The Collaborative Initial Glaucoma Treatment Study: Baseline Visual Field and Test–Retest Variability

Brenda W. Gillespie,1 David C. Musch,2,5 Kenneth E. Guire,1 Richard P. Mills,4 Paul R. Lichter,2 Nancy K. Janz,5 Patricia A. Wren,5 and the CIGTS Study Group6

PURPOSE. To compare the baseline Collaborative Initial Glaucoma Treatment Study (CIGTS) visual field (VF) score and mean deviation (MD), investigate test–retest variability, and identify variables associated with VF loss and VF measurement variability.

METHODS. Baseline data from a randomized clinical trial of 607 patients with newly diagnosed open-angle glaucoma were collected at 14 clinical centers. The CIGTS VF score and MD were obtained from 24-2 VF tests (Zeiss-Humphrey Systems, Dublin, CA) at two visits approximately 2 weeks apart.

RESULTS. Although most baseline CIGTS VF scores showed limited field loss, 15% (91/607) of patients showed a substantial deficit (VF score >10 on a 0–20 scale). A small but significant learning effect was seen over the two baseline measures for CIGTS VF score and MD. CIGTS VF score and MD correlate highly (r = −0.93); both have high test–retest correlation (0.83 and 0.91, respectively). Variables associated with greater baseline VF loss for both CIGTS VF score and MD include (probabilities for VF only): male sex (P = 0.018), black race (P ≤ 0.0001), lower visual acuity (P ≤ 0.0001), higher intraocular pressure if more than 30 mm Hg (P = 0.0177), poor field reliability score (P ≤ 0.0001), cardiovascular disease (P = 0.015), reduced patient-reported alertness (P = 0.023), and CIGTS clinical center (P ≤ 0.0001). Predictors of increased CIGTS VF score variability include a midrange VF score (P ≤ 0.0001), first-tested eye (P = 0.0027), reduced patient-reported alertness (P = 0.0177), increasing age (P = 0.004), current smoker (P = 0.0014), and CIGTS clinical center (P = 0.0215).

CONCLUSIONS. The CIGTS VF score provides a measure of VF strikingly similar to the MD. Variables associated with VF loss and VF variability may help identify patients who need greater clinical scrutiny. (Invest Ophthalmol Vis Sci. 2003;44:2613–2620) DOI:10.1167/iovs.02-0543

Measurement of visual field (i.e., differential light sensitivity over the central and peripheral regions) is an important aspect of evaluating patients with glaucoma. Use of automated perimetry has greatly improved efficiency of visual field (VF) testing. Yet the measurement of VFs continues to be plagued by variability,1–7 due to factors that include fatigue of the patient, learning effects, visual artifacts, measurement error, and perhaps inherent variability of the VF itself.

The Collaborative Initial Glaucoma Treatment Study (CIGTS) is a randomized multicenter clinical trial comparing initial treatment with trabeculectomy to initial treatment with medications in patients with newly diagnosed open-angle glaucoma. The primary outcome is deterioration of the VF as measured by a VF score developed by the CIGTS investigators. We present our experience with the CIGTS VF score, as well as traditional summary VF measures from the Humphrey 24-2 threshold test (Zeiss-Humphrey Systems, Dublin, CA), using the two baseline measurements obtained from each subject. We investigated baseline distributions, correlation between measures, test–retest variability, learning effects, and factors associated with both VF loss and VF measurement variability.

METHODS

Two baseline visits, conducted within 50 days of each other, were necessary to establish a candidate’s eligibility for the study before randomization. Baseline data collected included demographics, lifestyle variables (e.g., smoking), other health conditions, intraocular pressure (IOP), visual acuity (VA), CIGTS VF score, CIGTS VF reliability score, Humphrey VF global indices, and a battery of questionnaires measuring quality-of-life administered by telephone. Once eligibility was established, written informed consent was obtained. These procedures followed the tenets of the Declaration of Helsinki and were approved by the University of Michigan Institutional Review Board (IRB) as well as by the IRB at each of 14 clinical centers. Details of the CIGTS study design, eligibility criteria, and patient baseline characteristics are given in Musch et al.8 Before randomization, the clinician chose one eye (usually the more severely affected) as the first eye to be treated under the randomized treatment assignment. Only the baseline data for the studied eye are presented.

Patients were required to have had at least one threshold automated VF test on each eye before the first CIGTS baseline visit. The Humphrey 24-2 threshold test was used, which measures sensitivity to light at 52 points over a region within 24° of fixation in all directions except nasally, where the region extends to 27°. The right eye was always tested before the left eye, regardless of which eye was the studied eye. All Humphrey machines were equipped with Stataplus software (Zeiss-Humphrey Systems). If the two baseline visit VF scores (as defined below) differed by four or more points, a third VF test was required within 14 days. In addition, a VF test was considered invalid and repeated if the CIGTS reliability measure was greater than three, with points assigned as follows: one point for fixation losses of 33% or more (if at least 20 trials were performed); one point each for false-
positive responses of 33% or more and/or false-negative responses of 33% or more (if at least eight catch trials were performed), one point for short-term fluctuation (SF) greater than 4.0 but equal to or less than 6.0, two points for SF greater than 6.0 but equal to or less than 7.0, and three points for SF more than 7.0. Although the CIGTS reliability score has features similar to scores used in other studies, there is no currently accepted standard for measuring reliability. The recommendations of Johnson11 were followed in developing the protocol.

Although VF measures can be influenced by cataract, CIGTS baseline VF measures should be minimally affected by this problem because patients who were likely to need cataract surgery within 1 year of enrollment were excluded from the CIGTS, as were those with visual acuity less than 20/40.

The CIGTS VF Score

The CIGTS VF score was developed as a global measure to summarize the extent and depth of VF loss over the region of the field covered by the Humphrey 24-2 test. The score is based on the probabilities in the total deviation probability plot. These probabilities are percentiles empirically derived from the distributions of values at each of the 52 points from age-specific sets of normal subjects collected by the manufacturer.11 The proprietary distributions are built into the VF test software and are not available for inspection. The probability at each of the 52 points is reported as no defect, P ≤ 0.05, P ≤ 0.02, P ≤ 0.01, or P ≤ 0.005, meaning that the measured value at that point was at or below the respective percentile of the age-specific empiric distribution at that position of the field for normal subjects. Because artifacts may result in isolated points of defect in the field, we counted only defects forming clusters, as described later.

The VF score is calculated as follows: First, neighboring points are defined as those adjacent to a given point, whether on a side or a corner. Each of the 52 points in the field is individually graded. A point is called defective if its probability is 0.05 or less, and it has at least two neighboring points with probabilities of 0.05 or less in the same vertical hemiﬁeld (superior or inferior). A weight is assigned depending on the minimum depth of the defect at the given point and the two most defective neighboring points. A minimum defect of 0.05, 0.02, 0.01, and 0.005 is given a weight of 1, 2, 3, and 4, respectively. A point without two neighboring points all depressed to at least P ≤ 0.05 is given a weight of zero. For example, a point at P ≤ 0.01 with only two neighboring points of defect, both at P ≤ 0.05, would receive a weight of 1. The weights for all 52 points in the field are summed, resulting in a value between 0 and 208 (52 × 4). The sum is then scaled to a range of 0 to 20 (dividing the sum by 10.4), to yield values in the same range as the VF score previously developed by the Advanced Glaucoma Intervention Study (AGIS).1 The resultant score is a nearly continuous measure of VF loss. An illustration of a CIGTS VF score calculation from a hypothetical Humphrey VF deviation plot is given in Figure 1.

For a patient to be eligible for the CIGTS, both baseline VF scores had to be 16 or less.

Humphrey VF Global Indices

Measures of VF depression or variability presented on the Humphrey full-threshold printout include the mean deviation (MD), the short-term fluctuation (SF), the pattern standard deviation (PSD), and the corrected pattern SD (CPSD). The MD reflects the average VF depression over the whole VF and is negative when the VF is depressed compared with age-specific “normal” values, becoming more negative with increasing depression (AGIS). The SF measures test-retest variability by using the variance of 10 double-threshold determinations. The PSD reflects “unevenness” of depression across the VF. The CPSD adjusts the PSD to compensate for the SF level. The result of the glaucoma hemiﬁeld test, which assesses whether differences in overall sensitivity between the upper and lower hemiﬁelds are compatible with glaucoma, is also given. These measures are more fully deﬁned in the Statpac 2 user’s guide12 and are also described in Mills.13

Of the indices computed in the Humphrey software, the MD is the measure most similar to the CIGTS VF score, although the two measures are scaled in opposite directions (higher values of CIGTS and lower values of MD indicate greater defect). The primary difference between the scales is that the CIGTS VF score is based on extremes of VF defect (the score remains zero unless a point is at least below the fifth percentile), whereas the MD averages the actual age-adjusted decibel values, thereby capturing less severe defects.

AGIS Score

For 87 non-CIGTS patients with glaucoma, both CIGTS and AGIS VF scores were computed so that comparisons could be made. The sample was chosen to span the range of scores from both the CIGTS and AGIS in a representative way.

Alertness Scale

The quality-of-life telephone interview included the Sickness Impact Profile (SIP), a widely used measure of functional health status. The SIP provides 12 category subscales, including a 10-item alertness behavior scale. Examples of items in this scale are “I do not keep my attention on any activity for long” and “I react slowly to things that are said or done.” Each item receives a yes/no answer. The alertness behavior score used in the study is simply the number of items endorsed, and higher scores therefore represent more problems with attention-related behavior. Because the Humphrey VF test requires
steady concentration for periods up to half an hour, difficulty with alertness may contribute to a poor VF score.

### Statistical Methods

Descriptive statistics included means, standard deviations (SD), and correlation coefficients. Paired t-tests were used to assess learning effects. Variables associated with both baseline CIGTS VF scores and MD were investigated with regression analyses, using a nonautomated step-down procedure for variable selection. CIGTS clinical center was considered as a random effect in all models. A computer was used for determining the absolute difference between the first and second baseline values. Predictors of variability were investigated by using regression models, with square root transformations used in both cases to reduce skewness of the outcome measures. Because of the limited number of study participants self-identified as of a race other than black or white, all other races were grouped with whites for analysis. A significance level of 0.05 was used throughout.

### Results

Between October 1993 and April 1997, the CIGTS enrolled 607 patients from 14 clinical centers in the United States. Of the study participants, 55% were male, and the racial composition was 56% white, 38% black, and 6% Asian and other. The mean age was 57.5 years (range, 29–75). Approximately half (51%) had some education beyond high school, and a third (33%) reported a history of glaucoma in the immediate family. Hypertension was present in 37% of patients, and diabetes was present in 17%. Current smoking or other tobacco use at the time of enrollment was reported by 21% of patients. Slit lamp examination revealed that 51% of patients had some degree of lens opacity at baseline, although potential subjects likely to need cataract surgery within 1 year were excluded. The mean alertness score was 0.8 ± 1.7 (SD; range, 0–10) with 68% reporting no attention problems (score = 0) and 5% reporting noticeable problems (score, 5–10).

Mean IOP at baseline was 27.5 ± 5.6 mm Hg (range, 20–50). Baseline visual acuity scores ranged from 70 to 99, where a score of 70 (Early Treatment Diabetic Retinopathy Study [ETDRS] VA of 20/40) was the minimum permitted for eligibility into the trial. The mean VA score was 85.7 ± 5.7 (SD), which corresponds to an ETDRS VA of 20/20. CIGTS reliability scores were equal to zero (most reliable) for 89% of fields, although 9% had scores of one and 2% had scores of two or three (least reliable). Reliability scores for the first and second baseline fields correlated (r = 0.30, P < 0.0001). Within the average CIGTS baseline VF, the distribution of defects was as follows: Fifty-seven percent of points had no defect, 12% of points had a defect P ≤ 0.05; 7%, had P ≤ 0.02, 6% had P ≤ 0.01, and 18% had P ≤ 0.005. Approximately two thirds of patients (66%) had evidence of a central VF defect, as measured by the central four points of the Humphrey field. By the Humphrey glaucoma hemifield test, 9% of CIGTS patients were scored borderline and 70% were outside normal limits (indicating glaucoma), with the remaining 21% within normal limits (meeting CIGTS eligibility criteria with elevated IOP and a glaucomatous optic disc).

### Summary Statistics for the CIGTS VF Score and Humphrey VF Measures

Table 1 presents summary statistics for the CIGTS VF score, MD, SF, PSD, and CPSD for all CIGTS patients at baseline. First, the mean, SD, and range of each measure are given, followed by the correlation coefficient between each measure and the CIGTS VF score. A high correlation was observed with MD (r = −0.93; negative because of opposite scaling), and fairly high correlations with the other three measures (r = 0.64–0.75).

### Distribution of Baseline VF Values

Distributions of CIGTS VF scores and the MD at baseline for this study population of patients with newly diagnosed glaucoma are presented in Figure 2. Although a large proportion of patients had minimal to moderate defect (9% with a CIGTS VF score of zero, 50% with a score of 0.1–4.9, and 26% with a score of 5.0–9.9), a subset showed evidence of more substantial VF loss (15% with a score of 10 or greater). Both VF and MD distributions are skewed, reflecting little field loss for most people with newly diagnosed glaucoma, but substantial loss for a few.

Plots of the CIGTS VF score versus each of SF, PSD, and CPSD showed fairly linear relationships (not shown), although the PSD leveled off at the highest (12–16) CIGTS VF scores. The plot of CIGTS VF score versus MD shows some nonlinearity, with the MD showing a greater ability to spread values in the tails of the VF distribution than the CIGTS VF score, which is subject to the ceiling and floor effects (Fig. 3a).

### Learning Effects

Although all patients had completed at least one VF test for each eye before the first baseline measurement, learning effects between the first and second baseline measurements...
were still possible. The means of the first and second baseline VF scores were, respectively, 5.0 and 4.7, with a mean difference of 0.28 \((P = 0.0067)\) by paired \(t\)-test, Table 1). Although evidence of a learning effect was present (the second scores are lower than the first on average, indicating less defect), the magnitude of the effect is small. For MD, the first and second baseline scores were \(-5.6\) and \(-5.3\), respectively, for a difference of \(-0.26\) (representing improved VF; \(P = 0.0007\)).

We also observed a significant decay in the learning effect over time (\(P = 0.017\)), based on a linear regression of the difference in VF scores by the number of days (up to 30) between tests. When we tested again 3 days later, we observed a mean decrease of 0.61 in the VF score. After 20 days, essentially no decrease was observed. A similar pattern was observed for MD (\(P = 0.002\)), with a learning effect near \(-0.57\) VF units (i.e., improved VF) on day 3, increasing to zero by day 20.

### Variability within Subject

The difference between the first two baseline VF scores for each patient was used to characterize within-patient variability in the score. Most pairs of scores were very similar, with 44% of CIGTS baseline VF scores different by less than 1 unit, and 82% different by less than 3 units. However, 61 pairs (10%) differed by 4 or more (thus requiring a third VF measurement), and 5 pairs (1%) differed by 10 or more. MD scores were more consistent, with 54% differing by less than 1, 94% differing by less than 3, 97% differing by less than 4, and all differing by less than 9.

When the two baseline VF scores differed by 4 points or more and a third score was required \((n = 61)\), the third score was between the first two 75% of the time, above the higher score 3% of the time, and below the lower score 21% of the time. The small number of patients \((n = 2)\) with the third score

---

**Figure 2.** Relative frequency distributions of (a) CIGTS VF scores and (b) MDs at glaucoma diagnosis. For each patient, the average score from two VF tests (or the median of three if the first two scores differed by 4 or more) was used. Greater VF loss is represented by higher scores for CIGTS VF, but lower scores for MD.

**Figure 3.** Plots of CIGTS VF scores versus (a) MDs (CIGTS patients, \(n = 607\)), and (b) AGIS VF scores (non-CIGTS patients, \(n = 87\)).
higher than either of the first two scores may indicate that one of the first two scores was artificially high for some reason, rather than representing random variation.

The correlation coefficients between first and second baseline measurements are high (0.83 or greater) for all measures in Table 1 except SF (r = 0.36). The correlation is 0.83 for CIGTS VF scores and 0.91 for MDs. Although the correlation is not necessarily a good measure of test–retest agreement, because it measures the strength of the linear relationship, regardless of any differences between the two measures in location or scale, it can be useful when location or scale differences are negligible. In our case, the small location shifts in the second CIGTS VF or MD measures due to learning effects were negligible for practical purposes. Plots of the differences between visits for CIGTS VF and MD versus their respective averages revealed fairly symmetrical distributions above and below zero, with greater variability near the center for both measures than at the upper or lower ends of the distributions.

A different measure of test–retest variability is given by the pooled within-patient standard deviation estimates (pooling the 1-df standard deviation estimates from the first two baseline values from each patient). The pooled standard deviation estimates were 1.8 for CIGTS VF score and 1.4 for MD. These estimates reflect the variability of repeated scores in the same person around the person’s mean.

Factors Associated with Baseline VF Loss

Variables potentially associated with baseline CIGTS VF score were explored by using regression analysis. Variables tested included age, race, sex, education, baseline IOP, visual acuity, mean baseline reliability score, smoking (never smoked, former cigarette smoker, current cigarette smoker, current pipe or cigar smoker), family history of glaucoma, diabetes, hypertension and/or cardiovascular disease, evidence of cataract, right or left eye, and the alertness subscale from the SIP. In addition, CIGTS clinical center and technician effects were evaluated.

Several factors were found to affect baseline VF score significantly, as shown in Table 2. All variables together explained 23% of the variance in VF scores. A 1-point increase in the reliability score (worse VF reliability) was associated with a 2.7-point average increase in the VF score. VF scores of the men were higher than those of the women by 0.7 on average, and blacks were higher by 1.5 units than whites. A 10-letter decrease in VA score (equivalent to two Snellen lines) was associated with an average 1.5-unit increase in VF score. Cardiovascular disease was associated with a 1.1-unit increase in mean VF score, whereas diabetes was associated with a 1.8-unit decrease. Because patients with diabetes who had any evidence of diabetic retinopathy (≥10 microaneurysms in their retina) were excluded from the study, the eligible diabetic patients had higher visual function on average than nondiabetic patients. Thus, the diabetic effect is likely to be an artifact of this eligibility criterion and is not considered further.

IOP had a more complicated relationship with VF score, partly due to the study eligibility criteria that required qualifying VF loss for patients with IOPs less than 30 mm Hg (and later, 27 mm Hg), but did not require VF loss in patients with IOPs of 30 mm Hg or higher (later, ≥27 mm Hg) if glaucomatous optic disc damage was present. In the model, increasing IOP up to 30 mm Hg was associated with decreasing VF score, whereas an increase in IOP beyond 30 mm Hg was associated with increasing VF score. The IOP effects under 30 mm Hg are probably artifacts of the eligibility criteria, because patients with no qualifying VF defect entered only if they had elevated IOP. However, the magnitude of IOP effects at pressures higher than 30 mm Hg is not related to eligibility criteria. In that range, an increase of 10 mm Hg was associated with an increase of 1.6 units in VF score. An increase of 5 units on the alertness subscale (indicating more problems with attention) was associated with an increase of 1.2 units in VF score.

CIGTS clinical center effects were also significantly associated with baseline VF scores (P < 0.0001), although techni-
cians within clinical centers were not significantly different (P = 0.48). The proportion of variation explained increased from 16% to 23% after including clinical center in the model. The distribution of center effects (after adjusting for all other effects) was fairly normal with a SD of 1.0 VF unit; the three centers that varied the most from the mean had deviations of 2.0, −1.4, and −1.1 VF units from the adjusted mean of all centers. Although the regression assumption of normally distributed residual errors was not met because of the floor effect in the VF measurements, no transformation of the data could adequately correct the problem.

Similar modeling was performed to find variables associated with MD. Starting with the same initial list of variables used in the CIGTS VF score models, we arrived at a final model with the same variables as in the CIGTS VF score models shown in Table 2. The MD model explained 18% of the variation without including CIGTS clinical center effects, and 24% including center. All coefficients were in the same direction (taking the reversed scaling into account) and had magnitudes similar to those in Table 2. The result is not surprising, given the high correlation between MD and the CIGTS VF score. One advan-

Table 2. Results of Linear Regression Predicting Baseline CIGTS VF Scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient (SE)</th>
<th>P</th>
<th>Direction of Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability score</td>
<td>2.66 ± 0.48</td>
<td>0.0001</td>
<td>↑ Reliability score (less reliable) ⇒ ↑ VF score</td>
</tr>
<tr>
<td>Sex</td>
<td>0.69 ± 0.33</td>
<td>0.0390</td>
<td>Males ⇒ ↑ VF score</td>
</tr>
<tr>
<td>Race</td>
<td>1.52 ± 0.35</td>
<td>0.0001</td>
<td>Blacks ⇒ ↑ VF score</td>
</tr>
<tr>
<td>Visual acuity</td>
<td>−0.15 ± 0.03</td>
<td>0.0001</td>
<td>↓ VA ⇒ ↑ VF score</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>1.06 ± 0.45</td>
<td>0.0176</td>
<td>Cardiovascular disease ⇒ ↑ VF score</td>
</tr>
<tr>
<td>Diabetes*</td>
<td>−1.76 ± 0.44</td>
<td>0.0001</td>
<td>Diabetes ⇒ ↓ VF score</td>
</tr>
<tr>
<td>IOP ≤30*</td>
<td>−0.18 ± 0.05</td>
<td>0.0005*</td>
<td>↑ IOP up to 30 ⇒ ↓ VF score</td>
</tr>
<tr>
<td>IOP &gt;30</td>
<td>0.16 ± 0.06</td>
<td>0.0059</td>
<td>↑ IOP over 30 ⇒ ↑ VF score</td>
</tr>
<tr>
<td>SIP alertness†</td>
<td>0.24 ± 0.10</td>
<td>0.0129</td>
<td>↑ Alertness score ⇒ ↑ VF score</td>
</tr>
</tbody>
</table>

Proportion of variance explained ($R^2$) is 23%. The model also included CIGTS clinical center as a random effect.

* The significance of diabetes and IOP ≤30 are likely to be artifacts of the CIGTS eligibility criteria (see the Results section).
† SIP, Sickness Impact Profile. The alertness subscale is based on 10 questions and ranges from 0 to 10. A higher score indicates greater difficulty with alertness or attention.
Factors Associated with Baseline VF Variability

Modeling the variability of the two baseline VF scores was performed by using linear regression on the absolute difference between the two baseline measures (with the square root transformation to reduce skewness). Baseline covariates tested included all those tested in the baseline VF score model described earlier, plus the VF score itself and the number of days between baseline tests. For the CIGTS VF score, significant predictors of increased variability included right eye ($P = 0.0027$), increasing age ($P = 0.0040$), current smoking ($P = 0.0014$), an increased (worse) SLP alertness score ($P = 0.0177$), and a fourth-degree polynomial in the VF score itself (i.e., terms included for VF, VF$^2$, VF$^3$, and VF$^4$), reflecting lower variability for scores near zero and 16 and a plateau of constant variability between scores of approximately 3 and 13 ($P < 0.0001$). CIGTS clinical center effects were also significant ($P = 0.0160$). The proportion of variation explained by the model, $R^2$, was 59% on the transformed (square root) scale, but only 21% on the original scale of absolute differences.

Similar modeling of the variation of MDs found some of the same effects, but some differences. Variables predicting increased MD variability that were also significant in the CIGTS VF model included right eye ($P = 0.0017$), increasing age ($P = 0.0178$), current smoking ($P = 0.0249$), and a quadratic polynomial in the MD score itself ($P = 0.0001$), reflecting lower variability near the lower and upper limits of MD, and increased variability in the middle. New predictors of variability in this model included an increased (worse) reliability score ($P = 0.0079$) and high blood pressure (0.0176). CIGTS clinical center effects were also significant in this model ($P = 0.0101$). The number of days between baseline VF tests was not significantly associated with variability in either model.

Comparison with AGIS Scores

Figure 3b shows a plot of the CIGTS score versus the AGIS score for the sample of 87 non-CIGTS patients. Although the CIGTS and AGIS scores correlated highly ($r = 0.92$), the CIGTS score was on average 1 unit larger than the AGIS score (paired $t$-test $P = 0.0004$), with the difference increasing with the magnitude of the scores.

Discussion

Patients with newly diagnosed glaucoma are a group with diverse severity of VF defect. Although most patients in this study had minimal VF loss, 15% had more substantial VF loss (VF scores $> 10$). This result confirms that substantial VF loss can accrue before the diagnosis of glaucoma, unknown to the patient.

We found a small but significant learning effect in our two baseline measures, even though at least one previous VF test was required. Learning effects have been reported by others. Heijl and Bengtsson found substantial learning effects in MDs (of 2.8 dB) between first and second VF tests, but no statistically significant effect in the three subsequent tests. They tested only 25 patients, however, and the statistical power was fairly low and the results consistent with the small learning effect observed in this study, in which all patients had had at least one VF test before enrollment. The learning effects reported in Heijl and Bengtsson were present at approximately the same magnitude when the first and second tests were separated in time by weeks or even months, with similar results reported by Wild et al. The smaller learning effects that we observed, after a pre-enrollment VF test, diminished with time between the CIGTS baseline tests, with little or no learning effect left by 20 days after the first baseline test. These findings support the idea that requiring a preliminary VF test within the previous several months is sufficient to minimize substantial learning effects on subsequent VF testing.

Like the CIGTS score, the AGIS score is based on finding clusters of points of defect to avoid misclassifying artifacts as defects. The purpose of the AGIS score was to both diagnose glaucoma and monitor progression. For diagnostic purposes, certain patterns of defect were given greater weight, such as differences in VF loss above and below the midline or locations typical of glaucoma. The CIGTS score was developed only for monitoring progression, and therefore patterns indicative of glaucoma were not given special consideration.

The AGIS score is based on numbers of defect points falling in certain ranges, leading to a score that takes only integer values between 0 and 20. The CIGTS score allows for finer gradations. Despite these differences, the AGIS and CIGTS scores calculated from the same fields correlate highly ($r = 0.92$ in a sample of VF tests from non-CIGTS patients).

A drawback of the CIGTS VF score is the use of an ordinal scale to capture the depth of defect based on the probabilities. A logical consequence of the CIGTS VF score algorithm is that, for example, four points in a cluster depressed to the 0.05 level are “equivalent” with respect to defect to three points in a line (a single qualifying point) depressed to the 0.005 level. Whether this assumption has any justification in terms of visual function has not been tested. Furthermore, the interpretation of a sum of such ordinal values has no direct meaning other than as a general measure of defect.

The CIGTS VF score and the MD reported on the Humphrey printout also correlated highly ($r = -0.93$). A conceptual difference between the CIGTS VF score and the MD is that the MD reflects an average VF depression (in decibels) over the field, and even minor defects can depress the MD score. The CIGTS VF score is based on the probabilities from the total deviation probability plot, and only probabilities less than 0.05 can potentially increase the score. The CIGTS VF score has an advantage over MD, being based on the total deviation probability plot, that it may more accurately reflect field loss at points where the age-specific distribution is quite skewed. Also, the CIGTS VF score is not affected until at least three neighboring points are all depressed below the 5th percentile, potentially avoiding artifactual VF depressions at one or two points. The CIGTS VF score is somewhat less reproducible than the MD when comparing the first and second baseline measures. We conclude that the CIGTS VF score and the MD are quite comparable. The behavior of both scores over follow-up will provide a more complete basis of comparison.

A preferable VF score would use the actual empiric percentile of age-adjusted values at each point. Unfortunately, the data distributions required for such calculations were not available from Zeiss-Humphrey Systems, Inc. Although the MD uses a fairly continuous measure (in decibels) of defect to three points in a line of the field loss, and even minor defects can depress the MD score. The CIGTS VF score is based on the probabilities from the total deviation probability plot, and only probabilities less than 0.05 can potentially increase the score. The CIGTS VF score has an advantage over MD, being based on the total deviation probability plot, that it may more accurately reflect field loss at points where the age-specific distribution is quite skewed. Also, the CIGTS VF score is not affected until at least three neighboring points are all depressed below the 5th percentile, potentially avoiding artifactual VF depressions at one or two points. The CIGTS VF score is somewhat less reproducible than the MD when comparing the first and second baseline measures. We conclude that the CIGTS VF score and the MD are quite comparable. The behavior of both scores over follow-up will provide a more complete basis of comparison.
estimated percentile data would be an improvement over both the MD and the CIGTS VF score. However, the fact that the CIGTS VF score and the MD both give equal weight to all points of the field. The goal of each is to measure VF progression wherever it should appear, with measurement of function considered separately. We also acknowledge that a global score, such as either the CIGTS, MD, or the AGIS VF score, may miss VF progression if a worsening local scotoma is masked by minor improvement in other regions of the field. Also, counting only clusters of defect, as does the CIGTS score, may miss small but deep true points of defect.\textsuperscript{18}

We found several variables that were significantly associated with increased VF loss at baseline in CIGTS participants. These included poor reliability score, male sex, black race, decreased visual acuity, presence of cardiovascular disease, and high IOP. Most of these variables have been reported in previous studies (reliability,\textsuperscript{4,19} sex, race, visual acuity, and IOP\textsuperscript{20}). Although cardiovascular disease has not been explicitly associated with VF loss, VF loss has been associated with some risk factors for cardiovascular disease.\textsuperscript{20,21} Street et al.\textsuperscript{22} reported a weak association of atherosclerotic disease with visually significant cataract requiring surgery in Medicare beneficiaries. Vogel et al.\textsuperscript{23} reported a significant correlation between their measure of VF defect and initial IOP $(r = -0.26, \text{where lower VF scores indicate greater defect; } P = 0.0001)$, although the plot they present hints at a threshold rather than a linear relationship. Patients with initial IOPs less than 50 mm Hg had the whole range of VF scores, whereas patients with IOPs over 50 mm Hg had consistently poor baseline VF scores. The effects in our study of race and sex may be due to differential access to medical care or treatment-seeking behavior. Such effects are well documented in the literature.\textsuperscript{23,24} As in AGIS,\textsuperscript{1} we found no relationship between VF loss and baseline age.

We found short-term within-patient variability in VF scores to be slight in most patients, but found substantial variability in a small proportion of patients. This pattern of variability is consistent with the AGIS experience.\textsuperscript{1} Predictors of CIGTS VF score (and MD) variability included reliability of the VF score. A similar association between MD variability and reliability score was reported by McMillan et al.\textsuperscript{19} Katz et al.\textsuperscript{1} have noted that patients with glaucoma report greater difficulty in meeting the reliability criteria of the Humphrey software than normal subjects. Our finding that increased VF variability is associated with increased age has been reported previously by Katz and Sommer.\textsuperscript{25} but not Boeglin et al.\textsuperscript{2}

The increase in VF variability among smokers that we observed in both the CIGTS VF score and the MD has not been reported previously to our knowledge. Although this finding may be an artifact, smoking is a well-established risk factor for cataract formation.\textsuperscript{26,27} Although patients with clinically significant cataract were excluded from the CIGTS, it is possible that subclinical cataract was more common among the smokers. We have observed in CIGTS follow-up data that cataract formation is associated with worsening VF scores. The denser lens media among smokers may have globally suppressed the VF measurement, lowering sensitivity and increasing variability.

Although the effect of alertness on VF variability has not previously been directly tested, it is certainly plausible that lack of attention could lead to increased variability in VF scores. Related literature has shown that VF test results may be influenced by both alcohol consumption\textsuperscript{28} (decreased MD and sensitivity, and increased PSD, number of stimulus presentations, and false negatives) and use of antihistamines\textsuperscript{29} (higher SF). The observed effect of the subject’s alertness on VF variability could have implications for clinical practice. In patients with alertness problems, the effect may be partially diminished by scheduling the VF testing early in the day and before the clinical examination. The observed increase in variability with the right eye is probably associated with the fact that right eyes were always tested first. By the time the left eye was tested, the patient had settled into the routine of the test and was more consistent.

The significant effects of CIGTS clinical center in both the mean VF score and in the VF score variability found using both the CIGTS VF score and the MDs, may indicate differences in the patient populations at the various centers that were not captured in the other variables measured. This hypothesis is partially supported by our finding of significant clinical center effects for IOP $(P = 0.0017)$ and VA $(P = 0.0005)$, and even a marginally significant center effect on alertness scores $(P = 0.0501)$, where patient population differences are the likely causes. However, the center effects for VF and MD were stronger than those seen for the other effects tested. Because the centers’ VF machines are calibrated regularly, it is unlikely that the clinical center effects represent machine differences. However, other factors related to the setup may have more impact than previously considered. We tested for technician differences within clinical center among patients who were tested by the same technician at both baseline visits and found no significant effect.

In summary, this investigation has demonstrated that several measures of VF loss have very similar properties, and that learning effects are small when a pre-enrollment VF test is required. In addition, some variables associated with baseline VF loss and VF variability have been confirmed, and new associations have been found. These findings may be useful in explaining VF scores and variability among patients with glaucoma.

\section*{References}


