showed segregation of the Q368STOP mutation with disease.4

The Q368STOP mutation has mostly been detected in Caucasian, rather than Asian, populations.3–7 In the present study, all individuals with the Q368STOP mutation were Caucasians and shared the same alleles at the four genotyped markers previously defined for the Q368STOP disease haplotype4 in 15 Tasmanian POAG families. This finding further supports a global disease haplotype for the Q368STOP mutation in Caucasians rather than a series of de novo events. Additional genotyping of markers, including single nucleotide polymorphisms obtained from the HapMap project, should help define the approximate date of origin of the Q368STOP mutation in Caucasians.

REFERENCES


Test-retest Reproducibility of Optic Disk Deterioration Detected From Stereophotographs by Masked Graders

Richard K. Parrish II, MD,
Joyce C. Schiffman, MS, William J. Feuer, MS,
Douglas R. Anderson, MD,
Donald L. Budenz, MD, MPH,
Maria-Cristina Wells-Albornoz, MPH,
Ruth Vandenbroucke, BS, Michael A. Kass, MD,
Mae O. Gordon, PhD, and the Ocular Hypertension Treatment Study Group
PURPOSE: To assess the reproducibility of determining whether an eye has developed optic disk deterioration by the Optic Disc Reading Center (ODRC) in the Ocular Hypertension Treatment Study (OHTS).

DESIGN: Test-retest reproducibility study.

METHODS: Masked, certified graders at the ODRC determined the occurrence of optic disk deterioration in OHTS by comparing baseline with follow-up stereoscopic optic disk photographs. To assess reproducibility, regradings were obtained annually by inserting masked “quality control” photographs into the usual ODRC reading stream.

RESULTS: Agreement (kappa) ranged from 0.65 to 0.83 over 5 years. Specificity ranged from 98% to 100%, and sensitivity ranged from 64% to 81%.

CONCLUSIONS: The kappa statistic for test-retest agreement in OHTS is in the range considered good to excellent over 5 years. Consistency (specificity) in regrading optic disks that did not develop deterioration was particularly high. The sensitivity results show that detecting subtle deterioration of optic disks is challenging.

(NEWSPAPER TITLE) 2005;140:762–764. © 2005 by Elsevier Inc. All rights reserved.)

DEVELOPMENT OF PRIMARY OPEN-ANGLE GLAUCOMA (POAG) in the Ocular Hypertension Treatment Study (OHTS) is determined by confirmed visual field abnormality or confirmed optic disk deterioration, or both, that is attributed to POAG by a masked Endpoint committee. Of 125 POAG endpoints reported in the primary outcome article, 81 (65%) were detected by optic disk deterioration. We report the reproducibility of the OHTS protocol for determining whether an eye has developed optic disk deterioration by evaluation of stereophotographs by masked graders.

The institutional review boards of all participating institutions approved the protocol. Certified graders at the Optic Disc Reading Center (ODRC) determined the occurrence of optic disk deterioration by comparing baseline with follow-up optic disk stereophotographs masked to the order in which the photographs were taken, randomization, clinic, prior gradings, and information on the fellow eye. Each set of photographs was reviewed by two graders. Disagreement between graders was resolved by consensus and adjudication involving a third, senior grader. If the ODRC determined the presence of optic disk deterioration, a second set of confirmation photographs was taken for a second masked comparison with baseline. If this second grading confirmed disk deterioration, the Endpoint committee reviewed the optic disk photographs, visual fields, and ocular and medical history to determine whether deterioration was due to POAG. Because some eyes were randomized to observation without treatment, the goal was to detect small, but real change.

To assess reproducibility, masked regradings were obtained annually by inserting “quality control” photographs into the usual ODRC reading stream. Quality control sets included: (1) Baseline and follow-up photographs of 50 “normal” eyes not developing deterioration. An eye was correctly classified when the follow-up photograph was judged as not showing deterioration. (2) Baseline, follow-up, and confirmation photographs of 36 eyes developing “disk deterioration” due to POAG as determined by the Endpoint committee. An eye was correctly classified when both the follow-up and corresponding confirmation photograph were independently judged as showing deterioration from baseline.

Kappa, which estimates agreement adjusting for chance, is presented for each year for all eyes in the quality control sets (n = 86) (Table). For the 50 eyes in the “normal” quality control set, we report the percent correctly regraded as not showing deterioration (specificity). Similarly, for the 36 eyes in the “disk deterioration” set, we report the percent correctly regraded as showing “deterioration” (sensitivity).

OHTS findings compare favorably with both interobserver and intraobserver agreement (kappa) in the four other reports in the literature on reproducibility of determining glaucomatous optic disk change through side-by-side comparison of baseline and follow-up stereophotographs by masked graders. Because ODRC graders involved in consensus and adjudication were not always the same, the OHTS protocol more closely resembles interobserver than intraobserver gradings. Also, since both a follow-up and a second confirmation photograph had to be graded as deterioration independently, agreement was lower than with a single grading of only follow-up photographs (when only follow-up photographs were regarded, kappa = 0.85, 0.73, 0.73, 0.83, 0.95; sensitivity = 86%, 72%, 69%, 81%, 94% in 2000 through 2004.) The European Glaucoma Prevention Study Group reported interobserver agreement of 0.54 to 0.75 among three ophthalmologists grading 40 pairs of stereophotographs (intraobserver = 0.80 to 1.00 when regraded). Azuara-Blanco and associates reported interobserver agreement of 0.34 to 0.68 among six glaucoma specialists grading 40 sets of photographs (intraobserver = 0.55 to 0.78). The Glaucoma Screening Study reported interobserver agreement of 0.06 to 0.49 among four ophthalmologists grading six control eyes and 13 eyes with glaucomatous disk changes. Caprioli and associates reported interobserver agreement of 0.81.
among three experienced observers grading 75 eyes (intraobserver = 0.92).

Over 5 years, test-retest reproducibility for judging optic disk deterioration in OHTS (kappa = 0.65 to 0.83) was in a range considered good to excellent, in part due to excellent specificity (98% to 100%). The sensitivity (64% to 81%) shows that consistent recognition of very small changes, as ethically required in a study like OHTS with an untreated cohort, is challenging, even with skilled graders and a rigorous protocol.

REFERENCES


---

**TABLE.** Reproducibility (Kappa), Specificity, and Sensitivity of the Determination of the Development of Optic Disk Deterioration by Year in the Ocular Hypertension Treatment Study (OHTS)

<table>
<thead>
<tr>
<th>Year</th>
<th>Kappa Statistic for Agreement Between Original and Regrade Process (95% Confidence Interval)</th>
<th>Kappa Statistic for Agreement Between Original and Regrade Process</th>
<th>50 Normal Eyes</th>
<th>36 Eyes Developing Disk Deterioration</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>0.73 (0.58, 0.87)</td>
<td>0.62</td>
<td>49 (98%)</td>
<td>26 (72%)</td>
</tr>
<tr>
<td>2001</td>
<td>0.65 (0.49, 0.81)</td>
<td>0.60</td>
<td>49 (98%)</td>
<td>23 (64%)</td>
</tr>
<tr>
<td>2002</td>
<td>0.70 (0.55, 0.85)</td>
<td>0.79</td>
<td>50 (100%)</td>
<td>24 (67%)</td>
</tr>
<tr>
<td>2003</td>
<td>0.73 (0.58, 0.87)</td>
<td>0.81</td>
<td>50 (100%)</td>
<td>25 (69%)</td>
</tr>
<tr>
<td>2004</td>
<td>0.83 (0.71, 0.95)</td>
<td>0.89</td>
<td>50 (100%)</td>
<td>29 (81%)</td>
</tr>
</tbody>
</table>

*Kappa rescaled for percent of Optic Disc Reading Center gradings that showed deterioration each year.*