Detecting Early to Mild Glaucomatous Damage: A Comparison of the Multifocal VEP and Automated Perimetry

Donald C. Hood,¹ Phamornsak Thienprasiddhi,² Vivienne C. Greenstein,³,⁴ Bryan J. Winn,¹ Nitin Obri,¹ Jeffrey M. Liebmann,² and Robert Ritch²

PURPOSE. To gain better understanding of the relationship between abnormalities detected by the multifocal VEP (mfVEP) compared with those detected by static achromatic, automated perimetry in patients with glaucoma.

METHODS. Fifty patients were studied who had open-angle glaucoma that met the following criteria: (1) a mean deviation (MD) of better than −8 dB in both eyes on the 24-2 Humphrey visual field (HVF) test (Carl Zeiss Meditec, Dublin, CA); and (2) glaucomatous damage in at least one eye, as defined by a glaucomatous optic disc and an abnormal 24-2 HVF test result (pattern standard deviation [PSD] < 5% and/or glaucoma hemifield test [GHT] results outside normal limits). Monocular mfVEPs were obtained from each eye by using a pattern-reversal dartboard array, 4.45° in diameter, which contained 60 sectors. Recording electrodes were placed at the inion (I) and 1+4 cm, and also at two lateral locations up 1 cm and over 4 cm from I. Monocular and interocular mfVEP probability plots were derived by comparing the results with those of normal control subjects. For both the HVF and mfVEP probability plots, a hemifield was classified as abnormal if three or more contiguous points were significant at less than 5%, with at least one at less than 1%

RESULTS. Of the 200 hemifields tested (50 patients × two eyes × two hemifields), 75 showed significant clusters on the HVF, and 74 (monocular probability plot) and 93 (monocular or interocular plot) showed significant clusters on the mfVEP. Overall, the HVF and mfVEP results agreed on 74% of the hemifields, and 90 hemifields were normal and 58 were abnormal on both the mfVEP (interocular and/or monocular abnormal) and HVF cluster tests. Of the 52 disagreements, 35 hemifields had a significant cluster on the mfVEP, but not on the HVF, whereas the reverse was true of 17 hemifields. A case-by-case analysis indicated that misses and false-positive results occurred on both the HVF and mfVEP tests.

CONCLUSIONS. As predicted from a theoretical analysis, under these conditions (i.e., the signal-to-noise level) the HVF and monocular mfVEP tests showed a comparable number of defects, and, with the addition of the interocular test, the mfVEP showed more abnormalities than the HVF. However, although there were abnormalities detected by the mfVEP that were missed by the HVF, the reverse was true as well. (Invest Ophthalmol Vis Sci. 2004;45:492–498) DOI:10.1167/iovs.03-0602

Glaucoma is a progressive optic neuropathy characterized by a specific pattern of optic nerve head and visual field damage due to the death of retinal ganglion cells. Typically, the diagnosis is based on characteristic changes in the optic nerve head combined with a visual field defect determined with static automated, achromatic perimetry (SAP). However, substantial ganglion cell damage can take place before SAP detects functional deficits.¹,² Further, some patients have difficulty performing this test in a reliable and reproducible fashion. Because of these problems and the importance of assessing early glaucomatous damage, alternative tests of visual function have been proposed.

A new and particularly promising technique is the multifocal visual evoked potential (mfVEP). With this technique, many (typically 60) responses, each associated with a local region of the visual field (or retina), are recorded simultaneously.³,⁴ A number of studies have demonstrated that the mfVEP can detect glaucomatous damage.⁴–¹⁷ Compared with most electrophysiological tests of visual function, the mfVEP has the advantage that it produces a topographical measure of damage. Thus, mfVEP results can be compared with visual fields obtained with standard SAP, such as the 24-2 Humphrey visual field (HVF; Carl Zeiss Meditec, Dublin, CA). In fact, special procedures have been developed to make such comparisons.⁵,⁹–¹¹ With these procedures, probability plots like those used to summarize the HVF results are produced. Two kinds of probability plots have been developed based on either local monocular mfVEP amplitudes⁴–¹⁷ or the interocular ratio of the local monocular amplitudes.⁴,⁶–⁹

There is evidence that glaucomatous changes can be detected by the mfVEP technique before HVF losses occur. For example, Goldberg et al.¹³ studied patients with glaucomatous disc changes and normal HVF in at least one eye and found that the mfVEP was abnormal in more than 50% of the 29 fellow eyes that had normal fields on HVF. More recently, Thienprasiddhi et al.¹⁷ reported that the mfVEP detected deficits in hemifields with apparently normal HVF results in glaucoma patients with unilateral hemifield defects. In contrast, a recent study reported an unacceptably high false-positive rate for the mfVEP.¹⁰,¹⁸

Although there are clearly conditions under which the mfVEP detects early damage missed on the HVF, Hood and Greenstein³ argued that there are conditions under which the reverse also occurs. They developed a theoretical framework for determining when either SAP or the mfVEP technique
would be superior in detecting damage. This framework related mfVEP amplitude loss to HVF sensitivity loss and took into consideration the variability of each test. In particular, the analysis illustrated that the relative effectiveness of the two tests for detecting damage depends on both the signal-to-noise ratio (SNR) of the mfVEP response and on the variability in the mfVEP and HVF measures. The SNR depends in turn on the stimulus paradigm and on the quality of the recordings. For our recording conditions, Hood and Greenstein predicted that the two tests should detect approximately the same number of defects if only the results from the monocular mfVEP tests were considered. However, if both the interocular and monocular mfVEP tests were used, then the mfVEP technique would detect more defects.

To test these predictions and to gain a better understanding of the relationship between the mfVEP and HVF tests, we studied 50 patients with open-angle glaucoma. The purpose of this study was not to determine whether the mfVEP or the HVF would be better at detecting glaucomatous damage. Such a study would require a large group of control subjects tested with both techniques so that sensitivity and specificity rates could be determined. Rather, the purpose was to understand better how these two techniques compare. As both tests are likely to detect moderate to severe field defects, patients with relatively mild defects on HVF testing were chosen.

Methods

Subjects

Fifty patients with open-angle glaucoma (OAG) and with relatively mild visual field defects were identified for this study. They were selected from a larger group of patients tested with the mfVEP technique over a 3-year period. For inclusion in this study, at least one eye had to be classified as glaucomatous based on an abnormal appearing optic disc and an abnormal HVF in that eye. The HVF was considered abnormal if the pattern standard deviation (PSD) was significant at $P < 0.05$ and/or the glaucoma hemifield test (GHT) results were outside normal limits. To restrict this study to patients with relatively mild field defects, neither eye could have a mean total deviation (MD) worse than $-8\, \text{dB}$. The mean value of the MD for this group was $-2.72\, \text{dB}$ (range, 1.56 to $-7.84$). The patients ranged in age from 30 to 78 years (mean $\pm$ SD, 59.9 $\pm$ 11.5). No specific intraocular pressure was required for the diagnosis of OAG.

Patients more than 80 years of age, those with known diseases of the retina, and those who had no visual field tests within 6 months of mfVEP testing were excluded. Visual fields were measured with the 24-2 program of the HVF Analyzer (Carl Zeiss Meditec), or in 1 case with the 30-2 program. In the latter case, only the 24-2 matrix was analyzed. In 45 individuals the Swedish interactive threshold algorithm (SITA)-standard strategy was used, whereas the Full Threshold strategy was used in 7 individuals. If more than one HVF result was available, the results obtained closest to the date of mfVEP testing were analyzed. Patients were also excluded if the HVF closest to the mfVEP testing date was not reliable. A test was considered unreliable if false positives, false negatives, or fixation losses were greater than 33%. On average, the HVFs were measured within 55 days (range, 0–155) of the mfVEP recording.

Procedures adhered to the tenets of the Declaration of Helsinki, and the protocol was approved by the committee of the Institutional Board of Research Associates of Columbia University.

The mfVEP Procedures

The Stimulus. Figure 1A is a schematic of the scaled, dartboard display, which had a diameter of 44.5° and contained 60 sectors each with 16 checks, 8 white (200 cd/m²) and 8 black (<1 cd/m²). Figure 1B shows the spatial relationship between the locations of the 24-2 HVF test points for a left eye and the mfVEP display. The mfVEP display is a standard option produced by the visual evoked response imaging system (VERIS) software (Dart Board 60 With Pattern; Electro-Diagnostic Imaging [EDI], San Mateo, CA). The display appeared on a black-and-white monitor driven at a frame rate of 75 Hz. The checkerboard of each of the 60 sectors had a probability of 0.5 of reversing on any pair of frame changes.

Recording. Three channels of continuous VEP (EEG) records were obtained with gold cup electrodes. For the midline channel, electrodes were placed 4 cm above the inion (active), at the inion (reference), and on the forehead (ground). For the other two channels, the same ground and reference electrodes were used, but the active electrode was placed 1 cm up and 4 cm lateral to the inion on either side. By subtracting different combinations of pairs of channels, three additional “derived” channels were obtained resulting in effectively six channels of recording representing the six possible pairs of the four recording electrodes.

The records were amplified, with the high- and low-frequency cutoffs set at 3 and 100 Hz (1/2 amplitude; preamplifier PS11; Grass Telefactor, Quincy, MA), and were sampled at 1200 Hz (every 0.83 ms). The sequence of visual stimulation had $2^{15} - 1$ elements and approximately 7 minutes were required for a single mfVEP recording. In a single session, two 7-minute recordings were obtained for monocular stimulation of each eye. The two recordings from each eye were averaged, and the mfVEP responses extracted with the response imaging software (VERIS 4.x software; EDI). Technically, these mfVEP responses are the second-order kernels. The mfVEPs were low-pass-filtered using a Fourier transform technique and a sharp cutoff at 35 Hz. The low-pass filtering and all other analyses were performed with programs written in commercial software (MatLab; MathWorks Inc, Natick, MA). (For more details see Refs. 4,15,19,20.)

The mfVEP Probability Plot

As previously described, all analyses were performed on the “best” of the responses from the six channels.$^{3-14,19}$ Figure 1C shows the best arrays for monocular stimulation of the left (red) and right (blue) eyes of a patient. The patient's pattern deviation probability plots for this patient’s 24-2 fields are presented in Figure 1D. Probability plots, analogous to the total deviation probability plot of the HVF Analyzer (Carl Zeiss Meditec), were created based on tests that compared the amplitude of the response of each eye (monocular test), or the ratio of the amplitudes of the responses from the two eyes (interocular test), to group norms.$^{4,8,14,19}$ Probability plots are shown in Figure 1E for the records in Figure 1C. The points in these plots are positioned in the center of the 60 sectors of the display. A colored square indicates that the mfVEP was statistically significant at either the 5% (desaturated color) or 1% (saturated color) level, whereas the color indicated whether it was the left (red) or right (blue) eye that was significantly smaller than normal.

Results

mfVEP Probability Plots and Significant Clusters

Figure 1C shows mfVEP records for one patient and Figure 1E the associated interocular (left-hand side) and monocular (right-hand side) probability plots derived as just described. The monocular plot shows significant points in the lower hemifield of the left eye, whereas the interocular plot shows significant points above the midline as well as in the lower hemifield. Although by chance alone, the number of points in the mfVEP probability plot from a person without any defects will vary,$^{15}$ a cluster criterion, as suggested by Goldberg et al.,$^{15}$ was used to decide whether a local region of the probability plot was abnormal. The use of a cluster criterion technique was introduced for the 24-2 HVF, according to which a cluster of three points was defined as abnormal if they collectively met a set of criteria (For examples, see Refs. 21,22.) For this study, a cluster was defined as abnormal if, within a hemifield (i.e., the cluster could not cross the horizontal mid-
line), it included three contiguous points exceeding 5%, one of which had to exceed 1%. This yielded a false-positive rate of approximately 3% (monocular test) and 4% (both monocular and interocular test) in our normal control subjects. The red ellipses in Figure 1E show clusters that met these criteria.

A Comparison of Hemifield Defects on mfVEP and HVF Plots

To make a direct comparison of the local damage identified on the mfVEP and HVF plots, we defined clusters for the HVF PD plots in the same way, with one exception. As is the typical practice to avoid rim artifacts, the cluster could contain no more than one point from the outer ring of the 24-2 HVF points. According to the Hood-Greenstein framework, the monocular mfVEP test and the HVF should, on average, yield about the same number of abnormal hemifields for the recording conditions we use. The results are consistent with this prediction. For the monocular mfVEP test (Table 1), 74 (37%) of the 200 hemifields had abnormal mfVEP clusters, whereas 75 (37.5%) had abnormal HVF clusters. However, we know that each test can detect abnormalities missed by the other. Thus, the two test results will not necessarily agree for any given hemifield. Overall, the two results agreed for 75.5% of the hemifields with 101 hemifields classified as normal and 50 hemifields as abnormal on both tests. Of the 49 disagreements, 24 hemifields had a significant cluster on the mfVEP, but not on the HVF, whereas the reverse was true for 25 hemifields. Corrected for chance, the two tests show only a moderate degree of agreement ($\kappa = 0.48$). (Note that strictly speaking, the 200 hemifields studied cannot be considered independent samples because they include two hemifields per eye and two eyes per patient.)

Table 2 shows the results for the interocular mfVEP test. Fewer of the 200 hemifields were classified as abnormal with

<table>
<thead>
<tr>
<th>No Cluster on HVF</th>
<th>Cluster on HVF</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>No cluster on mfVEP</td>
<td>101</td>
<td>25</td>
</tr>
<tr>
<td>Cluster on mfVEP</td>
<td>24</td>
<td>50</td>
</tr>
<tr>
<td>Totals</td>
<td>125</td>
<td>75</td>
</tr>
</tbody>
</table>
the interocular test than in the monocular tests: 62 (31%) compared with 74 (37%). Overall, the HVF and interocular results agreed for 76.5% of the hemifields, with 112 hemifields classified as normal and 44 as abnormal on both tests. Of the 47 disagreements, 17 hemifields had a significant cluster on the mfVEP, but not on the HVF, whereas the reverse was true of 30 hemifields. Again, there is a moderate degree of agreement between the mfVEP and HVF test results.

For our recording conditions, the Hood-Greenstein framework predicts that the mfVEP will show more defects than the HVF if the interocular and monocular results are combined. Table 3 shows the results of defining an abnormal mfVEP as abnormal if either the monocular test or the interocular test showed a cluster. There were 93 abnormal mfVEP hemifields, 19 more than identified by the monocular test alone (Table 1). Of the 200 hemifields, 93 (46.5%) had abnormal mfVEP clusters, whereas 75 (37.5%) had abnormal HVF clusters. Overall, the two tests agreed for 74% of the hemifields, with 90 hemifields normal and 58 abnormal on both tests. (The $\kappa$ statistic, 0.47, was approximately the same as in the case of Tables 1 and 2.) Of the 52 disagreements, 35 hemifields had a significant cluster on the mfVEP, but not on the HVF, whereas the reverse was true of 17 hemifields.

**Individual Cases**

As can be seen in Table 3, there was disagreement in 52 of the hemifields. A disagreement can occur if one test misses a true defect (i.e., a false negative) and the other detects a defect (i.e., a true positive), or when one test falsely detects a cluster (i.e., a false positive) and the other correctly identifies the hemifield as normal (i.e., a true negative). By chance, false positives undoubtedly occurred in both tests. However, a closer look at the data suggests that at least some of the disagreements can be attributed to one test detecting a cluster missed by the other test. Figure 2 shows examples from two patients in whom the monocular test had a significant cluster of points, but the 24-2 HVF did not. The significant clusters on the mfVEP plots in Figure 2A (top row) do not appear on the 24-2 HVF (left side of bottom row). The 24-2 HVF showed significant points in the same region as the significant cluster in the mfVEP plot, but they did not meet our cluster criteria. The 10-2 HVF obtained 6 months earlier (Fig. 2A, right side of lower row) confirmed field defects in these regions. The HVF for the right eye is particularly interesting, as the mean PSD and GHT were within the normal range. As discussed later, the differences between the mfVEP and HVF findings in this case appear to be due, in large part, to the way the field is sampled by the two techniques. This is not the case in the second example (Fig. 2B). In this patient, the region containing a significant cluster on the mfVEP interocular plot (red ellipse) did not show a significant cluster on the 24-2 HVF obtained closest in time to the mfVEP (Fig. 2B, bottom left panels) even though there are numerous HVF test points in the region. However, the HVF (Fig. 2B, bottom right panels) obtained 17 weeks later showed a significant cluster in this region. In this case, the mfVEP detected damage missed on the original 24-2 HVF.

**Conditions Favoring One or the Other Test**

Hood and Greenstein listed some conditions that favor one or the other test. The patients' results in Figures 1, 2, and 3 illustrate some of these conditions. First, the visual field is not sampled in the same way by both tests (Fig. 1B). The mfVEP has an advantage when there is a relatively small defect in the central area of the field, because there are more test stimuli...
concentrated in this region on the mfVEP than in the 24-2 HVF (Fig. 2A).\textsuperscript{25} In fact, the central four rings of the mfVEP display (Fig. 1A) fall within the central 10° and contain 36 of the 60 sectors.

In contrast, the HVF can have an advantage in the periphery where three or four of the test points on the 24-2 HVF can fall within a single sector (Fig. 1B).\textsuperscript{4} This is probably one reason why the abnormality detected on the 24-2 HVFs in Figure 3A is missed on the mfVEP.

Second, the interocular mfVEP test is at its best when the damage is unilateral and the signal in the less affected eye is large. In fact, these are the conditions under which the Hood-Greenstein analysis\textsuperscript{4,14} predicted that the interocular tests would be more sensitive than either the 24-2 HVF or the monocular mfVEP test. The interocular comparison decreases the variability. The abnormality detected in the superior temporal visual field of the left eye on the mfVEP interocular plot in Figure 1E illustrates this point. Notice how large the responses were from the right eye in this region (Fig. 1C).

Third, bilateral damage can be missed on the monocular and the interocular mfVEP tests, as indicated in the case of the patient in Figure 3B. This would be especially true in regions with poor mfVEP signals in the control subjects. For example, bilateral damage in the superior portion of the upper visual field is not likely to be detected by the mfVEP test and can be missed by the monocular test as well.\textsuperscript{4} Likewise, the responses are small in some control subjects in the region shown by the rectangle in Figure 3A. This is one reason why the mfVEP missed the defect just below the horizontal midline in this patient.

The Importance of the Level of SNR

How well the mfVEP does relative to the HVF in detecting glaucomatous damage depends on the SNR of the mfVEP recordings. If the SNR of the recordings is poor, then the mfVEP will not do as well as the HVF. The SNR depends in turn on the stimulus paradigm and on the quality of the recordings. Shorter recordings, for example, decrease the SNR and decrease the number of abnormal clusters identified. Moreover, lower electrode resistance, additional electrodes, and less contamination from alpha and neck muscles increase the SNR and increase the number of abnormal clusters identified.

Similarly, the mfVEP performs better in the regions of the field with good SNRs. As discussed, according to our theoretical framework\textsuperscript{4,14} the interocular test is more likely to pick up a defect in a particular region when the better eye has a large SNR in that region. The data from the present study can be
used to test this hypothesis. In particular, the mean SNR for the better eye was calculated for the mfVEP locations corresponding to the clusters in the more affected eye. (See Ref. 14 for a description of the calculation of the SNR.) The SNR was, on average, 4.3 in the regions of the better eye when the mfVEP was abnormal compared with an average of 2.5 in all other regions. When we analyzed only those hemifields in which the interocular test identified clusters that were not detected on the HVF, then the mean SNR was even higher, 4.9.

Summary
The patients in this study had relatively mild 24-2 HVF losses in the more affected eye, whereas the less affected eye showed normal fields or mild loss. These criteria were chosen in an attempt to maximize the chance of disagreement between the tests. In particular, if only patients with severe field loss were included, then the agreement should have been very good. In spite of our selection criteria, the two tests agreed in approximately 74% of the hemifields. When the two tests disagreed, the mfVEP detected an equal number of abnormal hemifields with the monocular test and a greater number with the combined monocular and interocular test. Further, it is clear that although the mfVEP can detect damage missed on the HVF, the reverse is true. Finally, it is worth emphasizing that the relative advantage of the mfVEP varies with the SNR of the recordings. The mfVEP is an evolving technology and as the technique is improved (i.e., better SNRs) the performance of the mfVEP should be even better than reported in the present study. However, in its current form, although it does not reduce the need for static automated perimetry and its efficacy in detecting early glaucomatous damage is still under study, we find it is a useful clinical tool. In addition to evaluating patients with unreliable or questionable HVFs, the mfVEP can be used for ruling out nonorganic visual loss, diagnosing and observing patients with optic neuritis and multiple sclerosis, and observing disease progression.4,25,26

Acknowledgments
The authors thank Xiao Zhang for helpful discussions during various phases of the project.

References
2. Kerrigan-Baumrind LA, Quigley HA, Pease ME, Kerrigan DF, Mitchel RE. Number of ganglion cells in glaucoma eyes compared with threshold visual field tests in the same persons. Invest Ophthalmol Vis Sc. 2000;41:741–748.