Central Visual Field Assessment in Late Stage Glaucoma

by

Carmen Balian

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Vision Science

Waterloo, Ontario, Canada, 2006

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Glaucoma is defined as a progressive optic neuropathy, characterized by loss of visual function and often associated with high intra-ocular pressure. Testing the patients' visual function with Standard Automated Perimetry (SAP) is currently the clinical standard for detecting glaucomatous visual field loss. A new test algorithm using the Frequency Doubling illusion has been introduced on the Matrix perimeter (Humphrey Matrix; Carl Zeiss Meditech, Dublin CA) that measures the central 10˚ using a 2˚x 2˚ square flickering stimulus. This stimulus has the theoretical advantage of being both a large target, with good repeatability, and being perceptually selective, by preferentially stimulating the magnocellular projecting ganglion cells.

The purpose of this thesis was to determine the within-technique, between-visits repeatability and the within-visit, between-technique comparison of several techniques available to measure the central 10˚ visual field in patients with late stage glaucoma. In particular, to examine test-retest variability and compare sensitivity threshold values, visual field indices, and total and pattern deviation probability maps among the following techniques: Full Threshold SAP 10-2 size III (SAP III), Full Threshold SAP size V (SAP V), SITA SAP 10-2 size III (SS III), and Matrix 10-2 2˚ stimulus (M2).

Forty nine patients with advanced glaucomatous visual field defects attended 3 visits. During each visit, 1 eye was examined with each of the 4 techniques mentioned above. Data from the first visit was discarded to eliminate bias that may occur from the learning effect. Coefficient of Repeatability values of SAP III, SAP V, SS III, and M2 were calculated to be 10.33, 9.00, 9.90, and 12.04%dB respectively, relative to the average difference in threshold estimates between

visits. M2 had the most uniform test-retest characteristics across the full range of sensitivities; however the 90% confidence interval was the widest of all techniques in the normal to near normal range (24 to 38dB). Threshold estimates of SAP III and SS III were shown to be similar and slightly more variable than SAP V. M2 showed less severe defects than SAP III in the pattern deviation probability plots. Compared to SAP III and SS, M2 estimated sensitivity as less severe. Estimates of 20 dB and above on M2 were estimated at approximately 30 dB with SAP V. In the moderate to abnormal sensitivity range, Matrix estimated points to be shallower than that estimated by SAP V.

This thesis showed that M2 has lower sensitivity than SAP but shows fewer abnormal points than SS or SAP III and test-retest variability of the SAP techniques decreased with increasing sensitivity whereas; variability was constant throughout the dynamic range for M2 and smaller in the moderate to severe range. However M2 was worst in the normal to near-normal sensitivity range. This suggests that M2, compared to all SAP techniques, will be disadvantaged for the detection of early visual field loss but better positioned to repeatably detect and follow moderate to severe loss in the central 10˚ of patients with late stage glaucoma.

Acknowledgements

First and foremost I would like to thank my supervisor, Dr. John G. Flanagan, from whom I have learned so much, for giving me the opportunity to work with him and for extending much needed guidance and support.

Great thanks also go out to my committee members, Drs. Graham Trope and Natalie Hutchings, whose contributions to my studies have been invaluable.

Special thanks also to Dr. Dimitrios Kourkoutas who contributed to this study with both his expertise and time and was always more than willing to help.

The dedication of the patients involved in this study is greatly appreciated; if it were not for them, this study will not have been possible. Many thanks to the staff of the Glaucoma Department at Toronto Western Hospital who were prepared to help at anytime.

Special thanks to my fellow graduate students who were always available for assistance and encouragement. I wish you all success in your studies and future endeavors.

To Mom, Dad and Lina, I owe the most for standing by me at all times.

To Carl Zeiss Meditech for funding this project.

Words can not express the gratitude I feel towards you All.

Dedication

To Lina, who has always believed in me. I couldn't have done it without your love and support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AGIS	Advanced Glaucoma Intervention Study
ANOVA	Analysis of Variance
asb	apostilbs
CoR	Coefficient of Repeatability
CoV	Coefficient of Variability
dB	decibels
FD	Frequency Doubling
FDP	Frequency Doubling Perimetry
FL	Fixation Losses
FN	False Negative
FP	False Positive
FT	Full Threshold
GHT	Glaucoma Hemifield Test
GSS	Glaucoma Staging System
HFA	Humphrey Field Analyzer
LOCS	Lens Opacity Classification System
M2	Matrix using 2˚ stimulus
MD	Mean Deviation
MOBS	Modified Binary Search
PCA	Principal Curve Analysis
pdf	probability density function
PD	Pattern Deviation
PSD	Pattern Standard Deviation

RNFL	retinal nerve fibre layer
SAP	Standard Automated Perimetry
SAP III	Full Threshold Standard Automated Perimetry using Goldmann size III
SAP V	Full Threshold Standard Automated Perimetry using Goldmann size V
SD	standard deviation
SITA	Swedish Interactive Test Algorithm
SS	SITA-Standard Standard Automated Perimetry using Goldmann size III
SWAP	Short-wavelength Automated Perimetry
SWS	short-wavelength sensitivity
TD	Total Deviation
ZEST	Zippy Estimation of Sequential Testing

1 Introduction

1.1 Glaucoma and Automated Perimetry

Glaucoma is a general term that encompasses a range of ocular conditions that cause a specific neuropathy of the optic nerve[1]. It is the second leading cause of blindness in the world[2] and can occur in all age groups but is most common in the elderly[1]. Glaucoma is characterized as progressive optic nerve damage associated with visual function loss[1, 3, 4]. The most common clinical method of measuring the function of the visual system is the assessment of the eye's ability to detect the brightness of small points of light projected in both the central and peripheral areas of vision, also known as perimetry or visual field testing[5]. The detection of visual field abnormality in glaucoma is an indicator of optic nerve damage[6]. Testing the patients' visual field using standard automated perimetry (SAP) is currently considered to be the gold standard for detecting glaucomatous visual field loss[6, 7, 8]; SAP uses stationary or static targets of varying brightness, so called static perimetry, to determine functional glaucomatous changes[9].

The development of early glaucoma is a very gradual process[10]. Local depressions of sensitivity often appear and disappear before becoming stable defects in the patients' visual field[10]. The defects, commonly referred to as scotomas, then begin to enlarge and follow the arcuate pattern of the retinal nerve fibres. In the advanced stages of glaucoma, large arcuate scotomas from the superior and inferior field break through into the peripheral field and connect leaving only the central or temporal visual field intact[11].

Although SAP is used as the gold standard for perimetry testing, the stimulus is broadband in nature and not selective for any particular ganglion cell subtype. Newer techniques have been developed which are aimed at selectively testing subsets of ganglion cells that may be more prone to damage caused by glaucoma (selective loss hypothesis)[12] or better reflect generalized ganglion cell loss (selective testing hypothesis)[13]. For example, Short-Wavelength Automated Perimetry (SWAP) projects a blue target onto a high luminance yellow background[14] to selectively examine the short-wavelength sensitivity (SWS) pathway[14] which are believed to be damaged early in glaucoma[15, 16, 17, 18, 19]. Frequency Doubling (FD) Technology selectively tests magnocellular cells, which are stimulated by the FD illusion[20, 21]. It is believed that magnocellular projecting ganglion cells may be selectively damaged in glaucoma[12] as compared to parvocellular cells[22]. Previous studies have shown that the contrast sensitivity perceived with the FD illusion is reduced in patients with glaucoma[23, 24, 25] hence, selectively testing for magnocellular cells may be key in detecting and monitoring early glaucoma. An alternate to this theory is the Reduced Redundancy Hypothesis as proposed by Johnson[13]. This hypothesis states that early functional loss can be detected by testing subpopulations of ganglion cells that have reduced redundancy, or a sparse distribution throughout the retina. In such a situation defect may be detectable earlier when compared to that detected using a more general stimulus. It has been shown that FDT is effective at detecting patients with moderate to severe glaucoma[26, 27, 28]. For early glaucoma, sensitivity and specificity of the FDT has been reported as 85% and 90%, respectively; due to its short duration, resistance to blur and pupil size, it may be a useful screening tool for glaucoma[29]. The high temporal frequency and low spatial frequency required to perceive the frequency doubling illusion suggests that it is preferentially mediated by the magnocellular pathway[30]. This pathway is particularly responsive to high temporal frequency, low spatial frequency, and achromatic information[31].

1.2 Automated Perimetry

Automated perimetry is a diagnostic examination technique used for assessing visual function[9, 32, 33] including patients who have glaucoma, or are a glaucoma suspect[34]. Today, a person's visual field is tested with automated perimeters and standard thresholding algorithms (SAP)[35] which have improved our ability to quantify visual function[36]. Perimetry testing may be the only means of detecting progression in late stage glaucoma as the optic disc will no longer be a reliable or accurate indicator for progression[37].

The purpose of a visual field examination is to detect defects, determine the specific pattern of visual field loss for diagnostic purposes, and monitor patients for evidence of visual field progression[38, 39, 40]. In patients with glaucoma, the observed patterns of visual field abnormalities correspond to the anatomy of the nerve fiber layer of the retina and its projections to the optic nerve[41].

1.2.1 Humphrey Field Analyzer

1.2.1.1 Instrument Specifications

The Humphrey Field Analyzer (HFA; Carl Zeiss Meditec, Dublin, CA, USA) is most commonly used in clinical settings for assessing visual function; it uses contrast sensitivity testing to analyze a person's visual field sensitivity. Standard Automated Perimetry, also known as white-on-white perimetry, is the standard technique available on the HFA. Standard automated perimetry uses a white stimulus of a specific size (usually 0.43º diameter, Goldmann size III) on a white background of specific luminance (10cdm-² or 31.5 asb). This background illumination was

specifically chosen for the Goldmann perimeter, the previous clinical gold standard perimeter, because it is the minimum amount of light needed for photopic vision to function, i.e. for stimulation of the cones as well as the rods[10]. The subject is asked to respond to the flash of light by pressing a button. According to the intensity value(s) the patient responds to, the instrument will calculate the sensitivity at each stimulus location throughout the visual field. The result is expressed as a decibel (dB) value representing the minimum brightness the patient can see at each point[10]. The target(s) used by standard automated perimetry (SAP) can stimulate a wide spectrum of retinal ganglion cells within a tested retinal location[8, 42].

The 30-2 program on the HFA tests 76 points within the central 30 degrees, separated by 6˚ and offset from the horizontal and vertical meridians by 3˚[43]. The central 10-2 program tests 68 points in a grid within the central 10 degrees, each separated by 2˚ and offset from the horizontal and vertical meridians by 1˚[43, 11]. STATPAC is a statistical software package installed in the Humphrey perimeter which provides rapid analysis including a comparison of the patients' threshold values with that of normal age-based population, at the time of the exam[10, 44].

1.2.1.2 Goldmann Stimulus Size

The Goldmann stimulus size I-V targets are used with techniques available on the Humphrey Field Analyzer (HFA). These targets are smaller in size when compared to the blind spot (which is roughly 5 by 7 degrees)[10]. The Goldmann size III (0.43˚ diameter) stimulus is the most commonly used with SAP because it provides a valid assessment of neural loss[45]. The Goldmann size V (1.73˚) has been used with SAP in cases of severe glaucomatous visual field loss. The smaller the target size used with SAP, the deeper the visual field defect that will be recorded.

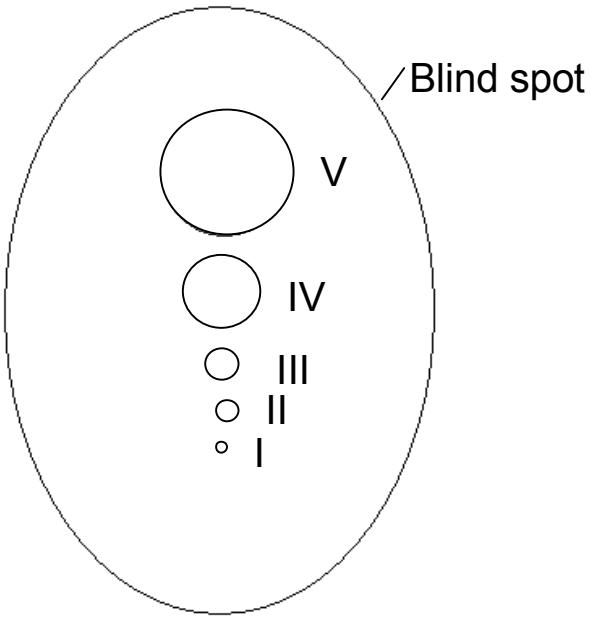

Figure 1.1: Relative comparison of Goldmann stimulus sizes I-V with the blind spot.

1.2.1.3 Test Procedures

Although the development of automated perimetry has greatly increased the precision and consistency of visual field results as compared to manual perimetry[46], new techniques are continuously being developed to help decrease test variability and increase sensitivity and specificity. A variety of test procedures are available which estimate threshold values[47]. The use of a standardized test procedure for automated perimetry testing will have a great impact on the validity of the test as well as its duration.

1.2.1.3.1 Bracketing Strategy

The commercially available Humphrey uses a bracketing strategy to estimate threshold levels[41]. In general, the threshold values for all Humphrey strategies are calculated from this repetitive up-and-down staircase technique[10]; the patient's response to the stimulus will determine whether the subsequent stimulus intensity will increase or decrease[48]. Therefore, if a patient responds to a stimulus, the subsequent stimulus at that location will be presented at a lower contrast level; this will continue until the patient does not respond to the presented stimulus. When this point is reached, the computer will increase the intensity of the target by a smaller step, if the patient responds then the following stimulus is presented at a lower intensity by an even smaller step. This reversal is repeated several times until consistent responses are obtained. The threshold is calculated with respect to the lowest intensity the patient responds to. As severity of the field loss increases, fewer stimulus presentations are needed to estimate the threshold at the given stimulus location[49]. The Humphrey perimeter uses a 4-2dB staircase.

1.2.1.3.2 Swedish Interactive Test Algorithm

The Swedish Interactive Test Algorithm (SITA) is a family of test algorithms designed to significantly reduce the test time for threshold estimation available on the HFA, without any reduction in data quality[7, 29, 50]. Swedish Interactive Threshold Algorithms use maximum likelihood methods to estimate threshold values[51]. Each test location has two likelihood functions each derived from normal and glaucomatous visual fields[52, 53, 54]. At the beginning of each test, threshold values for four predetermined test locations are calculated; these values are then used to calculate the initial intensity for the stimuli to be presented at adjacent test locations[51]. The

patient's responses to the stimuli at each location along with previously computed distributions[51] are used to calculate probability distributions; thus, the probability distribution at each location changes with each successive response[51]. The peak of the distribution represents the threshold value and the width determines its accuracy[51] hence, the narrower the distribution the more accurate is the estimate.

The SITA threshold testing algorithm has allowed perimetry testing to be more accurate and reliable with shorter test duration[7]. They are significantly faster than the standard full threshold algorithm, however, an increase in severity of the visual field decreases time saved[7].

1.2.2 Frequency Doubling Perimetry

1.2.2.1 Instrument Specifications

The Frequency Doubling Technology (FDT) is a perimetry technique which simultaneously exploits the utility of contrast sensitivity, spatial frequency and temporal modulation[55]. The commercially available FDT employs a sinusoidal grating of 0.25 cycles per degree undergoing counterphase flicker at 25 Hertz[56].

The phenomenon of FD is experienced when a low spatial frequency sinusoidal grating is combined with a high temporal frequency counter phase flicker[57, 58, 59, 60]. Thus, the subject perceives a stimulus with twice the number of bands spaced more closely[3]. For the frequency doubling effect to occur, the grating must have a spatial frequency <4 cycles per degree and counterphase at a temporal rate >15 Hertz[56]. The subject is asked to respond by pressing a button when he/she is able to see the target. The threshold of the visual field is determined by the

amount of contrast that the target had at the time a response was recorded. Figure 1.2 shows a schematic diagram of the FD illusion.

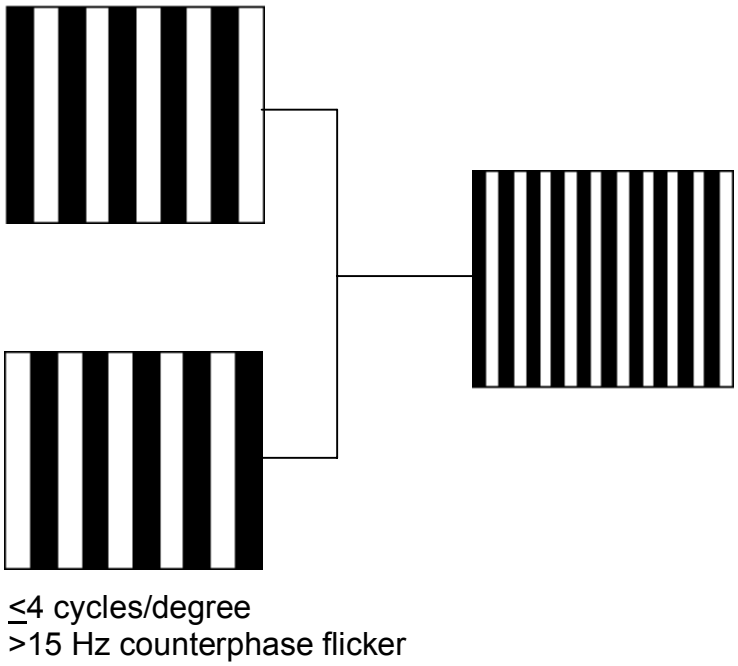

Figure 1.2: Schematic diagram of the Frequency Doubling Illusion.

Studies have shown that the frequency doubling perimetry is effective at detecting moderate to severe cases of glaucoma[12, 61] and shows promise as a screening tool for early glaucoma[12, 23, 61, 62] with very good sensitivity and specificity[23, 53].Because of the low spatial frequency of the target[53], the frequency doubling stimulus is resistant to blur up to 6 diopters; thus visual correction is not necessary with lower degrees of refractive error[12, 53]. Patients typically respond to the perception of any stimulus, irrespective of whether a frequency-doubled illusion is visible[64].

Frequency doubling perimetry is attractive in the clinical setting because the test is resilient to refractive errors and blur, it has a large dynamic range, and threshold test strategies are short in

duration[65]. Also, test-retest variability with increase in defect severity does not increase as much with FDT as compared with SAP[65].

A study by Cello et al[26] showed very high sensitivity and specificity values of the FDT in early, moderate, and advanced stages of glaucoma. The FDT 30-2 technique has shorter test duration as compared to conventional perimetry[66], however, the MD and PSD of the FDT results show a strong linear correlation with that of the Humphrey 30-2 technique[55]. These findings indicate that the performance level of the FDT is comparable with that of the full-threshold test strategies used in conventional automated perimetry. It is also possible that this technique can be used to detect and characterize the severity of glaucomatous visual field loss[26].

The Matrix is a new perimeter developed by Welch-Allyn to measure sensitivity of the visual field. It was developed as a means of improving efficiency and accuracy of the original Frequency Doubling Technology (FDT). In comparison to FDT, the Matrix (also known as FDT II (Welch Allyn, Skaneateles Falls, NY; Carl Zeiss, Meditech, Dublin CA) uses a smaller stimulus size allowing for more test locations to be tested thereby giving more detail on the spatial distribution of the visual field loss[67].

1.2.2.2 Testing Procedures

1.2.2.2.1 Modified Binary Search

The Modified Binary Search (MOBS) test procedure is used by the commercially available FDT for the Full Threshold test[68]. A range of possible thresholds sets the upper and lower thresholds for the patient at each test location. An average contrast value of the upper and lower threshold

limit is calculated as the target contrast for the initial presented stimulus. The patient's response to the target determines the interval from which the contrast for the proceeding stimulus at that location will be calculated. This process is continued until a certain number or response reversals are met and the difference between the upper and lower threshold is equal to or less than a predetermined interval[69]; this information is used to calculate the threshold values. An advantage of MOBS is that it can recover quickly from response error and it can make large jumps to remain close to the correct location of threshold[69].

1.2.2.2.2 Zippy Estimation of Sequential Testing

Threshold values for the Matrix are calculated using an adaptation of the Zippy Estimation of Sequential Thresholds (ZEST) procedure[67, 70]. At each test location, 4 stimuli are presented, each with a predetermined intensity and corresponding pdf curve[71, 72]; the pdf curve is modified for the next presentation with respect to the patients' responses ("seen" or "not seen")[67]. The 15 possible combinations of "seen/not seen" responses to the 4 stimuli presented determine the threshold estimates ranging from 0 to 38 dB[67] i.e. a frequency-of-seeing curve is obtained[71]. ZEST has been shown to be just as accurate and reproducible as the MOBS procedure with a 40% decrease in testing time in both normal and glaucomatous patients[72]. The time saved with this procedure helps decrease the effect of patient fatigue.

1.3 Testing Boundaries

Perimetry tests best suited for the early stages of glaucoma may not be well suited for later stages of the disease and its progression[42]. Testing the central 30˚ of the visual field in the early stages of glaucoma can give a good sense of the disease and its progression. However, as the disease

progresses into the later stages, when only a small portion of the central visual field is intact, switching to a technique which measures the visual function of the central 10˚ can provide a more detailed assessment of visual field progression. A retrospective study by Zalta[43] has shown that SAP 10-2 size V is able to measure visual function that is undetected by SAP 30-2 size III.

1.4 Sensitivity Values

Threshold and visual sensitivity, as measured in decibels (dB), are inverse functions. The dB scale is a logarithmic scale that is inversely related to luminance and each dB is equal to 0.1 log units with the SAP and approximately 0.05 log units with the Matrix[67]. The more sensitive a person is to a specific stimulus at a specific location, the lower his/her threshold is at that point of the visual field. Usually, neighboring points in the visual field have similar thresholds[48]. The most important aspect of the visual field lies in the accuracy of determining the sensitivity[73] at each determined location. The probability of seeing a stimulus presented at threshold is 50%[72].

1.5 Reliability Parameters

In order to use perimetric data to accurately evaluate visual field loss, it is crucial to know the reliability of the results which depend to a great extent on the patient's ability to consistently perform the perimetric task[29, 74, 75]. The reliability also depends strongly on the reproducibility of its results[75]. Fixation losses (FL), false positive (FP), and false negative (FN) calculations, as reported on the perimetry test printout, are an indication of the reliability and validity of the test. Fixation losses provide a relative idea of how well the patient kept his or her eye fixed on the fixation target during the test. Throughout the test, at random intervals, a stimulus is projected in the area of the blindspot (the instrument locates the area of the blindspot at the beginning of the

test)[76] at an intensity the patient is able to perceive; the number of times the patient reports seeing such stimuli is recorded alongside the other reliability indices[26, 48].

The false positive rate (FP) is presented as a ratio of the total number of times the subject responds without a stimulus being presented, divided by the total number of times the instrument pauses without presenting a stimulus[77]. A patient is termed "trigger happy" when he/she has a high false positive rate, i.e. frequently clicks the button when no stimulus is presented. False negative errors usually result when the subject fails to respond to a distinctly visible stimulus[10] in a location outside of the determined blind spot. The false negative rate (FN) is presented as a ratio of the total number of times the subject fails to respond when a stimulus of 9 dB higher than the previously determined threshold sensitivity at that location is presented, divided by the total number of such presentations[77], i.e. failing to respond to stimulus with 100% contrast[26]. In the presence of severe visual field loss, FN is not used to define reliability due to the low number of catch trials[43].

The SITA algorithm calculates FP and FN differently than described above. False positives are calculated by recording positive responses when none are expected, i.e. within the minimum reaction time interval after a stimulus is shown[78]. False negatives are calculated based on the patient's pattern of responses after the test is completed[79]. Different data from FP and FN are combined and the maximum likelihood method is used to calculate FP and FN responses as a percentage[80]. This method of estimating the frequency of FP and FN responses helps reduce testing time.

Vertex Monitoring and Gaze Tracking are features available on HFA II. The former ensures that the patient's eye is centered behind the lens at a proper distance, eliminating the trial lens as a possible source for unreliable results. The latter is used to determine the patients' fixation during

the test. This is done by using real-time image analysis. A gaze track graph is displayed on the printout.

1.6 Statistical Plots: Total Deviation, Pattern Deviation and Probability Maps

Total deviation (TD) values and its related probability plot are calculated on techniques which have a normal database available. The TD plot is composed of positive and negative integers which correspond to the difference in sensitivity between the subject and age-matched normal data at each point of the visual field[9, 10]. Total Deviation plots are useful because they accentuate areas of the visual field which fall outside the normal range[10]. Its corresponding probability map indicates how different the given results are from that of the normal[41, 81].

A pattern deviation (PD) plot and its related probability plot are also calculated with respect to a normal database. This particular plot allows for the field test results to be compensated with respect to the subject's height of the hill of vision[10], i.e. it eliminates defects caused by a generalized shift in sensitivity[9]. Thus, it signifies the difference in shape of the measured hill-of-vision as compared with that of the normal population[82]. This allows for differentiation of localized visual field loss from that resulting from age-related conditions such as small pupils and cataract formation[10].

Probability maps are used to evaluate the normality of the data[83]. It compares the threshold values of the patient with that of the age-matched normal database, if one is available for the technique.

1.7 Global Indices

Statistical analysis of visual fields has become a useful tool in interpreting on the results from automated perimetry[84]. Visual field indices are a statistical review of the retinal light sensitivities which are designed to recognize and evaluate the extent of visual field damage[36]. They are used to facilitate interpretation of the results from a single perimetric examination[36]. It assists the interpreter with defining visual field loss by summarizing the data obtained from the test[85, 86]. Visual field indices, Mean Deviation (MD) and Pattern Standard Deviation (PSD) are calculated based on previously acquired normal data[63].

Mean deviation is determined by averaging the deviation from normal for all points tested[74]. It quantifies overall change of visual field loss with respect to normal data of age-matched controls[34, 39, 63, 87]. Pattern standard deviation measures the extent to which the tested field deviates from the shape of the "normal hill of vision"[74]. It is an index for showing localized change in the visual field[34, 63, 87].

1.8 Glaucoma Hemifield Test

The Glaucoma Hemifield Test (GHT) is an algorithm that evaluates glaucomatous visual field loss from visual field sensitivity data of a single visual field test[88]; it is only applicable to glaucomatous defects[9], thus it should be disregarded if a visual defect other than glaucoma is suspected.

Each hemifield of the visual field is divided into 5 sectors as illustrated in the Figure 1.3, which correspond to the normal retinal nerve fibre layer (RNFL) anatomy. Each test location is

assigned an ordinal score with respect to the pattern deviation probability map probability score; the deeper the defect, the higher the score[88]. Every sector of the superior hemifield is compared to the mirror-image sector of the inferior hemifield, and a difference of the sum of probability scores is calculated. These differences are compared with the limits of normality to identify the "category" of the patient's visual field status. The GHT classifies visual fields into five categories: i) within normal limits, ii) borderline, iii) outside normal limits, iv) generalized reduction insensitivity, and v) abnormally high sensitivity[88, 89]. The last category is for tests with high false positive results, also known as "trigger happy patients".

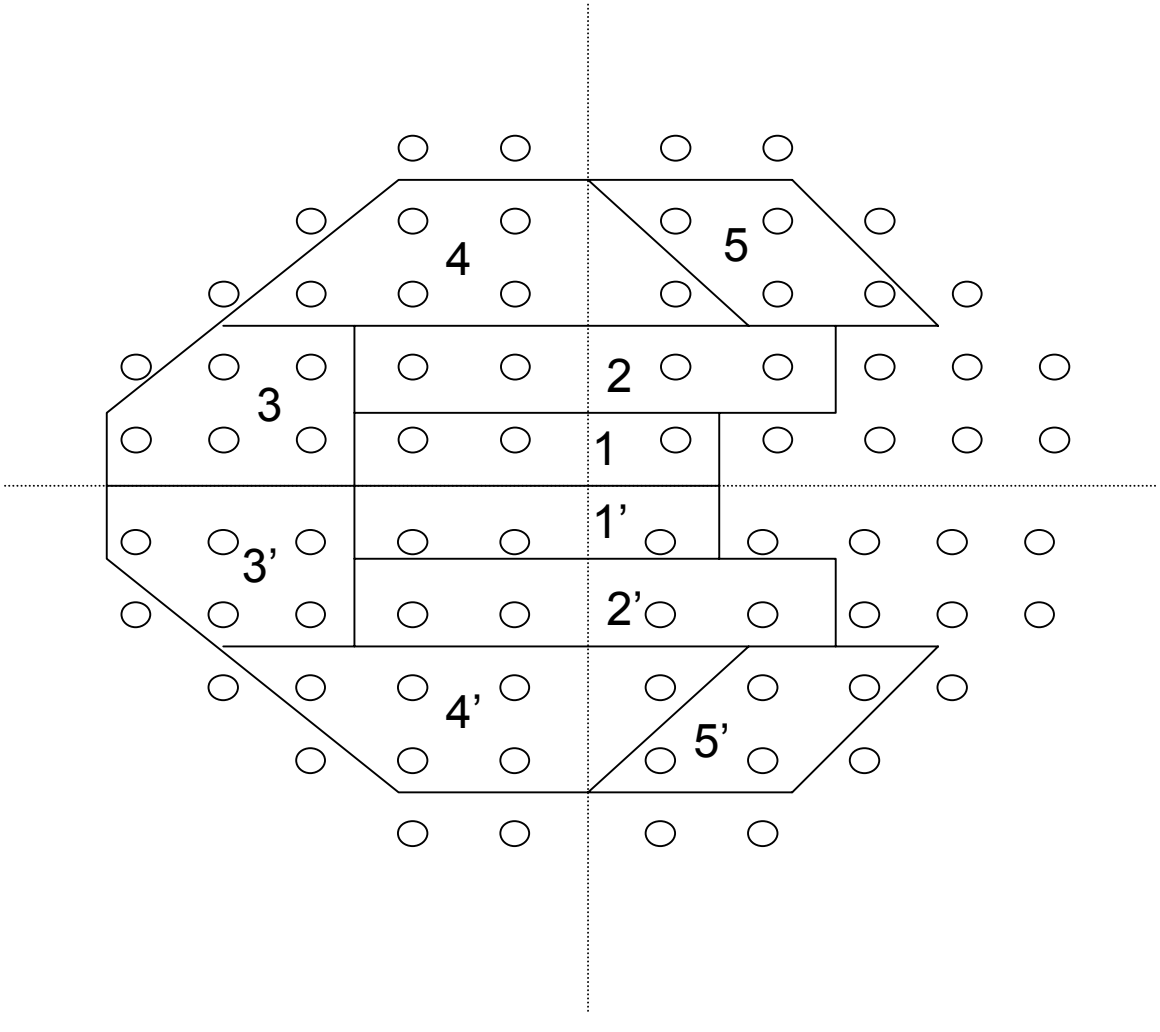

Figure 1.3: Partition of the visual field for Glaucoma Hemifield Test analysis. Each hemisphere is divided into 5 sections which correspond to the normal retinal nerve fiber layer anatomy.

Although the GHT may seem like a reasonable method of classifying glaucomatous visual field loss, it should be noted that, like other visual field indices, it will fail to reflect the true nature of the disease in far-advanced glaucoma cases[90]. This is explained by the actuality that in far-advanced glaucoma, both the superior and inferior hemifields of the visual field will be severely damaged and hence, very little or no difference will be noted between them.

1.9 Glaucoma Stage Classification

Many studies use the MD value from the Humphrey Field Analyzer to classify glaucomatous visual field loss. The most commonly used classification is that of Hodapp, Anderson and Parish[91]. They categorize glaucomatous visual field loss into the following: i) early, ii) moderate, and iii) late. The classifications are made using the Humphrey 30-2 technique, which measures 76 test locations in a 6˚ x 6˚ grid covering the central 30˚ of the visual field.

To classify a defect as "early", the MD must be worse than -6 dB; the PD plot should have fewer than 18 points depressed below the 5% probability level and fewer than 10 points below the 1% level and no point in the central 5˚ of the visual field with sensitivity less than 15 dB. A defect is classified as "moderate" when the MD is worse than -12 dB; the PD plot has fewer than 37 points depressed below the 5% level and fewer than 20 points depressed below the 1% level. No point in the central 5˚ of the visual field with sensitivity of 0 dB and only one hemifield within 5˚ of fixation may have a sensitivity less than 15 dB.

A severe defect consists of the following findings: MD worse than -12 dB, more than 37 points depressed below the 5% level on the PD plot, check if TD or PD *or* more than 20 points depressed below the 1% level on the PD plot. Any point in the central 5˚ visual field with

sensitivity of 0 dB and points within the central 5° with sensitivity less than 15 dB in both hemifields.

The Advanced Glaucoma Intervention Study (AGIS) classification of glaucoma uses the total deviation plot from the HFA 24-2 technique to assign a score from 0 to 20[92]; 0 indicating no visual field loss and 20 indicating end-stage disease. A total of 52 test locations are available on the total deviation plot: 6 locations in the nasal region and 23 in each hemifield. Each sector is scored based on the number and depth of defect and then the scores are summed to give the defect score[32].

The Glaucoma Staging System (GSS)[92] categorizes glaucoma into 6 stages, stages 0 (normal) to V (end-stage), based on Humphrey visual field results. A series of criteria must be met before classification into the appropriate stage. Mean Deviation is the primary measure for categorization; Pattern Standard Deviation, hemifield test, dB plot and pattern deviation plots are secondary factors. If the patient meets the criteria set for MD, but fails to meet the additional criteria for that stage, then he or she is placed in the preceding or succeeding stage depending on which additional criteria is met.

1.10 Frequency Doubling Perimetry in Glaucoma

The frequency doubling stimulus was found to be a promising stimulus for glaucoma testing[23]. A study by Spry et al[40] comparing SAP and FDT in normal individuals and patients with glaucoma showed that FDT exhibited significantly lower variability especially in areas of depressed sensitivity; the Matrix has been shown to have constant variability over its dynamic range in patients with early to moderate glaucoma[67]. The FDT has been shown to be effective at detecting moderate to severe glaucoma and is well suited as a screening tool for the disease[61]. Many

studies have shown the FDT to have good sensitivity and specificity for detecting glaucoma[24, 62, 164, 165]. A study by Cello et al[26] has shown FDT to have high sensitivity and specificity for all stages of glaucoma. These findings suggest that use of FD illusion may be suitable for monitoring patients with this disease.

1.11 Large Stimulus Testing in Glaucoma

Using larger stimuli in the later stages of glaucoma may yield additional information regarding the patients' visual field[95, 96]. The use of Goldmann size V, instead of the standard size III, in SAP has been shown to decrease variability in areas of moderately damaged to normal sensitivity in patients with glaucoma[97, 98]. Goldmann size V has been shown to measure visual function in areas where size III has measured as absolute defect[95]. Testing only the central 10˚ visual field and using Goldmann size V stimulus has been shown to be better at monitoring changes in the visual field of patients with only a small island of vision remaining[11, 43].

1.12 Rationale

Perimetry is essential for the management of glaucoma[10]. The developments of new perimetry techniques are aimed at providing shorter test duration, and optimum sensitivity and specificity. Different perimetry techniques are designed to test different aspects of retinal physiology. Standard Automated Perimetry is currently the gold standard for measuring retinal sensitivity of patients with glaucoma. Because of its inability to selectively test subsets of ganglion cells, it may not be the most effective technique for measuring defects in the visual field caused by the disease.

Previous studies[26, 27, 28] have shown that the FDT is very useful in measuring visual field depth in patients with moderate to severe stages of glaucoma. We want to see if this holds true for the Matrix in patients with late stage glaucoma using the 10-2 technique. We also want to determine if use of larger target size and different algorithms available on SAP provides a better means of evaluating visual field sensitivity in patients with late stage glaucoma.

1.13 Research Questions

Does choice of algorithm affect the repeatability and defect characteristics of perimetry testing in patients with late stage glaucoma?
Is the small FD stimulus, as available on the Matrix, more repeatable than SAP in late stage glaucoma?
Is the FD stimulus capable of measuring visual field depth and area when compared to the standard automated perimetry?

1.14 Objectives

The global aim of this thesis was to determine the capability of different perimetry techniques, available on the Humphrey Field Analyzer and Matrix instruments, to estimate retinal sensitivity in various locations within the central 10˚ visual field of patients with late stage glaucoma. The following techniques were used: i) Full Threshold SAP 10-2 size III (SAP III), ii) Full Threshold SAP 10-2 size V (SAP V), iii) SITA SAP 10-2 size III (SS III), and iv) Matrix 10-2 size 2˚ (M2).

The primary objective of this study was to determine which perimetry technique, the SAP III, SAP V, the SS III, or the M2, was most repeatable and better able to detect functional loss in the central 10˚ visual field of patients with late stage glaucoma.

1.15 Hypotheses

1. The Matrix shows more extensive and deeper field defect than Standard Automated Perimetry.

2. SAP V will show the least amount of visual field defect but will be the most repeatable technique.

4. No significant difference is expected between the results of SAP III and SS III.

5. The Matrix is more repeatable than SAP III or SS III.

2 Methods

2.1 Sample Size Calculation

The sample size for this study was calculated with the following formula:

$$\eta = \left[\frac{z_{\alpha/2}\sigma}{E} \right]^2$$

where η is the sample size, $z_{\alpha/2}$ is equal to the chosen confidence limit on a standard normal distribution, σ is the population standard deviation as obtained from previous studies, and E is the margin of error (the maximum difference between techniques).

For this study, we chose a 95% confidence interval therefore, $\alpha = 0.05$, and $z_{\alpha/2} = 1.96$ and $E = 1$. Similar studies performed previously have reported $\sigma = 3.30$. Substituting these values into the equation, we get:

$$\eta = [(1.96)(3.30)/1]^2$$
$$= 42$$

Although our sample calculation indicates that an η of 42 will provide statistically significant results; however 49 subjects were recruited to allow for dropout rate of 15%.

2.2 Study Sample Demographics

The study sample demographics are listed in Table 2.1.

Table 2.1: Study Sample Demographics

	Ratio	Average	Maximum	Minimum
Males: Females	30:19	----	----	----
Eyes (Right: Left)	31:18	----	----	----
Age (years)	----	68.37 $\pm$ 9.79	84	46
HFA Trial Lens Used (Diopters)	----	----	-9.25	+5.25
Pupil Size (mm)	----	3.55 $\pm$ 0.54	5.00	3.00
Visual Acuity	----	----	6/6	6/21
Intraocular Pressure (mmHg)	----	11.91 $\pm$ 3.28	20	4
Time Elapsed Between Visits (days)	----	17.43 $\pm$ 24.67	85	1
Time Elapsed Between Visits (weeks)	----	2.49 $\pm$ 3.52	12.14	0.14

Table 2.1 describes the patient demographics. The study sample consisted of 30 males and 19 females. 31 right eyes and 18 left eyes were included in the study. The mean age, standard deviation and range for this group was 68.4 $\pm$ 9.79 years and 46 to 84 years, respectively. The refractive error ranged from -9.25 to +5.25. The pupil sizes varied from 2 to 5 mm in diameter without dilation; tropicamide ½% was used to dilate pupils which were less than 3 mm in diameter to ensure that pupil size was not a contributing factor to visual field defect. Visual acuity ranged from 6/21 to 6/6. Intraocular pressure (Applanation Tonometry) ranged from 5 to 20 mmHg.

2.3 Inclusion/Exclusion Criteria

The subjects were recruited from the Glaucoma Department at the Toronto Western Hospital based on SITA-Standard 24-2 test results performed within six months of recruitment. Patients with advanced field loss were considered as possible candidates for participation in the study.

Advanced field loss was defined as a visual field with only the central or temporal portion of the central 10˚ visual field intact. Only patients who gave reliable results were recruited, i.e. false positive (FP), false negative (FN), and fixation losses (FL) less than 15%, 15%, and 30%, respectively[74].

Patients were excluded if they had one or more of the following: Visual acuity worse than 6/24; a defect in the visual field of the eye being tested that was explained by the patient's ocular status or history other than late stage glaucoma; history of disease or use of medication that may affect visual field reliability, or ability to undergo either perimetry test. If both eyes qualified for the study, then the eye to be included was chosen at random.

2.4 Definition of Disease Stage

Hodapp E et al[91] categorize glaucomatous visual field loss into the following: i) early, ii) moderate, and iii) late. The Humphrey 30-2 technique was used to make the classification. A total of 76 test locations are tested with this technique.

A severe defect consists of the following findings: MD worse than -12 dB, more than 37 points depressed below the 5% level *or* more than 20 points depressed below the 1% level using the PD plot. Any point in the central 5˚ visual field with sensitivity of 0 dB and points within the central 5˚ with sensitivity less than 15 dB in both hemifields.

For the purposes of this study, we defined late stage glaucoma as the condition with only the central or temporal portion of the central 10˚ visual field remaining on the grayscale plot of HFA 30-2 technique and MD worse than -12 dB.

2.5 Ethics

The study was approved by the University of Waterloo, Office of Research Ethics, and the University Health Network. Written consent was obtained from each of the subjects prior to enrollment in the study.

2.6 Procedures

Patients were given an ophthalmic examination consisting of the following: Goldmann applanation tonometry, slit lamp examination of the anterior segment of the eye, and dilated ophthalmic examination including LOCS II grading system for the evaluation of the crystalline lens.

This cross-sectional study compared perimetry results from the central 10˚ visual field of patients with severe glaucoma, using the following perimetry techniques: i) Full Threshold SAP 10-2 size III, ii) Full Threshold SAP 10-2 size V, iii) SITA SAP 10-2 size III, and iv) Matrix 10-2 2˚. We wanted to see whether or not the results would vary with the use of different stimulus variables and/or algorithms.

Each subject attended for three visits. During each visit, one eye from each patient was examined with each of the four perimetry techniques. The order of the testing for each patient was random but remained constant for each patient; this was done to eliminate any bias that may be caused from the "carryover effect". Subjects were given ample time to rest between the tests (at least 5 minutes) to eliminate the effect of fatigue on proceeding tests.

In this study, the global indices, sensitivity values, and probability plots among the various visual field techniques were compared.

2.7 Analysis

The results from Visit 1 were discarded to eliminate any bias that may result from the learning effect. The patient's age and severity of visual field were considered as separate between-subject factors and the order of perimetry testing were considered as within-subject factors; this allowed for direct comparison of the visual field data of patients with different patterns and severity of glaucomatous visual field loss. All threshold estimates with a value less than 0 dB were given a value of 0 dB. At locations with two threshold estimates the average value was used.

The Humphrey Field Analyzer 10-2 program tests 68 points whereas the Matrix 10-2 program tests only 44 points. In order to directly compare the results from both instruments, the test grids from each technique were superimposed and the threshold estimates from the HFA were recalculated to fit the test grid of the Matrix. Figure 2.1 displays the overlapping test grids from all techniques used, with respect to stimulus size. The squares are those of the Matrix, the small closed circles are those of SAP size III, and the larger open circles are those from SAP size V. All threshold estimates from SAP which overlapped onto a single Matrix test point were averaged; this was now the threshold value used for comparison purposes. The center co-ordinates of all overlapping points from SAP and Matrix were less than 2˚ apart. No threshold estimates were excluded from the analysis.

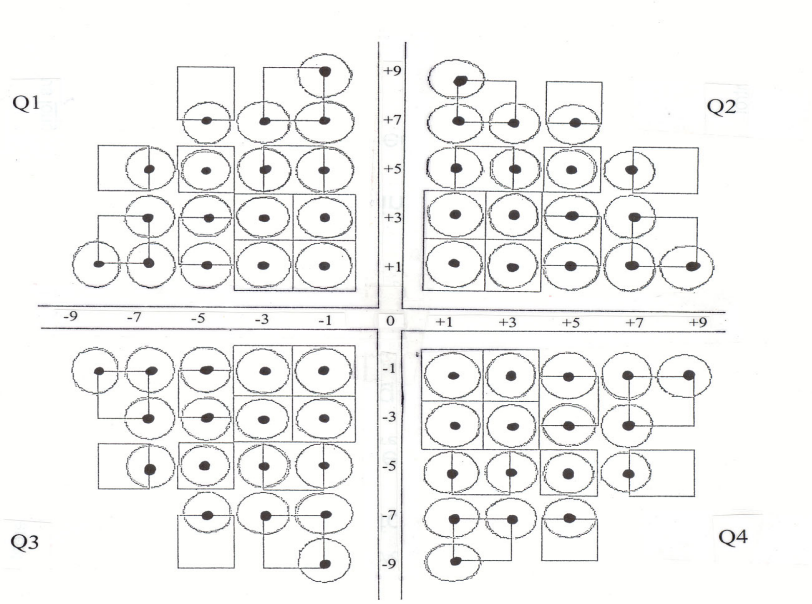

Figure 2.1: Overlapping Coordinates. Stimulus locations of the 10-2 algorithms for both right and left eyes: SAP size III (small closed circles), SAP size V (large open circles), and Matrix size 2° (open squares). For comparison purposes, the 68 test points from the SAP techniques used were recalculated to fit the 44 test grid of the Matrix. All points from SAP which overlapped onto a Matrix test point were averaged; this was now the threshold value used for comparison purposes. The centers of overlapping points from SAP and Matrix were less than 2° apart.

The data from Visits 2 and 3 were compared using ANOVA (Analysis of Variance), test-retest plots, and Bland-Altman plots for within-technique, between-visit data analysis. The same statistical analyses were used to determine the within-visit, between-technique comparison of the data obtained from Visit 3. ANOVA was used to compare the visual field indices and sum of quadrants for the techniques studied.

Plots of test-retest variability analyze the variability of follow-up data with respect to its baseline[77]. Test-retest can only be a valid measurement of variability if no actual change has taken place, i.e. no progression of the glaucomatous visual field. This analysis shows the range of values with the greatest variability upon repeated measures.

Bland-Altman[99] plots analyse the average of two paired variables versus their differences. These plots have been used for method-comparison and validity studies. Calculations of the 5th and 95th percent confidence limits define the interval at which 90% of the follow-up data fall. The narrower an interval is calculated to be, the less variable is the follow-up data, and hence the technique is regarded more repeatable.

The frequency of differences in threshold points between visits was calculated within each technique globally and for threshold values of 32dB, 27dB, 23dB, 18dB, 13dB, and 7dB at Visit 2. The frequency of data points was plotted as a function of decibel difference. The smaller the interval at which the fraction of data points reaches 100%, the less variable is that technique for the specified threshold estimate.

Principal curve analysis (PCA) was used to determine the relationships among the techniques. All data from test locations were excluded if at least one threshold estimate was 0 dB; this was done to avoid floor effects which may occur when the lower limit of any technique is reached[100].

3 Repeatability of Standard Automated Perimetry and Frequency Doubling Perimetry in the Central 10˚ Visual Field of Late Stage Glaucoma

Balian C[1,2], Kourkoutas D[1], Buys YM[1], Trope GE[1], and Flanagan JG[1,2]. [1]Department of Ophthalmology and Vision Science, University of Toronto, Toronto, ON, Canada; [2]School of Optometry, University of Waterloo, Waterloo, ON, Canada.

3.1 Abstract

Purpose: To determine test-retest characteristics of the central 10° visual field as measured by Standard Automated Perimetry (SAP, target size III and V) and Frequency Doubling Perimetry (Matrix) in patients with late stage glaucoma.

Methods: 49 patients with advanced glaucomatous visual field defects attended three visits. During each visit, one eye was examined with Full Threshold SAP 10-2 size III (SAP III), Full Threshold SAP 10-2 size V (SAP V), SITA Std 10-2 size III (SS III), and Matrix 10-2 2˚ stimulus (M2). The Coefficient of Repeatability (CoR) was calculated to determine overall repeatability within each technique. Threshold values from Visits 2 and 3 were averaged and plotted against their differences and the mean of differences and limits of agreement were calculated. Test-retest characteristics were also plotted and compared. Visual field defects were compared using the total and pattern deviation probability maps.

Results: CoR values of SAP III, SAP V, SS III, and M2 were calculated to be 10.33, 9.00, 9.90, and 12.04%dB respectively, relative to the average difference in threshold estimates between visits. M2 had the most uniform test-retest characteristics across the full range of sensitivities; however the 90% confidence interval was the widest of all techniques in the normal to near

normal range (24 to 38dB). M2 showed the greatest defects in both total and pattern deviation probability plots.

Conclusion: Test-retest variability increased with decreasing sensitivity in all SAP techniques. M2 showed constant variability throughout its dynamic range but it was also greatest in the normal to near normal range. This suggests that the M2 would have the most difficulty in following visual field progression in early disease but may be better than SAP techniques in moderate to late stage disease.

3.2 Introduction

Glaucoma is defined as a progressive optic neuropathy, characterized by loss of visual function and often associated with high intra-ocular pressure[4]. It has been suggested that in late stage disease tests of visual function may be the optimum means of detecting progression[101]. Testing the patients' visual function with Standard Automated Perimetry (SAP) is currently the clinical standard for detecting glaucomatous visual field loss[7, 8]. It has also been suggested that moderate to late stage disease should be followed by testing the central 10˚ of the visual field rather than the standard 30˚ field[3]. Furthermore, it has been proposed that the larger Goldmann size V stimulus offers advantages for the evaluation of central field defect and its progression[95]. However there has been concern that the large target size results in apparently shallower defects. A new test algorithm using the Frequency Doubling illusion has been introduced on the Matrix perimeter (Humphrey Matrix; Carl Zeiss Meditech, Dublin CA) that measures the central 10˚ using a 2˚x 2˚ square flickering stimulus[102]. This stimulus has the theoretical advantage of being both a large target, with good repeatability, and being perceptually selective, by preferentially stimulating the magnocellular projecting ganglion cells.

Standard Automated Perimetry (SAP), also known as white-on-white perimetry, projects a white Goldmann size III (0.43˚) stimulus of specific intensity onto a white background with a luminance of 10 cd/m2 (31.5 apostilbs)[10]. Standard AP provides a broadband stimulus that is non-selective with respect to retinal ganglion cell type[42]. The commercially available Humphrey uses a bracketing strategy to estimate threshold levels[41]. In general, the threshold values for all Humphrey strategies are calculated from this repetitive up-and-down staircase technique[10]; the patient's response to the stimulus will determine whether the subsequent stimulus intensity will increase or decrease[48]. This will continue until the patient does not respond to the presented stimulus. When this point is reached, the computer will increase the intensity of the target by a smaller step, if the patient responds then the following stimulus is presented at a lower intensity by an even smaller step. This reversal is repeated several times until consistent responses are obtained. The threshold is calculated with respect to the lowest intensity the patient responds to. The Humphrey perimeter uses a 4-2-2dB staircase, and records the last seen response as the differential light sensitivity.

The Swedish Interactive Test Algorithm (SITA) is a family of test algorithms designed to significantly reduce the test time for threshold estimation, available on the HFA, without any reduction in data quality[7, 29, 50]. Swedish Interactive Threshold Algorithms use maximum likelihood methods to estimate threshold values[51]. Each test location has two likelihood functions each derived from normal and glaucomatous visual fields[51, 53, 54]. At the beginning of each test, threshold values for four predetermined test locations are calculated; these values are then used to calculate the initial intensity for the stimuli to be presented at adjacent test locations[51]. The patient's responses to the stimuli at each location along with previously computed distributions[51] are used to calculate probability distributions; thus, the probability distribution at each location changes with each successive response[51].

The Matrix uses a Zippy Estimation of Sequential Testing (ZEST) algorithm for threshold estimation[67]. This procedure starts with a flat probability density function (pdf)[72]. At each location, a sequence of 4 stimuli are presented, the pdf curve is modified by the patients' responses ("seen" or "not seen") to each stimulus and the threshold is recorded as the mean of the final pdf curve[67]. There are 15 possible outcomes over a range from 0 to 38 dB[67].

The purpose of this study was to determine the within-technique, between-visits repeatability of several visual field techniques available to measure the central 10˚ visual field in patients with late stage glaucoma. In particular, to examine test-retest variability of sensitivity threshold values, visual field indices, and total and pattern deviation probability maps among the following techniques: Full Threshold SAP 10-2 size III (SAP III), Full Threshold SAP size V (SAP V), SITA SAP 10-2 size III (SS III), and Matrix 10-2 2˚ stimulus (M2).

3.3 Methods

49 subjects were recruited from the Glaucoma Service at Toronto Western Hospital, University Health Network, Toronto. The study sample consisted of 30 males and 19 females; 31 right eyes and 18 left eyes. The mean age was 68.4 $\pm$ 9.79 years ranging from 46 to 84 years. Subject recruitment was based on the results of a SITA-Std 24-2 test, performed within the six months prior to recruitment using the criteria of Hodapp et al[91]. The Hodapp criteria for late stage glaucoma states that the following is found when using the Humphrey 30-2 program: MD worse -12 dB, more than 37 points depressed below the 5% level on the PD plot, check if TD or PD *or* more than 20 points depressed below the 1% level on the PD plot. Any point in the central 5˚ visual field with sensitivity of 0 dB and points within the central 5˚ with sensitivity less than 15 dB in both hemifields. Patients with only a temporal or central field remaining in the central 10˚ grayscale plot of the SITA-SAP 24-2 size III test were considered as possible candidates for

participation in the study. Aside from the defect, the patient must have had a reliable test, i.e. false positive (FP), false negative (FN), and fixation losses (FL) must have been less than 20%, 20%, and 30%, respectively.

Patients were excluded if they had one or more of the following: Visual acuity worse than 6/24 (equivalent to 20/80), a visual field defect in the eye being tested that is explained by the patient's ocular status or history other than late stage glaucoma, any history of disease or use of medication that may affect visual field reliability, and inability to undergo any of the perimetry tests. The eye that best fit this inclusion criterion was included in the study, if both eyes were qualified, then the eye to be included in the study was chosen at random.

Refractive error ranged from -9.25 to +5.25 diopters, which was calculated from the patients current prescription. The pupil sizes varied from 2 to 5 mm in diameter without dilation; ½% tropicamide was used to dilate pupils which were less than 3 mm in diameter to ensure that pupil size was not a contributing factor to the visual field defect. Visual acuity ranged from 6/21 (20/70) to 6/6 (20/20). Intraocular pressure was measured by Goldmann applanation tonometry and ranged from 5 to 20 mmHg.

Each subject was scheduled for three visits during which the following four perimetry techniques were performed: i) SAP III, ii) SAP V, iii) SS III, and iv) M2. The order of the tests varied between subjects but remained constant for each subject. The patients were also given ample time between tests to rest in order to minimize the effect of fatigue. All patients were experienced with the SAP and SS III techniques but had no experience with SAP V, or M2. Consequently, data obtained from Visit 1 was discarded to eliminate any bias that may result from the learning effect.

In order to determine repeatability, comparisons were made using global indices, threshold estimates and probability plots. The study was approved by the institutional research ethics boards of both the University of Waterloo and the University Health Network, Toronto.

3.4 Analysis

All threshold estimates with a value less than 0 dB were given a value of 0 dB. At locations with two threshold estimates the average value was used. The patient's age and the severity of the visual field defect were considered as separate between-subject factors. Table 3.1 displays the mean and standard deviation for the visual field indices Mean Deviation (MD) and Pattern Standard Deviation (PSD), and test duration. Due to lack of a normal database, SAP V does not provide global indices. No significant differences were found between visits ($p>0.05$).

Table 3.1: Visual Field Indices and Examination Duration

Index	Visit	Full Threshold Size III	Full Threshold Size V	SITA Standard	SITA Matrix
Mean Deviation (dB)	2	-14.12 ± 7.22	----	-14.24 ± 7.29	-15.35 ± 6.82
	3	-14.23 ± 7.31	----	-14.38 ± 7.59	-15.30 ± 6.97
Pattern Standard Deviation (dB)	2	10.45 ± 2.85	----	10.79 ± 3.31	7.72 ± 2.54
	3	10.58 ± 2.77	----	10.59 ± 2.89	7.43 ± 2.67
Examination Duration (minutes)	2	13:37 ± 2:34	14:53 ± 1:46	7:50 ± 0:55	4:23 ± 0:12
	3	13:27 ± 2:26	14:53 ± 1:41	7:50 ± 1.18	4:21 ± 0:10

Values are the Group Means ± 1 SD.

The Student's t-test was used to compare quadrant sums between visits for each technique (Table 3.2). Calculations of Coefficient of Variability (CoV) and Coefficient of Repeatability (CoR) were made using the results from Visit 2, Visit 3 and Visit 2 & Visit 3 (Table 3.3). Test-retest variability within techniques was plotted as a function of visual field sensitivity (Figure 3.1).

Threshold estimates from Visits 2 and 3 were compared. This analysis illustrates the 5th and 95th confidence limits in which 90% of follow-up threshold estimates are likely to fall, provided no real change has occurred. The frequency of differences in threshold estimates between visits was calculated within each technique globally (Figure 3.2) and for threshold values of 32dB, 27dB, 23dB, 18dB, 13dB, and 7dB, (Figure 3.3) corresponding to available endpoints available when using the Matrix technique, at Visit 2. The frequency of data points was plotted as a function of decibel difference.

Table 3.2: Student's t-test: Visit 2 vs Visit 3; Sum of Quadrants and Mean Sensitivity (MS)

	Q1	Q2	Q3	Q4	MS
FT SAP 10-2, size III	0.68 (N)	0.90 (N)	0.84 (N)	0.79 (N)	0.94(N)
FT SAP 10-2, size V	0.95 (N)	0.85 (N)	0.93 (N)	0.99 (N)	0.74(N)
SITA SAP 10-2, size III	0.74 (N)	0.85 (N)	0.98 (N)	0.81 (N)	0.93(N)
Matrix 10-2, 2˚ stimulus (c. size V)	0.54 (N)	0.79 (N)	0.78 (N)	0.85 (N)	1.00(N)

p-values
R- Reject H_0
N- Not reject H_0

Table 3.3: CoV and CoR Values

Technique	CoV Visit 2	CoV Visit 3	CoR (excluding 0dB)
Full Threshold SAP 10-2, size III	74.67	75.92	10.33
Full Threshold SAP 10-2, size V	51.84	50.07	9.00
SITA SAP 10-2, size III	70.28	71.87	9.90
Matrix 10-2, 2 degree stimulus	79.71	78.28	12.04

All values are in decibels (dB).
CoV – Coefficient of Variability
CoR – Coefficient of Repeatability

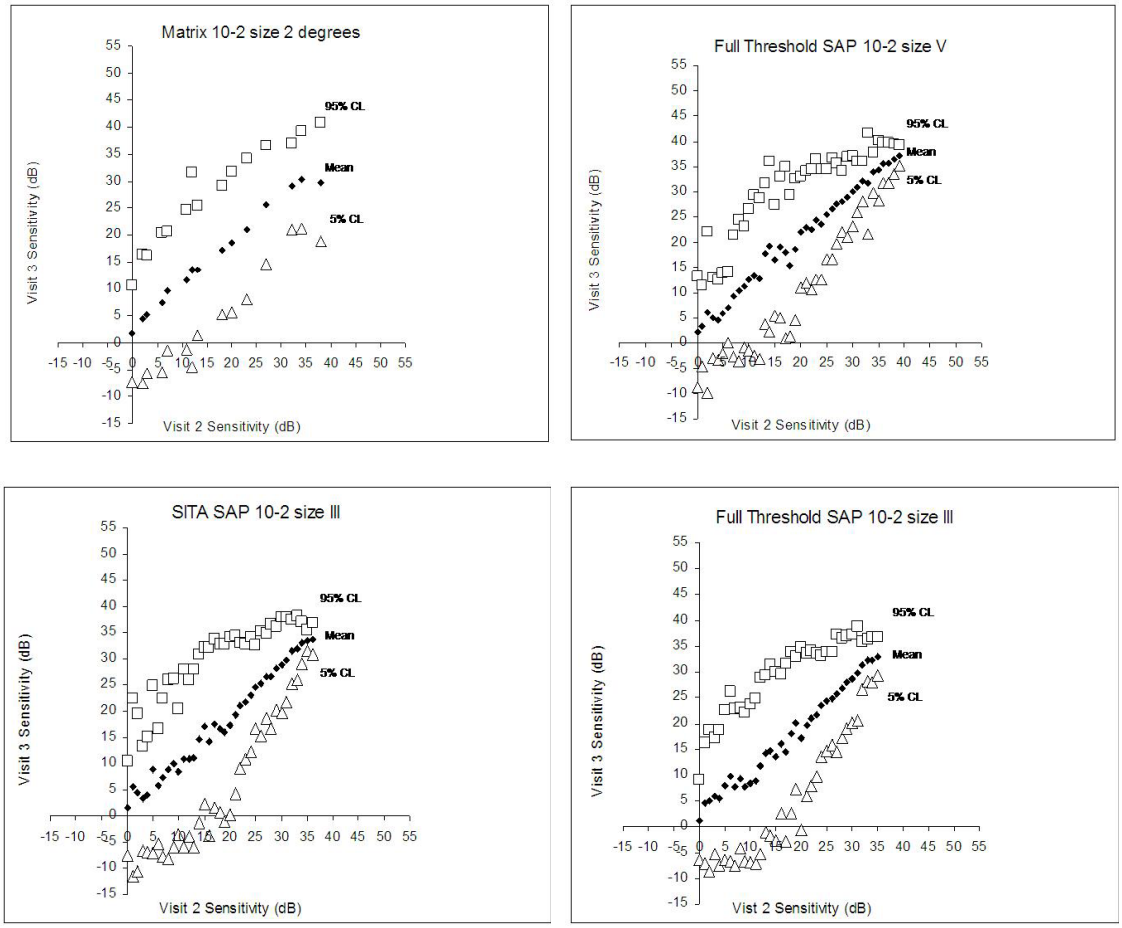

Figure 3.1: Test-retest plots showing the 95th, 50th, and 5th percentiles for the distribution of sensitivity across all test points for each technique. The threshold sensitivity from Visit 3 for each given location is plotted with respect to the threshold sensitivity value of Visit 2 (i.e. within-algorithm, between-visit analysis).

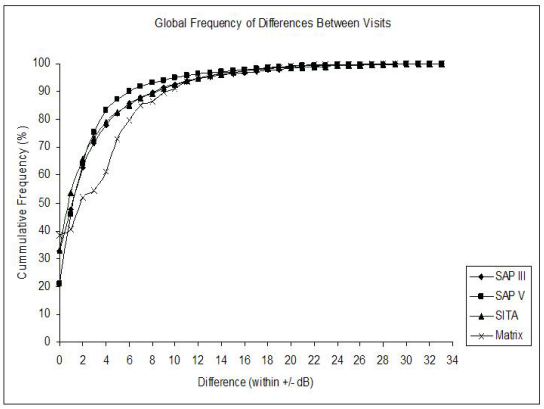

Figure 3.2: Global frequency of difference for all threshold points between Visits 2 and 3.

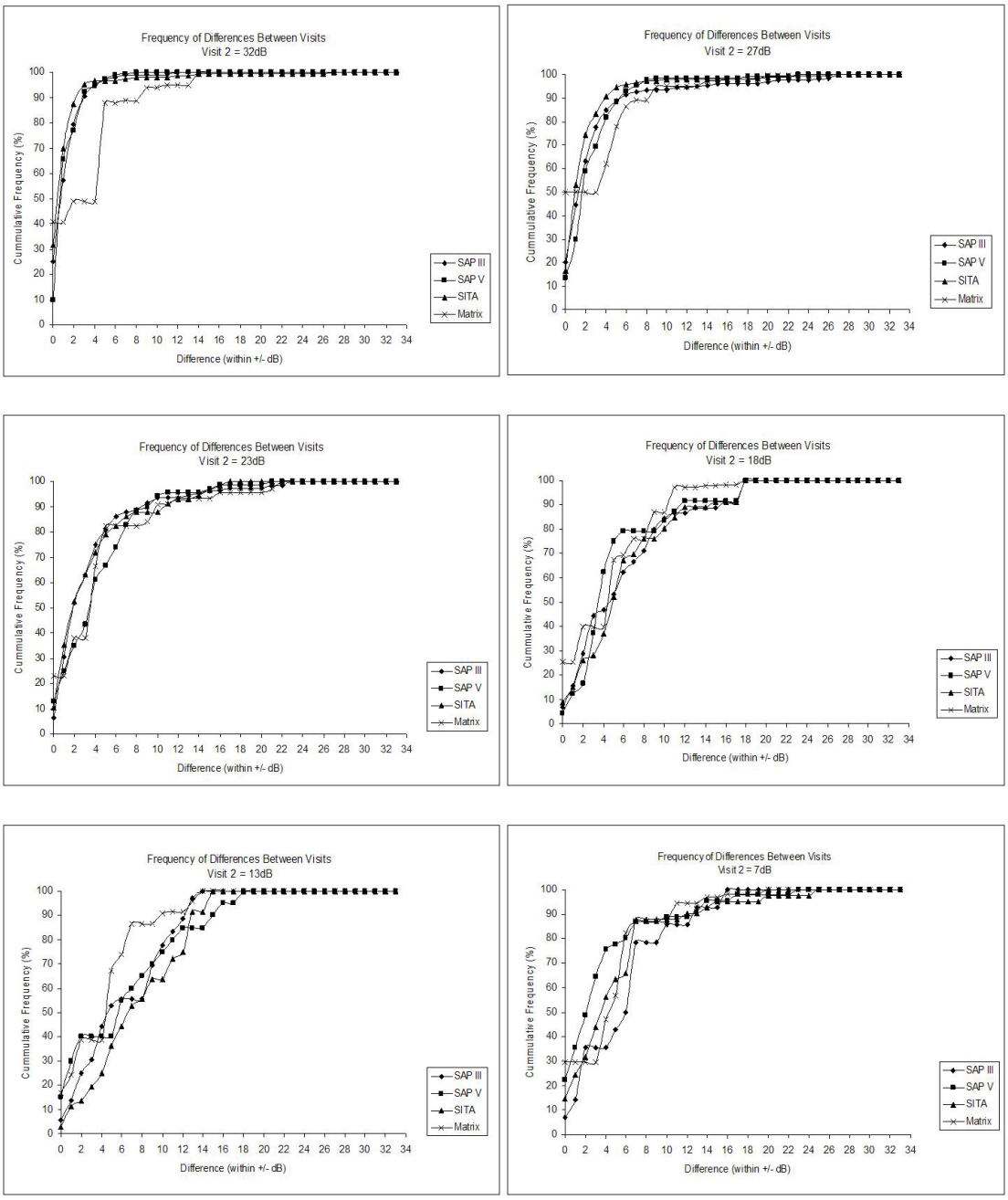

Figure 3.3: Frequency of differences for threshold estimates of 32 dB, 27 dB, 23 dB, 18 dB, 13 dB, and 7 dB at Visit 2.

Total Deviation (TD) and Pattern Deviation (PD) Probability plots from Visits 2 and 3 were compared for each technique (Figure 3.4). Because the SAP V algorithm does not have a normal database, this technique was excluded from the analysis. Each test location was assigned an

ordinal value corresponding to its probability value; p >5%, p <5%, p <2%, p <1%, and p <0.5% were given the values 0, 2, 5, 10, and 10, respectively. The scores were then summed across all test locations to give a defect score for each subject. This technique is similar to the approach used to calculate the Glaucoma Hemifield scores[88]. This defect score was used to compare the total and pattern deviation probability maps between the two visits.

Table 3.4: Statistics from Frequency of Differences Graphs. Maximum dB difference of 90% of data points.

Technique	Global	32 dB	27 dB	23 dB	18 dB	13 dB	7 dB
SAP III	9	3	5	8	15	12	12
SAP V	6	3	5	8	11	15	11
SS III	9	2	4	8	15	13	11
M2	10	8	8	10	10	10	10

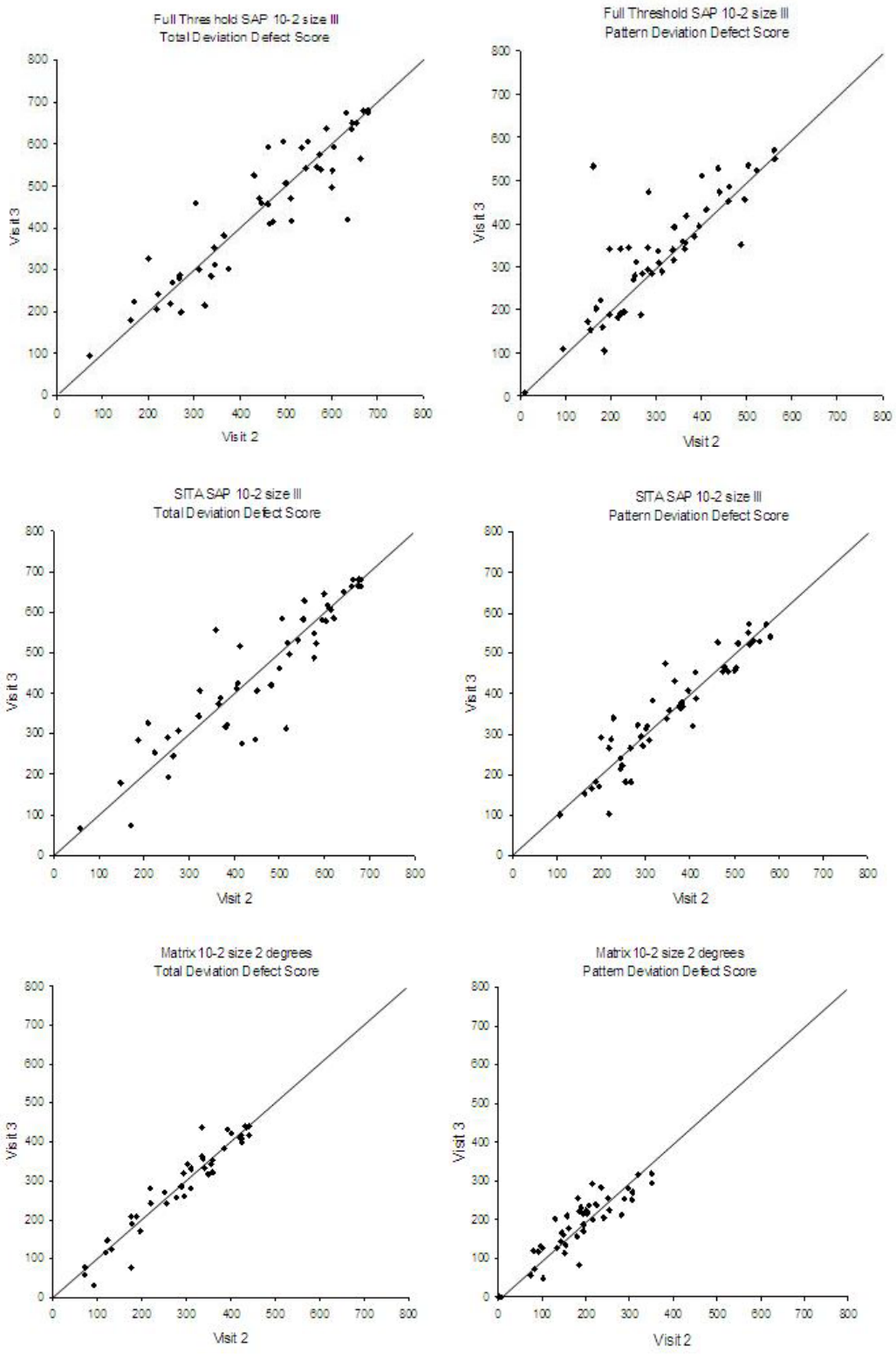

Figure 3.4: Defect scores of total and pattern deviation scores were tallied for each technique. Ordinal scores of 0, 2, 5, 10 and 10 were assigned to the probability calculations at each test location.

Table 3.5: Regression Analysis for Total and Pattern Deviation Probability Plots Ordinal Score Graphs

Technique	Total Deviation	Pattern Deviation
SAP III	0.8418	0.6912
SS III	0.8382	0.8688
M2	0.9359	0.8228

R^2 values

3.5 Results

Table 3.1 lists the global indices and test duration for each technique studied. No significant within test difference was found between Visits 2 and 3 for any of the parameters. Table 3.2 shows the Student's t-test for the sum of quadrants for each technique, and shows no within test difference between visits. CoV and the CoR for all techniques studied are displayed in Table 3.3. SAP V was calculated to have the least variability and best repeatability followed by SS III, SAP III, and then M2 with the highest overall variability and lowest repeatability.

Figure 3.1 displays the test-retest characteristics of each technique. The mean, 95% and 5% confidence limits were plotted for the range values recorded in Visit 3 for a specific stimulus level in Visit 2. If a certain threshold value appeared less than five times during Visit 2 it was excluded from this analysis. These graphs are in agreement with the CoV and CoR calculations shown in Table 3.1. The SAP techniques showed lowest variability in areas of normal to near normal sensitivity; however, variability increased as sensitivity decreased. SAP V was the least variable. The test-retest plot of M2 showed consistent variability throughout the instruments dynamic range.

Bland & Altman plots of repeatability comparing threshold estimates between visits for each technique are displayed in Figure 3.5. The average of the two corresponding points was plotted against their difference. The Limits of Agreement (90% confidence intervals) were least for SAP V and greatest for M2. The mean of differences was approximately zero for all techniques (Table 3.6).

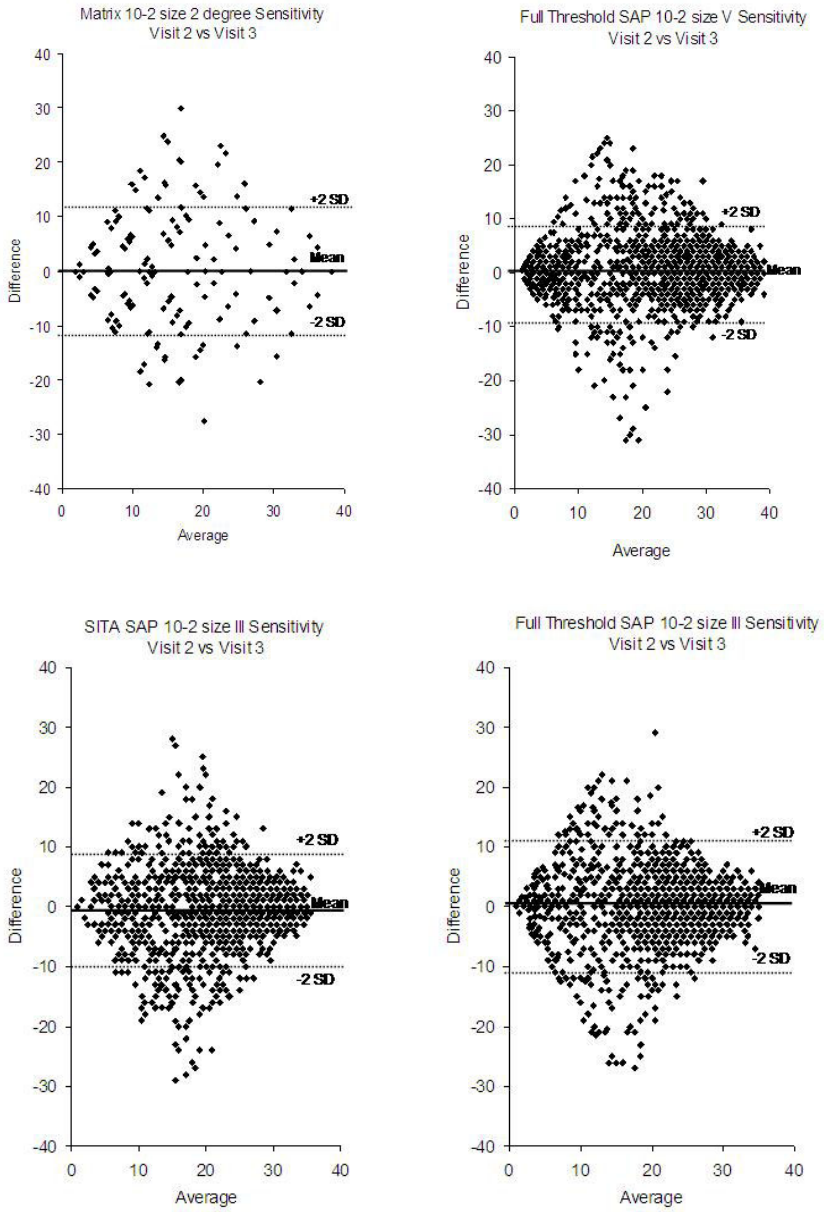

Figure 3.5: The sensitivity values with respect to stimulus location of Visit 2 vs Visit 3 for each technique with the Mean and ±2 standard deviations. The differences between Visits 2 and 3 are plotted with respect to the average of the two sensitivity values. All pairs of data with at least one value of 0dB were excluded from the analysis to avoid any floor effects.

Table 3.6: Statistics from Bland & Altman Plots

Technique	Mean of Differences	+2 SD	-2 SD	Limits of Agreement (90% CI)
SAP III	0.03	10.36	-10.30	20.66
SAP V	0.40	9.40	-8.60	18.00
SS III	-0.12	9.78	-10.02	19.80
M2	0.00	12.00	-12.00	24.00

All values are in decibels (dB).

Figure 3.2 illustrates the cumulative frequency for the difference in threshold estimates between visits. A similar analysis was performed for six threshold levels: 32dB, 27dB, 23dB, 18dB, 13dB, and 7dB (Figure 3.3). The Matrix had the highest percentage of datapoints with a difference of 0 dB between visits. The fraction of datapoints with 0dB difference between visits decreased with decreasing sensitivity for all techniques; this value was greatest for the M2 technique.

Figures 3.2 and 3.3 plot the frequency of data points as a function of dB difference between visits. Globally, M2 showed the greatest fraction of threshold locations with 0 dB difference between visits, but by 2 dB it was the worst performer. 90% of the data for all techniques showed, at most 10 dB difference between visits. The relative performance of each technique changes at different stages of the measurement dynamic range. At 32 and 27 dB, 90% of the data from SAP showed less than 3 and 5 dB difference, respectively; whereas 90% of data from M2 showed a maximum difference of 8 dB. In the moderate to moderately-severe threshold estimates, the dB difference between visits for 90% of the data started to increase. At 13 dB, M2 showed a greater number of points with less difference than SAP. The dB difference for M2 is fairly consistent throughout the threshold estimates.

Figure 3.4 illustrates the correlation between the total and pattern deviation probability map defect scores for each technique. All techniques showed good repeatability upon retest.

3.6 Discussion

Test-retest variability in perimetry is dependent upon stimulus characteristics such as target size and color[97], the stage of visual field defect being studied[77] and the algorithm used to estimate the sensitivity[52]. Previous reports have shown disturbingly high levels of test-retest variability, particularly in areas of moderate to severe defect, implying that the following of disease progression is problematic[77, 103]. It is therefore important to determine the variability within a technique to decide whether or not it might be suitable to measure visual field loss and/or progression. Previous studies have claimed that the Matrix has uniform test-retest variability across its dynamic range in patients with early to moderate glaucomatous visual field loss in comparison to standard automated perimetry[67]. The main objective of this study was to determine repeatability and the test-retest characteristics of M2 and SAP in techniques used to measure the central 10˚ visual field of patients with late stage glaucoma.

Test duration showed no significant difference between visits; the techniques with a normal database also showed no significant difference in any of the global indices. The CoV and CoR were calculated to give a global analysis of the variability and repeatability for each technique across the full range of data. The Matrix gave the highest scores in each measurement thereby suggesting that it was the most variable and least repeatable technique. Of the HFA techniques SAP V gave the most repeatable results due to its larger size and greater spatial summation[97], but surprisingly SAP III was more repeatable than SS III. However, these measures summate data from the entire dynamic range. The point-by-point analysis in Figure 3.1 describes in detail the repeatability at specific sensitivity estimates. Overall, variability of SAP techniques follow the same trend: it decreases as sensitivity increases. However, variability of this trend is seen among its techniques. SAP III has the widest variability and SAP V has the narrowest. Size V showed

least variability upon retest as it has the potential of stimulating a larger receptive field within the retina. This agrees with previous studies that Goldmann size V is more suitable for monitoring visual field progression in late stage glaucoma than either technique using size III stimuli[104].

The Bland and Altman graphs illustrate the large test-retest values in the mid-range of defect for all instruments. The Matrix gave the largest Limits of Agreement, but the smallest Mean of Differences. This tells us that there is no difference between the means of Visits 2 and 3; however, variability of M2 was greater than any SAP technique. Interestingly the Matrix also showed a peculiar clustering of points due to the truncated options for final threshold estimation. This is due to the nature of the 4 step Zest procedure, which results in 15 specific choices for the final threshold estimation (0, 2, 3, 6, 7, 11, 12, 13, 18, 20, 23, 27, 32, 34, 38)[72].The distribution of points on the Bland & Altman graphs restates that areas of near normal to normal sensitivity are more repeatable with SAP. All SAP techniques show upper limits of their range of repeatability, particularly in the 10 to 20 dB range, to be no better than chance for measuring the same sensitivity at repeated visits.

The test-retest graphs for the various SAP techniques in Figure 3.1 show good repeatability in the normal range but are greatest at locations of moderate to severe abnormality. In the normal range, M2 has the greatest test-retest characteristics but has the advantage of being constant throughout its dynamic range. This would imply that it would perform relatively poorly with respect to repeatability in the normal to near normal range but better than the SAP techniques in the moderate to severe defect range. This might be considered ironic given that frequency doubling is promoted as a technique optimized for early detection.

Frequencies of dB differences between visits are illustrated in Figures 3.2 and 3.3. Globally, M2 had the highest fraction of 0 dB difference, but the lowest proportion within ±2dB . This can be

attributed to the fact that M2 has a smaller number of final threshold estimates compared to SAP (i.e. 15 vs 39); hence the convergence of raw threshold values to the specific threshold estimates is greater than that of SAP. The graphs which show differences at specific threshold estimates confirm that variability increases as sensitivity decreases with SAP techniques. Size V showed least variability upon retest as it has the potential of stimulating a larger receptive field within the retina.

Comparison of defect scores from total and pattern deviation probability maps showed good overall repeatability within each technique. In comparison to each other, M2 calculated test points within defects to be shallower than the SAP techniques. However, this analysis has limited significance as the defect score of SAP techniques was calculated from 68 points and the Matrix' from 44.

In conclusion, this study showed that test-retest variability of the SAP techniques decreases with increasing sensitivity in patients with late stage glaucoma. SAP III has the widest variability and SAP V has the narrowest. This agrees with previous studies which show that Goldmann size V is more suitable for monitoring visual field progression in late stage glaucoma than either technique using size III stimulus[104]. However, variability was confirmed to be constant throughout the dynamic range for M2 but is worst in the normal to near-normal sensitivity range. This suggests that M2 will be disadvantaged for the detection of early visual field loss but better positioned to repeatably detect and follow moderate to severe loss, compared to SAP when considering the central 10˚ of patients with late stage glaucoma.

4 Comparison between Standard Automated Perimetry and Frequency Doubling Perimetry in the Central 10° Visual Field of Late Stage Glaucoma

Balian C[1,2], Kourkoutas D[1], Buys YM[1], Trope GE[1], and Flanagan JG[1,2]. [1]Department of Ophthalmology and Vision Science, University of Toronto, Toronto, ON, Canada; [2]School of Optometry, University of Waterloo, Waterloo, ON, Canada.

4.1 Abstract

Purpose: To compare results of Standard Automated Perimetry (SAP) and Frequency Doubling Perimetry (Humphrey Matrix; Carl-Zeiss Meditech, Dublin, CA) in patients with late stage glaucoma.

Methods: 49 patients with advanced glaucomatous visual field defects were enrolled in the study. Each patient attended three visits. During each visit, one eye was examined with Full Threshold SAP 10-2, size III (SAP III); Full Threshold SAP 10-2, size V (SAP V); SITA Std 10-2, size III (SS III); and Matrix 10-2, 2° stimulus (M2). Results from Visits 2 and 3 were compared. The M2 test grid pattern was used to compare results, with the SAP techniques being summated for spatial equivalence. Mean Deviation (MD) and Pattern Standard Deviation (PSD) were compared using the limits of agreement and regression analysis. Principal curve analysis compared threshold estimates between techniques. Total and pattern deviation probability maps were also compared.

Results: A correlation was noted among MD but not PSD values. Threshold estimates of SAP III and SS III were shown to be similar and slightly more variable than SAP V. M2 had the widest 90% confidence interval in the normal to near normal sensitivities and showed a constant variability throughout its dynamic range. SAP showed greater defects than M2 in pattern deviation probability plots.

Conclusion: M2 gives a lower sensitivity than the SAP techniques, but identifies fewer abnormal test locations than SS or SAP III. This is likely due to the relatively poor test-retest characteristics in the near normal range. However; it may better suited to following progression in the moderate to severe defect range.

4.2 Introduction

Glaucoma is characterized as a progressive optic neuropathy associated with visual function loss[4]. Measurement of visual function by static perimetry is an integral part of the diagnosis and management of the disease[9]. In patients with late stage glaucoma, treatment is aimed at prolonging the function of the central 10˚ of vision. Testing the patients' visual field using standard automated perimetry (SAP) is currently considered to be the gold standard for detecting glaucomatous visual field loss[7, 8]. Standard automated perimetry, also known as white-on-white perimetry, projects a white stimulus of specific intensity onto a white background with a known luminance, often the Goldmann standard 31.5 apostilbs ($10cdm^{-2}$)[10]. More recently there have been other types of stimuli available to test and monitor visual function in glaucoma. One such stimulus uses the frequency doubling (FD) illusion, in which a grating of low spatial frequency is flickered at a high temporal frequency[57], giving the percept of a doubling of spatial frequency[3]. It has been shown that the FD illusion stimulus used for perimetry is principally a flicker stimulus[105] and is processed cortically[64].

In the advanced stages of glaucoma, large arcuate scotomas from the superior and inferior field connect leaving only the central and/or temporal visual field intact[11]. Testing only the central 10˚ visual field and using a Goldmann size V stimulus have both been suggested as better methods for monitoring progression in these patients[43, 104].

The primary objective of this study was to determine the differences between perimetry techniques when measuring visual field sensitivity in patients with late stage glaucoma (within-visit, between-technique analysis).

4.3 Methods

The sample consisted of 49 patients with glaucoma recruited from the Glaucoma Department, Toronto Western Hospital, University Health Network, Toronto. There were 30 males and 19 females; 31 right eyes and 18 left eyes. The mean age was 68.4 $\pm$ 9.79 years ranging from 46 to 84 years. Subject recruitment was based on prior visual fields measured within 6 months (SITA-SAP 24-2). Patients were considered as possible candidates for participation in the study if they had only a temporal or central island of vision remaining within the central 10˚. Patients were also required to have a reliable test, defined as a false positive (FP) and false negative (FN) score of less than 20% and fixation losses (FL) of less than 30%.

Patients were excluded if they had one or more of the following: visual acuity worse than 6/24 (20/80), a defect in the visual field of the eye being tested other than glaucoma, that is explained by ocular status or history, any history of disease or use of medication that may affect visual field reliability, and inability to undergo any of the perimetry tests. The eye that fit these inclusion and exclusion criteria was included in the study, if both eyes qualified, then the eye to be included in the study was chosen at random.

Each subject was scheduled for three visits during which the following 4 perimetry techniques were performed: i) SAP III, ii) SAP V, iii) SS III, and iv) M2. The order of the tests varied

between subjects but remained constant for each subject to avoid any carryover effect that may occur amongst the perimetry techniques. The patients were also given a minimum of 5 minutes to rest between tests in order to minimize the effect of fatigue.

All patients had prior experience with SS III and/or SAP III, but no experience with either SAP V or M2. Data obtained from Visit 1 was discarded to eliminate bias that may result from the effect of learning. This study was approved by the institutional research ethics boards.

4.4 Analysis

All threshold estimates with a value less than 0 dB were given a value of 0 dB. At locations with two threshold estimates the average value was used. The age and severity of the visual field of each subject were considered as separate between-subject factors.

The Humphrey Field Analyzer tests 68 points with its 10-2 program whereas the Matrix tests only 44 points. To directly compare results from both instruments, the test grid from each were superimposed and the threshold estimates from the HFA were recalculated to fit the test grid of the Matrix. Figure 4.1 displays the overlapping test grids from all techniques used, with respect to stimulus size. All threshold estimates from SAP which overlapped onto a single Matrix test point were averaged; this value was used as the threshold estimate for that test location. The center co-ordinates of all overlapping points from SAP and Matrix were less than 2˚ apart. No threshold estimates were excluded from the analysis.

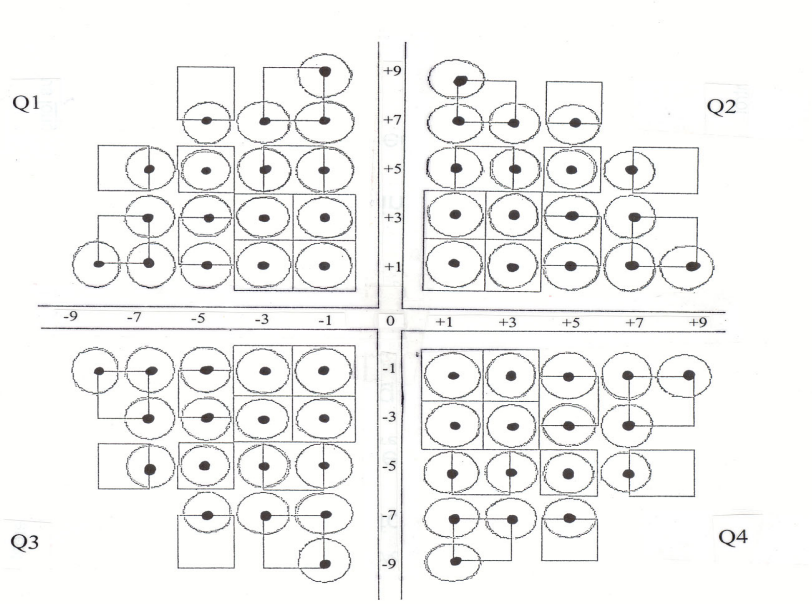

Figure 4.1: Overlapping Coordinates. Stimulus locations of the 10-2 algorithms for both right and left eyes: SAP III (small closed circles), SAP V (large open circles), and M2 (open squares). For comparison purposes, the 68 test points from the SAP techniques used were recalculated to fit the 44 test grid of M2. All points from SAP which overlapped onto a Matrix test point were averaged; this was the threshold value used for comparison among the techniques. The centers of overlapping points from SAP and M2 are less than 2° apart.

Analysis of Variance (ANOVA) was used to compare the mean deviation (MD) and pattern standard deviation (PSD) from Visit 3 of SAP III, SS III, and M2 stimulus techniques. Quadrant sums of sensitivity were also performed and included SAP V. Regression analysis was used to compare, MD and PSD, and test duration among the different techniques.

Threshold estimates among the techniques were compared using principal curve analysis[100] which calculates the line of best fit while minimizing residuals of both variables[67]. The average threshold estimates from Visits 2 and 3 were used for each test location. Data sets from test locations with a value of 0 dB were excluded to eliminate an artifactual decrease in test variability and to avoid a bottoming effect which occurs when the lower limits of the dynamic ranges are reached[67].

Point-by-point analysis was used to show the within-visit, between-algorithm variability between the techniques. This analysis outlines the 5th and 95th confidence limits within which 90% of the threshold estimates from the two techniques being compared are likely to fall. The intervals for each pair of techniques were calculated by comparing threshold estimates from Visit 3. Coefficient of Repeatability (CoR) and Coefficient of Variability (CoV) were also calculated, with respect to threshold estimates, for each of the techniques being studied.

Visual field defects as recorded on Total Deviation (TD) and Pattern Deviation (PD) probability plots were compared between the techniques. Because SAP V algorithm does not provide probability maps, this technique was excluded from this analysis. Each test location was assigned an ordinal value corresponding to its probability value; $p > 5\%$, $p < 5\%$, $p < 2\%$, $p < 1\%$, and $p < 0.5\%$ were given the values 0, 2, 5, 10, and 10, respectively. This technique is similar to the approach used to weight the Glaucoma Hemifield scores[88]. The 44 data points from the Matrix was recalculated to fit the 68 point grid from SAP; this was to endure that all data points were assigned one of the predetermined ordinal values. The ordinal values were summed for each visual field resulting in a defect score. This defect score was used to compare the TD and PD probability maps among the three techniques.

4.5 Results

Refractive error ranged from -9.25 to +5.25 diopters. The pupil sizes varied from 2 to 5 mm in diameter without dilation; tropicamide ½% was used to dilate pupils which were less than 3 mm in diameter to ensure that pupil size was not a contributing factor to the visual field defect. Visual acuity ranged from 6/21 to 6/6. Intraocular pressure ranged from 5 to 20 mmHg, by

Goldmann applanation tonometry. False positive and false negative values ranged from 0 to 20% while fixation losses ranged from 0 to 30%.

There was no significant difference between techniques for MD but there was a significant difference for PSD (Table 4.1). Table 4.2 displays the p-values from ANOVA for the quadrant sums of all the techniques. The results from Visits 2 and 3 were similar and show a strong correlation. The threshold values for all the techniques show similar results with respect to each quadrant, irrespective of stimulus type and/or size. There was a strong correlation for MD among the three techniques (Table 4.3). However, this relationship was not seen with PSD (Table 4.4). As might be expected, the strongest correlation was between SAP III and SS III (r^2= 0.8439, p<0.00), which were designed to give similar results. This relationship has been reported previously[1, 109]. SS III and M2 showed the greatest difference in their PSD calculations (r^2= 0.4925, p<0.00).

Table 4.1: ANOVA: Full Threshold SAP10-2 size III, SITA SAP 10-2 size III, and Matrix 10-2

	MD	PSD
Visit 2	0.6396 (N)	<0.0001 (R)
Visit 3	0.7348 (N)	<0.0001 (R)

P-values
R- Reject H_0
N- Not reject H_0

Table 4.2: ANOVA: Full Threshold SAP10-2 size III, Full Threshold SAP 10-2 size V, SITA SAP 10-2 size III, and Matrix 10-2

	Q1	Q2	Q3	Q4
Visit 2	<0.0001 (R)	<0.0001 (R)	<0.0001 (R)	<0.0001 (R)
Visit 3	<0.0001 (R)	<0.0001 (R)	<0.0001 (R)	<0.0001 (R)

P-values
R- Reject H_0
N- Not reject H_0

Table 4.3: Mean Deviation Regression Analysis

	SITA SAP 10-2 size III	Matrix 10-2 size 2˚
FT SAP 10-2 size III	$r^2 = 0.9434$ ($p<0.00$)	$r^2 = 0.7016$ ($p<0.00$)
SITA SAP 10-2 size III	n/a	$r^2 = 0.6835$ ($p<0.00$)

Table 4.4: Pattern Standard Deviation Regression Analysis

	SITA SAP 10-2 size III	Matrix 10-2 size 2˚
FT SAP 10-2 size III	$r^2 = 0.8439$ ($p<0.00$)	$r^2 = 0.4925$ ($p<0.00$)
SITA SAP 10-2 size III	n/a	$r^2 = 0.3514$ ($p<0.00$)

Principal curve analysis was used to determine the relationship of threshold estimates between techniques. In Figure 4.2 (top left), a fairly linear relationship was seen between the average threshold estimates of SAP III and SS III, where the correlation was strongest in the normal sensitivity values and weakest within the 5 to 20 dB range.

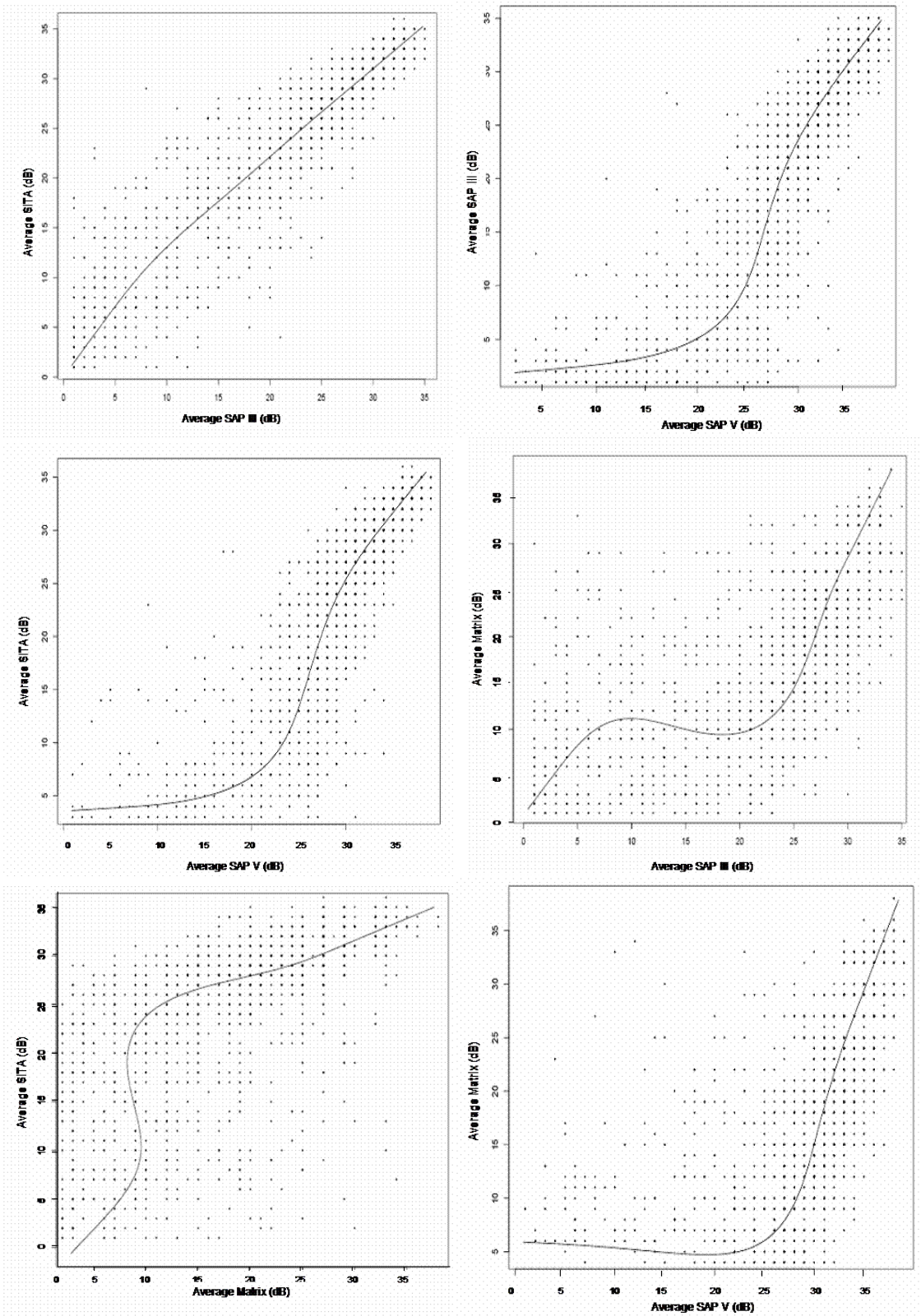

Figure 4.2: Principal curve analysis for comparison of thresholds among techniques. For each pair of thresholds, if at least one value was 0 dB then the pair was excluded from this analysis.

Differences in threshold estimates as a result of using different stimulus size with SAP, is seen in the graphs comparing SAP V with SAP III (Figure 4.2, top right) and SS (Figure 4.2, middle left). Both curves have slopes of approximately 1.5-2.0 at threshold estimates of >25dB as recorded by SAP V. At sensitivity values <25dB, the slope of the curve rapidly approaches 0. Hence, correlation between techniques decrease considerably as sensitivity is reduced.

Comparison of SAP V and M2 is seen in Figure 4.2, bottom right. With respect to this curve, M2 shows damage numerically three times more severe than SAP V in areas of normal sensitivity. However, at sensitivity estimates <30dB, the slope of the curve rapidly approaches 0. Hence, there is no agreement between the two techniques at sensitivity values <30dB.

In comparison to SAP III and SS, the distribution of M2 is divided into three sections (Figures 4.2 middle left and bottom right). In sensitivity values >25dB as estimated with SAP III and SS, the slope of the curves are between 2.0 and 3.0. Between 10 and 25dB the correlation is very weak, as the slope rapidly approaches 0. Sensitivity values below 10dB, correlation increases to 1.5 and 2.0 times for SAP III and SS, respectively.

Figure 4.3 displays the mean, 95% and 5% confidence limits of threshold sensitivity values for each pair of techniques being compared. This will show the variability in sensitivity obtained from one technique to another. If a certain threshold value appeared less than five times in the entire threshold estimates, it was excluded from this analysis. With techniques available on SAP, variability is reduced in areas of near-normal to normal sensitivity. In comparison to M2, variability is constant throughout the sensitivity range, hence greatest in the near-normal to normal range of sensitivity values.

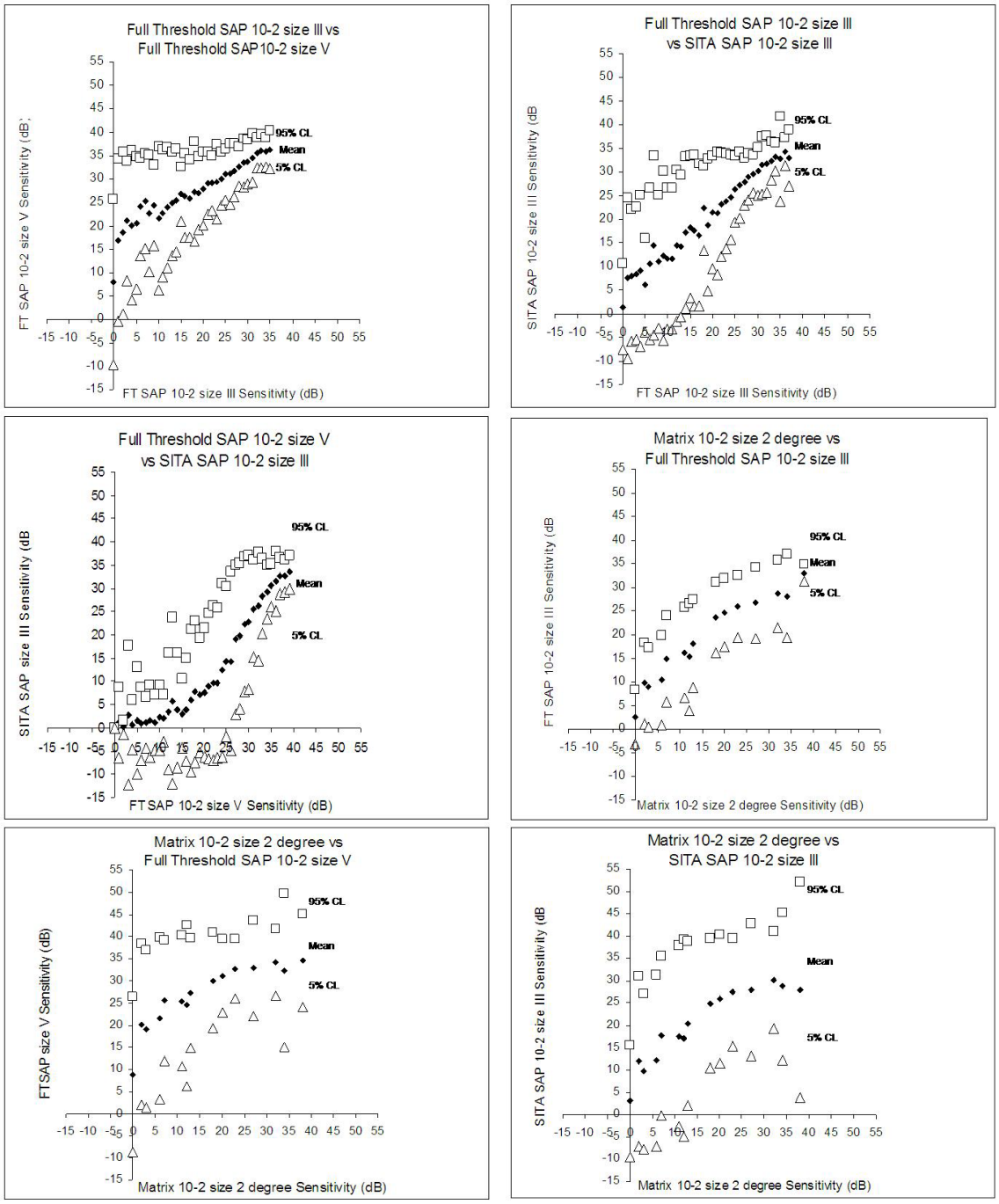

Figure 4.3: The 95th, Mean, and 5th percentiles for the distribution of sensitivity across all test points among the techniques. The sensitivity between each technique from Visit 3 for each given location is plotted with respect to the threshold value of another technique (i.e. within-visit, between-algorithm analysis).

Figure 4.4 displays the Mean and $\pm$ 2 standard deviations for comparing sensitivity values obtained within the techniques during Visit 3. The average of the two points is plotted against the

dB difference between them. The results from the Bland & Altman graphs are in compliance with the test-retest plots and CoR & CoV calculations (Table 4.5). Coefficient of Repeatability values of SAP III, SAP V, SS III, and M2 were calculated to be 10.33, 9.00, 9.90, and 12.04%dB respectively, relative to the average difference in threshold estimates between visits. Coefficient of Variability of SAP III, SAP V, SS III and M2 at Visit 3 were calculated to be 75.92, 50.07, 71.87 and 78.28, respectively.

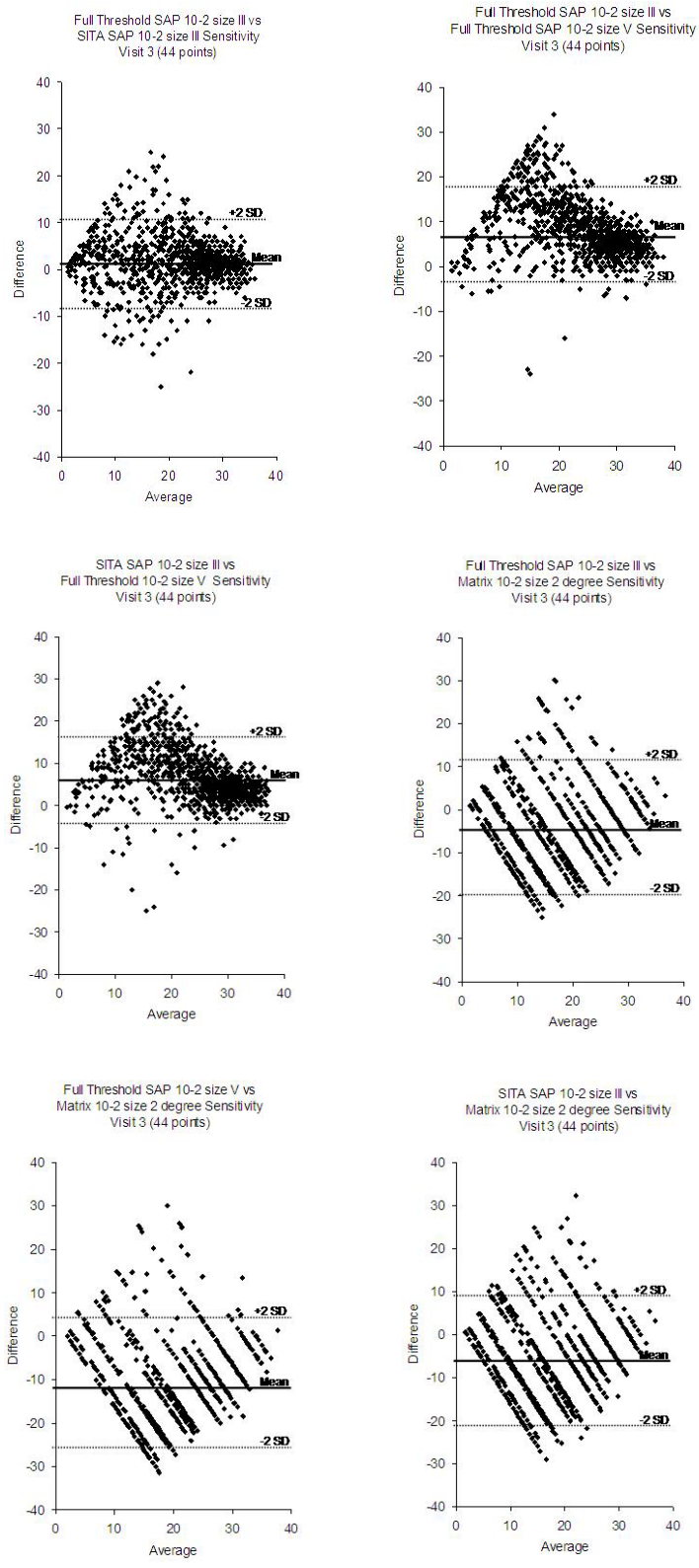

Figure 4.4: The sensitivity values with respect to stimulus location of Visit 3 within techniques with the Mean and ±2 standard deviations. The difference between threshold estimates of the techniques are plotted with respect to the average of the two sensitivity values.

Table 4.5: CoV and CoV Values

Technique	CoV Visit 2	CoV Visit 3	CoR	CoR (without 0dB)
Full Threshold SAP 10-2, size III	74.67	75.92	11.05	10.33
Full Threshold SAP 10-2, size V	51.84	50.07	9.50	9.00
SITA SAP 10-2, size III	70.28	71.87	10.95	9.90
Matrix 10-2, 2 degree stimulus	79.71	78.28	11.77	12.04

All values are in decibels (dB).
CoV – Coefficient of Variability
CoR – Coefficient of Repeatability

Total Deviation (TD) and Pattern Deviation (PD) probability plots were compared using ordinal scoring in Figure 4.5. Probability plots of SAP III and SS III are shown to give nearly identical results. In comparison to M2, both SAP III and SS III show more severe defects in the PD probability plots.

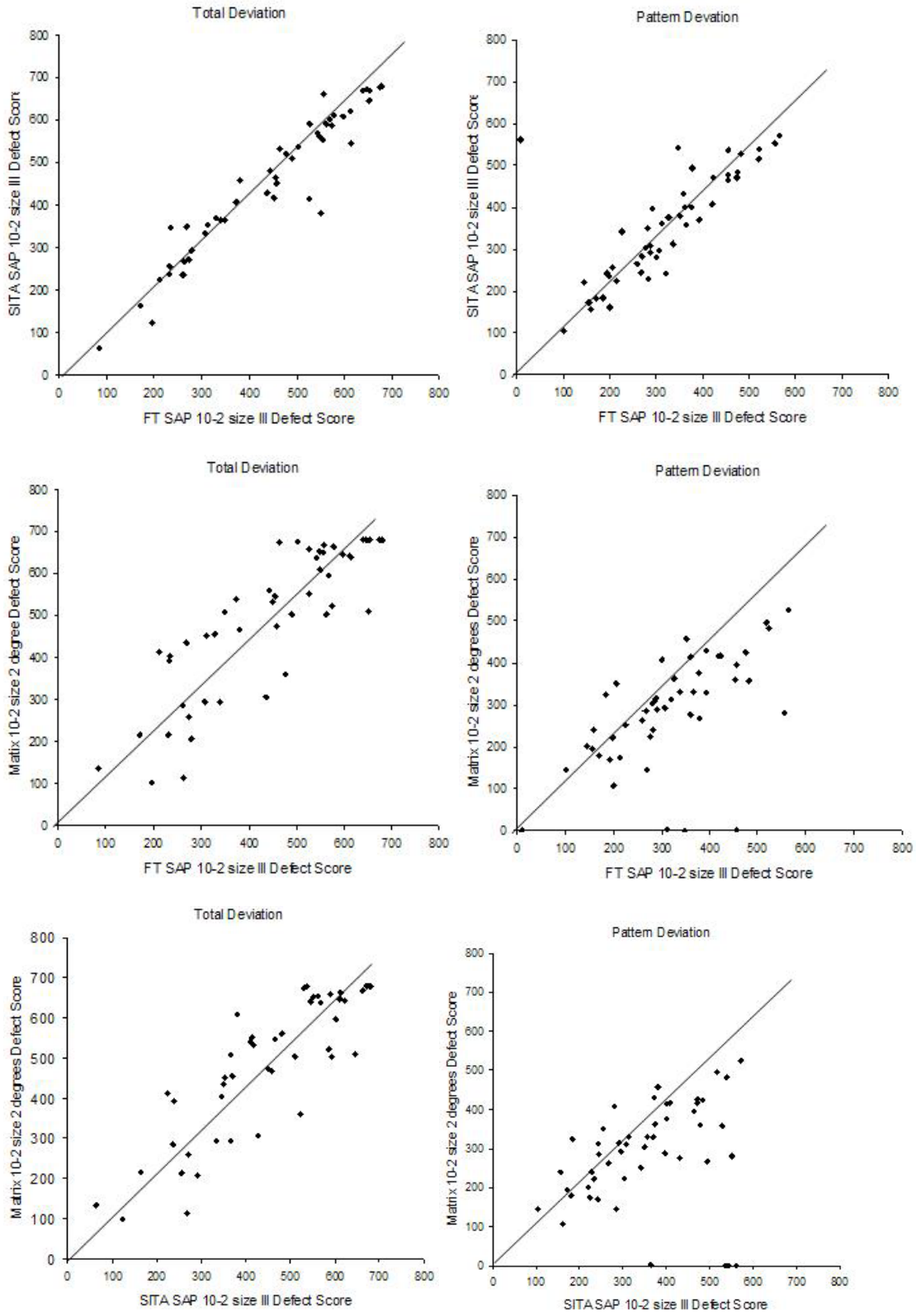

Figure 4.5: Defect scores of total and pattern deviation probability maps for SAP III, SS III, and M2. Ordinal scores of 0, 2, 5, 10 and 10 were assigned to the probability calculations at each test location.

Table 4.6 displays the regression analysis for test duration among the techniques studied. No significant difference was found between Full Threshold techniques (SAP III and SAP V) and the Matrix in comparison to SAP V and SS.

Table 4.6: Test Duration Regression Analysis

	FT SAP 10-2 size V	SITA SAP 10-2 size III	Matrix 10-2 size 2˚
FT SAP 10-2 size III	r^2= 0.0550 (p=0.1048)	r^2= 0.1301 (p=0.0109)	r^2 = 0.1273 (p=0.0118)
FT SAP 10-2 size V	n/a	r^2 = 0.1989 (p=0.0013)	r^2= 0.0387 (p=0.1754)
SITA SAP 10-2 size III	r^2= 0.1989 (p=0.0013)	n/a	r^2= 0.0176 (p=0.3640)

Table 4.6 displays the r^2 and p-values for test duration regression analysis.

4.6 Discussion

Previous studies have demonstrated uniform test-retest variability for M2 over its entire dynamic range in patients with early to moderate visual field loss[67]. We examined patients with late stage glaucoma using perimetry techniques which measure the central 10˚ visual field using both Frequency Doubling Perimetry and Standard Automated Perimetry.

Coefficients of variability and repeatability were calculated to give a global analysis for each technique across the full range of data. The Matrix gave the highest scores in each measurement thereby suggesting that it was the most variable and least repeatable technique. Of the HFA techniques SAP V gave the most repeatable results due to its larger size and greater spatial summation[97], but surprisingly SAP III was more repeatable than SS III as longer test duration is expected to result in patient fatigue. However, these measures summate data from the entire

dynamic range. The point-by-point analysis in Figure 4.3 describes in detail the repeatability at specific sensitivity estimates.

In order to compare the decibel scales of both instruments, it should be noted that a 1dB change in threshold estimate is equal to 0.05 log units of contrast sensitivity for the FDP and 0.1 log units for SAP[67]. Test-retest intervals for comparison between techniques are dependent upon the measurement scale, and therefore should be taken into account when comparing M2 results with that of SAP. Principal curve analyses comparing the techniques show just that. An approximately linear correlation is seen between SAP III and SS III. Hence, graphs comparing SAP V with SAP III and SS III are fairly similar. With sensitivity >25 dB, there is a strong correlation between the two techniques, in both comparisons, a slope of approximately 2.0, yet SS III and SAP III both calculate points as being more defective than SAP V at threshold estimates of <25 dB. The difference in threshold estimate is due to the difference in stimulus size used for each technique. The Goldmann stimulus size V is approximately 16 times larger than that of size III and therefore is capable of stimulating a larger receptive field in the retina allowing for the patient to see "more" of the points than when a Goldmann size III stimulus is used. Despite the fact that the stimulus size for M2 and SAP V are almost the same size, large variability is seen between threshold estimates of these two techniques; in areas of SAP V sensitivities of >25 dB there is a strong linear correlation with a slope of approximately 3.0. Figures 4.2 middle left and bottom right are in agreement with the study by Artes et al[67] where threshold estimates between SAP (techniques with size III stimulus) and Matrix are being compared; in points with high sensitivity thresholds (>25 dB) a strong linear correlation is seen between the instruments; a slope of approximately 2.0. This difference in sensitivity estimates is attributed to the difference in dB scale used by each instrument.

Point-by-point analysis (Figure 4.3) and Bland & Altman graphs (Figure 4.4) of threshold estimates showed variability among the techniques. Comparing SAP III with SAP V showed that variability between the two techniques decrease at locations with sensitivity values greater than 25 dB; SAP III shows more points outside normal limits with respect to threshold estimates less than 25 dB. This was as expected as the Goldmann size V stimulus is 16 times larger than Goldmann size III; hence, greater spatial summation. A similar trend is seen when SAP III and SS III are compared however, threshold estimates at the lower dynamic range are less variable than the first comparison. Threshold estimates between these two techniques are shown to be very similar; this was expected as previous studies have shown that the SITA algorithm decreased test time without decreasing test reliability[51]. SAP V vs SS III showed decreased variability at both ends of the dynamic range with SS III recording more test locations outside normal limits. Again, this is due to the difference in the size of the stimulus. Test-retest analysis of SAP techniques plotted against threshold estimates of the M2 showed consistent variability throughout the dynamic range. Compared to SAP III, the M2 estimates threshold points as less severe. The same is true for most points on the graph comparing the M2 with SITA SAP III. SAP V compared with the M2 showed that at threshold estimates of >20 dB M2 was estimated as approximately 30 dB.

Comparing the defect score graphs of the total and pattern deviation probability maps indicate that the techniques available on SAP graded more test locations outside normal limits than that of M2. This may be due to the difference in normal database, such as the inclusion criterion for the "normal" population[106], available for each technique and the difference in target size and properties between the Goldmann size III and the Matrix 2 degrees.

As expected, significant difference was seen in test duration between Full Threshold techniques when compared to SITA[7, 29, 50]. No difference was seen between SAP III and SAP V. Even with

the difference in stimulus size, no significant difference was noted between SAP III and SAP V; this was also seen with SITA and M2. Despite the fact that SAP III and SAP V had no significant difference in test duration, in comparison to M2, SAP V showed no difference while SAP III showed a significant difference. However, the results from this analysis are not reliable as the power values are <0.3 and <0.2, respectively.

With respect to the other techniques, M2 showed constant variability when measuring the central 10˚ visual field of patients with late stage glaucoma across the entire dynamic range but was greatest compared to SAP in the normal to near normal range of sensitivities. This study goes to show that the current algorithm available on M2 is better suited to measure visual field sensitivity in the later stages of glaucoma than SAP despite the fact that it was designed for early detection of the disease.

5 Discussion

The within and between test repeatability of techniques available on the Humphrey Field Analyser and Matrix perimeter for the measurement of the central 10˚ visual field, were compared in patients with late stage glaucoma. The following techniques were used: Full Threshold SAP 10-2 size III, Full Threshold SAP 10-2 size V, SITA SAP 10-2 size III and Matrix 10-2 size 2˚. Each patient was examined on three separate occasions with four techniques. The results of only the last two visits were used for analysis to eliminate learning effect bias. Test-retest variability in perimetry is dependent upon stimulus characteristics such as target size and color[97], the stage of visual field defect being studied[77] and the algorithm used to estimate the sensitivity[52]. It is therefore important to determine the repeatability within a technique to decide whether or not it might be suitable to measure visual field loss and/or progression. Previous studies have reported that the Matrix has uniform test-retest variability across its dynamic range in patients with early to moderate glaucomatous visual field loss in comparison to standard automated perimetry[67], which tends to have worse test-retest characteristics with deepening defect.

Overall, variability of SAP techniques follow the same trend: it decreases as sensitivity increases. However, variability of this trend is seen between techniques. The relative repeatability of techniques available on the HFA was as expected. Standard AP III had the widest variability and SAP V had the narrowest. This agrees with previous studies which show that SAP using a Goldmann size V stimulus is more suitable for monitoring visual field progression in late stage glaucoma than either technique using size III stimulus[104]. Increased test duration is reported to increase variability of test results[107, 108]. Hence, SAP III was found to be more variable than SS. Despite its short test duration, the Matrix showed greatest test-retest variability in the normal to

near-normal range of sensitivities. This may be due to the limited number of questions asked, 4 at each test location, and the limited number of end points available (15 as opposed to the 38 available on SAP) during the Matrix ZEST threshold estimation algorithm. Test duration showed no significant difference between visits, with the Full Threshold techniques taking the longest and M2 the shortest amount of time. The techniques with a normal database also showed no significant difference in any of the global indices. Coefficients of variability and repeatability were calculated to give a global analysis for each technique across the full range of data. The Matrix gave the highest scores in each measurement thereby suggesting that it was the most variable and least repeatable technique. However, these measures summate data from the entire dynamic range. The point-by-point analysis describes in detail the repeatability at specific sensitivity estimates.

Point-by-point test-retest analysis showed the variability of threshold estimates within and among the techniques. In the techniques available on SAP, variability increased with decreasing sensitivity. Variability of the Matrix was constant throughout its dynamic range and was greatest in the normal to near-normal sensitivity values in comparison to SAP. Full Threshold size V vs SITA SAP 10-2 size III showed decreased variability at both ends of the dynamic range with SS III recording more test locations as more severe. Compared to all of the SAP techniques, the Matrix estimated sensitivity as more severe. SAP 10-2 size V compared with the Matrix showed that estimated sensitivities of 20 dB and above on the Matrix were estimated at approximately 30 dB with size V. All SAP techniques show upper limits of their range of repeatability, particularly in the 10 to 20 dB range, to be no better than chance for measuring the same sensitivity at repeated visits.

Total and pattern deviation defect scores showed variability within each technique; SAP classified more test locations outside normal limits than that of the Matrix with respect to the

pattern deviation probability plots. This analysis has limited significance as total and pattern deviation probability calculations depend on the normal database which is unique to each technique, such as the inclusion criterion for the "normal" population[106]; differences in target size and stimulus type between the Goldmann size III and the Matrix size 2 degrees may also influence this outcome. One of the limitations of the study was that there was no normal database available for the Full Threshold SAP 10-2 size V technique. The lack of this data did not allow comparison of the total and pattern deviation probability plots of this technique with that of the others used in this study.

In conclusion, this study showed that test-retest variability of the SAP techniques decreased with increasing sensitivity whereas; variability was constant throughout the dynamic range for M2 and smaller in the moderate to severe range. However M2 was worst in the normal to near-normal sensitivity range. This suggests that M2, compared to all SAP techniques, will be disadvantaged for the detection of early visual field loss, as reflected in the reduced number of abnormal points in the TD and PD plots, but may be better positioned to repeatably detect and follow moderate to severe loss in the central 10˚ of patients with late stage glaucoma.

Previous studies have demonstrated a uniform test-retest variability of the Matrix over its entire dynamic range in patients with early to moderate visual field loss[67]; which is in agreement with this study. Despite the fact that the Matrix was designed to detect early visual field loss, threshold estimates in the normal to near-normal sensitivity range had the greatest test-retest variability. The study showed that the current algorithm for threshold estimation on the Matrix may not be suitable for the detection of early glaucoma, in spite of the technique being developed with this in mind. However it is possible that the Matrix would provide the best means by which to measure visual field progression in moderate to late stage. Altering the algorithm of the Matrix to increase the number of final threshold estimates, may help decrease the test-retest

characteristics in the near normal range, which in turn would help identify more abnormal points. This would require additional questions and therefore take more time, but could potentially make the Matrix the most obvious choice for a full scope technique for the detection and management of glaucoma.

For automated perimetry to be able to define, with its greatest efficiency, the status of visual field defect(s) in glaucoma, it may be most appropriate to change the technique used at different stages of the disease process. This study suggests the following: SS III be used for the early detection of glaucomatous visual field damage; SAP V should be considered for monitoring disease progression as it is the least variable technique in the moderate defect range. The Matrix as presently configured, is best suited to monitor progression in the later stages of the disease. It is theoretically possible that each technique could adopt refined threshold estimation techniques that would enable optimum full scope monitoring of glaucoma and its progression.

6 Limitations of the Study

Although the study was carried through as precisely as possible, there were several limitations to this study, most of which has to do with data analysis. First of which was the fact that the normal database is unique to each technique. This affects visual field indices and probability plot calculations, which were compared without accounting for this difference. Standard AP size V, did not have a normal database and was excluded from visual field indices and probability plot calculations.

Second was the difference in test location co-ordinates between the two instruments of which data points had to be recalculated for any comparison to be made.

Test-time for each technique was different for each technique, which may have resulted in variability of the data that was due to test duration and not technique algorithm.
And last but not least, all data that was analysed was subjective to each patient, although careful attention was given to ensure that each test was performed properly and only reliable tests were included for analysis.

References

1. Sassani JW. Ophthalmic Fundamentals: Glaucoma. Thorofare, NJ: SLACK, c1999.

2. Quigley HA & Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *British Journal of Ophthalmology*, 90(3), 262-267 (2006).

3. Yu JJ, Kiyosawa M, Nemoto N, Momose K, Mori H, Mochizuki M. Correlation Between Frequency Doubling Technology Perimetry and Temporal Frequency Characteristics in Early Glaucoma. *Documenta Ophthalmologica*, 107, 93-99, (2003).

4. Werner EB, Drance SM. Early Visual Field Disturbances in Glaucoma. *Archives of Ophthalmology*, 95, 1173-1175 (1977).

5. Anderson DR & Patella VM. Automated Static Perimetry: 2nd Edition. Mosby, Inc., 1999.

6. Sommer A, Katz J, Quigley HA, Miller NR, Robin AL, Richter RC, & Witt KA. Clinically Detectable Nerve Fiber Atrophy Precedes the Onset of Glaucomatous Field Loss. *Archives of Ophthalmology*, 109, 77-83 (1991).

7. Sekhar GC, Naduvilath TJ, Lakkai M, Jayakumar AJ, Pandi GT, Mandal AK, Honavar SG. Sensitivity of Swedish Interactive Threshold Algorithm Compared with Standard Full threshold Algorithm in Humphrey Visual Field Testing. *Ophthalmology*, 107(7), 1303-1308 (2000).

8. Bayer A & Erb C. Short Wavelength Automated Perimetry, Frequency Doubling Technology Perimetry, and Pattern Electroretinography for Prediction of Progressive Glaucomatous Standard Visual Field Defects. *American Academy of Ophthalmology*, 109(5), 1009-1017 (2002).

9. Bernhard J, Lachenmayer & Patrick MO Vivell. Perimetry and its Clinical Correlation. Thieme Medical Publishers, Inc., New York, 1993.

10. Heijl A & Patella VM. The Field Analyzer Primer: Essential Perimetry. Third Ed. Carl Zeiss Meditech Inc., 2002.

11. Weber J, Schultze T, and Ulrich H. The visual field in advanced glaucoma. *International Ophthalmology*, 13, 47-50 (1989).

12. Alward WLM. Frequency Doubling Technology Perimetry for the Detection of Glaucomatous Visual Field Loss. *American Journal of Ophthalmology*, 129(3), 376-378 (2000).

13. Johnson CA. Selective Versus Nonselective Losses in Glaucoma. *Journal of Glaucoma*, 3(Suppl. 1), S32-S44 (1994).

14. Johnson CA, Adams AJ, Casson EJ, Brandt JD. Blue-on-Yellow Perimetry Can Predict the Development of Glaucomatous Visual Field Loss. *Archives of Ophthalmology*, 111, 645-650 (1993).

15. Heron G, Adams AJ and Husted R. Central visual fields for short wavelength sensitive pathways in glaucoma and ocular hypertension. *Investigative Ophthalmology and Visual Science*, 29, 64-72 (1988).

16. Sample PA and Weinreb RN. Color perimetry for assessment of primary open glaucoma. *Investigative Ophthalmology and Visual Science*, 31, 1869-1875 (1990).

17. Johnson CA, Adams AJ and Lewis RA. Patterns of early glaucomatous visual field loss for short wavelength sensitive pathways. *Investigative Ophthalmology and Visual Science*, 30 (suppl): 56 (1989).

18. Sample PA and Weinreb RN. Progressive color visual field loss in glaucoma. *Investigative Ophthalmology and Visual Science*, 33, 2068-2071 (1992).

19. Hart WM, Silverman SE, Trick GL, Nesher R and Gordon MO. Glaucomatous visual field damage: luminance and color-contrast sensitivities. *Investigative Ophthalmology and Visual Science*, 31, 359-367 (1990).

20. Maddess T, Henry GH. Performance of nonlinear visual units in ocular hypertension and glaucoma. *Clinical Vision Science*, 7, 371-383 (1992).

21. Anderson AJ, Johnson CA. Mechanisms Isolated by Frequency-Doubling Technology Perimetry. *Investigative Ophthalmology & Visual Science*, 43(2), 398-401 (2002).

22. Quigley HA, Dunkelberger GR and Green WR. Chronic human glaucoma causing selectively greater loss of large optic nerve fibers. *Ophthalmology*, 95, 357-363 (1988).

23. Johnson CA, Samuels SJ. Screening for Glaucomatous Visual Field Loss With Frequency-Doubling Perimetry. *Investigative Ophthalmology & Visual Science*, 38(2), 413-425 (1997).

24. Maddess T, Goldberg I, Dobinson J, Wine S, Welsh AH & James AC. Testing for Glaucoma with the Spatial Frequency Doubling Illusion. *Vision Research*, 39, 4258-4273 (1999).

25. Maddess T & Severt WL. Testing for glaucoma with the frequency-doubling illusion in the whole, macular and eccentric visual fields. *Australian New Zealand Journal of Ophthalmology*, 27, 194-196 (1999).

26. Cello KE, Nelson-Quigg JM, & Johnson CA. Frequency Doubling Technology Perimetry for Detection of Glaucomatous Visual Field Loss. *American Journal of Ophthalmology*, 129(3), 314-322, (2000).

27. Patel SC, Friedman DS, Varadkar P and Robin AL. Algorithm for interpreting the results of frequency doubling perimetry. *American Journal of Ophthlamology*, 129, 323-327 (2000).

28. Burnstein Y, Ellish NJ, Magbalon M and Higginbotham EJ. Comparison of frequency doubling perimetry with Humphrey visual field analysis in a glaucoma practice. *American Journal of Ophthalmology*, 129, 328-333 (2000).

29. Delgado MF, Nguyen NTA, Cox TA, Singh K, Lee D. A., Dueker DK., Fechtner RD, Juzych MS, Lin SC, Netland PA, Pastor SA, Schuman JS, Samples JR. Automated Perimetry: A

Report by the American Academy of Ophthalmology. *Ophthalmology*, 109(12), 2362-2374 (2002).

30. Soliman MAE, de Jong LAM, Ismaeil AA, van den Berg TJTP, de Smet MD. Standard Automated Perimetry, Short Wavelength Automated Perimetry, and Frequency Doubling Technology for Detection of Glaucoma Damage. *American Academy of Ophthalmology*, 109(3), 444-454 (2002).

31. Girkin CA, Emadadi A, Sample PA, Blumenthal EZ, Lee AC, Zangwill LM, Weinreb RN. Short-Wavelength Automated Perimetry and Standard Perimetry in the Detection of Progressive Optic Disc Cupping. *Archives of Ophthalmology*, 118, 1231-1236 (2000).

32. The Advanced Glaucoma Intervention Study Investigators. Advanced Glaucoma Intervention Study, 2: Visual Field Test Scoring and Reliability. *Ophthalmology*, 101 (8), 1445-1455 (1994).

33. Flammer J, Drance SM, Augustiny L, & Funkhouser A. Quantification of Glaucomatous Visual Field Defects with Automated Perimetry. *Investigative Ophthalmology & Visual Science*, 26(2), 176-181 (1985).

34. Lindenmuth KA, Skuta GL, Rabbani R, Musch DC, & Bergstorm TJ. Effects of Pupillary Dilation on Automated Perimetry in Normal Patients. *Ophthalmology*, 97(3), 367-370, (1990).

35. Kerrigan-Baumrind LA, Quigley HA, Pease ME, Kerrigan DF, & Mitchell RS. Number of Ganglion Cells in Glaucoma Eyes Compared with Threshold Visual Field Tests in the Same Persons. *Investigative Ophthalmology & Visual Science,* 41(3), 741-748 (2000).

36. Chauhan BC, Drance SM, Douglas GR. The Use of Visual Field Indices in Detecting Changes in the Visual Field in Glaucoma. *Investigative Ophthalmology & Visual Science*, 31(3), 512-520 (1990).

37. Katsumori N, Okubo K and Mizokami K. The changes of visual field sensitivity accompanied with enlargement of optic cup. *Acta Societatis Ophthlamologica Japan*, 89, 928-933 (1985).

38. Spry PDG, Johnson CA, McKendrick AM & Turpin A. Variability Components of Standard Automated Perimetry and Frequency-Doubling Technology Perimetry. *Investigative Ophthalmology & Visual Science*, 42(6), 1404-1410 (2001).

39. Spry PDG & Johnson CA. Identification of Progressive Glaucomatous Visual Field Loss. *Survey of Ophthalmology*, 47(2), 158-173 (2002).

40. Spry PGD, Johnson CA, McKendrick AM, & Turpin A. Variability Components of Standard Automated Perimetry and Frequency-Doubling Technology Perimetry. *Investigative Ophthalmology & Visual Science*, 42 (6), 1404-1410 (2001).

41. Walsh TJ. Visual Fields: Examination and Interpretation. Ophthalmology Monographs 3. American Academy of Ophthalmology, Palace Press La Jolla, 1990.

42. Sample, PA, What Does Functional Testing Tell Us About Optic Nerve Damage? *Survey of Ophthalmology*, 45(supplement 3), S319-S324, (2001).

43. Zalta AH. Use of a central 10˚ field and a size V stimulus to evaluate and monitor small central islands of vision in end stage glaucoma. *British Journal of Ophthalmology*, 75(3) 151-154 (1991).

44. Morgan BK, Feuer WJ, & Anderson DR. Statpac 2 Glaucoma Change Probability. *Archives of Ophthalmology*, 109, 1690-1692 (1991).

45. Harwerth RS, Crawford MLJ, Frishman LJ, Viswanathan S, Smith III EL, Carter-Dawson L. Visual Field Defects and Neural Losses from Experimental Glaucoma. *Progress in Retinal and Eye Research*, 21, 91-125 (2002).

46. Werner EB, Bishop KI, Koelle J, Douglas GR, LeBlanc RP, Mills RP, Schwartz B, Whalen WR, & Wilensky JT. A Comparison of Experienced Clinical Observers and Statistical Tests in Detection of Progressive Visual Field Loss in Glaucoma Using Automated Perimetry. *Archives of Ophthalmology*, 106, 619-623 (1988).

47. Anderson AJ. Utility of a Dynamic Termination Criterion in the ZEST Adaptive Threshold Method. *Vision Research*, 43, 165-170 (2003).

48. Heijl A, Drance SM, & Douglas GR. Automatic Perimetry (COMPETER): Ability to Detect Early Glaucomatous Field Defects. *Archives of Ophthalmology*, 98, 1560-1563 (1980).

49. Flanagan JG, Wild JM, Trope GE. Evaluation of FASTPAC, a New Strategy for Threshold Estimation with the Humphrey Field Analyzer, in a Glaucomatous Population. *Ophthalmology*, 100(6), 949-954 (1992).

50. Coleman AL. Glaucoma. *The Lancet*, 354, 1803-1810 (1999).

51. Bengtsson B, Olsson J, Heijl A, & Rootzén H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmologica Scandinavica*, 75, 368-375 (1997).

52. Bengtsson B, Olsson J, Heijl A, & Rootzén H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmologica Scandinavica*, 75, 368-375 (1997).

53. Heijl A, Lindgren A and Lindgren G. Inter-point correlations of deviations of threshold values in normal and glaucomatous visual fields. In: Heijl (ed). Perimetry Update pp 177-183, Proceedings of the 8thh International Perimetric Society Meeting 1988. Krugler & Ghedini, Amsterdam (1989).

54. Heijl A, Lindgren A and Olsson J. A package for the statistical analysis of computerized visual fields. In: Greve & Heijl (eds). Documenta Ophthalmologica proceeding series 49:153-168, Proceedings of the 7th International Perimetric Society Meeting, 1988.

55. Sponsel WE, Arango S, Trigo Y & Mensah J. Clinical Classification of Glaucomatous Visual Field Loss by Frequency Doubling Perimetry. *American Journal of Ophthalmology*, 125(6), 830-836 (1998).

56. McKendrick AM, Anderson AJ, Johnson CA, & Fortune B. Appearance of the Frequency Doubling Stimulus in Normal Subjects and Patients with Glaucoma. *Investigative Ophthalmology & Visual Science*, 44(3), 1111-1116 (2003).

57. Kelly DH. Nonlinear Visual Responses to Flickering Sinusoidal Gratings. *Optical Society of America*, 71(9), 1051-1055 (1981).

58. Kelly DH. Frequency Doubling in Visual Responses. *Journal of the Optical Society of America*, 56(11), 1628-1633 (1966).

59. Richards W, and Felton TB. Spatial doubling; retinal or central? *Vision Research*, 13, 2129-2137 (1973).

60. Tyler CW. Observations on spatial-frequency doubling. *Perception*, 3, 81-86 (1974).

61. Trible JR, Schultz RO, Robinson JC, Rothe TL. Accuracy of Glaucoma Detection With Frequency-Doubling Perimetry. *American Journal of Ophthalmology*, 129, 740-745 (2000).

62. Quigley HA. Identification of Glaucoma-related Visual Field Abnormality With the Screening Protocol of Frequency Doubling Technology. *American Journal of Ophthalmology*, 125(6), 819-829 (1998).

63. Trible JR, Schultz RO, Robinson JC, Rothe TL. Accuracy of Glaucoma Detection With Frequency-Doubling Perimetry. *American Journal of Ophthalmology*, 129(6), 740-745 (2000).

64. Quaid PT, Simpson T and Flanagan JG. Frequency Doubling Illusion: Detection vs. Form Resolution. *Optometry and Vision Science*, 82, 36-42 (2005).

65. Spry PGD, Johnson CA, McKendrick AM, & Turpin A. Variability Components of Standard Automated Perimetry and Frequency-Doubling Technology Perimetry. *Investigative Ophthalmology & Visual Science*, 42 (6), 1404-1410 (2001).

66. Turpin A, McKendrick AM, Johnson CA, & Vingrys AJ. Development of Efficient Threshold Strategies for Frequency Doubling Technology Perimetry Using Computer Simulation. *Investigative Ophthalmology & Visual Science*, 43(2), 322-331 (2002).

67. Artes PH, Hutchinson DM, Nicolela MT, et al. Threshold and Variability Properties of Matrix Frequency-Doubling Technology and Standard Automated Perimetry in Glaucoma. *Investigative Ophthalmology & Visual Science*, 46(7), 2451-2457 (2005).

68. Turpin A, McKendrick AM, Johnson CS and Vingrys A. Performance of Efficient Test Procedures for Frequency-Doubling Technology Perimetry in Normal and Glaucomatous Eyes. *Investigative Ophthalmology & Visual Science*, 43(3), 709-715 (2002).

69. Tyrell RA and Owens DA. A rapid technique to asses the resting states of the eyes and other threshold phenomena: the modified binary search (MOBS). *Behavior research methods, instruments & computers*, 20, 137-141 (1988).

70. McKendrick AM. Recent developments in perimetry: test stimuli and procedures. *Clinical and Experimental Optometry*, 73-80, 2005.

71. King-Smith PE, Grigsby SS, Vingrys AJ, Benes SC, Supowit A. Efficient and Unbiased Modifications of the QUEST Threshold Method: Theory, Simulations, Experimental Evaluation and Practical Implementation. *Vision Research*, 34(7), 885-912 (1994).

72. Turpin A, McKendrick AM, Johnson CA, Vingrys AJ. Performance of Efficient Test Procedures for Frequency-Doubling Technology Perimetry in Normal and Glaucomatous Eyes. *Investigative Ophthalmology & Visual Science*, 43(3), 709-715 (2002).

73. Turpin A, McKendrick AM, Johnson CA, Vingrys AJ. Performance of Efficient Test Procedures for Frequency-Doubling Technology Perimetry in Normal and Glaucomatous Eyes. *Investigative Ophthalmology & Visual Science*, 43(3), 709-715 (2002).

74. Bickler-Bluth M, Trick GL, Kolker AE, Cooper DG. Assessing the Utility of Reliability Indices for Automated Visual Fields. *Ophthalmology*, 96(5), 616-619 (1989).

75. Bengtsson B. Reliability of Computerized Perimetric Threshold Tests as Assessed by Reliability Indices and Threshold Reproducibility in Patients with Suspect and Manifest Glaucoma. *Acta Ophthalmologica Scandinavic*a, 78, 519-522 (2000).

76. Drance SM & Anderson D. Automatic Perimetry in Glaucoma: A Practical Guide. Grune & Stratton, Inc., 1985

77. Heijl A, Lindgren A, & Lindgren G. Test-Retest Variability in Glaucomatous Visual Fields. *American Journal of Ophthalmology*, 108, 130-135 (1989).

78. de Boer RW, van den Berg TJ and Greve EL. Concepts for automatic perimetry, as applied to the Scoperimeter, and experimental automatic perimeter. *International Ophthalmology*, 5, 181-191, (1982).

79. Olsson J, Rootzén H and Heijl A. Maximum likelihood estimation of th e frequency of false positive and false negative answers from the up-and-down staircases of computerized threshold perimetry. In: Heijl (ed). Perimetry Update 1988'89 pp 245-251, proceedings of the 8th International Perimetric Society Meeting. Kugler & Ghedini Publications, Amsterdam (1988).

80. Olsson J, Bengtsson B, Heijl A & Rootzén H. An improved method to estimate frequency of false positive answers in computerized perimetry. *Acta Ophthalmologica Scandinavica*, 75, 181-183 (1997).

81. Werner EB, Petrig B, Krupin T, & Bishop KI. Variability of Automated Visual Fields in Clinically Stable Glaucoma Patients. *Investigative Ophthalmology & Vision Science*, 30(6), 1083-1089 (1989).

82. Heijl A., Lindgren G., Olsson J., & Åsman P. Visual Field Interpretation With Empiric Probability Maps. *Archives of Ophthalmology*, 107, 204-208 (1969).

83. Zeyen TG & Caprioli. Progression of Disc and Field Damage in Early Glaucoma. *Archives of Ophthalmology*, 111, 62-65 (1993).

84. Randall S & Argus WA. Fluctuations on the Humphrey and Octopus Perimeters. *Investigative Ophthalmology & Visual Science*, 28, 767-771 (1987).

85. Flanagan JG, Wild JM, Trope GE. The Visual Field Indices in Primary Open-Angle Glaucoma. *Investigative Ophthalmology & Visual Science*, 34(7), 2266-2274 (1993).

86. Heijl A, Lindgren G, Olsson J. The Effect of Perimetric Experience in Normal Subjects. *Archives of Ophthalmology*, 107, 81-86 (1989).

87. Katz J, Sommer A, Gaasterland AE, Anderson DR. Comparison of Analytic Algorithms for Detecting Glaucomatous Visual Field Loss. *Archives of Ophthalmology*, 109, 1684-1689 (1991).

88. Åsman P, Heijl A. Glaucoma Hemifield Test. *Archives of Ophthalmology*, 110, 812-819 (1992).

89. Katz J, Quigley HA & Sommer A. Detection of Incident Field Loss Using the Glaucoma Hemifield Test. *Ophthalmology*, 103(4), 657-663 (1996).

90. Blumenthal EZ & Sapir-Pichhadze R. Misleading Statistical Calculations in Far-advanced Glaucomatous Visual Field Loss. *Ophthalmology*, 110(1), 196-200 (2003).

91. Hodapp E, Parish II PK & Anderson DR. Clinical Decisions in Glaucoma. St Louis: Mosby, 1993.

92. Mills RP, Budenz DL, Lee PP, Noecker RJ, Walt JG, Siegartel LR, Evans SJ & Doyle JJ. Categorizing the Stages of Glaucoma From Pre-Diagnosis to End-Stage Disease. *American Journal of Ophthalmology*, 141(1), 24-30 (2006).

93. Casson R, James B, Rubinstein A & Haggai A. Clinical comparison of frequency doubling technology perimetry and Humphrey perimetry. *British Journal of Ophthalmology*, 85, 360-362 (2001).

94. Thomas R, Bhat S, Muliyil JP, Parikh R & George R. Frequency doubling perimetry in glaucoma. *Journal of Glaucoma*, 11, 188-197 (2001).

95. Wilensky JT, Mermelstein JR ans Siegel HG. The use of different-sized stimuli in automated perimetry. *American Journal of Ophthalmology*, 101, 710-713 (1986).

96. Choplin NT, Sherwood MB and Spaeth GL. The effect of stimulus size on the measured threshold values in automated perimetry. *Ophthalmology*, 97, 371-374 (1990).

97. Wall M, Kutzko KE & Chauhan BC. Variability in Patients With Glaucomatous Visual Field Damage is Reduced Using Size V Stimulus. *Investigative Ophthalmology & Visual Science*, 38(2), 4260435 (1997).

98. Gilpin LB, Stewart WC, Hunt HH & Broom CD. Threshold variability using different Goldmann stimulus sizes. Acta Ophthlmologica, 68, 674-474 (1990).

99. Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1, 307-310 (1986).

100. Hastie T and Stuetzle W. Principal curves. *Journal of American Statistics Association*, 84, 502-516 (1989).

101. Fankhauser F. Problems related to the design of automated perimeters. *Doucmenta Ophthalmologica*, 47, 89-138 (1979).

102. Humphrey Matrix: Operator's Manual. Carl Zeiss Meditech, Dublin, CA (2003).

103. Chauhan BC & Johnson CA. Test-Retest Variability of Frequency-Doubling Perimetry and Conventional Perimetry in Glaucoma Patients and Normal Subjects. *Investigative Ophthalmology & Visual Science*, 40(3), 648-656 (1999).

104. Wall M, Kutzko KE, & Chauhan BC. Variability in Patients With Glaucomatous Visual Field Damage Is Reduced Using Size V Stimulus. *Investigative Ophthalmology & Visual Science*, 38(2), 426-435 (1997).

105. Quaid PT, Simpson T and Flanagan JG. Monocular and Dichoptic Masking Effects on the Frequency Doubling Illusion. *Vision Research*, 44, 661-667 (2004).

106. Anderson AJ and Johnson CA. Anatomy of a Supergroup: Does a Criterion of Normal Perimetric Performance Generate a Supernormal Population? *Investigative Ophthalmology and Visual Science*, 44, 5043-5048 (2003).

107. Fujimoto N & Adachi-Usami E. Fatigue Effect Within 10˚ Visual Field in Automated Perimetry. *Annals of Ophthalmology*, 25, 142-144 (1993).

108. Searle AET, Wild JM, Shaw DE and O'Neil EC. Time-related variation in normal automated static perimetry. *Ophthalmology*, 98, 701-707 (1991).

109. Schimiti RB, Avelino RR, Kara-José & Costa VP. Full-threshold versus Swedish Interactive Threshold Algorithm (SITA) in Normal Individuals Undergoing Automated Perimetry for the First Time. *Ophthalmology*, 109(11), 2084-2092 (2002).