Comparison between Tendency-Oriented Perimetry (TOP) and Octopus Threshold Perimetry

José Morales, MD, 1 Marc L. Weitzman, MD, 2 Manuel González de la Rosa, MD 3

Objective: To compare the results obtained by a new ultra-short automated perimetry test known as Tendency-Oriented Perimetry (TOP), which is an algorithm based on estimation of thresholds from information gathered from adjacent points with those obtained by a standard bracketing approach. TOP is designed to save up to 1/5 of the time taken by standard strategy by presenting each stimulus once on each location (instead of 4 to 6 times per location with the standard technique) and reaching a final threshold estimate by gathering information from responses to adjacent locations.

Design: Prospective, multicenter, observational comparative case series.

Participants/Methods: Four academic institutions provided data from testing 57 subjects, 15 with a normal ocular exam and 42 with a variety of visual field abnormalities. A total of 228 visual fields were analyzed. Two examinations of standard thresholding testing (Octopus program 32) and two examinations of the TOP program were obtained in each subject the same day.

Main Outcome Measures: Comparison of global indices such as mean deviation (MD), square root of loss of variance (sLV), topographical defects, point by point differences, reproducibility, sensitivity/specificity, and time required to complete the test.

Results: Correlation coefficient of global indices between both tests was high, with mean deviation of $r = 0.97$ (SE[YX] ± 1.65 decibels) and square root of loss variance of $r = 0.93$ (SE[YX] ± 1.10 dB). Mean sensitivity tended to be 1 dB higher while MD tended to be 1 dB lower with TOP strategy. Reproducibility was equally good between both tests for threshold determination as well as for all global indices (MS, MD per quadrant, and LV). Cluster criteria for abnormality demonstrated TOP versus 32: sensitivity of 89/87; specificity of 90/77; positive predictive value of 96/91; negative predictive value of 75/68; and accuracy of 89/84. Mean time taken by this beta version of TOP was 4.05 minutes standard deviation ± 0.55 versus the standard 32 version taking 14.65 minutes standard deviation ± 3.75.

Conclusions: TOP was four times faster than the traditional full-threshold technique and was successful in detecting visual field abnormalities. Defects with TOP tended to be smaller, shallower, and with softer edges than with standard approach. TOP could prove an alternative to traditional perimetric techniques. Ophthalmology 2000;107:134–142 © 2000 by the American Academy of Ophthalmology.
Materials and Methods

TOP strategy tests each point only once, but each point is affected by the responses of the surrounding points to reach the final threshold approximation. In other words, instead of questioning every location four to six times as usually occurs in standard thresholding techniques, every location is adjusted four times—one by a direct question and three times by the results from the questions in neighboring locations. TOP program uses a mathematical algorithm to investigate the threshold through consecutive approximation, examining four intermingled grids (Fig 1). These matrices 1, 2, 3, and 4, are examined in sequential order while the individual thresholds of the other three grids are adjusted continuously to approximate further new values by linear interpolation. Within each matrix, the tested points are examined in random order. As a starting point, the program assumes that the patient’s threshold is equal to half of the normal value. The start value in fractions that would be employed in each sub-matrix to approximate the threshold in consecutive smaller steps for each sub-matrix. Note that the starting point of questioning is 1/2 of the expected normal threshold (8/16 dB).

Figure 1. Definition of TOP matrixes. The TOP program investigates the threshold through consecutive approximation, examining four intermingled grids. Each color indicates the assignment of each of the 76 points tested in a 32 matrix to four corresponding sub-matrixes. Notice how each sub-matrix point is in direct contact with the other three sub-matrixes. Each point response will determine its own threshold and that of the neighboring points.

dB bracketing strategy. The TOP program in this study tested the same 76 points used by the 32 program.

Fifty-seven patients were enrolled in four different academic centers (Halifax, Canada; Yale, Connecticut; Lubbock, Texas; and Tenerife, Spain). All patients were reviewed by the appropriate ethics committee; the tests were therefore performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. All patients gave their informed consent prior to their inclusion in the study. Inclusion criteria for patients were: age between 20 and 70 years, previous experience taking automated perimetry testing, normal eye exam (15 cases) or ocular pathology classified as early glaucoma (12 cases), advanced glaucoma (10 cases), neuro-ophthalmological (11 cases) or retinal disease producing visual field abnormalities (9 cases), and a complete eye examination by an ophthalmologist. Exclusion criteria for patients were: poor reliability on automated perimeter testing, visual acuity worse than 20/40 (0.5), and presence of multiple ocular pathology.

The examinations were performed with an Octopus 1-2-3 Perimeter (Interzeag, AC) connected to and controlled by an external personal computer (PC). The examination software TOP Version 1.01 was implemented as an application under Microsoft Windows. The “Remote Control” feature of the Octopus 1-2-3 was activated to run perimetric tests from the PC while functions such as automatic fixation control remained active in the same way as using the standard Octopus 1-2-3 software.

One eye of each patient was tested twice with the standard 32 program and twice with the TOP program. A period of 30 minutes between experimental and standard tests was allowed; 60 minutes were allowed for resting between the two sets of exams. The order of testing could be interchangeable, starting with either the TOP or the 32 program, but the alternation between the two tests was done in a way that the two possible sequences were TOP-32-TOP-32 or 32-TOP-32-TOP; the order was reverted depending on which sequence was chosen to start at random.

Data from each test were converted into PeriTrend Octopus database format and ASCII format for analysis of all the testing locations’ results. Analysis of data concentrated on comparison of results regarding global indices—mean deviation (MD), MD per quadrant, and loss of variance (LV); correlation; reproducibility; sensitivity/specificity indices; topographical diagnosis suggested by the defect; and correlation with clinical diagnosis. The statistical analysis of the results between both groups’ indices was performed using Student’s t-test for paired data.

The sensitivity/specificity comparison, which included objective as well as subjective criteria for abnormality, was done in three different ways.
1. A subjective assessment of the 228 fields was made by two clinicians (M.W. and J.M.) where information was masked regarding the program used and any clinical information. The evaluators were asked to divide the fields into normal and abnormal.

2. A strict numeric cut-off point between normality and abnormality was established to differentiate normal and abnormal by using the recommended parameters from the manufacturing company of the perimeter in question to suspect abnormality. The parameters consist of global indices (MD > 2 dB and/or LV > 6 dB). 

3. An adapted cluster criteria for abnormality was based on the presence of at least seven points with sensitivity decreased 5 or more dB, three of them being contiguous. These criteria were based on those used in a recent article dealing with perimetric screening in glaucoma but were modified to include other pathologies by not requiring the three adjacent points to be in the arcuate region.

Results

Because one eye of each patient was tested twice (first test called A and the second B) with each program, a total of 228 visual fields were available for comparison: 114 pairs each of 32 standard bracketing and TOP/32 testing.

Comparison of Global Indices

The mean sensitivity was significantly higher with TOP (20.5 dB) than with 32 (19.45 dB) (P < 0.001). The correlation coefficient of MD between both tests was 0.97 (SE[YX] ± 1.65 dB). The mean deviation was lower with TOP (6.31 dB) than with 32 (7.36 dB) (P < 0.001). The correlation coefficient of the mean deviation in each quadrant (superonasal, inferonasal, superotemporal, and inferotemporal) between both tests was 0.97, 0.97, 0.94, and 0.95, respectively (SE[YX] ± 2.13, 2.20, 2.42, 2.22 dB, respectively). The correlation coefficient of square root of loss variance (sLV) between both tests was 0.93 (SE[YX] ± 1.10 dB). The value of sLV was lower with TOP (4.76 dB) than with 32 (5.53 dB), with a difference of 0.77 dB (P < 0.001). The correlation coefficient for MD in each subgroup was as follows: normals, 0.74 (SE[YX] ± 1.15); glaucoma, 0.94 (SE[YX] ± 2.05); neuro-ophthalmology, 0.96 (SE[YX] ± 1.71); and retinal disease, 0.98 (SE[YX] ± 1.69).

Reproducibility

Reproducibility was comparable between both strategies (Table 1). Although there were small differences—such as TOP demonstrating slightly less fluctuation for threshold sensitivities and sLV values, whereas 32 had slightly less fluctuation for MD values—no statistically significant difference was noted.

Point by Point Analysis

If we label the first time that the examination was taken A and the second one B, there was more dispersion in the histogram comparing 32_A vs 32_B than in the histogram comparing TOP_A versus TOP_B (Fig 3); however, this difference did not reach statistical significance (P = 0.26).

Histogram analysis of TOP versus 32, whether comparing TOP_A versus 32_A or TOP_B versus 32_B, demonstrated a histogram centered around 0 dB but with a dispersion skewed to the right that was caused by the higher values of threshold sensitivity obtained by TOP (Fig 4). Differences between thresholds reached high statistical significance (P < 0.001).

Time Comparison

Test duration with this beta version of TOP was 4.05 minutes SD ± 0.55, versus the 32 program which took 14.65 minutes SD ± 3.75.

Sensitivity/Specificity and Associated Indices

Analysis by subjective assessment of two masked clinicians, global indices (MD > 2 dB and/or LV > 6 dB), and adjacent points criteria (greater than seven abnormal points with sensitivity higher than 5 dB and at least three contiguous points) resulted in the data presented in Table 2.

Clinical Assessment

Evaluation of the pairs of visual fields seemed to demonstrate good qualitative and pattern similarities in defects between both tests as evaluated by the different clinicians who participated in this study (Fig 5). Participants also observed a tendency for specific localized scotomas to be wider but shallower with the TOP strategy than with the 32 technique (Fig 6).

Finally, scotomas with very abrupt edges, especially those from a neurological etiology of retrochiasmatic origin causing hemianopic defects, seemed to show less distinct edges in the grayscale printout and mildly receding edges from the midline. These scotomas were still identifiable as hemianopic defects in the cases involved (Fig 7).

Discussion

The automated visual field examination programs that are used most frequently (program 32 by Octopus and 30-2 by Humphrey) include 76 points tested with the normal strategy (Octopus making 4-2-1 dB steps and Humphrey 4-2 dB steps, both with two reversals), but each of these strategies can take as long as 12 to 20 minutes per eye. These methods require a long examination time to obtain such limited information because traditional perimetric strategies rely on a standard staircase approach to assess the threshold of each individual point; they require four to six stimuli per position.

<p>| Table 1. Fluctuation Comparison of the Main Perimetric Indices of the TOP and 32 Strategies |
|--------------------------------------|------|---------------|-----------------|
| Thresholds fluctuation              | TOP  | 1.91 dB       | (SD ± 2.28)     |
|                                     | 32   | 1.92 dB       | (SD ± 2.45)     |
| Mean deviation fluctuation          | TOP  | 0.68 dB       | (SD ± 0.50)     |
|                                     | 32   | 0.67 dB       | (SD ± 0.76)     |
| Square root of loss variance        | TOP  | 0.38 dB       | (SD ± 0.29)     |</p>
<table>
<thead>
<tr>
<th>fluctuation</th>
<th>32</th>
<th>0.42 dB</th>
<th>(SD ± 0.42)</th>
</tr>
</thead>
</table>

dB = decibels; SD = standard deviation; TOP = Tendency-oriented perimetry.
to reach an approximation where the threshold is crossed at least once. The addition of stimuli to estimate false-positive and false-negative responses, as well as assessment of intratest reproducibility of the responses, adds even more time to the length of the test. This length represents an average of 400 to 600 stimuli needed to complete the assessment, which results in well-recognized fatigue and reluctance on the part of the patient to take the test.

The search for faster strategies of perimetry testing, which has been extensive in the past and has concentrated predominantly in the field of glaucoma, has included Bjerrum area testing and staircase technique modification. A reduction in the number of points tested while addressing the areas that are considered to produce the highest yield has produced strategies such as the G1 program from Octopus and its shortened versions, from Humphrey (54 points), a “reduced set of points,” and probabilistic deduction.

Fastpac has been one of the most popular “shortened” modalities to reduce the amount of time required to obtain

![Figure 3](image1.png)

**Figure 3.** Graph of histograms of $32_A$ versus $32_B$ and $TOP_A$ versus $TOP_B$. Histogram with pointwise differences comparing results from the first test with those from applying the same test a second time with each strategy.

db = decibels; TOP = Tendency-Oriented Perimetry.

![Figure 4](image2.png)

**Figure 4.** Graphs of histograms of $TOP_A$ versus $32_A$ and $TOP_B$ versus $32_B$. Histogram with pointwise differences comparing results from the first test with those from applying the same test a second time with each strategy.
Table 2. Analysis of Results between Both TOP and 32 Techniques

<table>
<thead>
<tr>
<th></th>
<th>TOP 32</th>
<th>TOP 32</th>
<th>TOP 32</th>
<th>TOP 32</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjective</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>92</td>
<td>92</td>
<td>94</td>
<td>95</td>
</tr>
<tr>
<td>Specificity</td>
<td>85</td>
<td>77</td>
<td>87</td>
<td>57</td>
</tr>
<tr>
<td>Positive pred.</td>
<td>95</td>
<td>92</td>
<td>95</td>
<td>86</td>
</tr>
<tr>
<td>Negative pred.</td>
<td>79</td>
<td>80</td>
<td>84</td>
<td>81</td>
</tr>
<tr>
<td>Accuracy</td>
<td>90</td>
<td>88</td>
<td>92</td>
<td>85</td>
</tr>
<tr>
<td>MD &gt; 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal points</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LV &gt; 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sLV &gt; 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Prevalence of disease in sample was 73.7%.  
- Sensitivity is percentage of correctly detected pathological cases.  
- Specificity is percentage of normal cases correctly identified.  
- Positive predictive value is, among patients with abnormal results, the proportion of subjects where pathology was correctly predicted. Negative predictive value, among patients with normal results, is the proportion of subjects where absence of disease was correctly predicted.  
- Accuracy is the overall rate of agreement between the diagnostic test and the gold standard.

In our study, mean sensitivity values were found to be approximately and consistently 1 dB higher with TOP strategy than with the traditional approach. Conversely, MD values were also approximately 1 dB higher with TOP than with the 32 strategy. We believe this disparity is based on the effect of the fatigue phenomenon. The fatigue phenomenon consists of deterioration of the responses proportional to the length of the test, and it has been well documented by several authors.3–9 A recent study5 has calculated the fatigue effect at a rate of 0.08 to 0.1 dB decrease of mean sensitivity (MS) and increase of MD per minute of testing. This calculated rate correlates well with the difference between strategies observed here, considering that the test with TOP reduces testing time 10 to 12 minutes from the typical testing time required by the 32 strategy.

A different rate of increase in MS values and decrease in MD or mean defect values has been observed with different shorter programs including Delphi,34,35 Fastpac,29–33 and SITA,36 but the tendency seems to be in the same direction for the differences in MS and MD values. This measured change in values also supports the hypothesis that decreased testing time will lead to improvement in these indices. Although there is a possibility that these indices could be underestimated with the shorter strategies, it is important to consider that the longer tests may be overestimating the real defects by producing a deterioration of the responses associated with the fatigue factor.

Even though the MS and MD values differed between the two strategies examined in this study, an excellent correlation level (0.97) was observed between those values obtained by TOP and the 32 strategy. The TOP values consistently corresponded closely to their counterparts obtained by the traditional technique, which supports the idea that TOP seems to be estimating the thresholds accurately but at a different level. This high correlation of results remained when each quadrant was analyzed individually.

The correlation for sLV was not so high (0.93) but was still very acceptable, very close to the 0.96 level demonstrated by 32 versus 32. The sLV values also were somewhat lower for TOP than for the longer 32 strategy.

When the individual subgroups of the study were analyzed for correlation of MD values, the only group that showed a low correlation was the subgroup of normals with a correlation coefficient of 0.72. Conversely, the standard error of the regression YX (SE[YX]) was the lowest in this group (0.77 dB). The value obtained for correlation coefficient, therefore, seems to be a consequence of minimal dispersion from the sample measurements and does not reflect a lack of concordance between the results of both tests.

Reproducibility was as good for the TOP short strategy as it was for the traditional 32 technique. This correlation is evidenced by the lack of statistical differences when the threshold values, MD, or sLV indices were compared between each test taken the first versus the second time. Fluctuation regarding these indices did not show significant differences, either. This distinction is important because good reproducibility is an essential factor in the ability to measure progression of visual field damage. For this reason, any newer, faster strategy must be assessed for its level of reproducibility.

Pointwise differences between TOP and 32 showed a mild shift to the right of the histogram, representing a normal distribution of the differences and demonstrating again the fact that the values obtained with TOP tend to be higher than those obtained with the 32 program.

TOP obtained the threshold approximation of the visual field in about 28% of the time required by the 32 strategy. Duration of the test was constant, regardless of whether it was testing normal or pathological fields. This TOP program was a beta version of the TOP algorithm and did not use the adaptation mechanism (usually available in the regular Octopus perimeter programs) for the frequency of the stimuli presentation according to the response of the patient. Because of the TOP version used in this study, even though the reduction in the number of stimuli required was 80%, the decrease in time was 72%.

Studies underway have demonstrated a test duration for the TOP method of approximately 2.5 to 3 minutes (Interzeag, unpublished data, 1997). Such reduction of time has been obtained by decreasing the interval between stimuli presentation, which was fixed in the strategy used for this study.

Sensitivity was comparable between both strategies, but specificity was clearly superior with the TOP algorithm than with the traditional 32 technique. These data indicate an enhanced reliability for TOP to designate that a normal field...
is indeed normal. It is generally known that, with perimetry, it is more likely that a normal field actually will be normal than an abnormal field result necessarily will be abnormal. This disparity with perimetry results from the low specificity of the traditional strategies, which tend to designate that a significant number of normal fields are either abnormal or borderline. This inaccuracy is the reason that repeated testing is typically needed to determine abnormality in some patients. In contrast, the reduced testing time of TOP makes it possible to obtain more reliable results in the case of normal individuals.

On the other hand, short strategies may be more sensitive to the “learning effect” than are traditional strategies. The general rule that a first perimetric test should not be accepted as absolute, especially if it is abnormal, still applies to short perimetric strategies. This point will need to be addressed in future work.

The difficulty with determining objective sensitivity and specificity measures in perimetry as well as glaucoma stems from the lack of a well-defined cutoff point between normality and abnormality. Attempts have been made to define abnormality based on different criteria, including global indices and abnormal cluster points. In reality, the clinician also relies on the rest of the clinical information to determine level of suspicion for abnormality versus normality of a visual field. It needs to be recognized that any cutoff point will necessarily be somewhat arbitrary. We tried to minimize the effect of this arbitrary division between normal and abnormal by employing both objective and subjective analysis.

Although sensitivity and specificity are important, they are considered “stable properties,” meaning that they do not change when different proportions of diseased and well patients are tested. Positive and negative predictive values are considered “unstable” and are also important to consider because they can change markedly when the prevalence of illness changes. The higher positive predictive value encountered with the TOP strategy offers distinct benefits.
because, when evaluating a patient’s disease with a new test, we don’t know what the standard (whole clinical assessment by a clinician) will show, and we want to know how well its results will predict the results of applying the gold standard. Negative predictive value of TOP was equivalent to that obtained by the 32 program, demonstrating the ability of TOP to predict the absence of pathology correctly. Accuracy reflects the overall rate of agreement between the diagnostic test in question and the clinical determination. Accuracy was higher for the TOP strategy than for 32.

Accuracy reflects the overall rate of agreement between the diagnostic test in question and the clinical determination. Accuracy was higher for the TOP strategy than for 32.

Clinical subjective assessment corroborated what was evident with the analysis of the different global indices, that TOP strategy tends to obtain fields with less pathological results than those obtained in the same patients by the 32 strategy. This result appeared to be true for extent and depth of scotomas and generalized damage. The question remains of whether the TOP strategy is underestimating defects and visual damage or if the longer 32 strategy is overestimating them.

A particularity of the TOP strategy when compared with the 32 clinical results is that those scotomas with sharp and abrupt edges tend to show smoother edges. It was noted in particular that those scotomas from hemianopic defects caused by retrochiasmatic lesions tend to recede slightly from the midline in the numeric scale and to advance over the midline in the grayscale. This observation seems to be a direct result of the strategy approach of TOP, which tends to combine more adjacent values due to the influence of surrounding points over each tested position. This effect did not seem to impact the capability of the clinician to identify the clinical pattern in question. This particularity has been alluded to in recent work addressing this issue.

The TOP algorithm is the fastest strategy reported in the current literature. It is capable of obtaining a full estimate of the visual field threshold in the 76 points commonly tested in glaucoma and in different pathological conditions of the visual field. It has the capability to require only 1/5 of the time taken by traditional strategies. This characteristic of TOP represents a tremendous advantage in time saving for the patient, the ancillary health personnel, and the practitio-
ner, offering a significant decrease in fatigue and possible improvement in the quality of the data obtained.

Further testing of larger samples of the general population to establish a new database for this type of test and larger samples of individual pathologies will increase our ability to draw conclusions regarding the test results that are obtained from these much faster perimetry strategies. Longitudinal studies testing the ability of the TOP program to detect visual field changes would also be valuable. Comparative studies with the Humphrey SITA program would provide useful information as SITA strategy is also aimed at reducing testing time while providing an estimation of the visual field threshold sensitivities.

Acknowledgments. Dr. Balwantray Chauhan (University of Dalhausie, Halifax, Canada) contributed data from 14 patients to this study. The authors also acknowledge Dr. David McCartney (Chairman, Department of Ophthalmology and Visual Sciences, Texas Tech University Health Sciences Center) Lubbock, Texas for providing useful suggestions regarding the manuscript and to Lynnetta Hibdon and Sherri Brouillette for assisting with the preparation of the manuscript.

References

8. Searle AET, Wild JM, Shaw DE, O’Neill EC. Time-related


