Misleading Statistical Calculations in Far-advanced Glaucomatous Visual Field Loss

Eytan Z. Blumenthal, MD, Ruthy Sapir-Pichhadze, MD

Objective: In this study, the capability of statistical analysis indices to characterize static automated visual fields (VFs) accurately in cases of far-advanced glaucoma was assessed.

Design: Retrospective observational case series.

Participants: Sixteen eyes of 15 patients with end-stage glaucoma and evidence of collapse of VF statistical analysis indices were included in the study.

Methods: End-stage glaucoma was defined as vertical cup-to-disc ratio of 0.9 or more, mean deviation less than −24 dB and with only a central or temporal island remaining in the VF gray scale. Collapse of statistical indices was defined as any of the following: pattern deviation probability plot without a single VF location showing $P < 0.5\%$; corrected pattern standard deviation (CPSD) and pattern standard deviation (PSD) probability less than 5% or within normal limits (WNL); short-term fluctuation (SF) probability WNL; glaucoma hemifield test (GHT) not outside normal limits (ONL); or presence of a low patient reliability comment triggered by 40% or more false-negative (FN) responses.

Main Outcome Measures: Visual field statistical indices.

Results: Of the 16 VFs showing misleading statistical calculations, 9 of 16 eyes had a normal pattern deviation probability plot. The PSD, SF, and CPSD parameters were normal or barely outside the normal range in 4 of 16, 10 of 16, and 5 of 16 eyes, respectively. The GHT was ONL in 7 of 13 eyes, borderline with generalized reduction of sensitivity (GRS) in three eyes, and only GRS in two additional eyes. Low patient reliability was triggered because of an FN score of 40% or more in 10 of 16 eyes.

Conclusions: Statistical indices are crucial for the interpretation of automated static VFs. However, in end-stage glaucomatous VF loss, both summary statistical indices and reliability indices may not detect abnormality, thus misleading the casual observer. *Ophthalmology 2003;110:196–200 © 2003 by the American Academy of Ophthalmology.*

Visual fields have undergone profound changes since the tangent screen and the arc perimeter. Automated static perimetry, the current gold standard, largely has replaced kinetic manual perimetry (i.e., the Goldmann perimeter).[1] The Humphrey Field Analyzer, combined with the STATPAC statistical analysis package (Humphrey-Zeiss, Dublin, CA), provides highly reproducible raw threshold values together with an in-depth complex statistical analysis that relies on a large normative database. Furthermore, STATPAC has the capacity to assess and quantify various characteristics unique to glaucomatous visual field (VF) loss.[1]

The single-field statistical analysis relies on several features typical of glaucomatous VF loss: (1) the focal nature of the disease, (2) typical patterns of loss (nasal defects, etc.), (3) a comparison of thresholds above and below the horizontal raphe, and (4) clustering in patterns consistent with retinal nerve fiber layer distribution.[2–4] The characteristics underlying the statistical analysis, though valid throughout most of the spectrum of glaucomatous VF loss, may fail to reflect accurately the VF loss of far-advanced glaucoma.

Statistically derived VF indices include:

1. Reliability indices, including fixation losses (FL), false-positive (FP) responses, and false-negative (FN) responses.
2. Statistical summary indices, including the glaucoma hemifield test (GHT), pattern standard deviation (PSD), short-term fluctuations (SF), corrected pattern standard deviation (CPSD), and mean deviation (MD). Attention should be drawn to the fact that MD is usually considered a "nonglaucomatous" parameter, flagging diffuse factors other than glaucoma that affect the VF (such as media opacity [cataract], a miotic pupil, or erroneous refraction).[5]
3. Location-specific analysis maps. Total deviation (presented both as numerical data and as a probability map) represents the age-corrected deviation in each VF location in comparison with a normative database. Because total deviation is analogous to the raw data (age correction matters little in far-advanced fields) and does not collapse in far-advanced glaucoma, it

Originally received: February 25, 2002.
Accepted: May 8, 2002. Manuscript no. 220145

From the Department of Ophthalmology, Hadassah University Hospital, and Hebrew University-Hadassah Medical School, Jerusalem, Israel.

The authors have no financial interest in the perimetry technology presented.

Correspondence to Eytan Z. Blumenthal, MD, Department of Ophthalmology, Hadassah University Hospital, P.O. Box 12000, Jerusalem 91120, Israel. E-mail: eblumenthal@md.huji.ac.il.

ISSN 0161-6420/03/$–see front matter
PII S0161-6420(02)01297-6

Table 1. Summary of Statistical Indices for Sixteen Visual Fields

Patient No.	FL	FP	FN	LPR	GHT	MD (dB)	PSD (dB)	PSD %	SF (dB)	SF %	CPSD (dB)	CPSD %	PD #
1(RE)	1/12	0/9	4/4	LPR	*	−28.71	5.31	<0.5%	1.27	WNL	5.11	<0.5%	0
1 (LE)	0/13	0/7	0/0	—	*	−27.68	8.71	<0.5%	0.00	WNL	3.71	<0.5%	0
2	0/19	1/13	4/6	LPR	ONL	−27.42	7.42	<0.5%	6.59	<0.5%	2.35	<0.5%	27
3	0/13	0/3	0/0	—	B,GRS	−29.05	0.00	WNL	0.00	WNL	2.45	<5%	0
4	0/14	1/13	5/6	LPR	—	−29.15	4.36	<1%	3.13	<10%	2.80	<0.5%	13
5	0/21	0/3	3/7	LPR	ONL	−24.09	7.73	<0.5%	2.24	<0.5%	7.35	<0.5%	26
6	1/17	0/14	6/7	LPR	ONL	−26.00	8.32	<0.5%	9.50	<0.5%	0.00	WNL	36
7	1/21	0/7	4/9	LPR	ONL	−25.08	10.61	<0.5%	1.00	WNL	10.55	<0.5%	34
8	0/18	0/6	0/0	—	ONL	−27.15	3.69	<5%	3.09	<2%	1.66	WNL	4
9	0/11	1/5	0/0	—	ONL	−30.17	3.88	<5%	0.00	WNL	3.88	<0.5%	0
10	0/13	0/7	0/0	—	GRS	−29.37	6.62	<0.5%	0.00	WNL	6.62	<0.5%	0
11	0/0	0/3	0/0	—	GRS	−29.75	5.36	<1%	0.00	WNL	5.36	<0.5%	0
12	2/18	0/13	4/5	LPR	*	−28.67	3.26	WNL	0.00	WNL	3.28	<2%	0
13	0/15	0/13	3/6	LPR	B,GRS	−28.81	7.50	<0.5%	1.94	WNL	7.17	<0.5%	0
14	1/14	0/5	2/5	LPR	ONL	−30.77	4.05	<2%	4.47	<1%	0.00	WNL	5
15	0/13	0/5	2/4	LPR	B,GRS	−27.34	8.26	<0.5%	1.42	WNL	8.12	<0.5%	0
"NLP VF"	0/11	0/2	0/0	—	GRS	−31.43	1.81	WNL	0.00	WNL	1.81	WNL	0

B = borderline; CPSD = corrected pattern standard deviation; FL = fixation losses; FN = false negative; FP = false positive; GHT = glaucoma hemifield test; GRS = general reduction of sensitivity; LPR = low patient reliability; MD = mean deviation; NLP VF = no light perception visual field; ONL = outside normal limits; PD # = number of black (<0.5%) locations in the pattern deviation probability map; PSD = pattern standard deviation; SF = short-term fluctuations; WNL = probability flag not raised (i.e. PSD ≥ 5%; SF ≥ 10%; CPSD ≥ 5%).
*No GHT analysis is available for FASTPAC.

will not be addressed in this manuscript. Pattern deviation (presented both as numerical data and as a probability map) represents the deviation in each VF location as compared with a normative database, after correcting for age and for the uniform depression component (MD) of the VF.

Material and Methods

Patients

Sixteen eyes (eight of them right eyes) of 15 patients with far-advanced glaucoma were identified, in which evidence of collapse of VF statistical analysis indices was present. Far-advanced glaucoma was defined as having a vertical cup-to-disc ratio of 0.9 or more, MD worse than −24 dB, having only a central or temporal island remaining in the VF gray scale, and clear evidence of glaucoma in the fellow eye (manifest as glaucomatous appearing discs and a glaucomatous VF). One patient (patient 9), however, manifested with only unilateral glaucomatous damage. Visual acuity ranged from 20/20 to finger counting. Nine of the 16 eyes had a visual acuity of 20/80 or better.

Collapse of statistical indices was defined as one of the following: (1) pattern deviation probability plot lacking black squares (not a single location showing $P < 0.5\%$); (2) CPSD probability less than 5% or within normal limits; (3) PSD probability less than 5% or within normal limits; (4) SF probability within normal limits; or (5) FN responses 40% or more, flagging the low patient reliability (LPR) comment in the absence of other abnormalities of reliability criteria, such as high FL or FP responses.

Visual Fields

All VFs in this study were obtained using the stimulus III, 24–2 program of the Humphrey Field Analyzer. All were tested using the full-threshold algorithm, except for patients 1 and 12, who were tested with the FASTPAC thresholding algorithm (for which GHT analysis is not available). Initially, VFs with statistical indices collapse (based on the above criteria) were identified. Then, each patient's chart was screened to verify that he complied with the far-advanced glaucoma criteria listed above. The total number of questions asked in each VF examination was (mean ± standard deviation): 235.6 ± 86.7. Each test lasted (mean ± standard deviation) 7.9 ± 2.4 minutes.

No Light Perception Visual Field

To challenge the Humphrey Field Analyzer STATPAC statistical analysis package under the most far-advanced VF defect, we ran the VF test without having an examinee present, so that the button was never pressed throughout the examination. This allowed us to create the worst possible VF that a patient with glaucoma may have at presentation, and we called it a *no light perception VF* (NLP VF). This was defined as a VF test in which absolutely no threshold values were detected. The Humphrey Field Analyzer records such thresholds as 'less than 0,' denoting that the subject did not respond even to the strongest possible stimulus.

One may wonder about the purpose of this exercise. It appears that a total collapse of most of the statistical indices occurs in an NLP VF. A similar collapse was observed, though to a somewhat lesser extent, in the genuine far-advanced glaucomatous VFs presented in this study.

Results

Table 1 presents the summary statistical indices for each of the 16 VFs included in this study, as well as the NLP VF (bottom row).

Figure 1 presents the raw threshold values, gray-scale map, and pattern deviation probability maps of two representative VFs, as well as for the NLP VF.

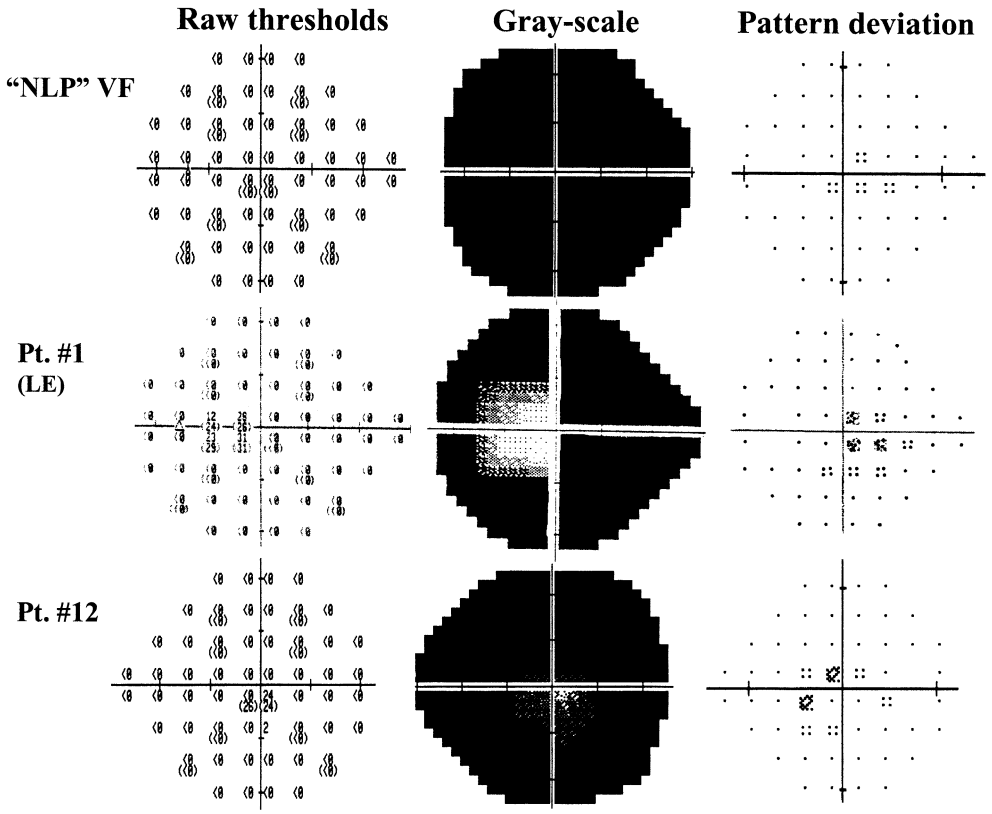

Figure 1. Raw threshold values, gray-scale representation, and pattern deviation probability maps of the no light perception visual fields in patients 1 and 12.

Reliability Indices

In 10 of 16 eyes, LPR was declared solely as a result of an FN score of 40% or more. This was in contrast to FL and FP responses that were present rarely (see Table 1). We consider high FN scores (often as high as more than 50%) to be a manifestation of advanced glaucomatous damage rather than lack of patient cooperation, as discussed below.

Statistical Summary Indices

The GHT was borderline with generalized reduction of sensitivity (GRS) in three eyes and GRS in two additional eyes. Only in 7 of 13 eyes did the GHT actually state 'outside normal limits.' A highly negative MD score was documented in all the eyes. For all eyes, with no exception, the MD probability was less than 0.5%. Conversely, PSD, SF, and CPSD were at the point of near normality (or barely outside the normal range) in 4 of 16 eyes, 10 of 16, and 5 of 16 eyes, respectively, as often happens with the progression of the VF in far-advanced glaucoma.

Pattern Deviation Probability Maps

In 9 of 16 eyes, the pattern deviation probability plot was normal, where normal was defined as not a single location showing $P <$ 0.5%.

Conclusions

This study supports the hypothesis that VF statistical indices, as calculated by the STATPAC software, may become misleading (collapse) and fail to reflect the true nature of the disease in far-advanced glaucomatous VF loss. We found considerable similarities between the STATPAC statistical analysis of far-advanced VFs with a remaining small central island and our hypothetical NLP VF.

Reliability Indices

Fixation losses, using the Heijl-Krakau method,[1] which are tested by projecting a stimulus on the presumed location of the blind spot, are detected if there is unsteady fixation during the test or if the blind spot was mapped incorrectly. According to Katz et al,[6] the main reason for unreliable tests in normal and ocular hypertensive patients is the result of fixation losses. It is our opinion that an important setting in which reliability indices fail to reflect reliability is far-advanced glaucomatous VF loss.

In far-advanced glaucoma, as clearly evident in Figure 2 (top), the "blind-spot" may grow to encompass almost the entire tested VF. For such a patient to generate an FL error, the projected stimulus need not only escape the true blind-spot (optic disc) location, but also accurately aim at the small remaining island of sight in the VF. The likelihood of this happening in eyes with far-advanced VF loss, such as those shown in Figure 2 (top), is extremely low.

To highlight further the assumption that a low FL score does not always represent steady fixation, a VF of a 37-year-old patient with marked nystagmus and a far-advanced VF damage from cicatricial retinopathy of prematurity is shown in Figure 2 (bottom). This VF of a patient with visual acuity of 20/400 in the better (left) eye produced reliability

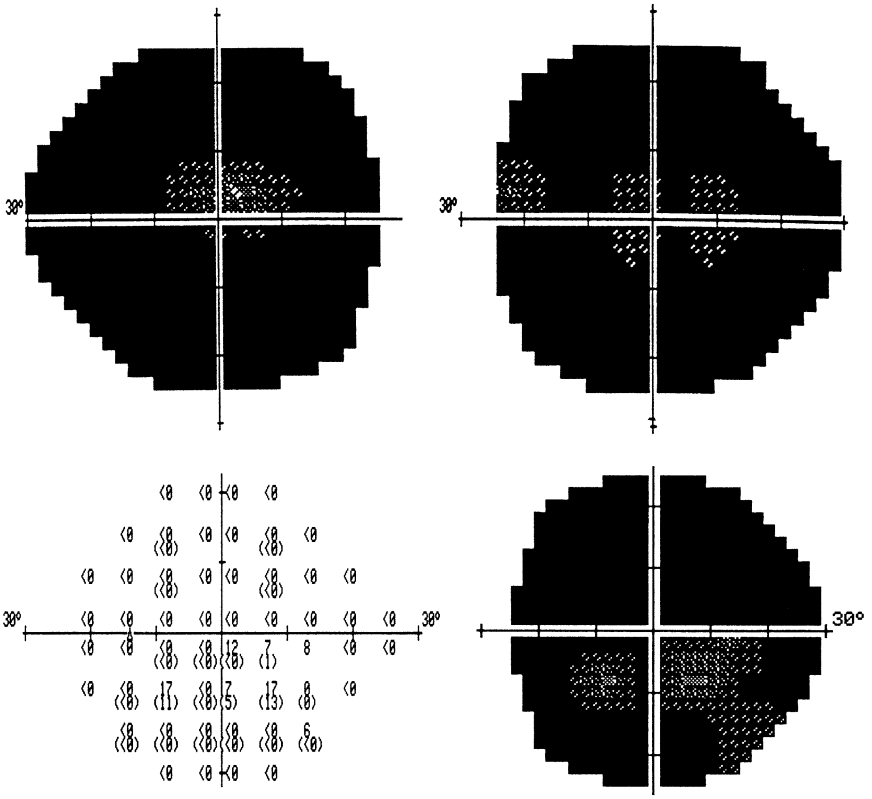

Figure 2. Top left, The gray-scale representation of patient 9. Fixation losses (FLs) for this visual field were 0/11. Regardless of true fixation, it is unlikely that this patient would respond to any of the FL stimuli. **Top right,** The gray-scale representation of patient 3. Fixation losses for this visual field were 0/13. **Bottom,** A stimulus-sized V visual field of a patient with marked nystagmus and cicatricial retinopathy of prematurity. Reliability parameters were: FL, 0/14; false positive (FP), 0/9; false negative (FN), 5/5.

parameters of: FL, 0/14; FP, 0/9; FN, 5/5. This VF, which is of questionable clinical value owing to the marked nystagmus, was ordered as part of a routine disability evaluation. Clearly, the perfect FL score (0/14) obtained failed to identify this patient's lack of ability to fixate. In addition, we believe that the high FN was also a poor marker of reliability in this case (see below). Of interest, this same FL issue occurs also with temporal hemianopsias. Note that the Humphrey Field Analyzer-II (700 series) field analyzers are equipped with, in addition to the Heijl-Krakau method of testing for FL, a gaze monitor[1] that is not prone to the FL artifact of far-advanced VFs discussed above.

False negative responses are used as indicators of inattention and are checked by presenting the patient with a stimulus 9 dB brighter than a stimulus previously seen in the same VF location. With the progression of glaucoma, FN scores often increase dramatically (as high as more than 50%), indicating advanced glaucomatous damage, rather than a lower degree of the patient's attentiveness. High FN rates among patients with glaucoma have been reported in previous studies; Katz et al[6] found higher rates of FN responses in repeated testing of glaucoma patients in comparison with normals or with those with ocular hypertension.

Patients with severe glaucomatous VF sensitivity loss exhibit higher variability in threshold values, especially in more affected visual field locations. In such patients, a higher prevalence of FN responses is encountered even in situations of full attentiveness.[7,8] Regardless of the reason, when the number of FN answers exceeds 33%, an unreliability flag is raised automatically. In addition, it should be noted that for severely damaged VFs, FN responses often are not tested for; instead, a "0/0" notation appears, as was the case in 6 of 16 VFs (see Table 1). Note, however, that in the new SITA algorithm "a high FN rate more reliably indicates inattention rather than disease."[1]

False positive responses occur when the patient presses the button despite the fact that no stimulus is presented. These FP responses probably remain a true indication of reliability, even in far-advanced VF loss. However, FP responses are a poor indicator of reliability because they are encountered infrequently in glaucoma patients experienced with VF testing.

Summary Statistics (Global Indices)

The sole summary statistical index indicating the extent of damage in far-advanced VFs probably is the highly negative MD value. It should be noted that MD is often considered a nonglaucomatous parameter, reflecting loss unrelated to glaucoma (such as from media opacity, miotic pupil, or erroneous refraction).[5]

Pattern deviation threshold values and the accompanying probability map provide an estimation of the localized component of VF sensitivity loss (the overall component minus the diffuse component). Localized VF defects, as commonly encountered in early glaucomatous fields, result in larger PSD values. In contrast, a diffuse reduction in VF sensitivity, which is more common in advanced glaucomatous VF loss, may induce a reduction, or even reversion, of the PSD value to normal; however, a high negative MD value alerts us that the VF is grossly abnormal.[5]

Similarly, SF and CPSD both can improve as the field deteriorates, to the point of normality (or barely outside the normal range). The SF is a measure of intratest variability and is measured by comparing the threshold values obtained in 10 predefined locations twice tested during the VF test. The SF is higher in glaucoma patients and is proportional to the amount of damage.[9] However, in far-advanced VFs such as those presented in this study, SF can revert back to the normal range. In 6 of 16 VFs, SF was calculated as "0.00." Because it is improbable for any human to perform a VF at zero variability, we consider this to be an artifact related to a "floor effect," meaning that locations with a threshold of less than 0 (stimulus not seen) are very likely to reproduce identical values on repeated testing.

The GHT is a robust parameter,[2,3] collapsing only in extremely advanced cases. The GHT compares five VF clusters and their mirror reflected zones above and below the horizontal raphe. Any difference found between the paired superior and inferior zones is compared with a normative GHT database.

As soon as the VF damage encompasses the vast majority of the VF locations, the focal components are largely replaced with a generalized overall reduction in sensitivity. The GHT, being a sensitive indicator of focal damage, is no longer capable of identifying this diffuse (generalized) damage.

In conclusion, it is important to realize that in far-advanced glaucomatous VFs, most STATPAC summary statistics may fail to quantify accurately the magnitude of VF damage. Moreover, in cases of very advanced damage, summary statistical indices paradoxically may revert to normal. On casual observation, if one fails to note the raw data (and total deviation plots), while focusing only on the statistical analysis, such a field may easily be dismissed as a nonreliable field, possibly secondary to marked media opacity, rather than a result of end-stage glaucoma. Incidentally, a far-advanced glaucomatous VF demonstrating misleading statistical indices appears in Anderson and Patella's textbook *Automated Static Perimetry*,[1] page 214.

Statistical indices in the Humphrey STATPAC software are based on formulas meant for use in more moderate cases, devised to evaluate the focal nature of glaucomatous VF loss. When far-advanced VFs lose this characteristic focal component (marked defects alongside relatively conserved areas of the VF), the equations are no longer predictive of glaucomatous damage. In fact, the indices for an end-stage glaucoma patient may "improve" over time. It is important to stress that the misleading results obtained in far-advanced VFs (with the exception of the reliability indices) are completely predictable from inspection of the formulas used for their calculations.

Because cases presented in this study were only those where collapse had occurred, data shown in Table 1 demonstrate the relative frequency of these findings, addressing the question of which parameters are more robust and which have a higher tendency to collapse.

How often does the phenomenon of statistical indices collapse occur, and how advanced a VF induces it? Our study was not set to address these questions. However, the fact that 14 of these 16 VFs were collected in a single tertiary glaucoma clinic during a 1.5-year period may imply that statistical indices collapse is not that rare a phenomenon.

References

1. Anderson D, Patella VM. Automated Static Perimetry, 2nd ed. St. Louis: Mosby, 1999;121–90.
2. Asman P, Heijl A. Glaucoma Hemifield Test. Automated visual field evaluation. Arch Ophthalmol 1992;110:812–9.
3. Asman P, Heijl A. Evaluation of methods for automated Hemifield analysis in perimetry. Arch Ophthalmol 1992;110:820–6.
4. Asman P. Computer-assisted interpretation of visual fields in glaucoma. Acta Ophthalmol Suppl 1992;206:1–47.
5. Smith SD, Katz J, Quigley HA. Analysis of progressive change in automated visual fields in glaucoma. Invest Ophthalmol Vis Sci 1996;37:1419–28.
6. Katz J, Sommer A, Witt K. Reliability of visual field results over repeated testing. Ophthalmology 1991;98:70–5.
7. Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability? Invest Ophthalmol Vis Sci 2000;41:2201–4.
8. Spry PG, Bates AB, Johnson CA, Chauhan BC. Simulation of longitudinal threshold visual field data. Invest Ophthalmol Vis Sci 2000;41:2192–200.
9. Flammer J, Drance SM, Fankhauser F, Augustiny L. Differential light threshold in automated static perimetry. Factors influencing short-term fluctuation. Arch Ophthalmol 1984;102:876–9.