CrossMark

American Academy of Ophthalmology
The Eye M.D. Association

Hydroxychloroquine and Chloroquine Retinopathy

A Systematic Review Evaluating the Multifocal Electroretinogram as a Screening Test

Adrian C. Tsang, BHSc,[1,2] Sina Ahmadi Pirshahid, MD,[1,2] Gianni Virgili, MD,[3] Chloe C. Gottlieb, MD,[1,2,4] John Hamilton, BSc,[1,4] Stuart G. Coupland, PhD[1,2,4]

Purpose: To determine the validity of multifocal electroretinography (mfERG) as a screening tool for detecting chloroquine (Aralen, Sanofi Aventis, Bridgewater, NJ) (CQ) and hydroxychloroquine (Plaquenil, Covis Pharmaceuticals, Inc, Zug, Switzerland) (HCQ) retinal toxicity in patients using these medications. To evaluate the sensitivity and specificity of mfERG when compared with automated visual fields (AVFs), fundus autofluorescence (FAF), and optical coherence tomography (OCT).

Clinical Relevance: The 2011 American Academy of Ophthalmology recommendations on screening for CQ/HCQ retinopathy recommended a shift toward more objective testing modalities. Multifocal electroretinography may be effective in detecting functional change before irreversible structural damage from CQ/HCQ toxicity.

Methods: We performed a search for records reporting the use of mfERG for screening CQ/HCQ retinopathy in MEDLINE (PubMed and OVID), EMBASE, and Web of Science, and assessed these using the QUADAS-2 risk of bias tool. We conducted an analysis of 23 individual studies and their reported individual patient data (449 eyes of 243 patients) published from January 2000 to December 2014.

Results: Multifocal electroretinography had the greatest proportion of positive test results, followed by AVF. The pooled sensitivity and specificity of mfERG were 90% (95% confidence interval [CI], 0.62−0.98) and 52% (CI, 0.29−0.74), respectively, with AVF as reference standard (13 studies). Sensitivity was high, but specificity was variable when OCT, FAF, and the positivity of 2 of 3 tests was used as the reference standard. When verified against AVF as the reference test, patients with a false-positive mfERG result received higher HCQ cumulative doses (1068 g) than patients with true-negative (658 g, $P < 0.01$) and false-negative (482 g, $P < 0.01$) results.

Conclusions: Multifocal electroretinography was shown to have a high sensitivity but variable specificity when verified against AVF, OCT, FAF, and a combination of tests. The greater average cumulative dose in the false-positive group compared with the true-negative group when mfERG was verified against AVF suggests that mfERG may have the ability to detect cases of toxicity earlier than other modalities. There is an unclear risk of bias in the available evidence, and future studies should adhere to Standards for Reporting of Diagnostic Accuracy reporting guidelines. *Ophthalmology 2015;122:1239-1251 © 2015 by the American Academy of Ophthalmology.*

Hydroxychloroquine (Plaquenil, Covis Pharmaceuticals, Inc, Zug, Switzerland) (HCQ) is a disease-modifying antirheumatic drug used for the treatment of rheumatic and dermatologic diseases.[1] Despite its favorable efficacy and safety profile in comparison with its more toxic predecessor, chloroquine (Aralen, Sanofi Aventis, Bridgewater, NJ) (CQ), retinal toxicity remains a widely recognized side effect of its long-term use. Although HCQ retinopathy has been reported at a cumulative dose as low as 57 g, the prevalence of this complication is low during the first 5 years of therapy.[2] The prevalence increases considerably to 1% after a cumulative dose of 1000 g.[3] The current recommendations advise baseline testing within the first 6 months of initiating therapy and yearly follow-up testing no later than the fifth year of continuous therapy, or when the 1000 g cumulative dose threshold is met.[3] The frequency of follow-up testing should be increased within the first 5 years if there are clinical findings suggestive of imminent toxicity.[3]

Early detection of CQ/HCQ retinopathy is important, because associated vision loss is thought to be irreversible and may progress despite discontinuation of therapy.[3] The 2011 recommendations of the American Academy of Ophthalmology (AAO) emphasized the need for a transition to objective testing modalities, such as multifocal electroretinography (mfERG), fundus autofluorescence (FAF), and spectral-domain optical coherence tomography (SD OCT) in addition to previous subjective screening

http://dx.doi.org/10.1016/j.ophtha.2015.02.011
ISSN 0161-6420/15

methods, such as automated visual fields (AVFs) and clinical fundus examination.[3] The sensitivity and specificity of these tests are not fully understood, particularly when determining which findings constitute the earliest stages of irreversible retinal toxicity. Addressing this gap in knowledge is critical in assisting clinicians to develop evidence-based guidelines to determine which test, or combination of tests, will best detect retinal toxicity. As a step toward addressing this gap, a systematic review following Meta-analysis of Observational Studies in Epidemiology (MOOSE) guidelines was conducted of studies published in the last 14 years, with patients who had undergone mfERG to screen for CQ/HCQ retinal toxicity.[4] Our goal was to evaluate the sensitivity and specificity of mfERG in comparison with the other AAO recommended screening tests for retinal toxicity.[3] Newer tests, such as peripapillary retinal nerve fiber layer thickness and microperimetric defects, have not been considered in this study because they are not part of the revised AAO recommendations.[5,6]

Methods

Search Strategy

We used relevant subject headings and keywords to search the following databases with language restricted to English and French: MEDLINE (PubMed and OVID), EMBASE, and Web of Science from January 2000 to December 2014 (Appendix 1). We consulted a research librarian for the search strategy. The abstracts and references obtained from this search were managed through RefWorks and read independently by 3 authors (A.C.T., S.A., C.C.G.) to include or exclude references for full-text article review. The Ottawa Health Science Network Research Ethics Board ruled that approval was not required for this study.

Selection Criteria for Full Text Review

The study populations in the reviewed articles were patients undergoing routine screening for CQ/HCQ toxicity. Studies were included if mfERG was used as a screening modality; included at least 1 of AVF, optical coherence tomography (OCT), or FAF; and allowed the extraction of aggregate sensitivity and specificity estimates or provided individual patient data (Table 1, available at www.aaojournal.org).[6–28] The use of individual patient data allows for the potential to determine the effect of covariates (e.g., age, treatment duration, and cumulative dose), identifies discordances, and when more than 1 reference standard is used, establishes a composite reference standard (2 of 3 reference tests). Records were excluded if they were duplicates, conference abstracts, or follow-up reports of previously published trials with no new relevant data, and if patients were not receiving CQ/HCQ therapy for rheumatologic disease (Fig 1). To be conservative, any relevant references that described screening methodology for CQ/HCQ toxicity underwent full-text review if approved by any 1 of the 3 reviewers. The full-text articles from the selected references were read independently to determine their eligibility for the meta-analysis (Fig 1). Individual authors were contacted regarding retrieval of individual patient data, but the response rate was poor. An assessment of the methodological quality of each included study was made using the QUADAS-2 tool.[29] The reference lists of each article were reviewed, and additional publications were considered according to the described criteria.

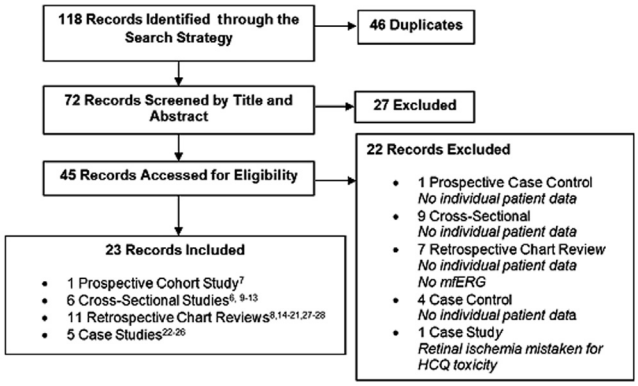

Figure 1. Flow diagram of literature search. HCQ = hydroxychloroquine; mfERG = multifocal electroretinogram.

Data Collection

The available results of mfERG, AVF, SD OCT, and FAF studies of individual patients were compiled. The results were categorized into positive or negative test findings, according to the parameters in Table 2, to enable comparison. Abnormal visual field data available in the publications were reevaluated according to criteria in the 2011 AAO recommendations. Interpretation using criteria based on the earlier 2002 AAO recommendations would underestimate the sensitivity of AVF, because not all changes were previously considered clinically significant.[3]

Statistical Analysis

The use of mfERG to aid in the diagnosis of CQ/HCQ retinopathy was considered in the context of the introduction of other objective testing methods over time. A shift in the approach to diagnostic testing occurred with the introduction of FAF and OCT within the past 10 years. The sensitivity and specificity of mfERG were estimated in comparison with AVF, the mainstay of screening, FAF, and OCT individually. In each comparison, any change in any of the 3 reference test results is a positive indication of toxicity. A reference standard was also developed to evaluate mfERG against multiple screening methods recommended by the AAO. A positive result in at least 2 of AVF, OCT, and FAF was established as the reference standard for the detection of toxicity. Any subject eye that underwent mfERG and at least 2 of AVF, OCT, or FAF was included in the analysis. A positive mfERG result (abnormal) was categorized as a true positive if at least 2 of AVF, FAF, or OCT results were also positive and was categorized as negative otherwise.

Review Manager 5.3 software (The Nordic Cochrane Centre, The Cochrane Collaboration, 2014) was used to prepare Forest plots to show the sensitivity and specificity of mfERG to detect CQ/HCQ retinopathy, as confirmed with each reference standard test. True diagnostic accuracy studies allow an estimate of both sensitivity and specificity of mfERG to detect CQ/HCQ retinopathy confirmed by at least 1 reference standard. Studies were identified if they included only all diseased patients or only all nondiseased patients because they do not allow for estimates of both sensitivity and specificity. Logistic regression was used to compare the proportion of positive results of each test using all studies including indirect comparisons and by restricting the analysis to studies that adopted all 4 tests, with studies and subjects as random effects to account for within-study correlation.