Scoring of Visual Field Measured through Humphrey Perimetry: Principal Component Varimax Rotation Followed by Validated Cluster Analysis

Jean-Philippe Nordmann,[1] *Mounir Mesbah,*[2] *and Gilles Berdeaux*[3,4]

Purpose. To extract unidimensional, well-separated latent scores that are anatomically and clinically valid from 52 standardized variables collected by Humphrey visual field (VF) perimetry (Carl Zeiss Meditec, Dublin, CA).

Methods. Visual field data of 437 patients were collected and classified by a glaucoma specialist into seven clinical groups: irregularities of VF (IVF), nasal step (NaS), arcuate scotoma (AC), paracentral scotoma (PCS), blind-spot enlargement (BSE), diffuse deficit (DD), and advanced deficit (AD). The number and content of constituent variable scores were identified by principal components analysis followed by Varimax Rotation and simple clustering, taking spatial distribution homogeneity and visual system anatomy into account. Unidimensionality was checked by a stepwise Cronbach α curve. Clinical predictability of the derived scores was checked by comparing clinical groups (ANOVA).

Results. Patients older than 60 years comprised 53.3% of the sample. The average mean deviation was −9.2 dB and pattern standard deviation was 6.5 dB. Six scores were identified: four peripheral scores (nasal superior, NS; nasal inferior, NI; temporal superior, TS; and temporal inferior, TI) and two paracentral scores (PCSs; superior, PCSS; and inferior, PCSI). Cronbach α was always >0.90. The six scores decreased sequentially from IVF to DD to AD. Scores of AC were lower in NS, NI, and TS; PCSS was less in PCS; BSE scores were less in TS and TI; NaS scores were less in NS and NI.

Conclusions. Six well-separated, optimal scores were obtained from the Humphrey perimetry matrix. Internal reliability was good. It was possible to discriminate between clinical subgroups. Further analyses, based on longitudinal data, must be performed to confirm these findings. (*Invest Ophthalmol Vis Sci.* 2005;46:3169–3176) DOI:10.1167/iovs.04-1214

Automated static perimetry is one of the methods used to screen and follow up patients who have glaucoma.[1–4] It consists of approximately 100 quantitative threshold measures that permit evaluation of retinal sensitivity. Each measure is standardized in a population free of ocular disease, and two simple statistics are calculated: mean deviation (MD) and pattern standard deviation (PSD). These indices are widely used in glaucoma clinical trials and patient follow-up.

Mean deviation (MD)[5] is the average of all differences between measures and their normal values, weighted by the variance observed in the general population (where X_i is the measured threshold, N_i is the normal reference threshold at point i, S^2_{1i} is the variance of normal field measurement at point i, and n is the number of test points).

$$\left[\frac{1}{n}\sum_{i=1}^{n}\frac{(X_i - N_i)}{S^2_{1_i}}\right] \Bigg/ \left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{S^2_{1_i}}\right].$$

PSD is a normalized distance, standardized with reference to the general population, calculated for each point.

$$\sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}S^2_{1_i}\right]\cdot\left[\frac{1}{n-1}\sum_{i=1}^{n}\frac{(X_i - N_i - MD)^2}{S^2_{1_i}}\right]}$$

Although there is a considerable body of literature about the properties of these indices as they relate to the clinical picture, some weaknesses can be identified a priori: (1) They do not take into account the spatial distribution of the points measured, the proximity of one point to another, and correlations between the points (e.g., switching one measure with another changes neither the MD nor PSD); (2) MD cannot be interpreted without knowing PSD, and vice versa; (3) MD is not a sensitive parameter in early stages of glaucoma; (4) PSD is not a sensitive parameter in late stages of glaucoma; (5) neither measure takes into account anatomic dimensions of the eye or visual system (horizontal threshold of retinal nerve fibers, the vertical threshold of vision cerebral hemispheres, and the retinal artery located centrally in the optic nerve).

Very little research has been performed to find algorithms that would help to identify visual field defects (VFDs) more precisely. Brigatti et al.[6,7] used computerized neural networks with some success to identify patients with early glaucomatous visual field loss, yielding sensitivity and specificity both >70%. To achieve this, they had to include information on the automated visual field index and other structural data.

Mandava et al.[8] identified 11 clusters by nearest-neighbor cluster analysis performed on Octopus visual fields (Haag-Streit, Köniz, Switzerland). A discriminant analysis was performed on the 11 scores used to classify patients with and without glaucoma. The sensitivity and specificity of this classification were very good (sensitivity and specificity >90%).

Brigatti et al.[7] and Mandava et al.[8] shared a common objective to develop a classifying algorithm that would help clinicians to detect new VFDs. Brigatti et al.[7] directly included global perimetry indicators that assume that MD, PSD, and short-term fluctuation are the optimal information that can be retrieved from this test. However, the model produces estima-

From [1]Hôpital des Quinze-Vingts, Centre Hospitalier National d'Ophtalmologie, Paris, France; [2]Laboratoire de Statistique Theorique et Appliquée, Université Pierre et Marie Curie, Paris, France; [3]Conservatoire National des Arts et Métiers, Paris, France; and [4]Alcon France, SA, Paris, France.

Supported by an unrestricted grant from Alcon France, SA, Rueil-Malmaison, France, conducted according to local laws, and contracted to Université de Bretagne-Sud, Vannes, France.

Submitted for publication October 13, 2004; revised February 18 and April 7, 2005; accepted April 28, 2005.

Disclosure: **J.P. Nordmann**, None; **M. Mesbah**, None; **G. Berdeaux**, Alcon France, SA (E, F)

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: Gilles Berdeaux, Alcon France, 4, Rue Henri Sainte-Claire Deville, F-92563 Rueil-Malmaison Cedex, France; gilles.berdeaux@alconlabs.com.

tors that are completely disconnected from clinical reality, and clinicians still must be persuaded by the conclusions of neural network modeling. Mandava et al.[8] used a cluster analysis with the sole purpose of optimizing information reduction before running a discriminant analysis. However, the clusters did not respect retinal anatomy very well, for example, they spanned the vertical meridian, and almost nothing was stated about either the construct validity or the internal consistency reliability of these clusters.

Although we agree with the approach developed by Mandava et al.,[8] we believe that the scores developed should be demonstrated to be clinically relevant before they are applied in the identification and monitoring of patients with glaucoma by using specific statistical techniques—for example, neural networks or discriminant function.

The present pilot study was designed to identify scores produced by automated static perimetry that would: (1) be structurally valid, (2) respect the anatomy of the eye and visual system, (3) follow the clinical evolution of glaucoma, and (4) be easily interpreted by physicians. This article addresses only the development and validation of the scores.

Materials and Methods

Data Collection

A cross-sectional experimental design was used. All patients were recruited from the glaucoma department of the Hôpital des Quinze-Vingts. Consecutive perimetry data from one eye (left or right) of each patient were selected from the medical records. Patients were of either sex with no age limit and were required to have medically confirmed glaucoma, presenting with at least one VFD. Patients with no VFD were not included because they did not contribute valuable information to the correlation matrix. As a general rule, any type of glaucoma was accepted (congenital, inflammatory, or neovascular, with narrow or closed angle, after cataract surgery), since all types contribute valuable information. Based on the available data, all patients had primary open-angle glaucoma. This research adhered to the tenets of the Declaration of Helsinki and also observed the guidelines in French law regarding data privacy.

Both eyes were taken into account when glaucoma was bilateral. To pool data from all eyes, a left mirror image of right eyes was computerized. A potential laterality effect on scores was checked by an analysis of variance (ANOVA).

All data used had been obtained according to the usual procedure of the glaucoma department. Patients, by default therefore, underwent a Humphrey perimeter 24-2 threshold test, with the FASTPAC algorithm (Carl Zeiss Meditec, Dublin, CA). The main characteristics of FASTPAC is the use of 3-dB steps. The fixation target was centralized on the blind spot used to monitor fixation. The stimulus was size III, white, and on a background of 31.5 asb. Patients were required to have experienced a previous visual field test. Only tests with normal reliability indices were included: fixation losses $<$20% and false-positive and -negative errors $<$30%.

On the basis of these data, a glaucoma specialist clinically classified patients into seven groups, according to VFD shape: (1) irregularities of visual field; (2) nasal step; (3) arcuate scotoma; (4) paracentral scotoma; (5) blind spot enlargement; (6) diffuse deficit; and (7) advanced deficit. Visual field irregularities corresponded to slightly depressed thresholds throughout the visual field, which did not amount to specific focal defects. Diffuse defects corresponded to a generalized loss of sensitivity throughout the whole field. Advanced glaucoma is defined by multiple scotoma surrounding the fixation point.

Data were entered into the study database with double-entry control. Analysis was conducted on the pattern deviation matrix since: (1) the data are standardized against a normal population; and (2) it reveals localized defects that may be masked by a generalized depression or an elevation of the hill of vision.

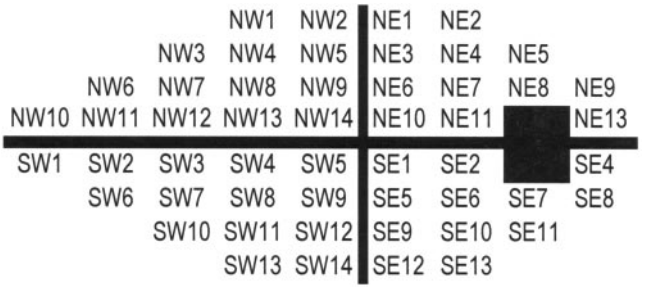

Figure 1. Variable names used for the clustering analysis.

Control patients without glaucoma were not included because the pattern deviation matrix is already standardized according to an age- and gender-matched normal population. In addition, the Varimax Rotation, used for optimization of the data, would have been affected by the presence of two (control and patient) subjects with different patterns of correlation.

Figure 1 describes the variable labels used in the analysis. Fifty-two variables were identified from the threshold data, as follows: 14 in the northwest quadrant (NW1–NW14); 12 in the northeast quadrant (NE1–NE11, NE13); 14 in the southwest quadrant (SW1–SW14); and 12 in the southeast quadrant (SE1, SE2, SE4–SE13). This nontraditional terminology is used to facilitate understanding of the Varimax Rotation used and the resultant clusters. MD and PSD were collected as the main parameters from the perimeter test.

Clustering of Variables and Construction of Scores

The statistical analysis was conducted on computer (SAS software, ver. 8.2; SAS Institute, Cary, NC). Two-sided tests were performed with α = 5%. Our sample size was fixed according to empiric rules used in factor analysis—that is, 5 to 10 patients should be included per variable in the analysis; hence, $>$400 patients were included for the 52 variables identified.

A principal components analysis (PCA) was performed on the 52 variables to retain all meaningful information. The number of factors retained was chosen after examining the plot of eigenvalues. Only factors with eigenvalues $>$1 were retained. Scores were extracted after Varimax Rotation followed by a cluster analysis with the number of clusters fixed and equal to the number of factors retained—that is, each of the 52 variables was classified into the most correlated rotated factor. An unweighted score was computed for each factor by summing the variables that correlated with it the most. Each variable was included in one score only. This clustering step was followed by a refining process, described later, that took into account the specificity of the data.

Correlations between each of the 52 variables and the scores were computed. A variable was regarded as correctly classified when its correlation with a score was higher than that with any other score. When this was not the case, the variable was moved to other scores to maximize its correlation. This iterative process continued until the system stabilized (i.e., no further movement was necessary). Some specific rules were added regarding the spatial homogeneity and clinical coherence of the clusters: (1) To maintain the spatial homogeneity of clusters (geographical continuity: a score had to relate to neighboring measurements), constraints were defined that restricted certain moves—that is, if a measurement was separated spatially from its cluster by another measurement, the former was moved to the nearest cluster. (2) To maintain the clinical coherence of clusters, specific constraints were defined that restricted certain moves—that is, if a variable was in a cluster not compatible with known medical constraints underlying eye function, it was moved to the nearest, clinically meaningful cluster. This rule applied when a cluster crossed the east–west or north–south axes.

Assessment of Unidimensionality and Clinical Validity of Scores

A stepwise Cronbach α curve[9–11] was plotted to check the unidimensionality of the variables yielding the score. This calculation made it possible to verify that a group of items measured the same underlying unidimensional concept (construct validity). The curvature should increase monotonically when all items belong to the appropriate score. Otherwise, items should be allocated to another score. However, because it is influenced by sample fluctuations, it should be interpreted cautiously, especially when scores contain few items.

Clinical validity can be estimated as the ability of a score to capture clinical relevance.[12] The mean of each score was compared across all seven clinical groups by ANOVA.

Results

In total, 437 consecutive visual fields were collected of which 229 (52.4%) were right eyes. Mean age was 61.2 ± 14.8 years (SD) and 53.3% of patients were older than 60. Patients manifested the following visual field abnormalities: irregularities of visual field (n = 34: 7.8%); nasal steps (n = 126: 28.8%); arcuate scotomas (n = 154: 35.2%); paracentral scotomas (n = 21: 4.8%); blind-spot enlargements (n = 13: 3.0%); diffuse deficits (n = 26: 5.9%); and advanced deficits (n = 63: 14.5%).

Pupil diameter was documented in 137 patients and, on average, was 4.3 mm ± 1.1 (SD). The average duration of the test was 8.0 minutes. The average MD was −9.2 ± 7.2 dB (SD) and the PSD was 6.5 ± 3.3.

Principal component analysis identified six factors that explained 61.09% of the total variance. After an abrupt decrease, the plot of the eigenvalues showed a clear break at the sixth eigenvalue, then a plateau, and again a new, but slighter, decrease. The sixth eigenvalue was also the first value less than unity,[1] and so the min-eigen criterion also retained six factors.

Figure 2 describes step-by-step how the scores were constructed. Examination of correlations with the six factors retained after Varimax Rotation indicated the first cluster pattern of the original variables. All measurements of the northwest quadrant plus NE3, NE6, and NE11 were correlated and constituted factor 1 (correlations: 0.53–0.85). Factor 2 comprised all measurements of the southwest sector (correlations: 0.59–0.85). Factor 3 included all measurements of the northeast sector (correlations: 0.40–0.75) except NE3, NE6, and NE11, already attracted by factor 1, and NE10, which alone represented factor 6 (correlation: 0.49). Factor 4 comprised measurements of the southeast sector (correlations: 0.54–0.70) except SE5 and SE6, which jointly made factor 5 (correlations: 0.53 and 0.63).

The six rotated factors gave us an initial set from which to build six simple scores. When examining empiric correlations between the original variables and these scores, we found that NE3 correlated more with score 3 (from factor 3) than with its own score 1 (from factor 1). Similarly, NE11 correlated more with score 6 than with score 1, and SW5 correlated more with score 5 than with its own score of 2 (Fig. 2: scores after the first correlation analysis). Moves were performed accordingly.

Correlations were calculated again, and we were obliged to move NE6 from score 1 to score 3, NW14 to score 6, and both SE1 and SE2 from score 4 to score 5. (Fig. 2: scores after the second correlation analysis). No additional moves were necessary after the second step.

Cronbach α curves were estimated for each subset to check that the scores were unidimensional. We were obliged to move NE13 and SE4 to another set to obtain unidimensional scores. If NE13 was moved to score 4, we produced a set crossing the north-to-south quadrants. Moreover, the new set (score 4 + NE13) was no longer unidimensional. A similar situation arose

Figure 2. Score construction. Step-by-step description: nasal superior (NW1–NW13); nasal inferior (SW1–SW4; SW6–SW14); temporal superior (NE1–NE5; NE8; NE9; NE13); temporal inferior (SE4; SE7–SE13); paracentral superior (NE6; NE7; NW14; NE10; NE11); paracentral inferior (SW5; SE1; SE2; SE5; SE6).

when we tried to move SE4 to score 3. NE13 and SE4 were therefore not moved. NE6 and NE7 could, however, be moved to score 6, in accordance with the adopted rules. In this way, spatial homogeneity was respected (no crossing from the north to the south quadrant), as was specificity (each variable correlated more with its own score than with any other score).

The six clusters from which scores were derived are described in Figure 2. Four scores were peripheral (nasal superior, NS; nasal inferior, NI; temporal superior, TS; temporal inferior, TI) and the remaining two central (paracentral superior, PCS; and paracentral inferior, PCI). The formulas were respectively:

$$NS = (NW1 + NW2 + NW3 + NW4 + NW5 + NW6 + NW7 + NW8 + NW9 + NW10 + NW11 + NW12 + NW13)/13$$

$$NI = (SW1 + SW2 + SW3 + SW4 + SW6 + SW7 + SW8 + SW9 + SW10 + SW11 + SW12 + SW13 + SW14)/13$$

$$TS = (NE1 + NE2 + NE3 + NE4 + NE5 + NE8 + NE9 + NE13)/8$$

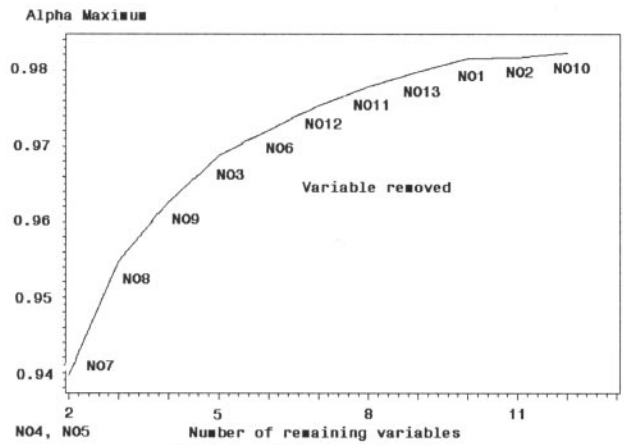

Figure 3a : Nasal superior score

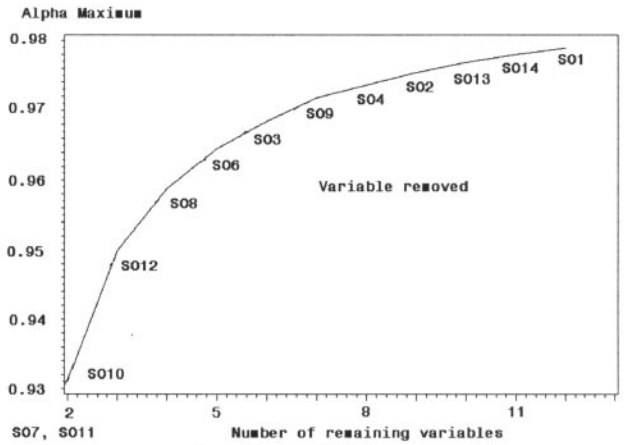

Figure 3b: Nasal inferior score

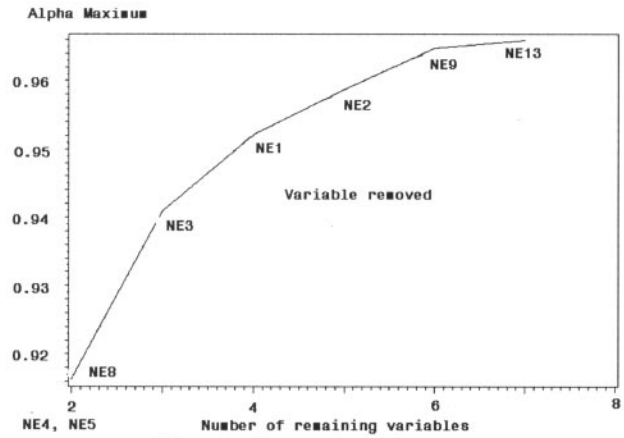

Figure 3c: Temporal superior score

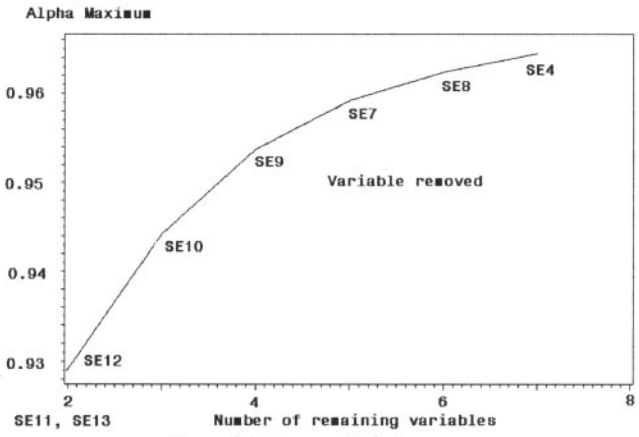

Figure 3d: Temporal inferior score

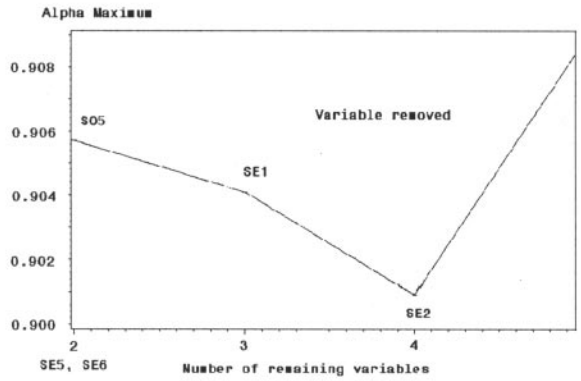

Figure 3e: Para-central inferior score

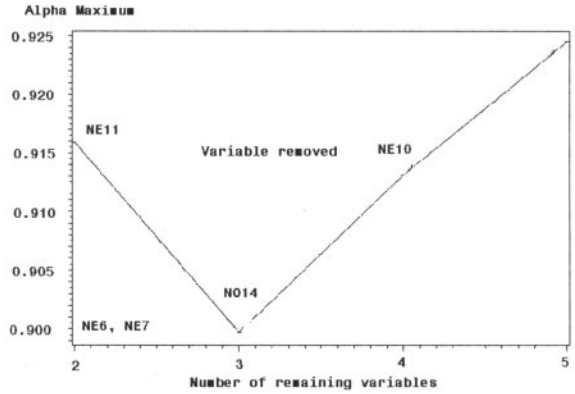

Figure 3f: Para-central superior score

FIGURE 3. Cronbach α curves.

$$TI = (SE4 + SE7 + SE8 + SE9 + SE10 + SE11 + SE12 + SE13)/8$$

$$PCS = (NE6 + NE7 + NW14 + NE10 + NE11)/5$$

$$PCI = (SW5 + SE1 + SE2 + SE5 + SE6)/5$$

For each score, Cronbach α[13] was always >0.90 (Fig. 3, the maximum value reported in the curve), demonstrating good internal reliability. Cronbach α curves for peripheral scores increased monotonically, supporting the requirement that all items should contribute to the score. However, this was not the case for central scores, although Cronbach α curves were always >0.90. Last, the decrease of curvature was small and limited to a single item.

Table 1 describes cross-correlations between the scores. Six of fifteen coefficients were >0.70 and most were between central and peripheral scores.

Table 2 and Figures 4 and 5 illustrate the six scores and the MD according to the type of clinical abnormality. As shown by the MD, all scores demonstrated highly significant differences across the seven groups of patients. At the extremes, patients

TABLE 1. Correlation Coefficients between Scores

Cluster	Nasal Superior	Nasal Inferior	Temporal Superior	Temporal Inferior	Paracentral Superior	Paracentral Inferior
Nasal superior	1.00					
Nasal inferior	0.58	1.00				
Temporal superior	0.80	0.53	1.00			
Temporal inferior	0.53	0.79	0.65	1.00		
Paracentral superior	0.45	0.76	0.52	0.77	1.00	
Paracentral inferior	0.80	0.51	0.78	0.53	0.56	1.00

All correlation coefficients were significantly different from zero ($P < 0.001$). According to Chassany *et al.*,[13] a correlation coefficient between 0.4 and 0.7 is a guarantee of no redundancy.

with advanced deficits produced lower MDs than those with VF irregularities. The MDs in Figures 4 and 5 were weighted averages of the six scores according to the size of the respective clinical groups. The scores showed significant variations around MD that were specific to clinical abnormalities. Patients with a nasal step had nasal superior and nasal inferior scores less than the MD, whereas the other four scores were greater than the MD. Patients with VFDs, or a diffuse deficit, did not demonstrate large differences between the six scores and the MD. Patients with an arcuate scotoma had nasal superior and temporal superior scores less than the MD, whereas the paracentral score was greater than the MD. Patients with a paracentral scotoma had temporal superior and temporal inferior scores greater than the MD, whereas the paracentral score was less than the MD. Patients with blind spot enlargement had temporal superior and temporal inferior scores less than the MD. Last, some variation was observed in even the most severe patients (advanced), as follows: the paracentral inferior and temporal inferior scores were greater than the MD, whereas the two nasal scores (superior and inferior) were less than the MD.

DISCUSSION

MD and PSD are often used in randomized clinical trials as primary end points for the evaluation of glaucoma treatments. A low MD or a high PSD is associated with moderate or advanced stages of glaucoma. Consequently, both parameters must be taken into account when making decisions at a population level, thus making a global decision rule difficult to formulate.

If homogenous subgroups of measures (i.e., a new set of MD scores) could be identified (ideally related, even empirically, to a clinical glaucoma classification), the follow-up of patients with glaucoma would be more clinically relevant. A vector of scores, each specific to a precise anatomic visual field area, starting with a high score at the onset of disease and decreasing with severity, would dramatically clarify treatment decisions and the follow-up of patients with glaucoma. In this manner, PSD would become an indirect measure of heterogeneity within scores.

A PCA followed by our clustering algorithm allowed us to reach this objective. We identified six scores that explained more than 60% of the observed variance. The scores had orthogonal properties meaning that their independence was maximized. They had also good construct validity, as demonstrated by Cronbach α curves—that is, with the four peripheral scores, at least; switching an item to another score did not improve its validity. In the case of the two paracentral scores, some switching did improve reliability. However, the loss of reliability, as measured by the Cronbach α curve, was very low and could be explained by stochastic sampling issues. We therefore preferred to keep scores close to retinal anatomy (proximity of points), instead of maximizing the mathematical properties of our scores. Finally, the scores were easy to calculate.

We used the pattern deviation matrix and performed statistical manipulations to achieve this result. We could have worked on threshold sensitivities or the total deviation matrix. The former would have required an adjustment for age. The latter provides an indirect standardization, based on a population-wise approach, which is better than local data-based adjustments. We used the pattern-deviation matrix because it emphasized localized defects and therefore would increase the correlation between points belonging to a same VFD. This correlation should stabilize the Varimax Rotation.

Although we used a rotation pattern that maximized score independence, we still found high correlations between certain scores. This indirectly, but strongly, supports the fact that the MD is a score with a high construct validity. In other words, each item contributes homogenously to the MD. In contrast, a possible use of MD as a single score would explain a much smaller part of the total variance; hence, much information would be lost.

TABLE 2. Scores According to VF Clinical Abnormalities

Clinical Group	Irregularities of Visual Field $n = 34$	Nasal Step $n = 126$	Arcuate Scotoma $n = 154$	Para-Central Scotoma $n = 21$	Blind Spot Enlargement $n = 13$	Diffuse Deficit $n = 26$	Advanced Deficit $n = 63$
Nasal superior	−3.50 (0.34)	−6.46 (0.45)	−13.30 (0.68)	−6.96 (1.14)	−3.42 (0.98)	−7.87 (0.98)	−23.98 (0.85)
Nasal inferior	−3.16 (0.33)	−6.10 (0.51)	−10.16 (0.62)	−6.47 (0.92)	−3.60 (0.87)	−6.52 (0.65)	−23.48 (0.92)
Temporal superior	−4.71 (0.46)	−4.00 (0.39)	−12.42 (0.62)	−4.68 (0.77)	−6.80 (1.77)	−7.94 (0.62)	−20.70 (0.85)
Temporal inferior	−3.25 (0.31)	−3.16 (0.34)	−8.56 (0.51)	−3.92 (0.71)	−6.39 (1.37)	−7.96 (0.79)	−19.88 (1.13)
Paracentral inferior	−2.42 (0.26)	−2.44 (0.29)	−5.80 (0.45)	−6.91 (1.40)	−5.65 (1.37)	−7.36 (0.71)	−18.58 (1.07)
Paracentral superior	−2.99 (0.34)	−3.13 (0.31)	−9.60 (0.63)	−10.61 (1.25)	−4.43 (0.91)	−6.75 (0.89)	−22.00 (1.01)
Humphrey MD (dB)	−3.19 (0.23)	−4.67 (0.31)	−10.04 (0.41)	−6.72 (0.65)	−4.49 (0.94)	−7.19 (0.61)	−22.17 (0.61)

Probability calculated according to ANOVA. Average scores by clinical groups were found to be statistically significant at $P < 0.0001$. Data are expressed as the mean ± SD.

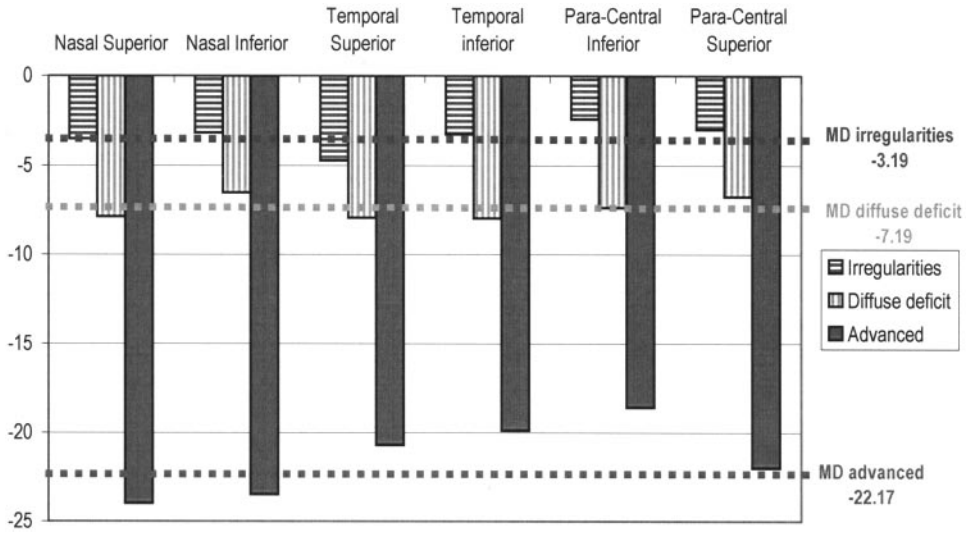

FIGURE 4. Scores according to clinical visual field abnormalities (irregularities, diffuse deficit, and advanced deficit). MD was the control.

Our algorithm was successful in producing scores that respected retinal anatomy (good and rapid convergence). Four quadrants, fully separated by the horizontal and vertical axes, were shown to be associated with two central scores. The coexistence of central and peripheral central scores could be interpreted as follows: Central scores may be sensitive to blood flow variations in the central retinal artery, whereas peripheral scores may be more sensitive to intraocular pressure's effects on nerve fibers at the disc junction. Additional data are needed to confirm this hypothesis.

The relationship between the localization of a clinical field defect and its corresponding score demonstrated that our scoring algorithm was clinically relevant and thereby possessed construct validity. Apart from visual field irregularities and diffuse deficit, our six scores assembled valuable data describing the localization of a VFD.

Nasal scores were involved in nasal step and arcuate scotoma, temporal and paracentral scores in blind-spot enlargement, and paracentral scores in paracentral scotoma. Even with advanced deficits, the six scores added information to the MD. The relationship between the localization of a clinical VFD and its corresponding score demonstrated that our scoring algorithm was clinically relevant and thereby possessed construct validity.

Our six scores differed from the eleven described by Mandava et al.[8] because of the algorithm used. We believe that fewer scores for retinal anatomy would be easier both to apply and understand in daily practice. Because our scores demonstrated good construct and external validity, they should assist in patient follow-up, although additional longitudinal data are needed to confirm this. Comparison with the work of Brigatti et al.[6,7] is not straightforward, since their main goal was to identify patients with glaucoma. Nonetheless, our scores could be used as entry parameters in a neural network to serve the same purpose.

AGIS (Advanced Glaucoma Intervention Study) scores[14] were not calculated for our sample of patients. Therefore, a head-to-head comparison with our six scores was not possible. The AGIS investigators decided that one single clinical score was appropriate to define the severity of glaucomatous VFDs. This assumption was somewhat contradicted by our findings. Our algorithm is also simpler than that used in AGIS and can be managed with a basic calculator. Finally, we demonstrated that our six scores brought additional information to the MD.

Our pilot study has limitations. A larger sample size may have increased the sensitivity of our algorithm to detect different or additional scores. The sample size of some patient groups was rather small, and certain analyses should be inter-

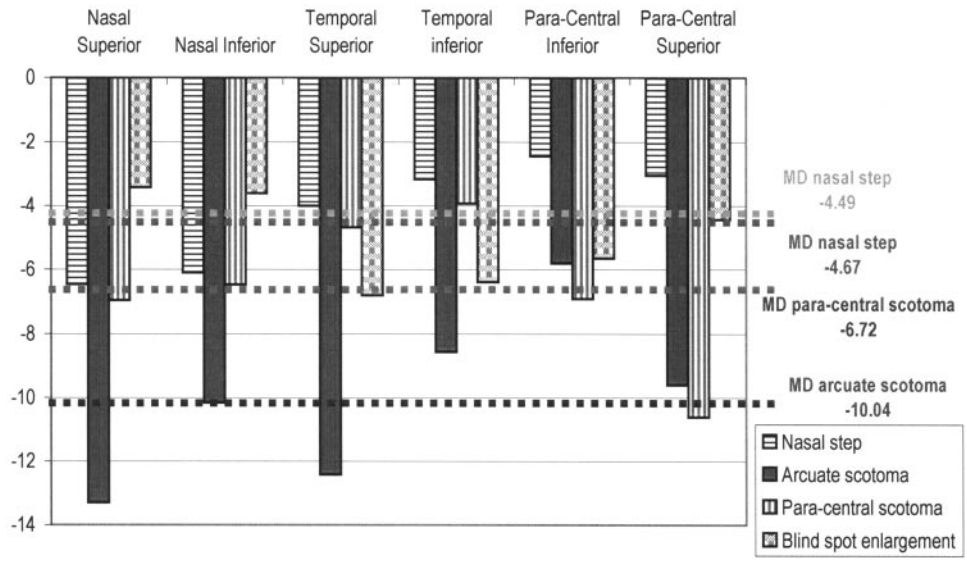

FIGURE 5. Scores according to clinical VFDs (nasal step, arcuate scotoma, paracentral glaucoma, and blind-spot enlargement). MD was the control.

preted cautiously. All patients came from a single center, which reduces external validity. We defined seven classes of clinical abnormality arbitrarily, but other clinical classifications may be of interest. We also used unweighted combinations of variables, whereas Humphrey algorithms use standardization according to several confounding factors. The use of weighted combinations of variables could improve the sensitivity our scores. Last, we restricted our algorithm to a linear approach. The use of nonlinear models may make some scores more accurate.

More development work is needed before our six scores can be used. Sensitivity to clinical changes should be explored, and the discrimination of different clinical abnormalities must be verified.

APPENDIX

Cronbach α and the Stepwise Cronbach α Curve

A Parallel Model Describing the Unidimensionality of a Set of Variables. Let $X_1, X_2, \ldots, X_k$, be a set of observed variables measuring the same underlying unidimensional latent (unobserved) variable. We define X_{ij} as the measurement of patient i, where $i = 1, \ldots, n$, given by a variable j, where $j = 1, \ldots, k$. The model underlying Cronbach α is a simple, mixed, one-way model: $X_{ij} = \mu_j + \alpha_i + \varepsilon_{ij}$, where μ_j is a variable fixed (nonrandom) effect and α_i is a random effect with zero mean and SE σ_α corresponding to patient variability. It produces the variance of the true latent measure ($\tau_{ij} = \mu_j + \alpha_i$); and ε_{ij} is a random effect with zero mean and SE σ corresponding to the additional measurement error. The true measure and the error are uncorrelated: $\text{cov}(\alpha_i, \varepsilon_{ij}) = 0$.

These assumptions are classic in experimental design. This model defines relationships between different kinds of variables: the observed score X_{ij}, the true score τ_{ij}, and the error ε_{ij}.

Reliability of an Instrument. A measurement instrument gives us readings that we call observed values. The reliability Δ of an instrument is defined as the ratio of the true over the observed measure. Under the parallel model, one can show that the reliability of any variable X_j (as an instrument to measure the true value) is given by

$$\rho = \sigma_\alpha/(\sigma_\alpha^2 + \sigma^2)$$

which is also the constant correlation between any two variables. This coefficient is also known as the intraclass coefficient. The reliability coefficient Δ can be easily interpreted as a correlation coefficient between the true and the observed measure.

When the parallel model is assumed, the reliability of the sum of k variables equals

$$\tilde{\rho} = k\rho/[k\rho + (1 - \rho)]$$

This formula is known as the Spearman-Brown formula. Its maximum-likelihood estimator, under the assumption of a normal distribution of the error and the parallel model, is known as the Cronbach α coefficient (CAC)[15]:

$$\alpha = [k/(k-1)]\left[1 - \left(\sum_{j=1}^{n} S_j^2/S_{tot}^2\right)\right],$$

where

$$S_i^2 = 1/(n-1)\sum_{i=1}^{n}(X_{ij} - \overline{X}_j)^2$$

and

$$S_{tot}^2 = 1/(nk-1)\sum_{i=1}^{n}\sum_{i=1}^{k}(X_{ij} - \overline{X})^2.$$

It is easy to show a direct connection between CAC and the percentage of variance of the first component in PCA, which in factor analysis is often used to assess unidimensionality.[16] The PCA is usually based on an analysis of the latent roots of the correlation matrix of k variables R, which, under the parallel model, looks as follows:

$$R = \begin{bmatrix} 1 & \rho & \ldots & \rho \\ \rho & 1 & \rho & \ldots \\ \ldots & \rho & 1 & \rho \\ \rho & \ldots & \rho & 1 \end{bmatrix}.$$

This matrix has only two different latent roots. The greater root is: $\lambda_1 = (k-1)\rho + 1$, and the other multiple roots are $\lambda_2 = \lambda_3 = \lambda_4 = \cdots = 1 - \rho = (k - \lambda_1)/(k-1)$. Thus, using the Spearman-Brown formula, we can express the reliability of the sum of variables as:

$$\tilde{\rho} = k/(k-1)[1 - (1/\lambda_1)].$$

This clearly indicates a monotonic relationship that is estimated by α (CAC) and the first latent root λ_1, which in practice is estimated by the corresponding value of the observed correlation matrix and thus the percentage of variance of the first principal component in a PCA. So, CAC is also considered as a measure of unidimensionality.

The Spearman-Brown formula indicates a simple relationship between CAC and the number of variables. It is easy to show that the CAC is an increasing function of the number of variables. This formula is obtained under the parallel model.

A step-by-step curve of CAC can be built to assess the unidimensionality of a set of variables.[11,16] The first step uses all variables to compute CAC. Then, at every successive step, one variable is removed from the scale. The removed variable is that which leaves the scale with its maximum CAC value. This procedure is repeated until only two variables remain. If the parallel model is true, increasing the number of variables increases the reliability of the total score, which is estimated by Cronbach α. Thus, a decrease of such a curve after adding a variable would cause us to suspect strongly that the added variable did not constitute a unidimensional set with the other variables.

References

1. Grodum K, Heijl A, Bengtsson B. A comparison of glaucoma patients identified through mass screening and in routine clinical practice. *Acta Ophthalmol Scand.* 2002;80:627–631.
2. Katz J, Tielsch JM, Quigley HA, Javitt J, Witt K, Sommer A. Automated supra-threshold screening for glaucoma: the Baltimore Eye Survey. *Invest Ophthalmol Vis Sci.* 1993;34:3271–3277.
3. Iwasaki A, Sugita M, Tatemichi M. Performance of glaucoma mass screening with only a visual field test using frequency

doubling technology perimetry. *Am J Ophthalmology*. 2002; 134:529–537.
4. Quigley HA. Current and future approaches to glaucoma screening. *J Glaucoma*. 1998;7:210–220.
5. Anderson DR, Patella VM. *Automated Static Perimetry.* 2nd ed. St. Louis: Mosby; 1999.
6. Brigatti L, Hoffman D, Caprioli J. Neural networks to identify glaucoma with structural and functional measurements. *Am J Ophthalmol*. 1996;121:511–521.
7. Brigatti L, Nouri Mahdavi K, Weitzman M, Caprioli J. Automatic detection of glaucomatous visual field progression with neural networks. *Arch Ophthalmol*. 1997;115:725–728.
8. Mandava S, Zulauf M, Zeyen T, Caprioli J. An evaluation of clusters in the glaucomatous visual field. *Am J Ophthalmol*. 1993;116:684–691.
9. Nunnaly J. *Psychometric Theory.* 2nd ed. New York: McGraw-Hill; 1978.
10. Hamon A, Mesbah M. Questionnaire reliability under the Rasch model. *Statistical Methods for Quality of Life Studies.* Amsterdam: Kluwer, 2002:155–168.
11. Curt F, Mesbah M, Lellouch J, Dellatolas G. Handedness scale: how many and which items? *Laterality*. 1997;2:137–154.
12. Chassany O, Sagnier P, Marquis P, Fullerton S, Aaronson N. for the European Regulatory issues on quality of life assessment group. Patient-reported outcomes: the example of health-related quality of life—a European guidance document for the improved integration of health-related quality of life assessment in the drug regulatory process. *Drug Inform J*. 2002;36: 209–238.
13. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297–334.
14. The Advanced Glaucoma Intervention Study Investigators. Advanced Glaucoma Intervention Study. 2. Visual field test scoring and reliability. *Ophthalmology*. 1994; 101:1445–1455.
15. Kristof W. The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. *Psychometrika.* 1963;28:221–238.
16. Moret L, Mesbah M, Chwalow J, Lellouch J. Validation interne d'une échelle de mesure: relation entre analyse en composantes principales, coefficient alpha de Cronbach et coefficient de corrélation intra-classe. [Internal validation of a measurement scale: relation between principal component analysis, Cronbach's alpha coefficient and intra-class correlation coefficient]. *Rev Epidemiol Sante Publique.* 1993;41:179–186.