



Published in final edited form as:

Acta Ophthalmol. 2014 August ; 92(5): e377–e381. doi:10.1111/aos.12353.

Validation of the Color Difference Plot Scoring System Analysis of the 103 Hexagon Multifocal Electroretinogram in the Evaluation of Hydroxychloroquine Retinal Toxicity

Gabrielle S. Graves, MD, Murtaza K. Adam, MD, Kimberly E. Stepien, MD, and Dennis P. Han, MD

Abstract

Purpose—To evaluate sensitivity, specificity, and reproducibility of color difference plot analysis (CDPA) of 103-Hexagon multifocal electroretinogram (mfERG) in detecting established hydroxychloroquine (HCQ) retinal toxicity.

Methods—23 patients taking HCQ were divided into those with and without retinal toxicity, and were compared with a control group without retinal disease and not taking HCQ. CDPA with two masked examiners was performed using age-corrected mfERG responses in the central ring (R_c ; 0 to 5.5 degrees from fixation) and paracentral ring (R_p ; 5.5 to 11 degrees from fixation). An abnormal ring was defined as containing any hexagons with a difference of 2 or more standard deviations from normal (color blue or black).

Results—Categorical analysis (ring involvement or not) showed R_c had 83% sensitivity and 93% specificity. R_p had 89% sensitivity and 82% specificity. Requiring abnormal hexagons in both R_c and R_p yielded sensitivity and specificity of 83% and 95% respectively. If required in only one ring, they were 89% and 80%, respectively. In this population, there was complete agreement in identifying toxicity when comparing CDPA using R_p with ring ratio analysis using R5/R4 P_1 ring responses (89% sensitivity, 95% specificity). Continuous analysis of CDPA with receiver operating characteristic analysis showed optimized detection (83% sensitivity, 96% specificity) when 4 abnormal hexagons were present anywhere within the R_p ring outline. Intergrader agreement and reproducibility were good.

Conclusions—CDPA had sensitivity and specificity that approached that of ring ratio analysis of R5/R4 P_1 responses. Ease of implementation and reproducibility are notable advantages of CDPA.

Corresponding author: Dennis P. Han, MD, Froedtert and Medical College of Wisconsin Eye Institute, 925 N. 87th St., Milwaukee, WI 53226, Telephone: 414-955-7875, Fax: 414-955-6300, dhan@mcw.edu.

All of the authors listed above meet the following criteria for authorship: 1) substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; 2) drafting the article or revising it critically for important intellectual content; and 3) final approval of the version to be published. None of the above authors have competing interests with respect to the material presented in this manuscript.

This study was presented as an abstract in a meeting previously: Association for Research in Vision and Ophthalmology, Poster Session on May 9, 2012.

Introduction

Hydroxychloroquine (HCQ), a drug often used to treat rheumatologic diseases, has a good safety and efficacy profile except for a low incidence of retinal toxicity [1]. An important goal has been to identifying HCQ retinal toxicity before the occurrence of visual loss, which is often characterized by paracentral or central visual field abnormalities and pigmentary macular disturbances observed by ophthalmoscopy. The American Academy of Ophthalmology recommendation for screening includes a comprehensive ophthalmologic exam and Humphrey visual field (HVF) 10-2 perimetry, along with one of the following: multifocal electroretinography, spectral domain optical coherence tomography, or fundus autofluorescence [2]. HVF 10-2 perimetry findings, such as paracentral scotomata, correlate with retinal damage and may appear before other clinical symptoms present [3]. So et al found that 103-hexagon mfERG showed decreased response amplitude in the pericentral region in both symptomatic and asymptomatic patients taking HCQ, and thus may demonstrate early signs of toxicity [4]. Lai et al showed asymptomatic loss of mfERG amplitudes of N1 and P1 in patients who took HCQ, with no further decrease of amplitude after 1-2 years of follow up in patients that continued to take HCQ, but an improvement in N1 and P1 amplitudes in patients who discontinued it. Lai et al also found that an increased cumulative dosage of HCQ correlated with decreased N1 and P1 amplitudes [5].

Several methods of evaluating mfERG results have been proposed, including quantitative and qualitative methods. These methods have analyzed ring amplitudes, ring ratios, and color difference plots [1,6-8]. This study utilized the 103-hexagon mfERG stimulus to evaluate patients taking hydroxychloroquine. The color difference plots were analyzed using Chang et al's Color Difference Plot Analysis (CDPA). Chang et al used this method in comparison with response amplitude measurements and had shown good statistical agreement [1]. We hypothesized that CDPA would also show good statistical agreement with ring ratio analysis. A goal of the study was to determine CDPA sensitivity and specificity in a group of patients diagnosed with HCQ toxicity by HVF 10-2 perimetry testing and fundus exam and/or OCT testing.

Materials and Methods

All data was collected during routine clinical practice. The study design was approved by the Institutional Review Board and is in accordance with the Health Insurance Portability and Accountability Act (HIPAA) regulations. The study population used in this study the same as described by Adam et al evaluating utility of ring ratio analysis for hydroxychloroquine toxicity detection [8].

Reference Controls

As previously reported by Adam et al [8], all of the eyes chosen as reference controls underwent mfERG for numerous reasons and were without known inherited retinal degeneration, injury, or disease. Data from one eye of 78 patients without confirmed bilateral retinal disease were obtained. Patients were aged 11 to 73 years (median 44 years, mean 41.8 years, standard deviation (SD) 23.4 years). Most patients were female (81%). All patients had Snellen acuity 20/30 or better.

Hydroxychloroquine Patients

Patients were recorded as taking hydroxychloroquine (HCQ) for systemic lupus erythematosus, rheumatoid arthritis, mixed connective disease, or Sjogren syndrome. The patients were split into HCQ-nontoxic and HCQ-toxic groups based on abnormal results on HVF 10-2 perimetry testing with confirmation via abnormal fundus exam and/or OCT findings. A detailed description for the process of group assignment was published by Adam et al [8]. Visual field evidence of retinal toxicity was confirmed by a review of automated visual fields by two study investigators (DPH and KES), each masked to the patients' clinical data, diagnosis and the other examiner's grades. In none of the cases did visual field reliability indices exceed error rates of 15%; 100% agreement between examiners was observed. Abnormalities that categorized patients as having retinal toxicity included arcuate, pericentral or central visual field loss, bullseye depigmentation of the retinal pigment epithelium on fundus exam, and loss of the photoreceptor ellipsoid line on spectral domain OCT. In all, 100% of these patients presented with one of the aforementioned visual field abnormalities while 75% had additional confirmatory findings on fundus exam or OCT. The patients within the HCQ-toxic group were aged 44-75 years (mean 60.8 years, SD 11.7 years). The patients within the HCQ-nontoxic groups were aged 17-76 years (mean 55.1 years, SD 15.2 years). Most patients in both groups were female (93%). There were 18 eyes in the HCQ-toxic group and 26 eyes in the HCQ-nontoxic group. A total of 44 eyes in 23 patients were analyzed.

Multifocal ERG testing

As described by Adam et al, ring ratio study [8] the Visual Evoked Response Imaging System v. 5.2.5X (VERIS; Electro-Diagnostic Imaging, Inc, Redwood City, CA, USA) was used following International Society for Clinical Electrophysiology of Vision guidelines for mfERG. Testing was done using 1% tropicamide and 2.5% phenylephrine to produce dilation. Proparacaine was applied to reduce discomfort. Burian Allen IR electrode was used and room lights were on during the duration of the testing. The 103-hexagon, 9 minutes 6 second test method was used following instructions by Maturi et al and Chang et al. Test stimuli spanned about 44 degrees across the central retina. Hexagons were scaled to form roughly equivalent mfERG amplitudes as a function of eccentricity. Spatial averaging for each of the 103 focal stimulus values was set at 17%. The VERIS spatial density plot setting of "refined" (default setting) was used for high resolution display of interpolated hexagons.

The mfERG color difference plot displayed color-coded differences (by standard deviation) in the first positive peak (P1) amplitudes between patient values and age-corrected normal values for each hexagon. The plot was displayed in 2-dimensional mode to best observe the ring boundaries (see Figure). The central ring (R_c) corresponded to the area within approximately 0 to 5.5 degrees from fixation. The paracentral ring (R_p) corresponded to the area between approximately 5.5 and 11 degrees from fixation. The R_c and R_p were evaluated by two independent, masked examiners (DPH and KES). With the CDPA method, if a ring contained a hexagon that was -2 or -3 standard deviations (colored blue or black in the color difference plot) than the ring was judged abnormal. If none of the hexagons in a ring were -2 to -3 standard deviations the ring was judged to be normal. If a hexagon straddled the circle dividing R_c and R_p , it was judged as being part of the ring in which the majority of the

hexagon was located. Therefore, a hexagon could only be judged as being within R_c or R_p , but never within both rings, as seen in the Figure. The results of both graders were combined, and discrepancies were resolved by using a third grader (MKA), allowing for final scores of R_c , R_p , combined R_c or R_p ($R_c \cup R_p$; i.e., an abnormal hexagon need only be present in one of the two regions, including if they were present in both), and the combination of both R_c and R_p ($R_c \cap R_p$, i.e., an abnormal hexagon had to be present in each of the two regions, not one or the other).

As previously published [8], but provided herein for comparison purposes, ring ratio analysis was performed by grouping the 103 response densities of each patient into six concentric circles. The VERIS software averaged the ring responses and placed cursors automatically on negative troughs and positive peaks of each averaged waveform. The amplitude was calculated between the first negative trough (N1) and the first positive peak (P1) yielding N1-P1 response density in nV/deg². R5 ring ratios were determined by utilizing R5 as the “internal reference ring” and dividing it by all other ring response amplitudes. The ring ratios were analyzed using ROC analysis. A 95% specificity cut-off threshold was used to determine toxicity, enabling specificities to be calculated.

Statistical Analysis

The sensitivity and specificity were determined for the final scoring of each ring and for both rings combined, with nontoxic and control groups defining “absence of disease” and the toxic group defining “presence of disease.” Statistical analysis was performed using GraphPad Software QuickCalcs (GraphPad Software Inc, La Jolla, CA, USA). The association between CDPA results and presence of toxicity was determined using Fisher exact test. Intergrader agreement was found using Cohen's kappa. The final scores obtained for R_p were compared to ring ratio R5/R4 acquired by Adam et al for discordance using McNemar's test [8]. R_p was used because HCQ toxicity was expected to be located within this region [1], and R5/R4 was used as it had the highest sensitivity and specificity⁸. The McNemar's test was performed to analyze “false positives” and “true negatives.” The McNemar's test was also used to analyze “true positives” and “false negatives.” Patient results were graded a second time by DPH and KES, again done independently and masked. Reproducibility between grading sessions was assessed using Cohen's kappa. The patient results were masked and randomized a third time and graded by DPH and MKA. The optimal number of abnormal hexagons that might define toxicity was determined using Receiver Operating Characteristic (ROC) analysis. In this analysis, each patient had three tallies of abnormal hexagons performed: (1) the number within the area of R_c ; (2) the number within the area of R_p (and outside R_c) and (3) the number in the central (R_c) and pericentral (R_p) regions combined as a single area (in the Table, termed “ R_c and R_p ”). In the latter, a single total count was taken from within the white ring that formed the outer border of R_p , ignoring the white ring that outlined R_c .

Results

Validation of grading

The intergrader agreement for ascribing abnormal (blue or black) hexagons to the region of R_c was “very good” ($\kappa = 0.918$, $SE = 0.046$). The intergrader agreement for R_p was also “very good” ($\kappa = 0.897$, $SE = 0.045$). The reproducibility of grader DPH was “perfect” ($\kappa = 1.000$, $SE = 0.000$) for R_c and “very good” ($\kappa = 0.934$, $SE = 0.037$) for R_p . The reproducibility of grader KES was “very good” ($\kappa = 0.946$, $SE = 0.038$) for R_c and “very good” ($\kappa = 0.961$, $SE = 0.028$) for R_p . Reproducibility of blue/black hexagon counts within regions was similarly good.

Categorical analysis of ring involvement

The intent of this analysis was to determine whether involvement of either or both of the regions R_c and R_p was better at predicting whether toxicity was present based upon the previously defined study criteria. Compared to R_c , R_p appeared to be more sensitive (89% vs 83%) but less specific (82% vs 93%) in detecting HCQ toxicity. When defining an abnormal state as having either or both rings involved ($R_c \cup R_p$) there was no increase in sensitivity (89%) but a slight decrease in specificity (80%) compared to R_p alone. The requirement for both rings to be involved ($R_c \cap R_p$) increased specificity (95%) but resulted in a relatively low sensitivity (83%). The Fisher exact test of the association of between CDPA results and presence of toxicity for R_c was $P = 0.0010$ and for R_p was $P = 0.0365$.

The above findings were compared to 103 hexagon ring ratio analysis as described by Adam et al [8] in this same data set. The ring ratio value of $R5/R4$ had 89% sensitivity with specificity set to 95%. The McNemar's test result demonstrated that R_p was equally sensitive as $R5/R4$ at detecting HCQ toxicity (89%), but at lower specificity (82%).

Continuous analysis—The intent of this analysis was to determine whether the severity of involvement of a region would be useful for discrimination between toxic and nontoxic groups. A larger number of blue/black hexagons within the regions were presumed to represent greater abnormality of the mfERG response. These data were submitted for receiver operating characteristic analysis to find an “optimal” point of discrimination.

The ROC analysis demonstrated that for R_c the highest sensitivity and specificity was obtained when setting threshold for toxicity to two hexagons, giving a sensitivity of 78% with 95% specificity. ROC analysis of CDPA demonstrated that for R_p the highest sensitivity and specificity were when the threshold was set to two hexagons, giving 83% sensitivity and 95% specificity. The ROC analysis for R_c and R_p areas combined gave the highest sensitivity and specificity when the threshold was set to four hexagons, giving a sensitivity of 83% and specificity of 96%. Therefore, the optimum number of blue or black hexagons to define eyes as having retinal toxicity when analyzing R_c was two hexagons, when analyzing R_p was two hexagons, and when analyzing R_c and R_p combined was four hexagons. The area under the curve (AUC) values and standard errors for R_c , R_p , and the combined area of R_c and R_p are described in the table.

Discussion

CDPA is an intuitively easy and reproducible way to assess HCQ toxicity when using mfERG, as judged by the high intergrader consistency. This study interpreted the results of CDPA in two different fashions: (1) a categorical fashion, in which R_c and R_p were categorized as being involved or not, irrespective of the number of abnormal hexagons within each ring, and (2) a continuous fashion, in which the number of abnormal hexagons within R_c and R_p were summated and the threshold of optimal utility was determined by ROC analysis. When compared to ring ratio analysis in this same cohort, the CDPA method was comparable in specificity (96 vs. 95%), but showed somewhat lower sensitivity (83 vs. 89%). Given that the mfERG is usually combined with a variety of tests used for evaluating patients with possible HCQ toxicity, a combination of which may add sensitivity to the diagnosis, this slight shortfall relative to ring ratio analysis is of uncertain significance but is one of which the clinician should be aware.

If a continuous method of counting abnormal hexagons in the regions of interest R_c and R_p is used, it may be appropriate to obtain counts from these two regions as a combined unit, defined as $R_c \cup R_p$. A larger number of abnormal hexagons is required within the areas to achieve “abnormal” status. Intuitively, this might lead to fewer false positive results relative to the requiring of only one or two hexagons for test positivity. Our results only hint at this possibility (see table). Notwithstanding, $R_c \cup R_p$ might be simpler to implement, since an evaluator is required only to attend to the outside perimeter of R_p and count everything within it. Notably, the two-dimensional display required for CDPA may be superior to the three-dimensional scalar display for viewing results. In the latter an intact foveal peak may obscure viewing of the perifoveal responses of the superior visual field that correspond to the inferior perifoveal region of the fundus, an area which can be preferentially involved in some cases of HCQ toxicity. Clinicians relying on a brief overview of a grossly normal three dimensional display without looking at the actual tracings would not be able to detect such abnormalities unless the two-dimensional color display were also presented.

The advantages of the CDPA method are that it is easy to implement and that it requires little expertise to grade the rings. CDPA also does not require synthesis of an analytical program, which is required for other methods (such as Adam et al ring ratio method) to be evaluated efficiently. However, the CDPA possesses intrinsic limitations. Unlike the ring ratio analysis, CDPA requires that the mfERG be compared against age-corrected normal values, a function that can be programmed into the on-board software for CDPA.

Development of a quantitative method of evaluating the color difference plots appears justified, with integration of such a method into the manufacturers' software having promise for clinical utility. It should be noted that the present study validates the CDPA method for the 103 hexagon test stimulus and not the 61 hexagon test stimulus that many laboratories use.

In conclusion, the color difference plot analysis [1] may be a useful method of detecting HCQ toxicity. Its sensitivity and specificity suggest that it must be used in conjunction with other validated tests when making clinical judgments. CDPA has the benefit of its relative simplicity and its having a high degree of intergrader agreement and reproducibility.

Acknowledgments

We thank Gayle Kremer for technical assistance and Aniko Szabo for statistical assistance

Grant funding: Supported in part by Research to Prevent Blindness, Inc., New York, NY; the Thomas M. Aaberg, Sr., Retina Research Fund, Milwaukee, WI; Jack A. and Elaine D. Klieger Professorship (DPH); Grant 1UL1RR031973 from the Clinical and Translational Science Award program of the NCRR, NIH

References

1. Change WH, Katz BJ, Warner JE, et al. A Novel Method for Screening the Multifocal Electroretinogram in Patients using Hydroxychloroquine. *Retina*. 2008; 28:1478–1486. [PubMed: 18997610]
2. Marmor MF, Kellner U, Lai TY, et al. Revised Recommendations on Screening for Chloroquine and Hydroxychloroquine Retinopathy. *Ophthalmology*. 2011; 118:415–22. [PubMed: 21292109]
3. 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). *Clinical Ophthalmology*. 2010; 4:1151–8. [PubMed: 21060664]
4. So SC, Hedges TR, Schuman JS, et al. Evaluation of Hydrochloroquine Retinopathy with Multifocal Electroretinography. *Ophthalmic Surg Laser Imaging*. 2003; 34:251–8.
5. 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]
6. Maturi RK, Yu M, Weleber RG. Multifocal Electroretinographic Evaluation of Long-term Hydroxychloroquine Users. *Archives of Ophthalmology*. 2004; 122:973–81. [PubMed: 15249360]
7. 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]
8. Adam MK, Covert DJ, Stepien KE, et al. Quantitative assessment of the 103-hexagon multifocal electroretinogram in detection of hydroxychloroquine retinal toxicity. *Br J Ophthalmol*. 2012; 96:723–9. [PubMed: 22328813]

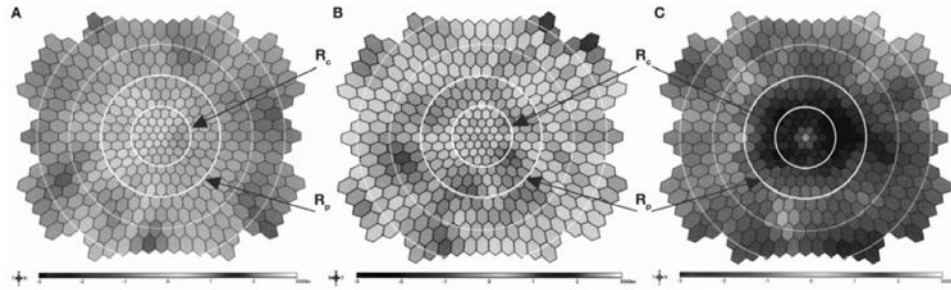


Figure 1. Color difference plots of (A) reference control, (B) HCQ-nontoxic, and (C) HCQ-toxic subjects. These consist of two dimensional displays of hexagons interpolated from the 103 hexagon test stimulus. Differences between patient value and age-adjusted normal value corresponding to each hexagon are color-coded, such that blue or black hexagons indicate patient value reductions that exceed 2 standard deviations from normal. Central (R_c) and pericentral (R_p) regions are outlined by white concentric circles.

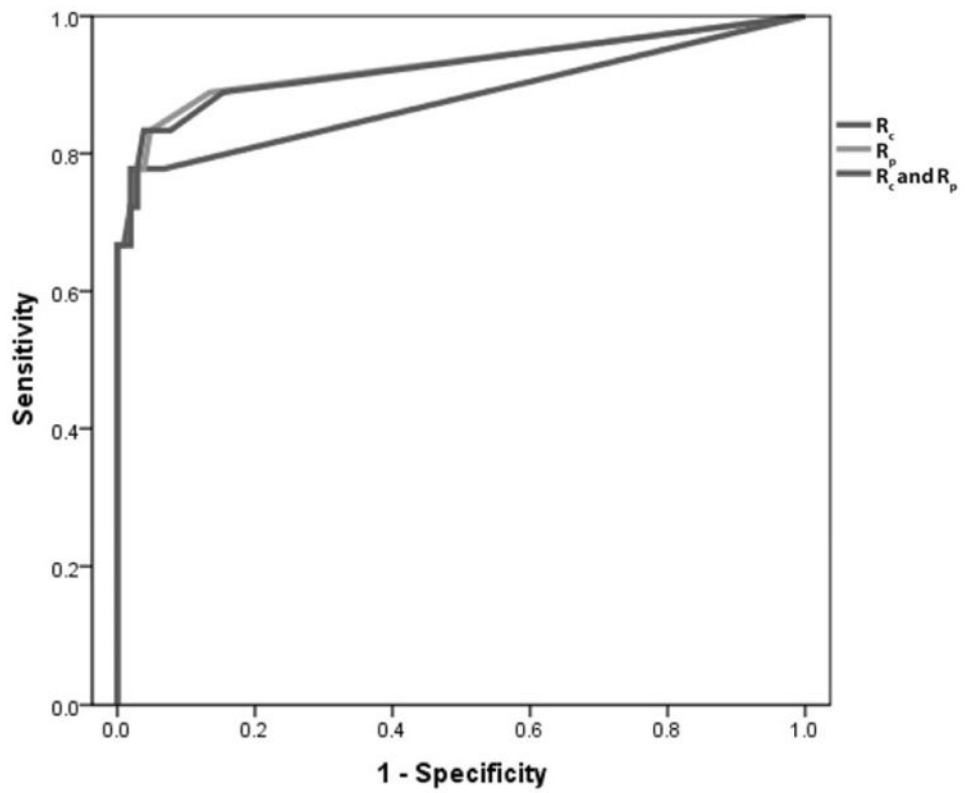


Figure 2.
The Receiver Operating Characteristic analysis of CDPA for R_c , R_p , and combination of R_c and R_p .

Table
Receiver Operator Characteristics AUC values by test parameter

Region	AUC	Optimum # of hexagons	Sensitivity	Specificity
R _c	.880	2	0.778	0.952
R _p	.927	2	0.833	0.952
R _c and R _p *	.925	4	0.833	0.962