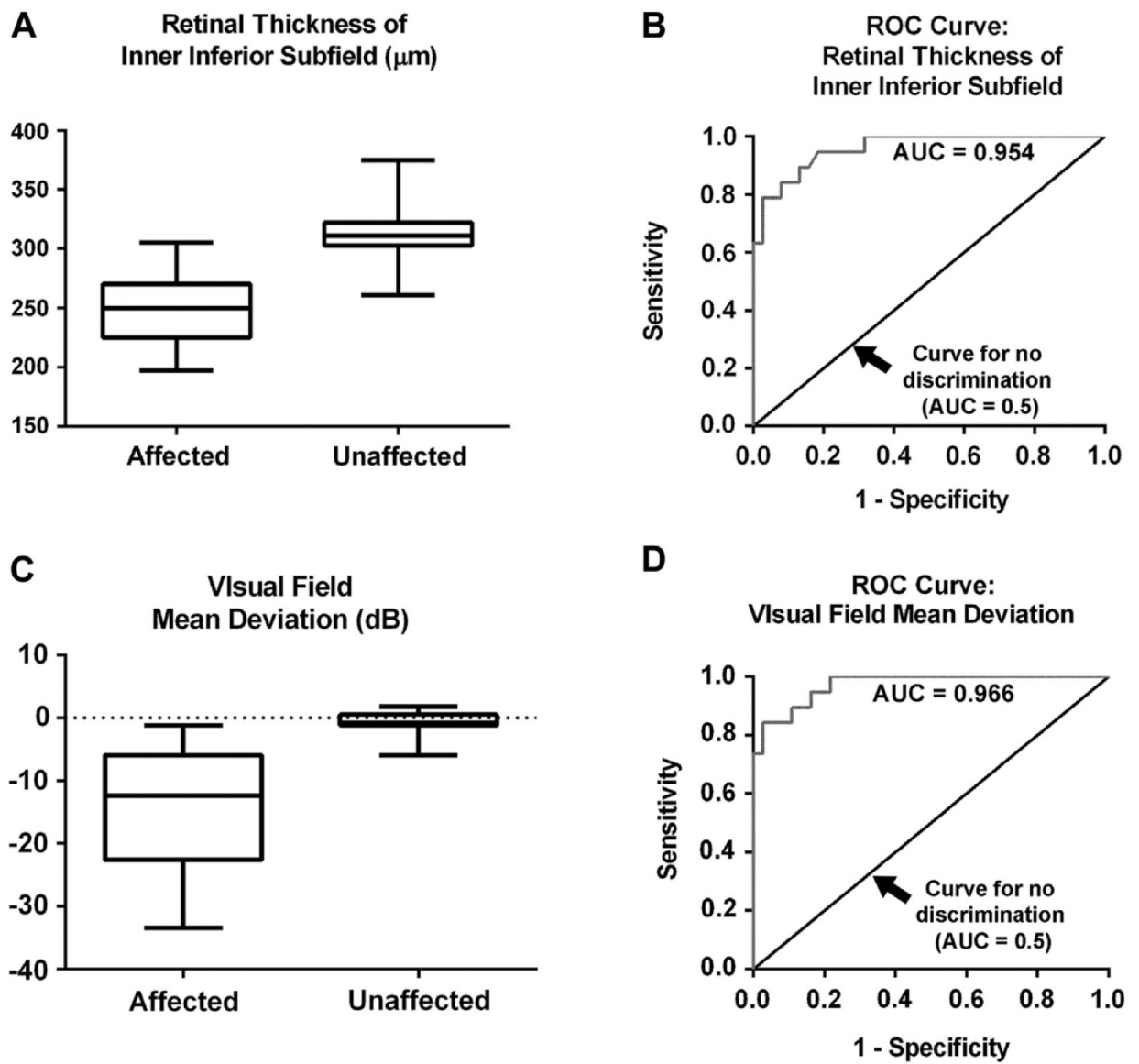


**Figure 2.**

Graph showing multifocal electroretinography (mfERG) data from the worse eye of all participants. R1 central amplitudes are reported from the output of the mfERG directly and R1/R2 plots ring ratios. All eyes categorized as unaffected are shown in solid red circles. Participants with both eyes meeting criteria for the affected category are shown in solid black squares. Open red circles indicate individuals with one eye meeting mfERG criteria for toxicity with the other eye not meeting criteria. One eye meeting criteria is sufficient in this analysis to categorize the person as affected.



**Figure 3.** **Left column,** Grading of Spectralis optical coherence tomography images of affected and unaffected participants with evaluation for disruption of the ellipsoid zone (EZ) in the paracentral retina. **Right column,** Examples of range of disruption in affected participants and images from affected participants without evidence of EZ disruption. Arrows indicate areas of disruption of the EZ.



**Figure 6.** Box-and-whisker plot for affected and unaffected patients showing (A) optical coherence tomography (OCT) inner inferior subfield retinal thickness and (B) the retinal thickness of inner inferior subfield from the OCT. Shaded boxes represent values from the 25th to 75th percentile with whiskers extending to the 1st percentile and 99th percentile. Receiver operating characteristic (ROC) curves for (C) OCT inner inferior subfield retinal thickness and (D) visual field mean deviation. AUC = area under the ROC curve.



Table 2.

## Baseline Patient Characteristics

	Affected (n = 19)	Unaffected (n = 38)	P Value
Age (yrs), mean ± SD	58.8±10.0	54.1±10.4	0.13
Patients older than 60 yrs			0.05
No. (%)	11 (58)	11 (29)	
Mean ± SD	66.5±3.3	66.3±3.0	
Gender, no. (%)			0.66
Female	18 (94.7)	34 (89.5)	
Male	1 (5.3)	4 (10.5)	
Height (inches), mean ± SD	62.8±2.6	62.6±3.9	0.51
Ideal body weight (%), mean ± SD	135.7±60.9	144.9±35.6	0.06
Total cumulative HCQ dose (g), mean ± SD	1871±927	2036±1141	0.71
No. of patients using >6.5 mg/kg daily of HCQ, no. (%)	13 (68.9)	29 (76.3)	0.54
Length of treatment (yrs), mean ± SD	15.3±7.1	14.9±8.3	0.54
Concomitant renal or liver disease, no. (%)	0 (0)	6 (15.8)	0.16
Obesity (BMI >30 kg/m <sup>2</sup> ), no. (%)	4 (21)	14 (36.8)	0.36
Indication for HCQ use, no. (%)			1.00
Lupus	12 (63.2)	24 (63.2)	
RA	6 (31.6)	12 (31.6)	
Sjögren syndrome	1 (5.3)	2 (5.3)	

BMI = body mass index; HCQ = hydroxychloroquine; RA = rheumatoid arthritis; SD = standard deviation.

**Table 3.**

**Baseline Ocular Characteristics**

	<b>Affected (n = 19)</b>	<b>Unaffected (n = 38)</b>	<b>P Value</b>
Mean Snellen VA			
Right eye			
No. of letters (range)	20/32 (20/16–20/200)	20/20 (20/12.5–20/63)	0.001
Mean ± SD	77.4±12.4	85.3±5.8	
Left eye			
No. of letters (range)	20/25 (20/12.5–20/250)	20/20 (20/12.5–20/63)	0.05
Mean ± SD	78.1±13.6	84.2±7.8	
Worse eye (study eye)			
No. of letters (range)	20/25 (20/16–20/250)	20/20 (20/12.5–20/50)	0.03
Mean ± SD	78.0±13.4	85.3±5.8	
Better eye (fellow eye)			
No. of letters (range)	20/32 (20/12.5–20/200)	20/20 (20/12.5–20/63)	0.04
Mean ± SD	77.5±12.7	84.2±7.8	
HVF mean deviation, mean ± SD			
Right eye	-12.3±8.8	-0.8±1.5	<0.0001
Left eye	-13.0±9.6	-0.9±1.4	<0.0001
Worse eye (study eye)	-14.0±10.1	-0.8±1.5	<0.001
Better eye (fellow eye)	-13.2±9.6	-0.9±1.4	<0.001
OCT photoreceptor IS/OS disruption, no. (%)			
Right eye	16 (84)	0 (0)	<0.0001
Left eye	16 (84)	0 (0)	<0.0001
Worse eye (study eye)			
Better eye (fellow eye)			

HVF = Humphrey Visual Field; IS/OS = inner segment–outer segment; OCT = optical coherence tomography; SD = standard deviation; VA = visual acuity.

**Table 4.** Mean Cirrus Optical Coherence Tomography Retinal Thickness in 9 Early Treatment Diabetic Retinopathy Study Subfields

Subfield	Affected (n = 19), Mean ± Standard Deviation (µm)	Unaffected (n = 38), Mean ± Standard Deviation (µm)	P Value	Absolute Mean Difference (µm)	% Difference
Center					
Right eye	208.3±49.3	248.5±37.4	0.009	40.2	16.18
Left eye	208.6±50.2	252.3±30.6	0.005	43.7	17.32
Study eye	208.9±49.9	248.5±37.4	0.001	39.6	15.94
Fellow eye	207.9±49.6	252.3±30.6	0.002	44.4	17.60
Inner superior					
Right eye	263.7±35.5	316.4±17.8	<0.0001	52.7	16.66
Left eye	262.5±39.9	317.0±18.4	<0.0001	54.5	17.19
Study eye	261.4±37.8	316.4±17.8	<0.001	55	17.38
Fellow eye	264.8±37.7	317.0±18.4	<0.001	52.2	16.47
Inner nasal					
Right eye	260.5±33.9	317.3±20.1	<0.0001	56.8	17.90
Left eye	262.5±36.4	317.3±18.4	<0.0001	54.8	17.27
Study eye	263.0±34.2	317.3±20.1	<0.001	54.3	17.11
Fellow eye	260.0±36.1	317.3±18.4	<0.001	57.3	18.06
Inner inferior					
Right eye	251.7±32.6	311.1±19.3	<0.0001	59.4	19.09
Left eye	251.0±32.4	311.2±17.4	<0.0001	60.2	19.34
Study eye	251.3±30.4	311.1±19.3	<0.001	59.8	19.22
Fellow eye	251.4±34.5	311.2±17.4	<0.001	59.8	19.22
Inner temporal					
Right eye	240.5±32.9	302.5±19.5	<0.0001	62	20.50
Left eye	238.9±36.0	302.3±18.3	<0.0001	63.4	20.97
Study eye	238.2±34.7	302.5±19.5	<0.001	64.3	21.26
Fellow eye	241.2±34.2	302.3±18.3	<0.001	61.1	20.21
Outer superior					
Right eye	232.8±30.5	277.6±15.7	<0.0001	51.6	16.14
Left eye	234.2±36.6	276.0±15.2	<0.0001	49.2	15.14

Subfield	Affected (n = 19), Mean ± Standard Deviation (µm)	Unaffected (n = 38), Mean ± Standard Deviation (µm)	P Value	Absolute Mean Difference (µm)	% Difference
Study eye	231.7±34.3	277.6±15.7	<0.001	49.5	16.53
Fellow eye	235.4±33.0	276.0±15.2	<0.001	51.3	14.71
Outer nasal					
Right eye	240.2±36.1	291.8±18.6	<0.0001	51.6	17.68
Left eye	242.8±36.9	292.0±17.2	<0.0001	49.2	16.85
Study eye	242.3±35.0	291.8±18.6	<0.001	49.5	16.96
Fellow eye	240.7±38.0	292.0±17.2	<0.001	51.3	17.57
Outer inferior					
Right eye	215.6±33.5	263.8±15.2	<0.0001	48.2	18.27
Left eye	215.5±34.3	263.3±15.5	<0.0001	47.8	18.15
Study eye	216.0±33.9	263.1±15.2	<0.001	47.1	17.90
Fellow eye	215.1±34.8	263.3±15.4	<0.001	48.2	18.31
Outer temporal					
Right eye	203.6±30.6	257.9±22.9	<0.0001	54.3	21.05
Left eye	204.3±33.5	259.9±17.7	<0.0001	55.6	21.39
Study eye	202.8±32.6	257.9±22.9	<0.001	55.1	21.36
Fellow eye	205.0±31.6	259.9±17.7	<0.001	54.9	21.12

**Table 5.**

## Fundus Photograph Consensus Grading

	Affected (n = 19), No. (%)	Unaffected (n = 38), No. (%)
Right eye		
0	6 (31.6)	38 (100)
1	2 (10.5)	0 (0)
2	6 (31.6)	0 (0)
3	4 (21)	0 (0)
4	1 (5.3)	0 (0)
Left eye		
0	8 (42.1)	38 (100)
1	0 (0)	0 (0)
2	6 (31.6)	0 (0)
3	4 (21)	0 (0)
4	1 (5.3)	0 (0)
Study eye		
0	7 (36.8)	38 (100)
1	0 (0)	0 (0)
2	7 (36.8)	0 (0)
3	4 (21.0)	0 (0)
4	1 (5.3)	0 (0)
Fellow eye		
0	7 (36.8)	38 (100)
1	2 (10.5)	0 (0)
2	5 (26.3)	0 (0)
3	4 (21)	0 (0)
4	1 (5.3)	0 (0)

0 = normal; 1 = patchy damage; 2 = bull's-eye damage; 3 = bull's-eye damage involving fovea or retinal pigment epithelium; 4 = diffuse posterior pole damage.

**Table 6.**

## Fundus Autofluorescence Consensus Grading

	Affected (n = 19)	Unaffected (n = 38 <sup>*</sup> ), No. (%)
Right eye		
0	5 (26.3)	34 (91.9)
1	1 (5.3)	2 (5.4)
2	4 (21)	0 (0)
3	3 (15.8)	1 (2.7)
4	6 (31.6)	0 (0)
Left eye		
0	5 (26.3)	33 (86.8) <sup>*</sup>
1	1 (5.3)	2 (5.4)
2	4 (21)	0 (0)
3	2 (10.5)	1 (2.7)
4	7 (36.8)	0 (0)
Study eye		
0	5 (26.3)	34 (91.9)
1	1 (5.3)	2 (5.4)
2	4 (21)	0 (0)
3	2 (10.5)	1 (2.7)
4	7 (36.8)	0 (0)
Fellow eye		
0	5 (26.3)	33 (91.7) <sup>*</sup>
1	1 (5.3)	2 (5.6)
2	4 (21)	0 (0)
3	3 (15.8)	1 (2.8)
4	6 (31.6)	0 (0)

0 = normal; 1 = patchy damage; 2 = bull's-eye damage; 3 = bull's-eye damage involving fovea or retinal pigment epithelium; 4 = diffuse posterior pole damage.

<sup>\*</sup> One participant without fundus autofluorescence images resulting from equipment malfunction during patient visit. Additional left eye of 1 participant unable to obtain fundus autofluorescence images.

**Table 7.** Receiver Operating Characteristic Curve Univariate Analysis of Random Half-Sampling of Data

Variable	Youden's J Statistic	Optimal Cut-Point	Area under the Receiver Operating Characteristic Curve	Sensitivity (%)	Specificity (%)
Best-corrected visual acuity	0.373	70.55 letters	0.739	67	71
Fundus photographs	0.700	—	0.850	70	100
Fundus autofluorescence	0.675	—	0.856	80	88
Visual field mean deviation	0.882	-3.30 dB	0.978	100	88
OCT subfield thickness (µm)					
Center	0.373	178.40	0.686	67	71
Inner superior	0.719	292.93	0.915	78	94
Inner temporal	0.830	269.80	0.954	89	94
Inner inferior*	0.882	278.50	0.980	100	88
Inner nasal	0.889	295.06	0.948	89	100
Outer superior	0.719	252.59	0.882	78	94
Outer temporal	0.824	205.82	0.941	100	82
Outer inferior	0.765	236.58	0.928	100	76
Outer nasal	0.660	261.69	0.863	78	88

OCT = optical coherence tomography.

\* Most strongly associated variable.

**Table 8.**

Cutoff Point Based on Percentiles for Affected and Unaffected Groups

Percentiles	Cirrus Optical Coherence Tomography Inner Inferior Subfield Thickness ( $\mu\text{m}$ )		Humphrey Visual Field 10–2 Mean Deviation (dB)	
	Affected ( $n = 17$ )	Unaffected ( $n = 38$ )	Affected ( $n = 16$ )	Unaffected ( $n = 37$ )
1	197*	261	-27.56*	-5.96
5	197	275	-27.56	-3.24
10	210	287	-22.60	-2.98
25	235	303	-14.33	-1.18
50	255	311	-10.09	-0.81
75	270	322	-4.94	0.49
90	294	328	-1.91	0.76
95	305	336	-1.22	1.24
99	305*	375	-1.22*	1.80

\* Insufficient numbers to distinguish this percentile estimate from the adjacent category.