CLINICAL SCIENCE

COMPARISON OF DIFFERENT ANALYTIC ALGORITHMS FOR INTERPRETATION OF THE SWEDISH INTERACTIVE THRESHOLD ALGORITHM STRATEGY

Gustavo S. Takahashi, Niro Kasahara

OBJECTIVE: To compare 4 analytic algorithms for interpretation of the Swedish Interactive Threshold Algorithm.

INTRODUCTION: Analytic algorithms were initially developed for interpretation of standard automated perimetry (using a full threshold strategy). The Swedish interactive threshold algorithm is a novel strategy that was developed to shorten test duration.

METHODS: One hundred forty-three printouts of normal and glaucomatous patients were analyzed using Caprioli's (strict, moderate and liberal) criteria and Anderson's modified criteria for perimetric defect. Areas under the receiver operator characteristics (ROC) curves, sensitivity, and specificity for each criteria were calculated.

RESULTS: Caprioli's strict and Anderson's modified criteria presented similar sensitivity (94.5% and 92.3%, respectively) and specificity (63.5% and 61.5%, respectively). Caprioli's liberal criteria were more sensitive (98.9%) and less specific (42.5%) than the other three criteria.

CONCLUSION: Both Caprioli's and Anderson's modified criteria can be used for interpretation of the Swedish interactive threshold algorithm.

INTRODUCTION

Automated perimetry is the gold standard method to measure the functional status of the optic nerve in glaucoma patients. In addition to diagnosis, automated perimetry is used to stage the severity of disease and indicate the progression of glaucomatous damage to the visual pathway. To recognize a glaucomatous visual field defect, several analytic algorithms have been suggested.[1,2] These algorithms were initially developed for a white size III stimulus on a white background, as determined by full threshold test strategies on the Humphrey perimeter.

Department of Ophthalmology, Santa Casa de São Paulo, School of Medical Sciences - São Paulo/SP, Brazil.
nirokasahara@hotmail.com

The Swedish interactive threshold algorithm (SITA) is a new automated perimetry test strategy that was developed to shorten test duration without compromising its sensitivity.[3,4] The SITA strategy is based on a forecasting procedure that employs Bayesian statistics,[5] but it has the disadvantage of not determining the short-term fluctuation (SF), an intra-test fluctuation of threshold sensitivity, hindering the use of Anderson's criteria for visual field interpretation.

A number of clinical studies have shown that the thresholds returned by SITA and the full threshold strategy are very similar in terms of test-retest variability, sensitivity and specificity.[6,7] Nevertheless, it has not been established whether analytic algorithms can also be used for interpretation of the SITA strategy.

The purpose of this study is to compare 4 different analytic algorithms for the detection of a glaucomatous visual field defect using the SITA strategy to ascertain the best criteria for SITA interpretation.

Comparison of different analytic algorithms for interpretation of the swedish interactive threshold algorithm strategy
Takahasho GS et al.

CLINICS 2008;63:333-8

MATERIALS AND METHODS

The medical records of 1478 patients from the Santa Casa Central Hospital, Glaucoma Service were reviewed. The Institutional Board Review approved the study, which conformed to the provisions of the Declaration of Helsinki in 1995 (as revised in Edinburgh 2000). The charts of consecutive patients with diagnoses of glaucoma (open angle, normal tension, pseudoexfoliative and chronic angle closure), those suspected of having glaucoma, and normal individuals with at least 3 visual field tests were selected for this retrospective study. All visual field testing were done by a trained technician. Only the charts of patients with reliable tests were included. A third test of one eye was randomly chosen for the study. Reliability was defined as less than 20% fixation loss and less than 33% false positive and false negative errors; these are the cut-off levels used in full threshold strategies to 'flag' unreliable exams. Since the SITA strategy uses a different method for the determination of false-positive and -negative errors, it is not clear whether the 33% value is appropriate. In the absence of other information, however, this may be a useful clinical cutoff.[8] Patients with refractive errors > 5 diopters, aphakic eyes and visual field defects other than glaucoma were excluded. Visual fields were assessed using the central 24 - 2 program (SITA standard strategy) with the Humphrey Field Analyzer II, model 750 (Zeiss Humphrey Systems, Dublin, CA) with the appropriate correction of refractive error. Each selected visual field was analyzed by a pair of experienced observers and classified in common agreement as glaucomatous or not by intuitive interpretation of the defect and its correlation with the optic disc aspect, as recorded in the patient's chart. The numeric and gray scale printouts, probability maps derived from total and pattern deviation plots, MD, PSD and glaucoma hemifield tests were analyzed and included for the interpretation of each exam. Typical glaucomatous optic disc changes included concentric enlargement of the cup, vertical cupping, notch, and focal narrowing of the neural rim. This work allowed us to establish an independent reference standard. Printouts with suspicious defects (points with P < 5% in the standard and pattern deviation plots) in patients with "healthy" looking optic discs were not included in the study, as they may be artifacts. Caprioli's strict, moderate and liberal criteria and Anderson's modified criteria were applied to each visual field printout (Table 1).[1,2] Caprioli's criteria were defined for the central 30 degrees, and the superior and inferior rows were excluded from analysis. In our study, all visual fields were analyzed with the central 24 - 2 program, which does not test the additional superior and inferior rows in the central 30 degree region. Hence, all 54 points of the visual field tested with the central 24 - 2 program were

Table 1 - Criteria for visual field abnormality

Criteria
Caprioli strict
≥ 4 adjacent points of ≥ 5 dB loss each
≥ 3 adjacent points of ≥ 10 dB loss each
Difference of ≥ 10 dB across nasal horizontal meridian at ≥ 3 adjacent points
Caprioli moderate
≥ 3 adjacent points of ≥ 5 dB loss each
≥ 2 adjacent points of ≥ 10 dB loss each
Difference of ≥ 10 dB across nasal horizontal meridian at ≥ 2 adjacent points
Caprioli liberal
≥ 2 adjacent points of ≥ 5 dB loss each
≥ 1 adjacent points of ≥ 10 dB loss each
Difference of ≥ 5 dB across nasal horizontal meridian at ≥ 2 adjacent points
Anderson modified
≥ 3 adjacent points in an expected location of the central 24° field that have p < 5% on the pattern deviation plot, one of which must have p < 1%
Glaucoma hemifield test "outside normal limits"
PSD with a p value < 5%

included in the analysis, except those surrounding the blind spot. One of Anderson's criteria was changed to replace corrected pattern standard deviation (CPSD) with the pattern standard deviation (PSD), as the SITA strategy does not calculate short-term fluctuation.[9] The application of criteria was done with an *"or"* operator rather than with an *"and"* operator; i.e., the presence a single criterion (either 'cluster points', 'PSD' or 'GHT') was enough to tag a visual field as abnormal and not all of the three criteria concurrently.

The sensitivity and specificity of each analytic algorithm, as well as the positive predictive value and the negative predictive value, were calculated. A receiver operator characteristics (ROC) curve was plotted, and the area under the curve for each analytic algorithm was calculated. A pairwise comparison of areas under ROC curves for each analytic algorithm was performed, and the differences were compared using a univariate z-score test. A *P* value of less than 0.05 indicated significance. A sub-analysis of the sensitivity for eyes with early glaucoma was done after stratifying each visual field exam according to its severity, as proposed by Hoddap et al.[10]

RESULTS

One hundred forty-three patients were selected, providing a study sample of 143 visual fields printouts. Of the 143

patients, 52 were considered normal, and 91 were considered glaucomatous (36 early, 24 moderate and 31 severe). Most of them (74.2%) had open angle glaucoma. The mean age of the study population was 63.7 ± 15.2 years. Sixty-nine patients were male (48.2%), and 74 were female (51.8%). Snellen visual acuity ranged from 20/20 to 20/400, and 76.4% of the patients had visual acuity of 20/40 or better.

The sensitivity, specificity, and positive and negative predictive values for each analytic algorithm are shown in Table 2. Anderson's criteria and Caprioli's strict criteria presented similar sensitivity (92.2% and 94.4%, respectively) and specificity (60.4% and 62.3%, respectively). Caprioli's liberal criteria were the most sensitive (98.9%) and the least specific (41.5%). When stratified by visual field severity, 36 eyes were classified as early glaucoma. The sensitivity was 83.3% for Anderson's modified criteria, 86.1% for Caprioli's strict, 88.9% for Caprioli's moderate and 97.2% for Caprioli's liberal criteria in this sub-group. The specificity

was 61.5% for Anderson's modified criteria, 63.4% for Caprioli's strict, 51.2% for Caprioli's moderate and 42.3% for Caprioli's liberal criteria.

The area under the ROC curve for each analytic algorithm is shown in Table 3, and the pairwise comparison of the areas under ROC curves is presented in Table 4. The differences in these values did not reach statistical significance ($P > 0.05$).

DISCUSSION

The SITA strategy has been shown to be as sensitive and specific as the full threshold strategies for glaucoma detection and has replaced older strategies for automated perimetry in clinical practice. Sekhar et al. found a sensitivity of 95% of SITA when compared to full threshold strategy.[6] Budenz found a 98% sensitivity of SITA as compared to full threshold.[7] SITA is able to shorten test time by 50% and the

Table 2 - Sensitivity, specificity and predictive values of different analytic algorithms for SITA interpretation

	Sensitivity	Specificity	+PV	-PV
Anderson	92.2%	60.4%	80.8%	82.2%
Caprioli strict	94.4%	62.3%	81.9%	86.8%
Caprioli moderate	95.0%	52.9%	77.7%	87.1%
Caprioli liberal	98.9%	41.5%	75.0%	95.7%

+PV: positive predictive value; -PV: negative predictive value

Table 3 - Area under the ROC curve for different analytic algorithms for SITA interpretation

	Area under ROC curve	95% CI	Standard error
Anderson	0.763	0.685 to 0.830	0.039
Caprioli strict	0.784	0.707 to 0.848	0.038
Caprioli moderate	0.732	0.652 to 0.803	0.041
Caprioli liberal	0.702	0.620 to 0.776	0.043

Table 4 - Pairwise comparison of ROC curves

Pair of criteria	Difference between areas	95% CI	Standard error	*P* value
A / C liberal	0.061	-0.025 to 0.147	0.044	0.162
A / C moderate	0.031	-0.038 to 0.099	0.035	0.381
A / C strict	0.021	-0.052 to 0.094	0.037	0.581
C liberal/C moderate	0.031	-0.042 to 0.103	0.037	0.408
C liberal/C strict	0.082	-0.009 to 0.172	0.046	0.078
C moderate/C strict	0.051	-0.014 to 0.117	0.033	0.127

A: Anderson; C: Caprioli

number of presented stimuli by 30%.[3,4] This is accomplished by use of a number of features, including the probability density function, likelihood function, dynamic monitoring of patient response times to interactively pace the test and comparison of adjacent locations to adjust the final threshold estimate.[5]

In spite of the trend towards replacement of full threshold strategies for SITA in the management of glaucoma patients, the application of analytic algorithms for interpretation of this new strategy has received little attention. The criteria for minimal perimetry abnormality in glaucoma suggested by Anderson were originally developed for the full threshold strategy. One criterion was a corrected standard pattern deviation with a p value of less than 5%.[2] SITA strategy, however, eliminates the retest trials for the 10 points used in full threshold strategies to determine short-term fluctuation such that CPSD is not calculated.[3] Thomas et al. evaluated the use of PSD instead of CPSD, as part of Anderson's criteria, to categorize a single field printout using the full threshold strategy on the Humphrey Field Analyser. They found almost perfect agreement between the two indices (0.77, kappa statistic).[9] The authors suggested that the substitution of CPSD by PSD seemed valid for the full threshold programs.[9]

In our study, we wanted to validate the use of Anderson's analytic algorithm, replacing the CPSD criterion by PSD, in a series of SITA visual field printouts and to compare this modified criteria with Caprioli's criteria for perimetric defect. The results of the current study showed similar sensitivity and specificity amongst Anderson's modified criteria and Caprioli's strict, moderate and liberal criteria for glaucoma detection. The similarity between Anderson's and Caprioli's strict criteria was remarkable (Table 2). Caprioli's strict criteria presented the largest area under the ROC curve (0.784), followed by Anderson's modified criteria (0.763). Nevertheless, these differences were not statistically significant. Caprioli's liberal criteria were the most sensitive (98.9%) and the least specific (41.5%) for the detection of glaucomatous perimetric defect, and, conversely, Caprioli's strict criteria were the least sensitive (94.4%) and the most specific (62.3%). These results were expected because, when strict criteria are used, specificity for glaucomatous defects will be high and sensitivity will be low, whereas, when one uses liberal criteria, sensitivity will be high and specificity will be low.[1] In the early glaucomatous group, sensitivity was lower for Anderson's modified criteria (83.3%) but similar to Caprioli's strict criteria (86.1%). Specificity values for each algorithm were similar to those calculated for the 143 subjects.

The specificity of the four criteria found in our study was lower than that shown in previous reports. In our study, specificity ranged from 41.5% (Caprioli liberal) to 62.3%

(Anderson modified). Using the same specific Anderson modified criteria for glaucoma defect, Budenz reported a specificity of 96% for both SITA standard and SITA fast.[7] In their prospective observational study, the authors used program 30-2 and patients with unreliable test exams were excluded from the study. Our study, however, was retrospective in design, and we used program 24-2. We selected only reliable test exams and, to prevent possible bias from any learning effect, we selected the third reliable exam from each patient included in the study. Although we are unsure of why the specificity was so divergent in these two studies, a possible explanation is that the studies used different gold standards. Budenz compared SITA to the full-threshold strategy while this study compared SITA to clinical impression plus optic disc appearance. We set the bar high by including abnormal optic nerves in the analysis, which is likely to systematically reduce specificity.

One caveat of this study is that we excluded abnormal visual fields with normal optic discs. These fields were rejected because these "defects" were most likely artifacts and not true defects. Nevertheless, removing these patients may have "pre-screened" the visual field, affecting the results. Elimination of false positives would have increased the specificity and predictive values, but it may also have undermined the clinical validity of the study. The purpose of specific criteria for the interpretation of visual fields is to separate normals from abnormals. Removal of some abnormal samples may give certain criteria an advantage, with some criteria being better able to identify false positives and being superior, even with similar sensitivity.

Another limitation of this study is our gold standard. We defined the internal gold standard based upon intuitive interpretation of the defect by correlating it with the optic disc aspect. Only subjects with typical glaucomatous optic disc changes were included. Recent literature, however, demonstrates that structural and functional loss may not be closely related in early glaucoma such that some subjects will demonstrate structural damage first, while others will demonstrate functional loss first.[11] We have assumed that all individuals with early field loss and normal discs were false positives. This particular assumption may alter the interpretation of this study by shifting focus to the utility of the Anderson and Caprioli criteria for classifying "glaucoma" in subjects whose diagnosis of glaucoma is largely based on optic disc evaluation. Some may not be convinced that this is not an ideally useful analysis and, indeed, some of the differences in the specificity observed in this study and previous studies may be due to inaccuracies in optic nerve head analysis, despite the experience of the observers. Use of full threshold visual fields as the gold standard would have been optimal; however, this was not possible in this retrospective study.

This study was also limited by the method of the ROC curve construction. In this study, the ROC curves were constructed for a categorical variable with only two possible outcomes: normal versus abnormal. This approach can lead to an underestimation of the area under the ROC curve.[12] Evaluation of the visual field printouts along with the optic disc aspect permitted only two possible outcomes: abnormal glaucomatous visual field defect or normal visual field. Suspicious visual field exams, as represented by single points depressed at $P < 5\%$ on the standard and pattern deviation plots in eyes with healthy looking optic discs, were excluded from the analysis.

In conclusion, the results of this study showed that Anderson's modified criteria can be used for SITA interpretation with sensitivity and specificity similar to those of Caprioli's strict, moderate and liberal criteria for perimetric defect. The four criteria are essentially identical when it comes to interpreting visual fields. None of these criteria achieved the level of specificity that is necessary for their use as the only determinant of glaucoma. Therefore, we recommend that any abnormal visual field should be confirmed before it is classified as abnormal. Furthermore, we recommend that this analysis be done in conjunction with structural evaluation of the optic disc.

REFERENCES

1. Caprioli J. Automated perimetry in glaucoma. Am J Ophthalmol. 1991; 235-9.

2. Anderson DR, editor. Automated Static Perimetry. 1st edition. St. Louis: Mosby; 1992. p.123.

3. Bengtsson B, Olsson J, Heijl A, Rootzén H. A new generation of algorithms for computerized threshold perimetry, SITA. Acta Ophthalmol Scand. 1997;368-75.

4. Bengtsson B, Heijl A. Evaluation of a new perimetric threshold strategy, SITA, in patients with manifest and suspect glaucoma. Acta Ophthalmol Scand. 1998;268-72.

5. Johnson CA. Recent developments in automated perimetry in glaucoma diagnosis and management. Curr Opin Ophthalmol. 2002;77-84.

6. Seckar GC, Naduvilath TJ, Lakkai M. Sensitivity of swedish interactive threshold algorithm in Humphrey visual field testing. Ophthalmol. 2000;1303-8.

Comparison of different analytic algorithms for interpretation of the swedish interactive threshold algorithm strategy
Takahasho GS et al.

CLINICS 2008;63:333-8

7. Budenz DL, Rhee P, Feuer WJ. Sensitivity and specificity of the swedish interactive threshold algorithm for glaucomatous visual field defects. Ophthalmol. 2002;1052-8.

8. Johnson CA. Update on developments. New developments. In: Nussdorf JD, editor. Glaucoma in the Millennium. Proceedings of the 50th Annual Symposium of the New Orleans Academy of Ophthalmology. The Hague: Kluger; 2003. p. 207.

9. Thomas R, Paul P, Muliyil J. Use of pattern standard deviation instead of corrected pattern standard deviation in Anderson's criteria. J Glaucoma. 2000;480-82.

10. Hodapp E, Parrish RK II, Anderson DR, editors. Clinical decisions in glaucoma. 1st edition. St. Louis: Mosby; 1993. p.53-9.

11. Artes PH, Chauhan BC. Longitudinal changes in the visual field and optic disc in glaucoma. Prog Retin Eye Res. 2005;333-54.

12. van den Hout WB. The area under an ROC curve with limited information. Med Decis Making. 2003;160-6.