Variability in perimetry and optical coherence tomography: its importance in glaucoma detection, diagnosis and management

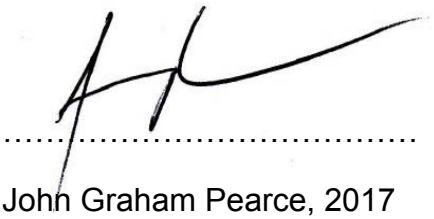
John Graham Pearce

A thesis submitted for the degree of Doctor of Philosophy of The Australian National University

Submitted for examination
February 2017

I, John Graham Pearce, do hereby declare that the thesis submitted for examination is entirely my own original work. I hereby declare that I have been involved in the construction of this thesis from the outset, and have been involved in the study design and methodology, data collection, statistical analysis and the writing up of the various components of the thesis, both for thesis submission as well as for submission of various chapters to journals for publication.

Signed,



……

John Graham Pearce, 2017

Presentations and publications from this thesis

Publications

Pearce JG, Maddess T. Retest Variability in the Medmont M700 Automated Perimeter. *Optom Vis Sci* (2016) Mar;(3): 272-280. doi: 10.1097/OPX.0000000000000798.

Pearce JG, Maddess T. Inter-visit test-retest variability of the Topcon 3D OCT-2000 in glaucoma. *Optom Vis Sci* 2016 Nov 19 [Epub ahead of print] doi: 10.1097/OPX.0000000000001022

Conference Presentations

Pearce JG, Maddess T. (2016). Synchronous diurnal variation in the test-retest variability of standard automated perimetry and spectral domain optical coherence tomography. ARVO - Invest Ophthalmol Vis Sci 2016; 57: e-abstract 4228.

Pearce JG, Maddess T. (2015) Variability of peripheral test points, central test points and visual field indices in the Medmont M700 perimeter. ARVO - Invest Ophthalmol Vis Sci 2015; 56: e-abstract 1056

Pearce JG, Pearce ME, Maddess T. (2013). Variability of peripapillary retinal nerve fibre layer measurements in four spectral domain optical coherence tomographs. AOVSM - Clin Exp Ophthalmol. Volume 41, Issue S1 p 131

Pearce JG, Maddess T. (2012). Daily variations in spectral domain OCT measurements. AOVSM - Clin Exp Ophthalmol. Volume 40, Issue S1:p 121

Acknowledgements

This research was supported by the Australian Research Council through the ARC Centre of Excellence in Vision Science (CE0561903).

John Graham Pearce received funding from the Australian Federal Government through the Nursing and Allied Health Scholarship Support Scheme. The views expressed in this paper do not necessarily represent those of the NAHSSS, its administrator, Services for Australian Rural and Remote Allied Health or the Australian Government Department of Health.

The Australian Federal Government National Eye Health Initiative provided a grant to John Graham Pearce, which enabled the purchase of the Topcon 3D-OCT 2000 used in this study.

Abstract

The detection, diagnosis and monitoring of glaucoma requires numerous investigative techniques and strategies in order to provide definitive diagnosis and management of glaucoma. Two investigative techniques used extensively for this purpose today are standard automated perimetry (SAP) and spectral domain optical coherence tomography (SD-OCT).

The presence of test-retest variability (TRV) for repeat measures in these instruments can impact on our ability to differentiate TRV from true progressive change. In this thesis, we examine two popular devices which, at present, have little or no published TRV criteria available. The instruments under consideration are the Medmont M700 automated perimeter and the Topcon 3D OCT-2000 SD-OCT. The purpose of this study is to quantify TRV for both instruments in both healthy and glaucoma patients, and identify any modifiable external components of TRV.

Chapters 1 and 2 determine the test-retest variability (TRV) of both instruments in glaucoma patients. We were able to develop event based progression criteria which could be easily applied by clinicians using these instruments. We also established that TRV for several summary measures, structural and functional, was dependent on the level of glaucomatous damage.

We reported that for the Medmont, retest variance did not increase with eccentricity when comparing points of equal decibel value in the central 10° of the 30-2 test with points of equal decibel value in the outer 20° of the same test. This finding had not been reported previously, and was counter to that of other studies (involving healthy eyes) which had reported that variability increased with eccentricity.

There are few published studies which have examined diurnal variations in SAP and SD-OCT summary measures, and none that had investigated diurnal

variations in TRV in either instrument. In Chapter 3, we therefore investigated whether diurnal variation existed in either summary measures or TRV.

Whilst we found no detectable changes to mean summary measures over the course of a normal working day, we were able to detect statistically significant diurnal variations in TRV for several SAP and SD-OCT summary measures. This was now evidence to suggest that reduced TRV, and therefore enhanced progression detection, might be available by testing subjects on both instruments at consistent and specified times of day.

In Appendices 1 and 2, we investigated whether any circadian variations in SAP and SD-OCT summary measures were detectable over a 24 hour period. We were unable to discern any significant differences in SAP or SD-OCT summary measures, nor were we able to discern any consistent cycles in any summary measures. We were able to detect statistically significant circadian variation in the peripapillary retinal nerve fibre layer using SD-OCT, and we were able to achieve the first recorded 24 hour test session results for both SAP and SD-OCT.

The final part of our study, Appendix 3, involved generating TRV values for healthy subjects on the Topcon, which could be used clinically to differentiate true progression from TRV.

We hope that the overall results generated from our studies have enhanced disease and progression detection for possibly millions of patients worldwide, and that we have stimulated the development of future studies to further investigate some of the preliminary findings from this study.

Table of contents

Glossary

Average Defect: global index used in the earliest Medmont perimeter, the M600, to indicate the level of generalised depression (or elevation) of a subjects visual field sensitivity in comparison to an expected age normal result

Confocal scanning laser ophthalmoscope: retinal imaging method using laser light rather than white light to illuminate the retina.

Cup/disc ratio: the ratio of the area of the central depression (cup) present in an optic nerve head (disc) to the overall area of the disc. This figure has no units.

Event based criteria: are one method used to define the test-retest variability of a diagnostic method. Any subsequent measurement is compared with a first, or a baseline, measurement. If the change in the new measurement recorded exceeds the level of test-retest variability (an “event”), true significant change is deemed to be present

Frequency doubling technology: refers to the selective method of perimetry which uses the frequency doubling illusion (where subjects see twice as many grating patterns as in the original stimulus) to look for cell losses in a relatively sparse sub-population of retinal ganglion cells (the y like ganglion cells)

Ganglion cell complex: the inner retinal layers, consisting of the nerve fibre layer, ganglion cell layer and inner plexiform layer, grouped together under the title of ganglion cell complex

Glaucomatous optic neuropathy: term used to describe changes to the optic nerve head due to glaucoma, as opposed to other diseases such as multiple sclerosis, etc

Humphrey Field Analyser: static automated bowl perimeter manufactured by Carl Zeiss Meditec. As one of the earliest automated perimeters, it became a gold standard in visual field testing and is therefore often referred to in comparative studies of other perimeters

Intra-ocular pressure: is the pressure exerted by the fluids in the eye against the external structures of the eye. It is measured in millimetres of mercury.

Mean Deviation: global index used in the Humphrey Field Analyser to indicate how much, on average, an individual subjects' field differs from that of an age matched expected value

Medmont Automated Perimeter: static automated bowl perimeter manufactured by Medmont P/L, Nunawading, Victoria, Australia. Differs from the Humphrey Field Analyser in stimulus layout and stimulus projection method.

Open angle glaucoma: refers to the various types of glaucoma where the anterior angle remains open but aqueous drainage is still impeded. Differentiates these glaucoma sub-types from those where aqueous outflow is impeded due to iris blockade (angle closure glaucoma) or other structural impediments such as in neovascular glaucoma. Also includes the subgroup of normal tension glaucoma, where intra-ocular pressure is not abnormal.

Optic nerve head: also referred to as the disc or the optic disc, it represents the exit point of the combined retinal ganglion cell axons from the retina

Overall Defect: global index used in the Medmont M700 automated perimeter to indicate the general level of depression or elevation of an individuals' field from an expected age normal value. This global index is calculated in a different way to its predecessor, Average Defect as used in the M600, and these two indices are not comparable or interchangeable

Pattern Defect: global index used in the Medmont M700 to quantify localised field defects by qualifying the extent to which deviations from normal are clustered

Pattern electroretinogram: the electroretinogram measures the electrical responses of retinal cells to a light stimulus, while the pattern electroretinogram measures the retinal responses to a grating or chequerboard stimulus

Pattern Standard Deviation: global index used in the Humphrey Field Analyser to quantify localised sensitivity losses and scotomas

Peripapillary atrophy: refers to outer retinal atrophic changes adjacent to the optic nerve head, which may or may not entirely surround the optic nerve head

Peripapillary retinal nerve fibre layer: refers to the retinal nerve fibre layer surrounding the optic nerve head within a short distance of the optic nerve head

Primary open angle glaucoma: refers to glaucoma present when the anterior angle is unobstructed and there is no other underlying disease or pathophysiology present (e.g. steroid induced glaucoma has an open angle, but occurs as a result of a secondary process and is therefore a secondary open angle glaucoma)

Retinal ganglion cells: transmit photoreceptor and intra-retinal cellular outputs to intra-cranial structures for further processing

Retinal pigment epithelium: pigmented layer of cells underlying the neuro-retinal structures and lying above the vascular choroid

Scanning laser polarimetry: investigative technique for measuring the thickness of the retinal nerve fibre layer utilising the birefringence properties of the retinal nerve fibre bundles

Short wavelength automated perimetry: selective perimetric testing protocol which utilises a blue stimulus and a yellow background to isolate blue cones and blue -yellow coding retinal ganglion cells. These cells are relatively sparse, with cellular losses therefore being detected earlier than other test protocols which stimulate large numbers of retinal ganglion cells.

Spectral domain optical coherence tomography: imaging technique for optically transparent tissue which uses interferometry to investigate the amount and location of backscattered light reflected from an object to determine the internal structure of the object.

Standard automated perimetry: refers to static white on white automated perimetry to differentiate it from other forms of perimetric techniques such as short wavelength automated perimetry, edge perimetry, frequency doubling perimetry etc.

Test-retest variability: refers to the variation in results obtained from repeated measurements as a result of subject variability, instrument variability or a combination of both.

Time domain optical coherence tomography: utilises interferometry to determine the time delay of backscattered light reflected from an object. Predecessor of spectral domain optical coherence tomography

Trend Analysis: Statistical technique which uses regression based analysis of a number of tests carried out over time to determine whether statistically significant change has taken place

World Health Organisation: is the public health arm of the United Nations, whose function is to assess, direct and co-ordinate international health programs

Background

This thesis is directed towards improving the diagnosis and management of glaucoma using *Standard Automated Perimetry* (**SAP**) and *Spectral Domain Optical Coherence Tomography* (**SD-OCT**). To do this, we will be investigating factors that may limit the ability of the relevant diagnostic instruments to differentiate *Test Retest Variability* (**TRV**) from true progression.

The term glaucoma is generally used to refer to a variety of ocular conditions that result in a clinically characteristic *Glaucomatous Optic Neuropathy* (**GON**) (Casson *et al.*, 2012), which is due to the death of the retinal ganglion cells whose axons make up the optic nerve. Although considered to be an ocular condition, recent research suggests that glaucoma patients have detectable changes in cortical brain structure, which are correlated with disease severity (Williams *et al.*, 2013). There are also indications that at a molecular level, the pathogenesis of glaucoma can be compared to neurodegenerative diseases such as Alzheimer's and Parkinsonism (Jindal, 2013). Perhaps the most important aspect of glaucoma is that no matter what the underlying pathophysiology may be, it remains one of the few eye conditions which, if left untreated, can result in the total and irreversible loss of vision in one or both eyes.

In developed countries, increased life expectancies are accompanied by an increased prevalence of glaucoma in the ageing population (Guedes *et al.*, 2011). The prevalence of *Open Angle Glaucoma* (**OAG**) in the United States was estimated to be 1.86% of the population aged 40 years and older (Friedman *et al.*, 2004), 3.9% of the population aged over 40 years in Japan (Iwase *et al.*, 2004), while in Australia, a prevalence of 3.0% has been reported in a population aged 49 years or older (Mitchell *et al.*, 1996). The conflicting figures for prevalence are exemplified in two separate studies that looked at the prevalence of glaucoma in indigenous Australians. One study looked at the prevalence of glaucoma in indigenous Australians living in Central Australia, and determined a prevalence of 0.52% for those aged ≥ 40 years (Landers *et al.*, 2012), whilst another study looked at Aboriginal and Torres Strait Islander adults aged ≥ 40 years from randomly selected areas or communities, and obtained a prevalence of 2.2% (Chua *et al.*, 2011).

Perhaps more critically, the prevalence of undetected glaucoma was found to be 1.53% of the population aged 49 years or older in Australia (Mitchell *et al.*, 1996), and 1.23% in a population aged 55 to 79 years in Sweden (Heijl *et al.*, 2013a), with this latter study finding that unilaterally blind subjects were present in all age groups. Vision loss from glaucoma can impact on a patients' health related quality of life even in the early stages (Varma *et al.*, 2011), with *Primary Open Angle Glaucoma* (**POAG**) also entailing a significant economic burden in Australia (Dirani *et al.*, 2011). With an exponential rise in prevalence with age (Mitchell *et al.*, 1996), the importance of refining current methods of glaucoma detection to enable early intervention becomes critical.

Early glaucoma detection becomes vitally important for patient quality of life, and the economic costs to society, when we consider that worse visual fields at baseline were associated with an increased risk of bilateral blindness (according to the *World Health Organisation* (**WHO**) criteria) in POAG and pseudoexfoliative glaucoma (Peters *et al.*, 2013a). Among 19.3% of patients who progressed to legal blindness in one eye over a 34 year period, visual field loss at the time of diagnosis posed the greatest risk for blindness (Oliver *et al.*, 2002).

Although some older patients may have mild disease and slow progression, the detection of true progression in these patients remains a critical component of their ongoing management strategy. Detecting the presence of true progression is vital to the clinical decision making employed in the management of any glaucoma patient.

Some studies have indicated that the most patients in glaucoma clinics may not be at high risk of progressing to statutory blindness (Saunders *et al.*, 2014), other studies have found that at their last visit prior to death (median age at death 88 years), 16.4% of glaucoma patients were bilaterally blind according to the WHO criteria, and 42.2% were unilaterally blind (Peters *et al.*, 2013b).

Whilst the median age for developing bilateral blindness due to glaucoma was 86 years in that study, (Peters *et al.*, 2013b), it should be borne in mind that some studies have forecast the number of Australians aged over 80 to be 2.8 million by 2050 (Kowal *et al.*, 2014).

Whilst glaucoma is often a slowly progressive disease, this may not always the case. Progression rates of visual field loss in glaucoma have been reported to be highly variable, with progression rates rapid enough to influence quality of life being common (Heijl *et al.*, 2013b). Progression rates may also vary with the type of glaucoma, with pseudoexfoliation glaucoma patients reported to progress considerably faster than high-tension and normal tension glaucoma patients (Heijl *et al.*, 2009). Perhaps more importantly, individual patients may demonstrate highly variable progression rates, with some patients exhibiting quite rapid progression in visual field defects (Susanna, 2009).

The early detection of glaucoma, and the earliest detection of glaucomatous progression in patients diagnosed with the disease, is therefore of paramount importance in a society which has increasing life expectancies and a concomitant increasing incidence of glaucoma. Not only do those individuals who are living longer wish to enjoy an acceptable quality of life by retaining their independence and pursuing interests and leisure activities, society as a whole benefits from the reduced health care costs engendered by the early detection of glaucoma and glaucomatous progression.

Modern glaucoma detection and monitoring is a multi-faceted discipline, involving the integration of a multitude of factors. These include *Intra-Ocular Pressure* (**IOP**), central corneal thickness, alterations to visual field sensitivity and the identification of risk factors associated with glaucoma, such as family history. It also includes the observation of ocular structures such as the optic nerve head and retinal reflectivity by direct and non-direct ophthalmoscopy, slit lamp biomicroscopy and retinal photography. Optical coherence tomography is a more recent method of investigating the structure of features such as the optic nerve head, the retinal nerve fibre layer surrounding the optic nerve head, and the inner retinal layers surrounding the macula which make up the *Ganglion Cell Complex* (**GCC**).

Subjective assessment of ocular structures can result in difficulties in detecting change over time, and automated instrument measurements also have certain levels of retest variability. *Differentiating variability from true change can be difficult in some instruments currently in use, simply because there is no*

published information to give clinicians any guidelines for interpreting different measurements from their instruments.

The use of trend based progression detection protocols is often thought to be the preferred method for early detection of progressive field loss and structural change. However, a recent study examining visual fields in glaucoma reported that there was little difference in progression detection between trend-based and event analysis (Rao *et al.*, 2013), while another reported that trend analysis, event based analysis and expert visual field evaluation detected similar numbers of progressing cases (Anton *et al.*, 2013). Event based analysis has also been reported to show earlier and greater sensitivity for detecting progression than trend based analysis (Casas-Llera *et al.*, 2009).

In many clinical situations, clinicians may only have two (or fewer) fields or scans per year with which to determine the presence or otherwise of progressive glaucomatous loss. Event based progression criteria may therefore be useful in clinical practice, and are easily accessed and understood by clinicians. Typically, event based criteria are applied to a first, or baseline measurement. In perimetry, the baseline is usually derived by taking the mean values of the first two tests undertaken. In SD-OCT, the baseline may be simply the results of the first scan undertaken, or again, the mean values of the first two scans undertaken.

Whilst physiological alterations to structure and function may also take place over time, they do not necessarily confound event based criteria applied over time, and trend analysis also has to contend with physiological change over time.

In perimetry, physiological change over time is compensated for by comparisons to the age based normative data base, and physiological decreases in sensitivity are thus taken into account when calculating the instruments summary indices.

In SD-OCT, a similar strategy is employed, whereby physiological change is catered to by the categorisation (in terms of percentiles) of the patients' results in comparison to the normative database for most parameters. Many

parameters have their results is colour coded as to what percentile the subjects' thickness lies in.

For those within the 95th percentile (coded green) and those in the 95th to 99th percentile (coded yellow), a change in a parameter that does not induce a change percentile ranking may thus be more likely to be interpreted as physiological change.

Significant change with an associated change in percentile ranking may be more likely to represent pathological change. For those patients already ranked outside the 99th percentile, the clinician would need to take into account the patients overall presentation to determine whether any significant change is physiological or pathological.

In the Topcon 3D OCT-2000, average peripapillary retinal nerve fibre layer (pRNFL) does not have an age based comparative percentile assigned to it. The detection of event based change in this parameter therefore needs to be considered in the overall context of the patient presentation. The event based criteria generated in this thesis for this parameter are currently the only means of detecting what constitutes true progression in this device, and therefore provides important information to be used in clinical decision making.

The central purpose of this thesis is to therefore generate some event based criteria in several instruments commonly used in Australia and around the world. These instruments currently have little or no information available regarding the TRV inherent in their test protocols.

In addition, this study intends to examine whether there are any modifiable external factors which may affect retest variability in these instruments. Specifically, we will be looking at the effect of time of day on TRV as regards structural measurements using SD-OCT and functional change using SAP.

The following is an introduction to the development of some of the techniques currently used extensively in glaucoma detection and management throughout the world, several of which will be examined in depth in the current study.

1. Introduction

Glaucoma detection may have begun with the observation that elevated IOP, noted as firmness of the eye, was associated with vision loss as far back as the 10th century AD (Stamper, 2011), although intra-ocular pressure was unable to be properly assessed until von Graefe invented the first experimental instrument capable of doing so in 1862 (Kniestedt *et al.*, 2008). The ability to measure IOP was a significant improvement in the detection of glaucoma, given that most glaucoma patients do not report any symptoms (Fernandez Lopez *et al.*, 2014).

Extremely high IOP's can occur in several glaucoma presentations. Primary angle closure glaucoma can cause acute elevations in IOP, and has been reported to be responsible for approximately half of the glaucoma blindness worldwide (Wright *et al.*, 2015). Steroid induced glaucoma, which results in an accumulation of basement membrane type material in the trabecular meshwork (Tektas & Lutjen-Drecoll, 2009), can result in significant elevations in IOP induced in the short term by topical (Ang *et al.*, 2012), or systemic (Carnahan & Goldstein, 2000) use of corticosteroids. Neovascular glaucoma can occur secondary to retinal vascular pathologies such as ischemic central retinal vein occlusion, with the resultant neovascular blockade of the anterior angle causing highly elevated IOP's. Whist these clinical entities may result in rapid and irreversible vision loss, the acutely elevated IOP's are usually painful enough to be symptomatic, causing the patient to seek treatment.

Pseudoexfoliative glaucoma occurs due to the accumulation of an abnormal protein (the material which gives the crystalline lens its pseudoexfoliative appearance) in the trabecular meshwork (Hollo *et al.*, 2015). Pseudoexfoliative glaucoma typically presents with higher IOP's and greater visual field loss than POAG (Konstas *et al.*, 1997).

Pigmentary glaucoma is a much more challenging clinical entity, since it has been reported as typically occurring in young myopic patients aged 20 to 40 years (Niyadurupola & Broadway, 2008). Although a later study suggested that the average age at diagnosis was 48.9 years (Musch *et al.*, 2012), it remains a form of secondary OAG which does occur in a younger demographic than other forms of OAG, making diagnosis simultaneously more difficult and more critical.

The IOP becomes elevated due to pigment released from irido-zonular contact being transported to the trabecular meshwork (Campbell & Schertzer, 1995). This liberated pigment results in the loss of trabecular meshwork cells, causing the fusion of trabecular meshwork lamellae and the collapse of the trabecular meshwork (Tektas & Lutjen-Drecoll, 2009).

In conditions such as POAG, the presenting IOP may often not be elevated to extreme levels in early disease, with mean IOP's on diagnosis being reported to be 19.5 mmHg (SD 4.5) in one study (Garway-Heath *et al.*, 2013). As a result, POAG may often be being pain-free and asymptomatic in the early stages. Although elevated IOP is significantly associated with the onset and progression of POAG (Miglior & Bertuzzi, 2013), other conditions such as ocular hypertension exist, where the optic nerve head and visual fields show no damage (Wahl, 2011). In contrast, normal tension glaucoma also exists, where IOP is not abnormal (Anderson, 2011).

Other systemic factors, such as decreased optic nerve head perfusion pressure, may contribute to RGC axonal losses in glaucoma (Caprioli *et al.*, 2010), and optic nerve head perfusion has been reported to be significantly lower in glaucoma patients, and correlated with disease severity.(Chen *et al.*, 2016) As a result, the presence or absence of elevated IOP cannot solely be used as a diagnostic criteria for glaucoma diagnosis (Stamper, 2011), other means of determining the glaucoma status of a patient are required.

Early assessment of the optic nerve head

The invention of the direct ophthalmoscope by Hermann von Helmholtz in 1850 (Armour, 2000; Tan & Shigaki, 2006) or 1851 (Koehler, 2002; Berendschot *et al.*, 2003) enabled direct observation of the fundus (Grewe, 1986). Diagnostically, the most important structure initially assessed by direct observation was the *Optic Nerve Head* (**ONH**), which is the point where the optic nerve exits the eye.

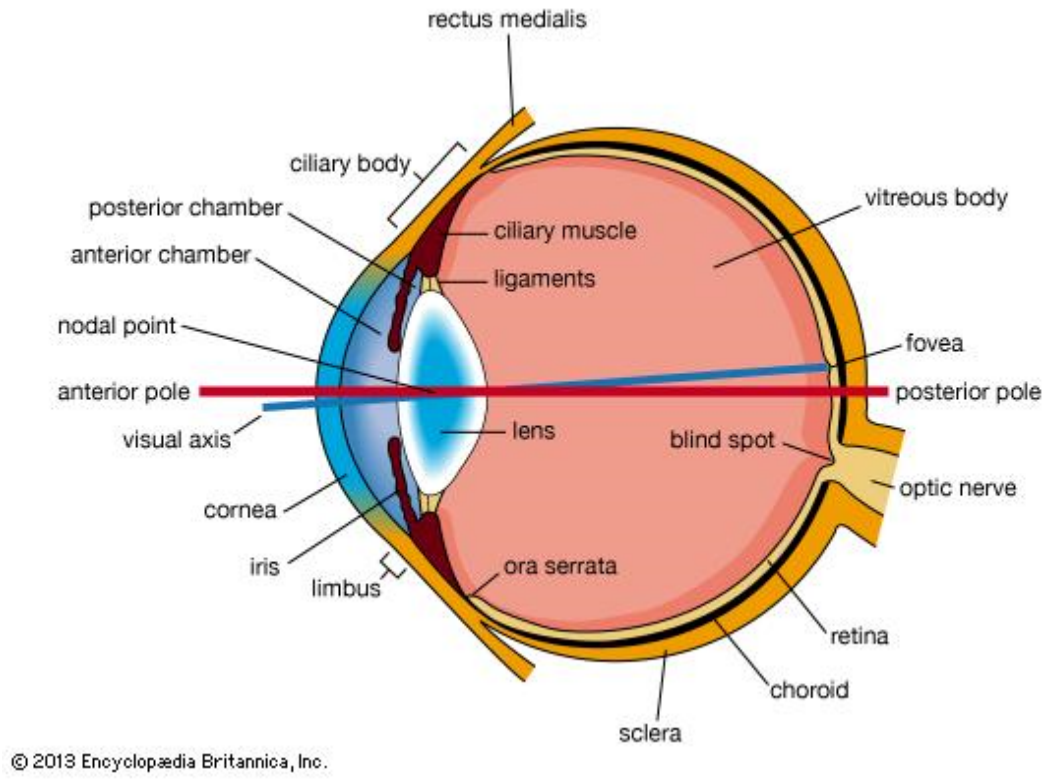


Figure 1.

Cross section of the human eye showing the relative positions of the major structures. In this figure, the optic nerve head can be seen to be located at the rear of the globe.

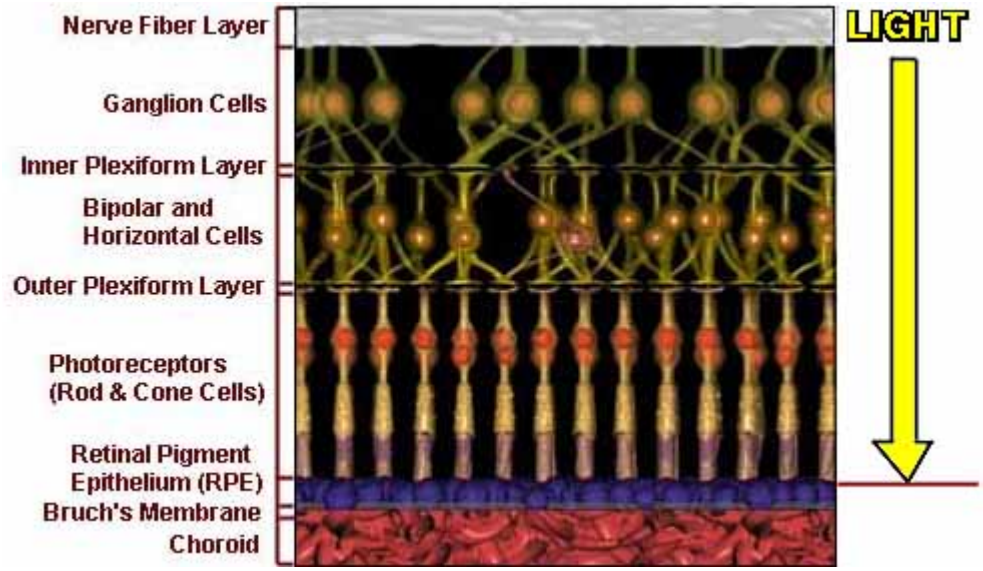


Figure 2.

Cross section of the retina showing the relevant structural layers and neural and non-neural components. Note the bipolar and horizontal cells often connect to more than one ganglion cell and more than one photoreceptor outside the fovea. At the fovea, the midget bipolar cells only connect to one retinal ganglion cell.

The retina is the location of the transduction of photons into electrical impulses by photoreceptors, which travel via intra-retinal synapses to eventually reach the *Retinal Ganglion Cells* (**RGC's**), which then relay this information through the optic nerve to intracranial structures for further processing. As RGC's are responsible for retinal output to higher processing centres, their loss represents a loss of signal available for processing and hence the loss of perception of light falling on the area of retina sub-served by those cells.

Histologically, the highest density of RGC's are found in an elliptical ring around the fovea with an average density of 35,100 cells/mm², with the result that approximately 50% of the total number of RGC's are located within 16° of the foveal centre (Curcio & Allen, 1990). In terms of visualising this area, we can approximate the radius of this circle as extending from the centre of the fovea and encompassing the nasal outer edge of the optic nerve (Fig. 1). With increasing eccentricity, RGC populations decrease significantly, with

approximately 2415 cells/mm² at 33.7° (Sjostrand *et al.*, 1999), and densities of around 170-350 cells/mm² near the ora serrata (Curcio & Allen, 1990).

In addition, Puchella et al. (2005) found that as individual RGC's generally do not transmit unique visual information, there is substantial redundancy in the encoding of retinal information. In the central visual field, RGC receptive field density reportedly exceeds 30,000 per solid degree (Drasdo, 1989). The value of approximately 30,000 receptive fields per solid degree is much larger than the mathematical value which would be arrived at if a single layer of central ganglion cell receptive fields were spread over one solid degree, and is a reflection of the fact that RGC's have receptive fields that overlap extensively (Puchalla *et al.*, 2005). The figure of 30,000 is thus an indication of an approximately 10-fold overlap in receptive fields, resulting in a significant amount of RGC redundancy.

These central RGC population densities, in contrast to the much more sparse peripheral RGC population densities with reduced redundancy, led to the historical view that peripheral visual field loss in glaucoma was thought to precede changes to the central visual field (Stamper, 1984).

The inability of perimetric techniques to detect early changes to the central visual field are most likely due to the dense central RGC populations as described above, as well as the issue of under-sampling. Under-sampling occurs when perimetric stimuli are too widely spaced to reconstruct rapid variations in sensitivity across the retina and also results in much of the field not being tested (Maddess, 2011b). For example, the *Humphrey Field Analyser* (**HFA**) utilises a 6° grid pattern in the most commonly used test, the 30-2 (Fig. 13). This reduced central sampling rate has meant that small but deep retinal nerve fibre layer defects may not be detected using this sampling strategy (Burk *et al.*, 1998; Asaoka, 2014).

In healthy eyes, the optic nerve itself is composed of the axons emanating from approximately one million RGC's (Tatham *et al.*, 2013), which are enclosed and segmented by extracellular matrix (Goldbaum *et al.*, 1989). An underlying element in all forms of glaucoma is the death of RGC's (Almasieh *et al.*, 2012), with the resultant loss of axons in the optic nerve causing specific changes in

optic nerve head morphology (Mardin, 2012), which results in glaucomatous cupping of the ONH.

Direct observation of the optic nerve head allows for the clinical evaluation of any changes which might be associated with the onset or progression of GON. There are numerous diagnostic clinical signs which can be observed at the optic nerve head and surrounding retina, such as the area, configuration and colour of the neuro-retinal rim, depth and shape of the optic cup, cup to disc ratio, disc size, point of exit of the central retinal vessels on the surface of the lamina cribrosa, the appearance of the lamina cribrosa itself, the presence or absence of disc haemorrhages, aspects of *Peripapillary Atrophy* (**PPA**) such as size and location, changes to the diameter of the retinal arterioles and the degree of reflectivity of the *Peripapillary Retinal Nerve Fibre layer* (**pRNFL**) surrounding the disc and extending to the posterior pole (Jonas *et al.*, 1999).

It should be remembered that some of these clinical entities, such as PPA, can also be found in healthy subjects (Fig.4), and disc haemorrhage in healthy subjects has been reported to be significantly associated with PPA (Sugiyama *et al.*, 1999). PPA is an area of outer retinal atrophy adjacent to the ONH, which may or may not entirely surround the ONH (Ehrlich & Radcliffe, 2010). PPA is further subdivided into alpha (α) and beta (β) sub-types, with the αPPA being peripheral to βPPA. In βPPA, the sclera and choroidal vessels can be seen due to the total loss of *Retinal Pigment Epithelium* (**RPE**) cells and the incomplete loss of photoreceptors (Kubota *et al.*, 1993). αPPA retains some disarrayed RPE cells which may cause irregular hypo- or hyper-pigmentary changes in appearance (Ehrlich & Radcliffe, 2010). Early studies reported βPPA to be more frequent and larger in early glaucoma than in normal subjects, and the size and frequency of PPA was correlated with the level of glaucoma present (Jonas *et al.*, 1989). Later studies indicated that the addition of PPA findings to existing commonly assessed variables did not improve the discrimination of OAG patients from glaucoma suspects (Ehrlich & Radcliffe, 2010), and an example of PPA without glaucoma is exemplified in Figure 4.

Perhaps even more critically, it needs to be remembered that intra-cranial pathologies can result in optic atrophy not associated with glaucoma (Trobe *et*

al., 1980b). Non-glaucomatous optic atrophy may be frequently misdiagnosed as glaucoma even by ophthalmologists experienced in optic nerve head assessment (Trobe *et al.*, 1980a), and chiasmal compression due to pituitary tumours may be misdiagnosed as normal tension glaucoma (Drummond & Weir, 2010).

Although some 160 years have passed since the invention of the ophthalmoscope, many disc features we use today to detect glaucoma and glaucomatous progression have only been discovered relatively recently. Optic disc splinter haemorrhages (also known as Drance haemorrhages) were first flagged as acute ischaemic disc changes by Drance and Begg (1970), while the importance of vertical elongation of the optic cup as a pathological change associated with glaucoma was noted by Kirsch and Anderson (1973). The role of radial peripapillary capillaries in providing nourishment to the retinal nerve fibres surrounding the optic nerve head was observed by Henkind (1967), with atrophic changes to these vessels in glaucoma specimens found by Kornzweig et al. (1966). Peripapillary atrophy was found to be associated with glaucoma by Primrose (1971).

Direct observation of the optic nerve head is subjective in nature, resulting in significant scope for inter-observer differences in interpretation of any or all disc features. Photographic assessment of the optic disc provided an improvement in the ability to record disc features such as *Cup/Disc Ratio* (**CDR**), vessel configurations and peripapillary atrophy. Perhaps the most important diagnostic feature observed has traditionally been CDR.

The ONH contains of approximately one million retinal nerve fibres (Tatham *et al.*, 2013), and the size of the average ONH can vary depending on ethnicity (Lee *et al.*, 2013). In a normal population, there may be an area of the ONH that does not contain any RGC axons. This area is known as the physiological cup of the optic disc. The ratio of the size of the cup to the size of the disc is known as the cup/disc ratio, and as such it has no units of measurement.

Average optic disc vertical and horizontal diameters have been reported as 1.92 mm and 1.76 mm respectively (Jonas *et al.*, 1988), and 1.88 mm and 1.77 mm respectively (Quigley *et al.*, 1990), with an average vertical CDR of 0.4 in 6

year old children (Samarawickrama *et al.*, 2012). The wide range of normal values for optic disc size means that the range of naturally occurring CDR's in the population limits the diagnostic ability of CDR on its own (Tatham *et al.*, 2013). This natural variability in ONH size is compounded by the finding that the ONH in myopic eyes differs from that of normal eyes (Hyung *et al.*, 1992).

Figure 3A.

Large optic disc (2.39 mm vertically, 1.97 mm horizontally), CDR 0.53

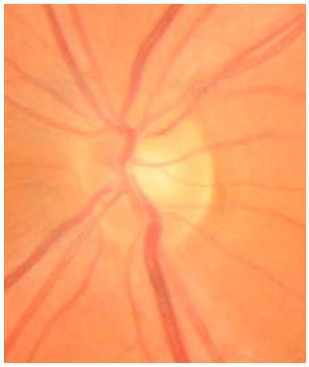


Figure 3B.

Small optic nerve (1.63 mm vertically, 1.39 mm horizontally), CDR 0.44

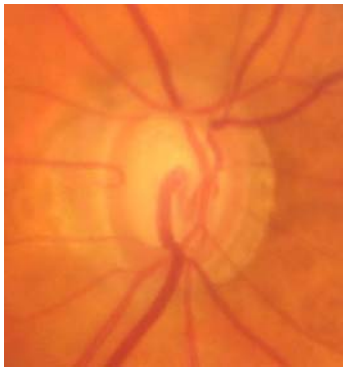


Figure 4.

Large CDR with normal IOP and no structural or functional glaucomatous change. Note the circumpapillary atrophy associated with this optic nerve head.

The effect of disc size on CDR and cup size is shown above (Figs. 3A and 3B). Although markedly different in appearance, CDR differs only marginally. The difficulty of using CDR as a diagnostic indicator is illustrated in Figure 4, where a very large physiological CDR is present in a patient with no structural change using SD-OCT, no functional change using SAP and normal IOP's.

The loss of RGC's due to disease reduces the number of axons exiting via the ONH, and will cause an increase in the size of the optic disc cup. By extension, this will alter the CDR, since although changes to optic disc size in glaucoma suspects has been reported (Toda *et al.*, 2010), disc size generally remains relatively stable. The observation and recording of changes to CDR over time is an important component in determining the possible presence of a disease process, and the need for further investigation with other diagnostic modalities

Retinal photography not only provided a permanent record of a patients' optic nerve head and clinically relevant features, it also provided some evidence of glaucomatous changes (e.g. via loss of retinal reflectivity) to the retinal nerve fibre layer, not only at the disc, but also extending to the macula (Tuulonen *et al.*, 1993). The advent of increasingly sophisticated assessment methods has enabled a much more extensive investigation of the retinal changes associated with glaucoma.

Stereophotography enhanced the photographic technique of assessing optic disc topographic features (Kottler *et al.*, 1975), and was found to be helpful in assisting in the early diagnosis of glaucoma (Odberg & Riise, 1985). Despite this improvement in optic nerve head evaluation, stereophotography still engendered poor inter-observer agreement. In one study, using glaucoma specialists to determine progressive disc change in glaucoma, in 40 % of cases judged to have progressed, the "worse/progressed" disc photo was actually taken first (Jampel *et al.*, 2009). Moreover, inter-observer experience was found to influence inter-observer agreement in the assessment of cup/disc ratio (Hanson *et al.*, 2001), and inter-observer agreement in assessing stereo-slides for glaucomatous change between non-expert ophthalmologists was significantly lower than that of glaucoma specialists, whose own inter-observer agreement was only moderate (Breusegem *et al.*, 2011). A knowledge of the

chronological order stereophotographs were taken in was also found to provide considerably different interpretations of glaucomatous change when compared to observations made without this knowledge (Altangerel *et al.*, 2005).

Technological advances

Clearly, an improvement in objective assessment of optic nerve head topography and retinal structure would be vital to the improvement in detection of glaucomatous changes to the optic nerve head. One of the earliest methods was scanning confocal microscopy, the principles of which were first described in 1961 (White, 1987). The *confocal Scanning Laser Ophthalmoscope* (**cSLO**) incorporates a focussed laser beam to scan the retina, obviating the need for mydriasis and improving the resolution of some fundus structures in comparison to retinal photography (Woon *et al.*, 1992).

Improvements in cSLO technology have seen cSLO incorporated into modern SD-OCT instruments, most notably in the Heidelberg Engineering Spectralis, where it provides a complimentary fundus imaging strategy to SD-OCT. It is able to provide images using infra-red reflectance, blue reflectance, multiple colours, blue peak autofluorescence fluorescein angiography, which can image defects found with SD-OCT, as well as defining some pathologies which are not delineated by SD-OCT. The alignment of the near infra-red image and the tomograph scan in the Spectralis has been reported to vary according to temperature and the number of scans per patient (Barteselli *et al.*, 2013). Although this difference was not sufficient to require realignment in clinical settings, it was suggested that for research purposes, the realignment should be checked after every second patient.

Another method used to detect changes to retinal structure was *Scanning Laser Polarimetry* (**SLP**), with the first commercial nerve fibre layer analyser appearing in 1993 (Da Pozzo *et al.*, 2009). This technique makes use of the birefringence properties of the RNFL, and measures RNFL thickness by interpreting the changes in polarisation occurring when light is reflected from the RNFL (Da Pozzo *et al.*, 2009). Although recent studies have reported similar specificity for SLP and SD-OCT (Garas *et al.*, 2012), the introduction of SD-OCT instruments having greater clinical versatility has meant that clinicians,

often forced to choose one piece of equipment for RNFL assessment, have preferred SD-OCT due to its greater range of scan protocols and its use in a variety of ocular conditions.

The pre-eminence of SD-OCT in the clinical detection of glaucoma has therefore, to some extent, arisen due to the fact that it is also a vital component in the detection and ongoing management of many other retinal conditions. These include serous maculopathy, dry age-related macular degeneration, cystoid macular oedema, central serous retinopathy, drusenoid macular degeneration, diabetic macular oedema, macular telangiectasia and numerous other retinal pathologies. The importance of SD-OCT in diagnosing and managing these conditions areas has led to a significant uptake of the technology by optometrists and ophthalmologists.

Optical coherence tomography

One of the more recent advances in glaucoma detection and monitoring has been the ability to perform non-invasive cross sectional imaging of the posterior eye using optical coherence tomography (Huang *et al.*, 1991). The earliest instruments applied to glaucoma detection and analysis were *Time Domain Optical Coherence Tomographs* (**TD-OCT**), which operated by scanning the reference path length by moving the reference mirror in a raster pattern (Podoleanu, 2012). The Stratus OCT (Carl Zeiss Meditec Inc, Dublin, Ca) is a good example of this technology, and is able to acquire 512 A scans in 1.3 seconds (Sakata *et al.*, 2009).

The technological progression of optical coherence tomography has led to the development of SD-OCT, which either utilises a spectrometer as a detector or varies the wavelength of the light source (van Velthoven *et al.*, 2007). This alteration in signal acquisition enables modern instruments to obtain 18,000 to 40,000 A scans per second (Sakata *et al.*, 2009), not only increasing the amount of data available for processing but also reducing the effects of eye movements due to much shorter scan times.

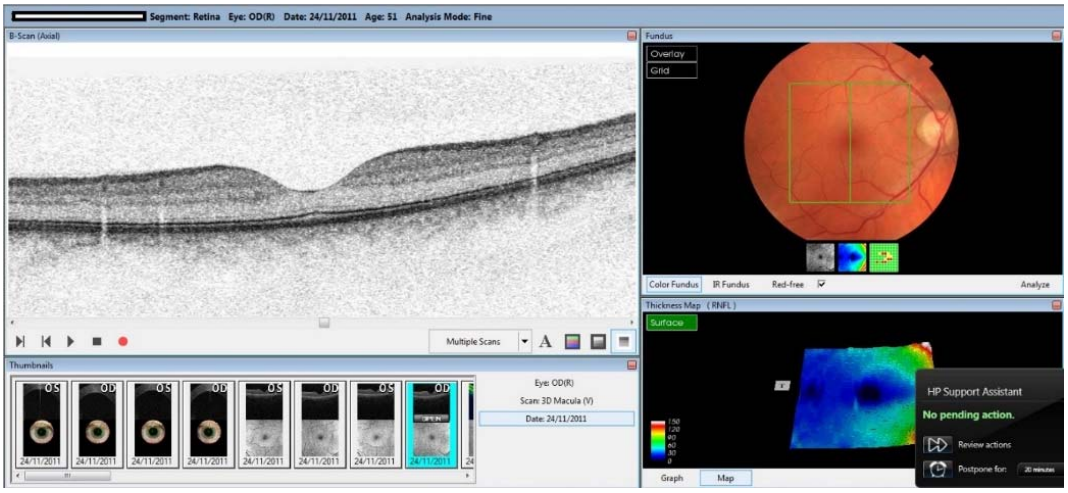


Figure 5.

Topcon SD-OCT display for the Macula V scan, which provides information on the macular inner retinal layer thicknesses. The scan is centred over the macula as per the grid overlying the fundus photo in the top right hand corner. Individual retinal layers can be seen in the scan at left. Clear vertical lines are artefacts caused by retinal blood vessels overlying the neurosensory retina. The subject of this scan was a healthy patient.

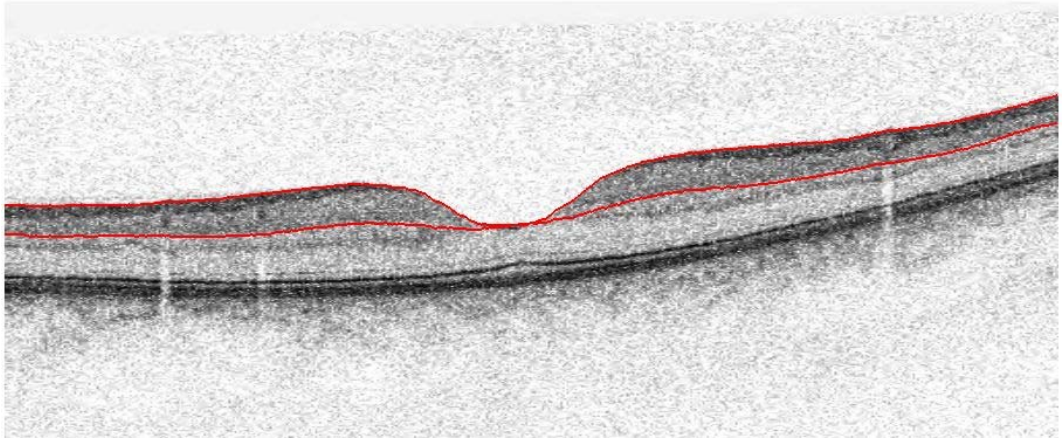


Figure 6.

Scan from the healthy subject in Figure 5, with the ganglion cell complex delineated by the red lines. The ganglion cell complex comprises the inner retinal layers surrounding the macula, and consists of the nerve fibre layer, the ganglion cell layer and the inner plexiform layer.

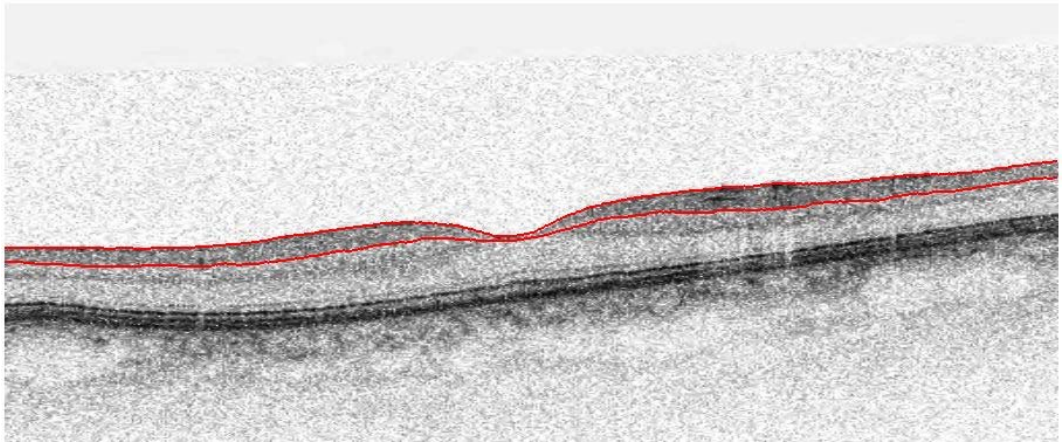


Figure 7.

Macula V scan of a glaucoma patient, illustrating the decreased thickness of the ganglion cell complex as compared to the healthy subject in Figure 6.

Technological improvements also led to different SD-OCT scan protocols, enabling the measurement of pRNFL parameters. ONH topography, pRNFL thickness, disc dimensions, and the individual components of the GCC and the overall GCC thickness (Figs. 5, 6, 7 and 8) were all able to be assessed. The ability to assess structural changes in the central retina that had previously been difficult to detect and quantify with functional psychometric tests (see the Perimetry section below) has become an important additional diagnostic criteria to be used in the assessment, diagnosis and management of glaucoma patients and glaucoma suspects. Measurement of the GCC is now an integral component of glaucoma diagnosis and assessment.

Although some studies have suggested GCC measurements have not demonstrated greater diagnostic ability than pRNFL measurements (Kim *et al.*, 2010), other studies have found that the GCC is significantly thinner in glaucoma suspects and pre-perimetric glaucoma patients (Kotera *et al.*, 2011), and that 38% of glaucoma suspects had an abnormal GCC thickness while only 13% had an abnormal pRNFL thickness (Ganekal, 2012).

Objective measurement of the ONH removed much intra- and inter-observer variability. Although this was of course replaced by instrument retest variability, the reduction of the subjective element in detecting structural change provided an opportunity for significant reduction in inter-visit variance in the assessment of ONH morphology.

In order to improve on current subjective levels of variability involving cup/disc ratios, instrument variability had only to improve beyond levels observed in some studies, such as Varma et al. (1992), where inter-observer differences in cup/disc estimation was as high as 0.16 when viewing stereoscopic disc photos. More importantly, many other optic disc parameters could now be quantified objectively, including disc area, cup area, rim area, cup/disc area ratio, linear and vertical cup/disc ratios, cup volume, rim volume, horizontal disc diameter and vertical disc diameter. Any or all of these parameters could now be monitored for progressive changes in glaucoma patients and glaucoma suspects, representing a substantial improvement in the structural assessment of the optic nerve head in glaucoma.

Scans of the optic nerve head are also able to incorporate measurements of the pRNFL, including average, quadrant and clock hour thicknesses. The loss of RGC's is reflected in axonal losses and consequent thinning of the pRNFL, with the thickness characteristics of the losses able to be used in conjunction with other clinical observations to determine the presence or progression of glaucoma. This was done by comparing the results from any individual patient with the normative database of the instrument, and assigning a percentile value to the data to indicate whether the result was within normal limits (generally within the 95th percentile) or outside normal limits. The data generated in this regard could also be used diagnostically to monitor progressive changes in glaucoma patients.

Interpretation of SD-OCT scans

Clinicians need to be careful when interpreting pRNFL results, as other conditions can also produce thinning of the pRNFL. Neurological conditions such as optic tract syndrome can produce preferential superior and inferior thinning of the optic disc rim and surrounding pRNFL in the ipsilateral eye, and preferential temporal and nasal thinning of the optic disc rim and surrounding pRNFL in the contralateral eye (Kanamori *et al.*, 2013), giving rise to pRNFL thickness patterns which might be misinterpreted as glaucomatous rather than neurological.

The macula scan protocols incorporated into SD-OCT's has enabled the objective measurement of central retinal structures associated with changes to central visual function in glaucoma, particularly the inner retinal layers comprising the inner plexiform layer, the ganglion cell layer and the retinal nerve fibre layer, which together comprise the GCC. Although the GCC scan protocols are designed to target the inner retinal layers directly affected by glaucoma in the area of their greatest density (Kim *et al.*, 2010), it should be remembered that a variety of other disease processes can also affect the inner retina, such as diabetic retinopathy, retinal vein occlusion and retinopathy of prematurity (Margalit & Sadda, 2003). Decreased hemiretinal GCC thickness corresponding to hemifield losses has been found in patients with a history of unilateral

posterior cerebral artery infarction (Yamashita *et al.*, 2012), and GCC thinning has been found in non-arteritic anterior ischaemic neuropathy (Gonul *et al.*, 2013). Perhaps more critically, patients with non-arteritic ischaemic optic neuropathy can present with superior, inferior or bihemispheric field loss (Aggarwal *et al.*, 2012), which may mimic those encountered in glaucoma and also result in a GCC loss pattern consistent with the visual field loss (Aggarwal *et al.*, 2012).

The clinical application of SD-OCT has consequently led to an increasing number of manufacturers supplying these instruments, as well as generating the requirement for clinicians to be able to interpret the results obtained from these instruments. In terms of glaucoma detection and monitoring, some of the earliest OCT's, such as the Stratus TD-OCT and the Cirrus SD-OCT (Carl Zeiss Meditec Inc, Dublin, Ca) have been extensively utilised in studies on both glaucoma patients and healthy patients. This has resulted in the availability of a wide range of clinically applicable guidelines for interpreting both change over time in SD-OCT scans of optic nerve head morphology, pRNFL and measurements of the nerve fibre layer, ganglion cell layer and inner plexiform layer (which together comprise the ganglion cell complex). These studies have also been able to provide indications as to the presence of absence of disease at the initial consultation, and over time, with these instruments (Huang *et al.*, 2011; Kotera *et al.*, 2011; Moreno *et al.*, 2011; Mwanza *et al.*, 2011; Ganekal, 2012).

SD-OCT objectives of this study

One of the problems for clinicians utilising SD-OCT instruments produced by more recent entrants to the market is the lack of research data with which to use in the interpretation of the data obtained from these instruments. ***One of the aims of the current study has been to develop clinically applicable diagnostic criteria for one of the newer instruments on the market***, the Topcon 3D OCT-2000 (Topcon Corporation, Tokyo, Japan). The Topcon is a widely used SD-OCT, with over 30 published papers reporting on this instrument. The Topcon also incorporates a high resolution fundus camera, and

is able to carry out pachymetry and anterior ocular scans as well as the posterior macular and optic nerve head scans.

The Topcon 3D disc scan (Fig. 8) does incorporate some methods of analysis to assist clinicians in determining change over time in average pRNFL and superior and inferior quadrants via trend analysis. The slope is given in μm/year, but there is no significance calculated. The problem with this strategy is that it can be difficult to differentiate a progressive trend (in any direction) from the variability inherent in the instrument itself if the limits of instrument TRV for the parameter under consideration are not known.

This difficulty in the Topcon is compounded by the fact that the average pRNFL figure for this instrument does not provide any indication as to which population percentile any particular measured value falls into (Fig. 8), and the fact that relatively small changes in micrometre thickness for some parameters can have a dramatic impact on the change in population percentile. There is no trend analysis in the instrument for GCC scans, making the detection of true progressive loss in this parameter difficult.

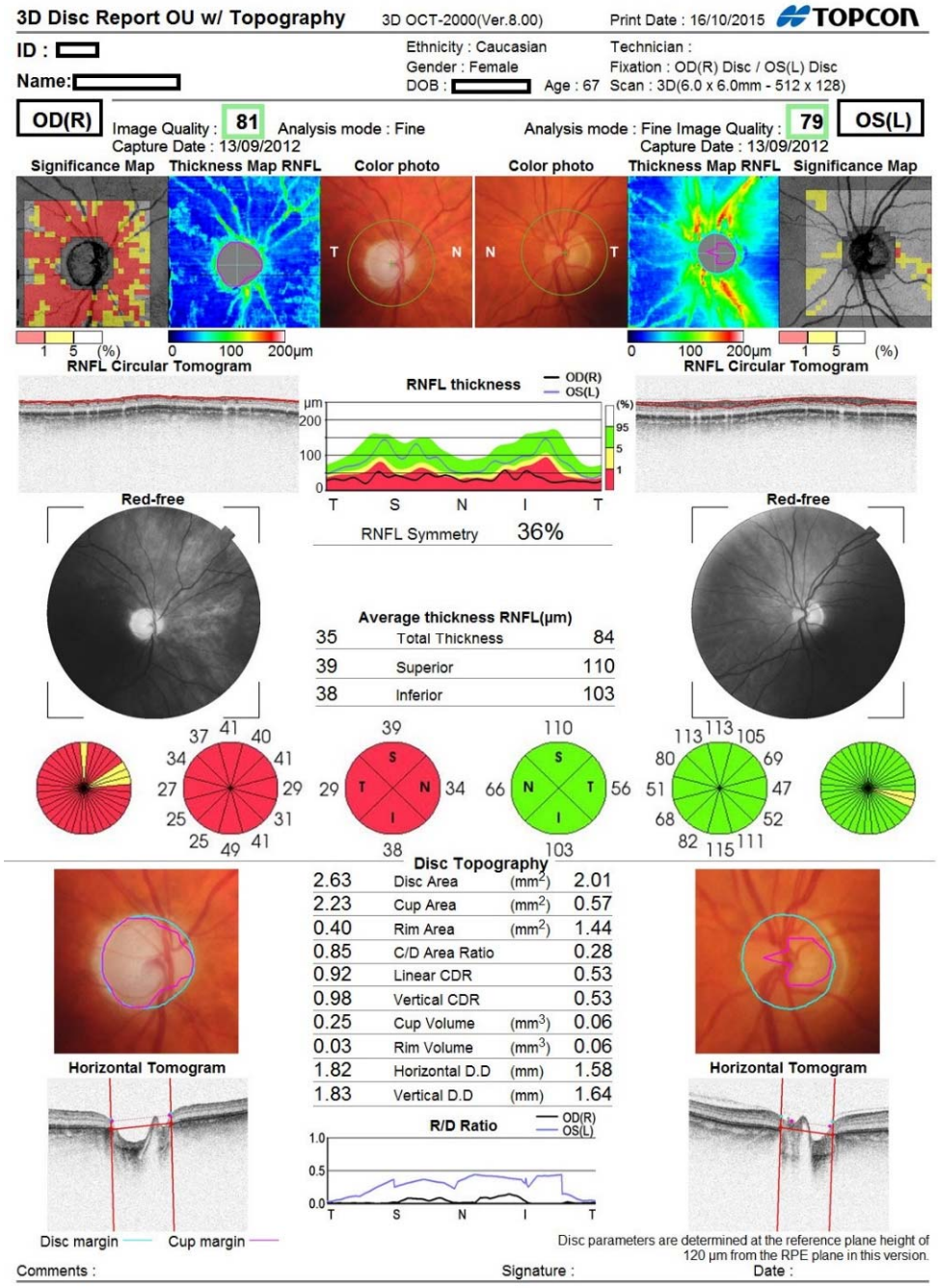


Figure 8.

Topcon 3D Disc scan results display. Note that although clock hours and quadrants are colour coded for the degree of abnormality, average RNFL thickness makes no categorisation as to the deviation from an age normal expected result. The right eye scan in Fig. 8 is the same eye used for the scan in Fig. 7

At present, only one other study has developed any repeatability criteria for the Topcon 3D-OCT 2000 average pRNFL measurements (Pierro *et al.*, 2012), and no studies have as yet developed any TRV criteria for GCC scan results. In that study (Pierro *et al.*, 2012), the standard deviation for average pRNFL for this instrument was calculated, and compared to six other SD-OCT's. The clinical applicability of this measurement seems minimal, and was only determined for healthy patients. At present, no studies have as yet developed any TRV criteria for GCC scan results.

Due to the current lack of TRV criteria for the Topcon, the first part of our study consisted of the measurement and comparison of inter-visit TRV of the Topcon and three other SD-OCT's in healthy subjects in average pRNFL (Appendix 3, Part 1). We also undertook a comparison of GCC scan results of the Topcon with another well studied SD-OCT, the RTVue 100 (Appendix 3, Part 2).

The retest variance (σ) encountered in any instrument is the sum of any elements which might induce retest variance. This can be represented as

(1) $\sigma^2_{total} = \sigma^2_{retest} + \sigma^2_{machine} + \sigma^2_{subject\ (n)} + \sigma^2_{operator} + \sigma^2_{diurnal} + \sigma^2_{annual} \ldots + \sigma^2_{error}$.

We wanted to reduce this to

(2) $\sigma^2_{total} = \sigma^2_{retest} + \sigma^2_{machine} + \sigma^2_{subject\ (x2)} + \sigma^2_{operator\ (n = 1\ or\ 2)} + \sigma^2_{error}$

which we achieved by testing two very well characterised subjects (decreasing subject-wise σ), at the same time of day (removal of diurnal σ) over a period of four weeks (removal of seasonal σ). We also conducted a large number of scans per eye (eight per instrument) over a four week period (to generate clinically meaningful inter-visit confidence intervals) to make this comparative study quite robust.

This study gave us TRV information for several scan protocols, which could be used in later parts of the study when determining whether any circadian variations in SD-OCT parameters could be detected (Chapter 3 and Appendix 2).

In this thesis, we generate 95% confidence intervals for a number of pRNFL and GCC parameters for both healthy eyes and glaucoma patients. These event based criteria can then be easily applied clinically. The diagnostic criteria developed in the current study is intended to be applied by clinicians involved in patient assessments on a daily basis, such as optometrists and ophthalmologists.

Extrinsic factors which might impact on SD-OCT retest variability were also investigated here. At present, there are no studies regarding circadian or diurnal variations in pRNFL measurements with SD-OCT. In this thesis, we investigate the influence of any circadian or diurnal influences on TRV using healthy patients. Given that glaucoma may cause disturbances to circadian rhythms *(Wang et al., 2013)* and induce greater asymmetry in paired 24 hour rhythms of IOP (Liu & Weinreb, 2014), our findings are therefore not directly applicable to glaucoma patients. Since ageing may also cause disturbances to the circadian pacemaker (Hofman & Swaab, 2006), this factor may also affect the application of our findings to older healthy subjects.

This aspect was examined in the current study by conducting SD-OCT scans over a 24 hour period to determine whether retest variability might be influenced by testing being carried out at different times of day (Appendix 2). In an adjunctive study, SD-OCT TRV was also assessed during normal office hours to determine whether any clinically significant effects on TRV could be attributed to the time of day testing was undertaken (Chapter 3).

The results obtained from this part of the study were extremely important, as they have indicated that although there were no significant changes to mean SD-OCT parameters at different times of day, there were significant changes to TRV in some parameters. This finding may have implications not only for the detection of progression in glaucoma using serial SD-OCT scans, but may also have implications for the results of other studies which may have included scans from subjects at different times of day.

SD-OCT is an important technological improvement in the assessment of structural change to the optic nerve and the retina in the detection of glaucoma and glaucomatous progression. At present, it cannot detect other diagnostic

indicators such as neuro-retinal rim colour or the presence of disc haemorrhages, and therefore clinical observation of the disc remains an essential component of glaucoma diagnosis and management. As an adjunctive measure, it can provide confirmatory evidence to assist clinicians in challenging cases, such as those where field defects are present without cupping (Armaly, 1969), retinal nerve fibre layer loss is present in conjunction with glaucomatous visual field defects without disc cupping (Sherman *et al.*, 2004), and may assist clinicians in differentiating glaucomatous and non-glaucomatous cupping of the optic disc (Gupta *et al.*, 2011).

Perimetry

History of perimetry

Although structural assessment of the ONH and GCC are extremely important for diagnosing and monitoring glaucoma, the assessment of functional losses in visual field sensitivity are still critically important, and in some cases may in fact precede the structural changes detected by SD-OCT (Hood & Kardon, 2007).

Perimetry, in terms of visual fields, refers to the assessment and measurement of the sensitivity, function and extent of the visual fields. An early record of visual field loss dates from the late fifth century B.C, when Hippocrates observed and described a hemianopia, while the first estimate of the extent of the normal visual field was reported by Thomas Young in the early 1800's (Johnson *et al.*, 2011). The arc perimeter (Fig. 9) was introduced in the 1860's by Richard Forster (1825-1902) as a means of maintaining equidistant stimulus presentations to all retinal points using a target moved along the arc. Although the arc perimeter allowed an accurate assessment of the size of the visual field, it was not suited to assessing small central visual field defects (Simpson & Crompton, 2008). To address this shortcoming, in 1890 Jannik Bjerrum (1851-1920) introduced improvements in the quantitative method of utilising the tangent screen as a means of investigating the central 30° of the visual field more intensively (Hardy le, 1931).

Figure 9.

Arc perimeter showing the two arcs along which stimuli are moved. The arcs can then be rotated to examine different parts of the visual field.

Figure 10.

Goldmann bowl perimeter showing the evolution from two moveable arcs being replaced by a solid bowl. The Goldmann perimeter still required manual testing protocols to be carried out by the examiner.

Arc perimeters and tangent screens continued to be the main methods of visual field assessment until 1945, when Hans Goldmann (1899-1991) introduced the Goldmann Cupola (Bowl) perimeter (Fig. 10), which enabled the standardisation of both background and target luminance (Gloor, 2010). It also incorporated a movable optical projection system which allowed both static and kinetic perimetry and enabled the variation of stimulus size, colour and luminance characteristics (Johnson *et al.*, 2011). The Goldmann perimeter still required manual positioning of the test targets, and the next goal was to reduce the time taken for testing by automating stimulus presentation.

One of the first instruments to achieve this commercially was the OCTOPUS, introduced in 1975 (Gloor, 2009), which utilized threshold static perimetry

(Portney & Krohn, 1978). Although the OCTOPUS was found to detect over 90% of patients found to have a visual defect using Goldmann or Tubingen perimeters, it was found to be inferior to the Goldmann perimeter in providing clinically useful information in neurological disease (Li *et al.*, 1979).

Standard Automated Perimetry

By 1980, the OCTOPUS automated perimeter was able to detect greater field loss in glaucoma and glaucoma suspect patients than the manual Goldmann perimeter, and was also able to detect losses in patients where no losses had been found with the Goldmann perimeter (Schmied, 1980). In 1985, the Humphrey Field Analyser (HFA) became commercially available, and due to its extensive use throughout the world, both in office and in many glaucoma studies, became the de facto gold standard for comparisons with other perimeters. As a result, many of the studies in the literature regarding retest variability in visual fields, for example, those by Heijl et al. (1989), Chauhan et al. (2008) and Junoy Montolio et al. (Junoy Montolio *et al.*, 2012) have used the HFA.

With improvements in computer technology, new test algorithms, such as the Swedish Interactive Thresholding Algorithm (SITA) in the HFA, were developed which significantly decreased the time taken for testing while maintaining the test quality (Bengtsson *et al.*, 1998).

With the advent of a variety of different perimetric techniques, the term SAP is now used to define achromatic white-on-white static perimetry, generally using a Goldmann Size 3 stimulus. There are now numerous alternative perimetric techniques that have been developed to try and provide earlier detection of initial glaucomatous field loss as well as earlier detection of progressive field deterioration.

Short Wavelength Automated Perimetry (**SWAP**) was designed to have the stimulus detected by short wavelength cones, with the test then targeting the blue-yellow ganglion cell population (Sample *et al.*, 2000). *Frequency Doubling Technology* (**FDT**) attempts to isolate the Y-like ganglion cells which, being larger and having less overlap, meant that cell losses may be more easily

detected (Maddess, 2011a). To what degree FDT is able to do this is unknown (Sample *et al.*, 2000). Although some studies suggest that FDT may be superior to SAP in detecting early visual field loss in glaucoma suspects (Ferreras *et al.*, 2007), recent studies have reported that FDT does not provide any improvement in the detection of visual field deterioration when compared to SAP (Redmond *et al.*, 2013). These alternative perimetric techniques are now often referred to as "*selective perimetry*" due to their attempts to isolate particular retinal cell populations that may have some advantages in detecting specific RGC losses.

Although an increasing range of test modalities are now available, SAP remains the gold standard for detecting functional losses in glaucoma and progressive damage (Alencar & Medeiros, 2011). Improvements in computer technology, examination protocols and advances in analysis techniques have helped to maintain the importance of SAP in glaucoma detection and analysis (Tonagel *et al.*, 2012).

Despite these advances, the sensitivity of SAP is often regarded as questionable, with some studies indicating that a 3 dB loss of sensitivity requires the loss of 50% of RGC's at 21° eccentricity, and 70% at 4° eccentricity (Garway-Heath *et al.*, 2000). Estimated RGC counts in eyes with early visual field defects show on average a 28.4% loss of RGC's compared to healthy eyes at the earliest detection of a SAP visual field defect (Medeiros *et al.*, 2013). Some of the repeatability and sensitivity issues regarding SAP revolve around the fact that RGC densities, redundancy and dense overlapping of receptive fields can lead to under-sampling and aliasing using a 6° grid (Maddess, 2011b, 2014), as in the HFA.

Despite this apparent shortcoming, SAP remains an important clinical method of measuring functional loss. Although structural measures of RGC loss have become increasingly important with the advent of SD-OCT, functional assessment remains critical given that RGC's apparently pass through a period where there are indications of neuronal damage, but the death of the cells is not a foregone conclusion (Kisiswa *et al.*, 2010).

Electroretinography can be used to measure the electrical activity of the retina, and a flash of light can generate this electrical response. More recently, stimulation with gratings or chequerboard patterns generated responses called *Pattern Electroretinograms* (**PERG**) (Berninger & Arden, 1988) have been used to assess RGC function. Studies have indicated that the loss of RGC function with **PERG** may precede the detection of structural loss with OCT by several years (Banitt *et al.*, 2013), which reinforces the importance of detecting RGC dysfunction at the earliest possible stage.

The importance of early detection of functional loss is highlighted by the fact that alterations in the PERG in early glaucoma may be reversed by lowering the IOP (Porciatti & Ventura, 2012), and improvements in visual field sensitivity have been noted in up to 31% of adult patients treated for glaucoma (Katz *et al.*, 1989). This apparent restoration of RGC function may lie in the fact that dysfunctional, but viable, RGC populations may exist in early glaucoma (Ventura *et al.*, 2006). The early detection of functional loss becomes even more important when we consider that detectable improvement in RGC function with the PERG occurs with IOP reduction in eyes with early visual field loss, whereas eyes with severe visual field loss demonstrate little improvement in RGC function (Ventura & Porciatti, 2005).

Test-Retest Variability in Standard Automated Perimetry

A significant problem encountered with subjective psychometric functional tests is TRV, which creates difficulties in clinical decision making when trying to differentiate variability from true progression in glaucoma suspects and glaucoma patients. Sources of variability with SAP are myriad, with undersampling and aliasing with some test protocols (Maddess, 2011b), seasonal changes affecting mean deviation (Gardiner *et al.*, 2013), small fixational eye movements interacting with the visual field gradient (Wyatt *et al.*, 2007; Maddess, 2014), stimulus size (Wall *et al.*, 2013), reduced sensitivity in damaged areas of the visual fields (Wall *et al.*, 2009) and technician experience, time of day and time of year (Junoy Montolio *et al.*, 2012). It can be seen that some sources of variability can be addressed clinically, and it is

important for clinicians to identify those sources and minimise their impact on test reliability.

A possible contribution to SAP TRV is performing tests at different times of day, and this has been noted by Junoy-Montolio et al. (2012), with the finding of a 0.2 dB reduction in Mean Defect (Humphrey Field Analyser, 30-2 Swedish Interactive Test Algorithm) for tests performed after lunch. Diurnal visual field testing has been proposed as part of glaucoma assessment and management protocols for improving the diagnosis of early stage glaucoma by Radzikhovskii et al. (1974). The implications from these studies is that diurnal and circadian variations in visual field test repeatability, sensitivity and diagnostic indices may provide a clinically relevant amount of variability in SAP testing.

In this thesis, we have attempted to reduce the sources of variance contributing to the total variance in any test regime. If we consider total variance to be

(1) $\sigma^2_{total} = \sigma^2_{retest} + \sigma^2_{machine} + \sigma^2_{subject\ (n)} + \sigma^2_{operator\ (m)} + \sigma^2_{diurnal} + \sigma^2_{annual} + \sigma^2_{error}$

we can reduce this to

(2) $\sigma^2_{total} = \sigma^2_{\ retest} + \sigma^2_{machine} + \sigma^2_{\ subject\ (n\ is\ small)} + \sigma^2_{operator\ (m\ =\ 1\ or\ 2)} + \sigma^2_{\ error}$

which we have done when investigating the TRV of four different SD-OCT's (Appendix 3, Parts 1, 2 and 3).

To date, no studies have investigated the possibility of variations in visual field sensitivity over an entire 24 hour period, which would enable a greater range of information to compare with normal office hour field test results. Any circadian variability could then be compared to reported nocturnal changes in IOP (Kida *et al.*, 2006; Read *et al.*, 2008; Chakraborty *et al.*, 2011) and optic nerve head perfusion (Pemp *et al.*, 2009). Any correlations between these findings could then be further investigated regarding the mechanisms involved in RGC death and dysfunction in glaucoma. One component of this thesis has therefore been the undertaking of visual field tests using the M700 perimeter over a 24 hour period, on two separate occasions, to determine whether any circadian

variations exist in visual field sensitivity, and the clinical significance of these variations should they exist (Appendix 1).

The results from these 24 hour test sessions indicated the possibility of changes to visual field sensitivity during normal office hours, when patients might be reasonably expected to undergo visual field testing. Since the Medmont perimeter (Fig. 12) has a significantly different stimulus presentation pattern to the HFA (Fig. 13) (as used by Junoy-Montolio et al. (2012)), an investigation of changes to sensitivity and repeatability with this instrument may provide an indication as to whether a different spatial arrangement of test points may indicate a different level of circadian alterations visual field sensitivity, variability and diagnostic indices.

An additional investigation (Chapter 3) was therefore added to the original study component to ascertain the effect of time of day testing on visual field sensitivity and retest variability over the course of a normal working day. The results from this part of the study have indicated that although visual field index mean values did not alter significantly over the course of the day, TRV for most parameters did. These results indicated that visual field testing in the late afternoon may be able to reduce TRV in the M700.

Although SAP is regarded as a relatively insensitive measure of ganglion cell loss (Garway-Heath *et al.*, 2000), SAP defects are sometimes found prior to detectable structural losses with OCT (Harwerth *et al.*, 2007). There are suggestions that although RNFL thickness may be a more sensitive measure of early glaucomatous changes, SAP may be a better method of detecting progression in moderate to advanced glaucoma patients (Harwerth *et al.*, 2007). It is for this reason that clinically applicable measures to detect statistically significant changes in visual field indices have been developed for the M700 as part of the current study.

The Medmont Perimeter

As automated perimetry became increasingly important in the detection of glaucomatous field loss and progression, the range of instruments and programs also increased substantially. As with SD-OCT, many of the newer

instruments were not used as extensively in comparative studies, and nor was there information available about the TRV characteristics of these instruments.

An example in this regard is the *Medmont Automated Perimeter* (**MAP**) (Medmont P/L, Nunawading, Victoria, Australia), the original version of which, the M600, was introduced in 1989 (Fig. 11). A later version, the M700, was introduced in 2000, with the M600 being able to be upgraded to use M700 software. With around 3,700 units in total sold worldwide, the Medmont automated perimeter plays an integral role in the detection and monitoring of glaucoma for numerous patients in many countries.

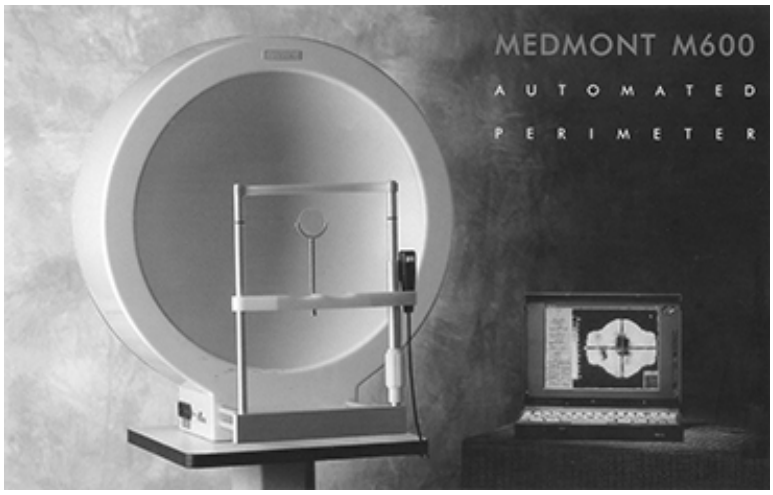


Figure 11.

Medmont M600 perimeter still retains a similar bowl design to the Goldmann, but stimulus presentations are now fully automated and carried out by software programs. The perimetrist can still make adjustments to the test protocols as required for individual patients.

The M700 automated perimeter utilises light emitting diodes as test stimuli that match the Goldmann Size 3 angular subtense. This strategy means that although target size and position cannot be altered as it can in the HFA, there are 164 test positions that are utilised differently in the various test strategies available in the Medmont. The HFA 30-2 and Medmont Central 30° tests differ substantially in the distribution and spacing of test points. The HFA has test points arranged in a grid pattern with each point separated by 6°, with the

central 10° of the 30-2 test only having 12 test points (Fig.13). The M600 and M700 central 30° test has test points arranged in circular patterns around the fixation point, with an increasing number of test points in each circle (Fig.12). The test points are at 3, 6, 10, 15, 22 and 30° from fixation. The result of this strategy is that the M600 and M700 have 44 points in the central 10° of the Central 30° test strategy, resulting in a significantly increased sampling rate centrally compared to the HFA. This is a significant decrease in test area sampled at each point, which may result in better detection of functional loss, given that small but deep retinal nerve fibre layer defects may not necessarily be associated with a scotoma using a 6° SAP grid (Burk *et al.*, 1998).

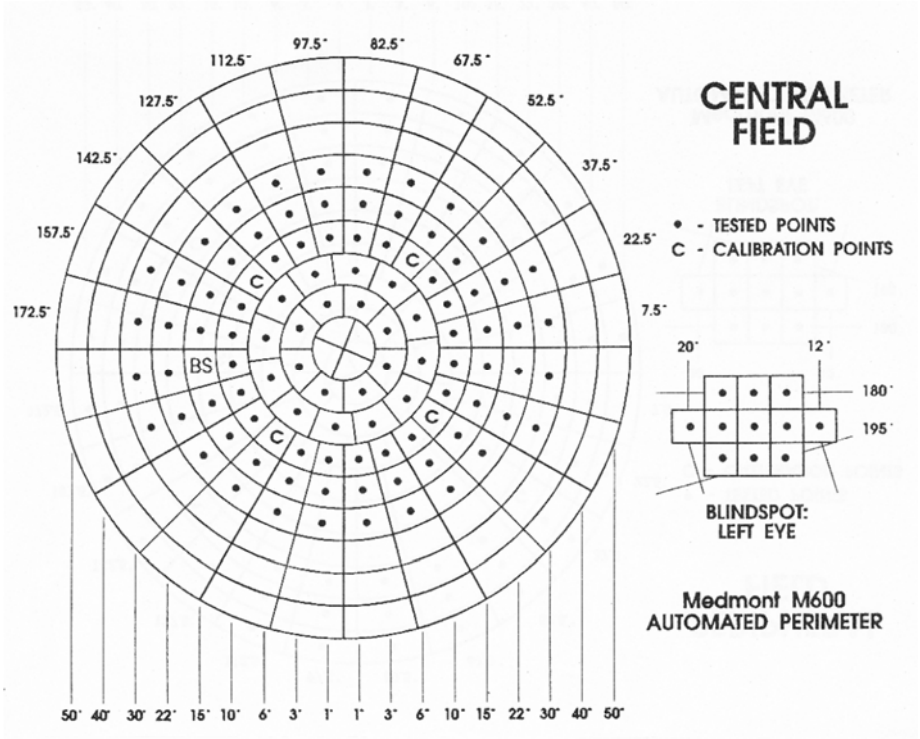


FIGURE 12.

Medmont M600 and M700 Central (30°) field test pattern layout, left eye shown. Reprinted with the permission of Medmont P/L, Australia.

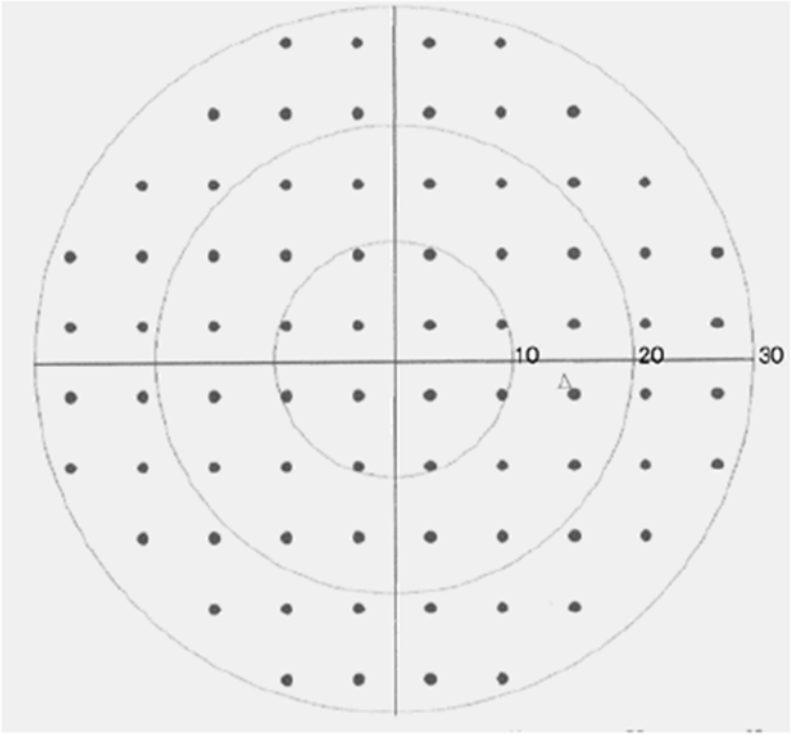


FIGURE 13.

Stimulus pattern layout of the Humphrey Field Analyser 30-2 test, reprinted with the permission of Carl Zeiss Meditec. Note the significantly reduced central sampling density in the central 10 °of the HFA compared to the Medmont (Fig. 12)

Relatively few studies have been carried out on the Medmont perimeters to provide comparative information with other perimeters, and those that have are often based on the older M600 programs rather than the newer M700 model and their associated programs. An example is the study by Landers et al. (2007), which, although published in 2007, used the original M600 in a comparison of global indices with the HFA II. This study found that the *Average Defect* (**AD**) and *Pattern Defect* (**PD**) from the M600 could be substituted, after appropriate conversion, for the *Mean Deviation* (**MD**) and *Pattern Standard Deviation* (**PSD**) obtained with the HFA II.

Critically, the Medmont global index of AD as generated by the M600 was replaced in the newer M700 by a new index: the *Overall Defect* (**OD**). The OD index was calculated in a significantly different way to AD. The results from studies with the M600 for AD were therefore not transferable to the OD index in the M700. This also resulted in OD not being directly comparable with the MD of the HFA II. ***One of the aims of the current study is to determine clinically applicable event based criteria for the M700 indices of OD and PD to improve the early detection of visual field deterioration with this instrument.***

Like the Topcon 3D OCT-2000, the M700 does have some trend analysis protocols incorporated into its software. Although this trend analysis does include OD and PD, there are no values given for slope and significance. Interpretation of these trend based analyses is further constrained by the lack of 95% confidence intervals for these indices to enable clinicians to differentiate variability from progression. Given that the M700 is a commonly used perimeter in Australia (Landers *et al.*, 2007), one aim of the current study was to generate 95% confidence intervals for OD and PD in the M700 which could be easily applied clinically when assessing progressive field loss in the visual field test results of glaucoma suspects and confirmed glaucoma patients.

The fact that the stimulus layout incorporated in the M700 cannot be altered enabled the investigation of an additional component of variability that has not been previously studied. The M700 has a *Central 30°* (**30°-test**) (Fig. 12) and a *Macula 10°* (**10°-test**). The 10°-test only differs from the central 10° degrees of

the 30°-test by the addition of four central test points. It is therefore possible with the M700 to compare the sensitivity and variability of the central 10° visual field under differing test conditions (i.e. in the 30°-test, where more peripheral points are also tested, and in the 10°-test, where no points outside the central 10° are tested).

In the current study, this aspect of TRV was investigated using a cohort of glaucoma patients with varying degrees of visual field loss to determine whether a given test protocol for the central 10° may provide lower TRV than other test strategies (Chapter 1). Given that it has been suggested that a 20% reduction in SAP test variability would be required to generate a clinically detectable improvement in the assessment of visual field change (Turpin & McKendrick, 2011), a possible decrease in TRV through better test selection for certain clinical populations might improve the glaucoma management of these patients.

Dividing the 30°-test into the central 10° and outer 20° also gave us the opportunity to assess the effect of eccentricity on TRV. The results we obtained in this regard indicated that with the M700, TRV did not increase with eccentricity. For points of equal decibel value, TRV in the central 10° did not differ from TRV in the outer 20°. This finding has significant implications for clinicians interpreting results from serial M700 visual field tests, as consideration now needs to be given to the sensitivity of a point, rather than its eccentricity, when differentiating TRV from true progressive change. This finding may also have implications for progression detection protocols in other perimeters, which utilise a smaller change in decibel value in the central 10° as being indicative of significant change compared to the change in sensitivity required to be significant in the outer 20° of the same test.

The integration of structural and functional testing results has also become an important component in the detection and management of glaucoma. Although SAP (functional) and SD-OCT (structural) testing are important diagnostic clinical entities in their own right, the integration of these two test paradigms may lead to earlier diagnoses of progression in glaucoma and glaucoma suspects. Hood and Raza (2014) suggested that the sensitivity and specificity of OCT imaging could be improved by visually examining high-quality images and

combining topographical information from visual fields and OCT scans. Medeiros et al. (2012) suggested that the nonlinear relationship between HFA MD and RGC counts indicated that face value interpretations of rates of MD loss over time can be misleading, reinforcing the need to combine structural and functional elements to improve progression detection in glaucoma.

Although none of our studies were specifically directed at investigating any relationships between structure and function, we did report some findings in this regard. In Chapter 3, I have investigated diurnal variations in the correlations between structure and function in the instruments under consideration, with the results indicating that there is the possibility that structure/function relationships may not be consistent at different times of day.

At present, the absence of clinically applicable confidence intervals for the instruments under investigation makes definitive confirmation of statistically significant progressive field loss difficult to determine. There are obviously implications for ongoing patient management, with the possibility that very early changes may be regarded as variability rather than true progression. The converse is also true, whereby variability may be misinterpreted as progression, resulting in patients perhaps being treated unnecessarily or current treatment regimes altered unnecessarily.

Synopsis

Glaucoma is an enigmatic disease, with numerous underlying pathophysiological processes contributing to the loss of retinal ganglion cells and subsequent structural and functional changes which characterise the condition. The key to effective glaucoma treatment protocols remains the early detection of the disease itself and the earliest detection of any progression subsequent to the instigation of topical and/or surgical therapies.

Current clinical practice for the detection of glaucoma and glaucomatous progression is also multifactorial, with IOP measurement, central corneal thickness measurement, direct observation of the optic nerve head, retinal reflectivity, structural assessment using SD-OCT and functional investigations using SAP providing a variety of inputs to assist in detection and treatment.

SD-OCT and SAP are critically involved in detecting structural and functional change in glaucoma, but both suffer some degree of TRV. In the current study, we have developed easily applicable event based criteria to differentiate variability from progression in two widely used instruments, the Topcon 3D OCT-2000 SD-OCT and the Medmont M700 automated perimeter. We have also compared the Topcon with three other SD-OCT's to compare its TRV with other commercially available instruments.

Here, I have also investigated the extent to which circadian and diurnal variations in structure, function and TRV exist in normal subjects, and whether they impart clinically significant additional variability. The results from the current study have not only provided additional impetus for conducting visual field tests at specific times of day, they have also indicated that consistent time of day testing (preferably late in the afternoon) may also be applied to SD-OCT scan protocols.

I have also generated 95% confidence intervals and clinical practice protocols for the instruments under investigation that can be rapidly and easily applied to test results obtained in clinical practice. To this end, this thesis has made a valuable contribution to the detection, diagnosis and management of glaucoma patients worldwide who are currently, and will be in the future, assessed with these instruments and protocols.

The integration of structural and functional testing results has also become an important component in the detection and management of glaucoma. By examining the results from Chapters 1 and 2, I have been able to assess the correlations between structure and function in the instruments under investigation. Crucially, I have also been able to examine the results from Chapter 3 to assess diurnal variations in these correlations, indicating that perhaps structure/function relationships may not be consistent at different times of day.

References

Aggarwal D, Tan O, Huang D & Sadun AA. (2012). Patterns of ganglion cell complex and nerve fiber layer loss in nonarteritic ischemic optic neuropathy by Fourier-domain optical coherence tomography. *Invest Ophthalmol Vis Sci* **53,** 4539-4545.

Alencar LM & Medeiros FA. (2011). The role of standard automated perimetry and newer functional methods for glaucoma diagnosis and follow-up. *Indian J Ophthalmol* **59 Suppl,** S53-58.

Almasieh M, Wilson AM, Morquette B, Cueva Vargas JL & Di Polo A. (2012). The molecular basis of retinal ganglion cell death in glaucoma. *Prog Retin Eye Res* **31,** 152-181.

Altangerel U, Bayer A, Henderer JD, Katz LJ, Steinmann WC & Spaeth GL. (2005). Knowledge of chronology of optic disc stereophotographs influences the determination of glaucomatous change. *Ophthalmology* **112,** 40-43.

Anderson DR. (2011). Normal-tension glaucoma (Low-tension glaucoma). *Indian J Ophthalmol* **59 Suppl,** S97-101.

Ang M, Ho CL, Tan D & Chan C. (2012). Severe vernal keratoconjunctivitis requiring trabeculectomy with mitomycin C for corticosteroid-induced glaucoma. *Clin Experiment Ophthalmol* **40,** e149-155.

Anton A, Pazos M, Martin B, Navero JM, Ayala ME, Castany M, Martinez P & Bardavio J. (2013). Glaucoma progression detection: agreement, sensitivity, and specificity of expert visual field evaluation, event analysis, and trend analysis. *Eur J Ophthalmol* **23,** 187-195.

Armaly MF. (1969). Cup-disc ratio in early open-angle glaucoma. *Doc Ophthalmol* **26,** 526-533.

Armour RH. (2000). Manufacture and use of home made ophthalmoscopes: a 150th anniversary tribute to Helmholtz. *BMJ* **321,** 1557-1559.

Asaoka R. (2014). Mapping glaucoma patients' 30-2 and 10-2 visual fields reveals clusters of test points damaged in the 10-2 grid that are not sampled in the sparse 30-2 grid. *PLoS One* **9,** e98525.

Banitt MR, Ventura LM, Feuer WJ, Savatovsky E, Luna G, Shif O, Bosse B & Porciatti V. (2013). Progressive loss of retinal ganglion cell function precedes structural loss by several years in glaucoma suspects. *Invest Ophthalmol Vis Sci* **54,** 2346-2352.

Barteselli G, Bartsch DU, Viola F, Mojana F, Pellegrini M, Hartmann KI, Benatti E, Leicht S, Ratiglia R, Staurenghi G, Weinreb RN & Freeman WR. (2013). Accuracy of the Heidelberg Spectralis in the alignment between near-infrared image and tomographic scan in a model eye: a multicenter study. *Am J Ophthalmol* **156,** 588-592.

Bengtsson B, Heijl A & Olsson J. (1998). Evaluation of a new threshold visual field strategy, SITA, in normal subjects. Swedish Interactive Thresholding Algorithm. *Acta Ophthalmol Scand* **76,** 165-169.

Berendschot TT, DeLint PJ & van Norren D. (2003). Fundus reflectance--historical and present ideas. *Prog Retin Eye Res* **22,** 171-200.

Berninger TA & Arden GB. (1988). The pattern electroretinogram. *Eye (Lond)* **2 Suppl,** S257-283.

Breusegem C, Fieuws S, Stalmans I & Zeyen T. (2011). Agreement and accuracy of non-expert ophthalmologists in assessing glaucomatous changes in serial stereo optic disc photographs. *Ophthalmology* **118,** 742-746.

Burk RO, Tuulonen A & Airaksinen PJ. (1998). Laser scanning tomography of localised nerve fibre layer defects. *Br J Ophthalmol* **82,** 1112-1117.

Campbell DG & Schertzer RM. (1995). Pathophysiology of pigment dispersion syndrome and pigmentary glaucoma. *Curr Opin Ophthalmol* **6,** 96-101.

Caprioli J, Coleman AL & Blood Flow in Glaucoma D. (2010). Blood pressure, perfusion pressure, and glaucoma. *Am J Ophthalmol* **149,** 704-712.

Carnahan MC & Goldstein DA. (2000). Ocular complications of topical, peri-ocular, and systemic corticosteroids. *Curr Opin Ophthalmol* **11,** 478-483.

Casas-Llera P, Rebolleda G, Munoz-Negrete FJ, Arnalich-Montiel F, Perez-Lopez M & Fernandez-Buenaga R. (2009). Visual field index rate and event-based glaucoma progression analysis: comparison in a glaucoma population. *Br J Ophthalmol* **93,** 1576-1579.

Casson RJ, Chidlow G, Wood JP, Crowston JG & Goldberg I. (2012). Definition of glaucoma: clinical and experimental concepts. *Clin Experiment Ophthalmol* **40,** 341-349.

Chakraborty R, Read SA & Collins MJ. (2011). Diurnal variations in axial length, choroidal thickness, intraocular pressure, and ocular biometrics. *Invest Ophthalmol Vis Sci* **52,** 5121-5129.

Chauhan BC, Garway-Heath DF, Goni FJ, Rossetti L, Bengtsson B, Viswanathan AC & Heijl A. (2008). Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol* **92,** 569-573.

Chen CL, Bojikian KD, Gupta D, Wen JC, Zhang Q, Xin C, Kono R, Mudumbai RC, Johnstone MA, Chen PP & Wang RK. (2016). Optic nerve head perfusion in normal eyes and eyes with glaucoma using optical coherence tomography-based microangiography. *Quant Imaging Med Surg* **6,** 125-133.

Chua BE, Xie J, Arnold AL, Koukouras I, Keeffe JE & Taylor HR. (2011). Glaucoma prevalence in Indigenous Australians. *Br J Ophthalmol* **95,** 926-930.

Curcio CA & Allen KA. (1990). Topography of ganglion cells in human retina. *J Comp Neurol* **300,** 5-25.

Da Pozzo S, Marchesan R & Ravalico G. (2009). Scanning laser polarimetry - a review. *Clin Experiment Ophthalmol* **37,** 68-80.

Dirani M, Crowston JG, Taylor PS, Moore PT, Rogers S, Pezzullo ML, Keeffe JE & Taylor HR. (2011). Economic impact of primary open-angle glaucoma in Australia. *Clin Experiment Ophthalmol* **39,** 623-632.

Drance SM & Begg IS. (1970). Sector haemorrhage--a probable acute ischaemic disc change in chronic simple glaucoma. *Can J Ophthalmol* **5,** 137-141.

Drasdo N. (1989). Receptive field densities of the ganglion cells of the human retina. *Vision Res* **29,** 985-988.

Drummond SR & Weir C. (2010). Chiasmal compression misdiagnosed as normal-tension glaucoma: can we avoid the pitfalls? *Int Ophthalmol* **30,** 215-219.

Ehrlich JR & Radcliffe NM. (2010). The role of clinical parapapillary atrophy evaluation in the diagnosis of open angle glaucoma. *Clin Ophthalmol* **4,** 971-976.

Fernandez Lopez E, Karaca EE, Ekici F, Waisbourd M & Spaeth GL. (2014). Symptoms reported by patients with varying stages of glaucoma: review of 401 cases. *Can J Ophthalmol* **49,** 420-425.

Ferreras A, Polo V, Larrosa JM, Pablo LE, Pajarin AB, Pueyo V & Honrubia FM. (2007). Can frequency-doubling technology and short-wavelength automated perimetries detect visual field defects before standard automated perimetry in patients with preperimetric glaucoma? *J Glaucoma* **16,** 372-383.

Friedman DS, Wolfs RC, O'Colmain BJ, Klein BE, Taylor HR, West S, Leske MC, Mitchell P, Congdon N & Kempen J. (2004). Prevalence of open-angle glaucoma among adults in the United States. *Arch Ophthalmol* **122,** 532-538.

Ganekal S. (2012). Ganglion cell complex scan in the early prediction of glaucoma. *Nepal J Ophthalmol* **4,** 236-241.

Garas A, Vargha P & Hollo G. (2012). Comparison of diagnostic accuracy of the RTVue Fourier-domain OCT and the GDx-VCC/ECC polarimeter to detect glaucoma. *Eur J Ophthalmol* **22,** 45-54.

Gardiner SK, Demirel S, Gordon MO & Kass MA. (2013). Seasonal changes in visual field sensitivity and intraocular pressure in the ocular hypertension treatment study. *Ophthalmology* **120,** 724-730.

Garway-Heath DF, Caprioli J, Fitzke FW & Hitchings RA. (2000). Scaling the hill of vision: the physiological relationship between light sensitivity and ganglion cell numbers. *Invest Ophthalmol Vis Sci* **41,** 1774-1782.

Garway-Heath DF, Lascaratos G, Bunce C, Crabb DP, Russell RA, Shah A & United Kingdom Glaucoma Treatment Study I. (2013). The United Kingdom Glaucoma Treatment Study: a multicenter, randomized, placebo-controlled clinical trial: design and methodology. *Ophthalmology* **120,** 68-76.

Gloor BP. (2009). Franz Fankhauser: the father of the automated perimeter. *Surv Ophthalmol* **54,** 417-425.

Gloor BR. (2010). Hans Goldmann (1899-1991). *Eur J Ophthalmol* **20,** 1-11.

Goldbaum MH, Jeng SY, Logemann R & Weinreb RN. (1989). The extracellular matrix of the human optic nerve. *Arch Ophthalmol* **107,** 1225-1231.

Gonul S, Koktekir BE, Bakbak B & Gedik S. (2013). Comparison of the ganglion cell complex and retinal nerve fibre layer measurements using Fourier domain optical coherence tomography to detect ganglion cell loss in non-arteritic anterior ischaemic optic neuropathy. **97,** 1045-1050.

Grewe R. (1986). [The history of glaucoma]. *Klin Monbl Augenheilkd* **188,** 167-169.

Guedes G, Tsai JC & Loewen NA. (2011). Glaucoma and aging. *Curr Aging Sci* **4,** 110-117.

Gupta PK, Asrani S, Freedman SF, El-Dairi M & Bhatti MT. (2011). Differentiating glaucomatous from non-glaucomatous optic nerve cupping by optical coherence tomography. *Open Neurol J* **5,** 1-7.

Hanson S, Krishnan SK & Phillips J. (2001). Observer experience and Cup:Disc ratio assessment. *Optom Vis Sci* **78,** 701-705.

Hardy le GH. (1931). Scotometry: History and Technique; with a Scotometric Tangent Screen and Scales. *Trans Am Ophthalmol Soc* **29,** 486-586.

Harwerth RS, Vilupuru AS, Rangaswamy NV & Smith EL, 3rd. (2007). The relationship between nerve fiber layer and perimetry measurements. *Invest Ophthalmol Vis Sci* **48,** 763-773.

Heijl A, Bengtsson B, Hyman L, Leske MC & Early Manifest Glaucoma Trial G. (2009). Natural history of open-angle glaucoma. *Ophthalmology* **116,** 2271-2276.

Heijl A, Bengtsson B & Oskarsdottir SE. (2013a). Prevalence and severity of undetected manifest glaucoma: results from the early manifest glaucoma trial screening. *Ophthalmology* **120,** 1541-1545.

Heijl A, Buchholz P, Norrgren G & Bengtsson B. (2013b). Rates of visual field progression in clinical glaucoma care. *Acta Ophthalmol* **91,** 406-412.

Heijl A, Lindgren A & Lindgren G. (1989). Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol* **108,** 130-135.

Henkind P. (1967). Radial peripapillary capillaries of the retina. I. Anatomy: human and comparative. *Br J Ophthalmol* **51,** 115-123.

Hofman MA & Swaab DF. (2006). Living by the clock: the circadian pacemaker in older people. *Ageing Res Rev* **5,** 33-51.

Hollo G, Katsanos A & Konstas AG. (2015). Management of exfoliative glaucoma: challenges and solutions. *Clin Ophthalmol* **9,** 907-919.

Hood DC & Kardon RH. (2007). A framework for comparing structural and functional measures of glaucomatous damage. *Prog Retin Eye Res* **26,** 688-710.

Hood DC & Raza AS. (2014). On improving the use of OCT imaging for detecting glaucomatous damage. *Br J Ophthalmol* **98 Suppl 2,** ii1-9.

Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, Hee MR, Flotte T, Gregory K, Puliafito CA & et al. (1991). Optical coherence tomography. *Science* **254,** 1178-1181.

Huang JY, Pekmezci M, Mesiwala N, Kao A & Lin S. (2011). Diagnostic power of optic disc morphology, peripapillary retinal nerve fiber layer thickness, and macular inner retinal layer thickness in glaucoma diagnosis with fourier-domain optical coherence tomography. *J Glaucoma* **20,** 87-94.

Hyung SM, Kim DM, Hong C & Youn DH. (1992). Optic disc of the myopic eye: relationship between refractive errors and morphometric characteristics. *Korean J Ophthalmol* **6,** 32-35.

Iwase A, Suzuki Y, Araie M, Yamamoto T, Abe H, Shirato S, Kuwayama Y, Mishima HK, Shimizu H, Tomita G, Inoue Y & Kitazawa Y. (2004). The prevalence of primary open-angle glaucoma in Japanese: the Tajimi Study. *Ophthalmology* **111,** 1641-1648.

Jampel HD, Friedman D, Quigley H, Vitale S, Miller R, Knezevich F & Ding Y. (2009). Agreement among glaucoma specialists in assessing progressive disc changes from photographs in open-angle glaucoma patients. *Am J Ophthalmol* **147,** 39-44 e31.

Jindal V. (2013). Glaucoma: an extension of various chronic neurodegenerative disorders. *Mol Neurobiol* **48,** 186-189.

Johnson CA, Wall M & Thompson HS. (2011). A history of perimetry and visual field testing. *Optom Vis Sci* **88,** E8-15.

Jonas JB, Budde WM & Panda-Jonas S. (1999). Ophthalmoscopic evaluation of the optic nerve head. *Surv Ophthalmol* **43,** 293-320.

Jonas JB, Gusek GC & Naumann GO. (1988). Optic disc, cup and neuroretinal rim size, configuration and correlations in normal eyes. *Invest Ophthalmol Vis Sci* **29,** 1151-1158.

Jonas JB, Nguyen XN, Gusek GC & Naumann GO. (1989). Parapapillary chorioretinal atrophy in normal and glaucoma eyes. I. Morphometric data. *Invest Ophthalmol Vis Sci* **30,** 908-918.

Junoy Montolio FG, Wesselink C, Gordijn M & Jansonius NM. (2012). Factors that influence standard automated perimetry test results in glaucoma: test reliability, technician experience, time of day, and season. *Invest Ophthalmol Vis Sci* **53,** 7010-7017.

Kanamori A, Nakamura M, Yamada Y & Negi A. (2013). Spectral-domain optical coherence tomography detects optic atrophy due to optic tract syndrome. *Graefes Arch Clin Exp Ophthalmol* **251,** 591-595.

Katz LJ, Spaeth GL, Cantor LB, Poryzees EM & Steinmann WC. (1989). Reversible optic disk cupping and visual field improvement in adults with glaucoma. *Am J Ophthalmol* **107,** 485-492.

Kida T, Liu JH & Weinreb RN. (2006). Effect of 24-hour corneal biomechanical changes on intraocular pressure measurement. *Invest Ophthalmol Vis Sci* **47,** 4422-4426.

Kim NR, Lee ES, Seong GJ, Kim JH, An HG & Kim CY. (2010). Structure-function relationship and diagnostic value of macular ganglion cell complex measurement using Fourier-domain OCT in glaucoma. *Invest Ophthalmol Vis Sci* **51,** 4646-4651.

Kirsch RE & Anderson DR. (1973). Clinical recognition of glaucomatous cupping. *Am J Ophthalmol* **75,** 442-454.

Kisiswa L, Dervan AG, Albon J, Morgan JE & Wride MA. (2010). Retinal ganglion cell death postponed: giving apoptosis a break? *Ophthalmic Res* **43,** 61-78.

Kniestedt C, Punjabi O, Lin S & Stamper RL. (2008). Tonometry through the ages. *Surv Ophthalmol* **53,** 568-591.

Koehler PJ. (2002). The historical roots of the visual examination. *Semin Neurol* **22,** 357-366.

Konstas AG, Stewart WC, Stroman GA & Sine CS. (1997). Clinical presentation and initial treatment patterns in patients with exfoliation glaucoma versus primary open-angle glaucoma. *Ophthalmic Surg Lasers* **28,** 111-117.

Kornzweig AL, Eliasoph I & Feldstein M. (1966). The retinal vasculature in macular degeneration. *Arch Ophthalmol* **75,** 326-333.

Kotera Y, Hangai M, Hirose F, Mori S & Yoshimura N. (2011). Three-dimensional imaging of macular inner structures in glaucoma by using spectral-domain optical coherence tomography. *Invest Ophthalmol Vis Sci* **52,** 1412-1421.

Kottler MS, Drance SM & Schulzer M. (1975). Simultaneous stereophotography. Its value in clinical assessment of the topography of the optic cup. *Can J Ophthalmol* **10,** 453-457.

Kowal P, Towers A & Byles J. (2014). Ageing across the Tasman Sea: the demographics and health of older adults in Australia and New Zealand. *Aust N Z J Public Health* **38,** 377-383.

Kubota T, Jonas JB & Naumann GO. (1993). Direct clinico-histological correlation of parapapillary chorioretinal atrophy. *Br J Ophthalmol* **77,** 103-106.

Landers J, Henderson T & Craig J. (2012). The prevalence of glaucoma in indigenous Australians within Central Australia: the Central Australian Ocular Health Study. *Br J Ophthalmol* **96,** 162-166.

Landers J, Sharma A, Goldberg I & Graham S. (2007). A comparison of global indices between the Medmont Automated Perimeter and the Humphrey Field Analyzer. *Br J Ophthalmol* **91,** 1285-1287.

Lee RY, Kao AA, Kasuga T, Vo BN, Cui QN, Chiu CS, Weinreb RN & Lin SC. (2013). Ethnic variation in optic disc size by fundus photography. *Curr Eye Res* **38,** 1142-1147.

Li SG, Spaeth GL, Scimeca HA, Schatz NJ & Savino PJ. (1979). Clinical experiences with the use of an automated perimeter (Octopus) in the diagnosis and management of patients with glaucoma and neurologic diseases. *Ophthalmology* **86,** 1302-1316.

Liu JH & Weinreb RN. (2014). Asymmetry of habitual 24-hour intraocular pressure rhythm in glaucoma patients. *Invest Ophthalmol Vis Sci* **55,** 7398-7402.

Maddess T. (2011a). Frequency-doubling technology and parasol cells. *Invest Ophthalmol Vis Sci* **52,** 3759; author reply 3759-3760.

Maddess T. (2011b). The influence of sampling errors on test-retest variability in perimetry. *Invest Ophthalmol Vis Sci* **52,** 1014-1022.

Maddess T. (2014). Modeling the relative influence of fixation and sampling errors on retest variability in perimetry. *Graefes Arch Clin Exp Ophthalmol* **252,** 1611-1619.

Mardin CY. (2012). [The most important signs of a glaucomatous disc]. *Klin Monbl Augenheilkd* **229,** 112-118.

Margalit E & Sadda SR. (2003). Retinal and optic nerve diseases. *Artif Organs* **27,** 963-974.

Medeiros FA, Lisboa R, Weinreb RN, Liebmann JM, Girkin C & Zangwill LM. (2013). Retinal ganglion cell count estimates associated with early development of visual field defects in glaucoma. *Ophthalmology* **120,** 736-744.

Medeiros FA, Zangwill LM, Bowd C, Mansouri K & Weinreb RN. (2012). The structure and function relationship in glaucoma: implications for detection of progression and measurement of rates of change. *Invest Ophthalmol Vis Sci* **53,** 6939-6946.

Miglior S & Bertuzzi F. (2013). Relationship between intraocular pressure and glaucoma onset and progression. *Curr Opin Pharmacol* **13,** 32-35.

Mitchell P, Smith W, Attebo K & Healey PR. (1996). Prevalence of open-angle glaucoma in Australia. The Blue Mountains Eye Study. *Ophthalmology* **103,** 1661-1669.

Moreno PA, Konno B, Lima VC, Castro DP, Castro LC, Leite MT, Pacheco MA, Lee JM & Prata TS. (2011). Spectral-domain optical coherence tomography for early glaucoma assessment: analysis of macular ganglion cell complex versus peripapillary retinal nerve fiber layer. *Can J Ophthalmol* **46,** 543-547.

Musch DC, Shimizu T, Niziol LM, Gillespie BW, Cashwell LF & Lichter PR. (2012). Clinical characteristics of newly diagnosed primary, pigmentary and pseudoexfoliative open-angle glaucoma in the Collaborative Initial Glaucoma Treatment Study. *Br J Ophthalmol* **96,** 1180-1184.

Mwanza JC, Durbin MK & Budenz DL. (2011). Interocular symmetry in peripapillary retinal nerve fiber layer thickness measured with the Cirrus HD-OCT in healthy eyes. *Am J Ophthalmol* **151,** 514-521 e511.

Niyadurupola N & Broadway DC. (2008). Pigment dispersion syndrome and pigmentary glaucoma--a major review. *Clin Experiment Ophthalmol* **36,** 868-882.

Odberg T & Riise D. (1985). Early diagnosis of glaucoma. The value of successive stereophotography of the optic disc. *Acta Ophthalmol (Copenh)* **63,** 257-263.

Oliver JE, Hattenhauer MG, Herman D, Hodge DO, Kennedy R, Fang-Yen M & Johnson DH. (2002). Blindness and glaucoma: a comparison of patients progressing to blindness from glaucoma with patients maintaining vision. *Am J Ophthalmol* **133,** 764-772.

Pemp B, Georgopoulos M, Vass C, Fuchsjager-Mayrl G, Luksch A, Rainer G & Schmetterer L. (2009). Diurnal fluctuation of ocular blood flow parameters in patients with primary open-angle glaucoma and healthy subjects. *Br J Ophthalmol* **93,** 486-491.

Peters D, Bengtsson B & Heijl A. (2013a). Factors associated with lifetime risk of open-angle glaucoma blindness. *Acta Ophthalmol.*

Peters D, Bengtsson B & Heijl A. (2013b). Lifetime risk of blindness in open-angle glaucoma. *Am J Ophthalmol* **156,** 724-730.

Pierro L, Gagliardi M, Iuliano L, Ambrosi A & Bandello F. (2012). Retinal nerve fiber layer thickness reproducibility using seven different OCT instruments. *Invest Ophthalmol Vis Sci* **53,** 5912-5920.

Podoleanu AG. (2012). Optical coherence tomography. *J Microsc* **247,** 209-219.

Porciatti V & Ventura LM. (2012). Retinal ganglion cell functional plasticity and optic neuropathy: a comprehensive model. *J Neuroophthalmol* **32,** 354-358.

Portney GL & Krohn MA. (1978). Automated perimetry: background, instruments and methods. *Surv Ophthalmol* **22,** 271-278.

Primrose J. (1971). Early signs of the glaucomatous disc. *Br J Ophthalmol* **55,** 820-825.

Puchalla JL, Schneidman E, Harris RA & Berry MJ. (2005). Redundancy in the population code of the retina. *Neuron* **46,** 493-504.

Quigley HA, Brown AE, Morrison JD & Drance SM. (1990). The size and shape of the optic disc in normal human eyes. *Arch Ophthalmol* **108,** 51-57.

Radzikhovskii BL, Titenko KS, Lovlia GD & Luchik VI. (1974). [Importance of diurnal perimetry, tonometry and diurnal fluctuations of corneal sensitivity in the early diagnosis of glaucoma]. *Oftalmol Zh* **29,** 489-494.

Rao HL, Kumbar T, Kumar AU, Babu JG, Senthil S & Garudadri CS. (2013). Agreement between event-based and trend-based glaucoma progression analyses. *Eye (Lond)* **27,** 803-808.

Read SA, Collins MJ & Iskander DR. (2008). Diurnal variation of axial length, intraocular pressure, and anterior eye biometrics. *Invest Ophthalmol Vis Sci* **49,** 2911-2918.

Redmond T, O'Leary N, Hutchison DM, Nicolela MT, Artes PH & Chauhan BC. (2013). Visual field progression with frequency-doubling matrix perimetry and standard automated perimetry in patients with glaucoma and in healthy controls. *JAMA Ophthalmol* **131,** 1565-1572.

Sakata LM, Deleon-Ortega J, Sakata V & Girkin CA. (2009). Optical coherence tomography of the retina and optic nerve - a review. *Clin Experiment Ophthalmol* **37,** 90-99.

Samarawickrama C, Pai A, Tariq Y, Healey PR, Wong TY & Mitchell P. (2012). Characteristics and appearance of the normal optic nerve head in 6-year-old children. *Br J Ophthalmol* **96,** 68-72.

Sample PA, Bosworth CF, Blumenthal EZ, Girkin C & Weinreb RN. (2000). Visual function-specific perimetry for indirect comparison of different ganglion cell populations in glaucoma. *Invest Ophthalmol Vis Sci* **41,** 1783-1790.

Saunders LJ, Russell RA, Kirwan JF, McNaught AI & Crabb DP. (2014). Examining visual field loss in patients in glaucoma clinics during their predicted remaining lifetime. *Invest Ophthalmol Vis Sci* **55,** 102-109.

Schmied U. (1980). Automatic (Octopus) and manual (Goldmann) perimetry in glaucoma. *Albrecht Von Graefes Arch Klin Exp Ophthalmol* **213,** 239-244.

Sherman J, Bass SJ & Slotnick S. (2004). Glaucoma without cupping. *Optometry* **75,** 677-708.

Simpson DA & Crompton JL. (2008). The visual fields: an interdisciplinary history I. The evolution of knowledge. *J Clin Neurosci* **15,** 101-110.

Sjostrand J, Olsson V, Popovic Z & Conradi N. (1999). Quantitative estimations of foveal and extra-foveal retinal circuitry in humans. *Vision Res* **39,** 2987-2998.

Stamper RL. (1984). The effect of glaucoma on central visual function. *Trans Am Ophthalmol Soc* **82,** 792-826.

Stamper RL. (2011). A history of intraocular pressure and its measurement. *Optom Vis Sci* **88,** E16-28.

Sugiyama K, Tomita G, Kawase K, Onda E, Shinohara H, Hayakawa T & Kitazawa Y. (1999). Disc hemorrhage and peripapillary atrophy in apparently healthy subjects. *Acta Ophthalmol Scand* **77,** 139-142.

Susanna R, Jr. (2009). Unpredictability of glaucoma progression. *Curr Med Res Opin* **25,** 2167-2177.

Tan SY & Shigaki D. (2006). Hermann von Helmholtz (1821-1894): inventor of the ophthalmoscope. *Singapore Med J* **47,** 359-360.

Tatham AJ, Weinreb RN, Zangwill LM, Liebmann JM, Girkin CA & Medeiros FA. (2013). The relationship between cup-to-disc ratio and estimated number of retinal ganglion cells. *Invest Ophthalmol Vis Sci* **54,** 3205-3214.

Tektas OY & Lutjen-Drecoll E. (2009). Structural changes of the trabecular meshwork in different kinds of glaucoma. *Exp Eye Res* **88,** 769-775.

Toda R, Mochizuki H, Sone T & Kiuchi Y. (2010). Changes in optic disc shape and size in two patients with suspected glaucoma during a two- and three-year follow-up period. *Jpn J Ophthalmol* **54,** 94-96.

Tonagel F, Voykov B & Schiefer U. (2012). [Conventional perimetry. Antiquated or indispensable for functional glaucoma diagnostics?]. *Ophthalmologe* **109,** 325-336.

Trobe JD, Glaser JS, Cassady J, Herschler J & Anderson DR. (1980a). Nonglaucomatous excavation of the optic disc. *Arch Ophthalmol* **98,** 1046-1050.

Trobe JD, Glaser JS & Cassady JC. (1980b). Optic atrophy. Differential diagnosis by fundus observation alone. *Arch Ophthalmol* **98,** 1040-1045.

Turpin A & McKendrick AM. (2011). What reduction in standard automated perimetry variability would improve the detection of visual field progression? *Invest Ophthalmol Vis Sci* **52,** 3237-3245.

Tuulonen A, Lehtola J & Airaksinen PJ. (1993). Nerve fiber layer defects with normal visual fields. Do normal optic disc and normal visual field indicate absence of glaucomatous abnormality? *Ophthalmology* **100,** 587-597; discussion 597-588.

van Velthoven ME, Faber DJ, Verbraak FD, van Leeuwen TG & de Smet MD. (2007). Recent developments in optical coherence tomography for imaging the retina. *Prog Retin Eye Res* **26,** 57-77.

Varma R, Lee PP, Goldberg I & Kotak S. (2011). An assessment of the health and economic burdens of glaucoma. *Am J Ophthalmol* **152,** 515-522.

Varma R, Steinmann WC & Scott IU. (1992). Expert agreement in evaluating the optic disc for glaucoma. *Ophthalmology* **99,** 215-221.

Ventura LM & Porciatti V. (2005). Restoration of retinal ganglion cell function in early glaucoma after intraocular pressure reduction: a pilot study. *Ophthalmology* **112,** 20-27.

Ventura LM, Sorokac N, De Los Santos R, Feuer WJ & Porciatti V. (2006). The relationship between retinal ganglion cell function and retinal nerve fiber thickness in early glaucoma. *Invest Ophthalmol Vis Sci* **47,** 3904-3911.

Wahl J. (2011). [Arguments for prophylactic therapy of ocular hypertension]. *Ophthalmologe* **108,** 1011-1015.

Wall M, Doyle CK, Zamba KD, Artes P & Johnson CA. (2013). The repeatability of mean defect with size III and size V standard automated perimetry. *Invest Ophthalmol Vis Sci* **54,** 1345-1351.

Wall M, Woodward KR, Doyle CK & Artes PH. (2009). Repeatability of automated perimetry: a comparison between standard automated perimetry with stimulus size III and V, matrix, and motion perimetry. *Invest Ophthalmol Vis Sci* **50,** 974-979.

Wang H, Zhang Y, Ding J & Wang N. (2013). Changes in the circadian rhythm in patients with primary glaucoma. *PLoS One* **8,** e62841.

White JG. (1987). Confocal microscopy comes of age. *Nature* **328,** 183-184.

Williams AL, Lackey J, Wizov SS, Chia TM, Gatla S, Moster ML, Sergott R, Spaeth GL & Lai S. (2013). Evidence for widespread structural brain changes in

glaucoma: a preliminary voxel-based MRI study. *Invest Ophthalmol Vis Sci* **54,** 5880-5887.

Woon WH, Fitzke FW, Bird AC & Marshall J. (1992). Confocal imaging of the fundus using a scanning laser ophthalmoscope. *Br J Ophthalmol* **76,** 470-474.

Wright C, Tawfik MA, Waisbourd M & Katz LJ. (2015). Primary angle-closure glaucoma: an update. *Acta Ophthalmol*.

Wyatt HJ, Dul MW & Swanson WH. (2007). Variability of visual field measurements is correlated with the gradient of visual sensitivity. *Vision Res* **47,** 925-936.

Yamashita T, Miki A, Iguchi Y, Kimura K, Maeda F & Kiryu J. (2012). Reduced retinal ganglion cell complex thickness in patients with posterior cerebral artery infarction detected using spectral-domain optical coherence tomography. *Jpn J Ophthalmol* **56,** 502-510.

Chapter 1

Retest variability in the Medmont M700 automated perimeter

Abstract

Purpose: To investigate the level of test-retest variability in the Medmont M700 automated perimeter. We compare the retest variability of the outer 20° test points of one test method to test points in the inner 10° of two test methods to determine whether test points from different tests and regions exhibit different retest variability. We also generate some clinically applicable Coefficient of Repeatability (CoR) values for M700 Overall Defect (OD) and Pattern Defect (PD) indices.

Methods: Twenty four glaucoma patients with varying degrees of field loss were enrolled, and 21 patients (40 eyes) had usable results. A Central (30°) and Macula (10°) test were performed on each eye on the same day. To determine retest variability, the tests were repeated one week later at the same time of day.

Results: Test points from 5 dB to 20 dB in the outer 20° of the 30° test showed lower retest variance than points of equal decibel value in the central 10° of the same test. For the 30° test, the OD CoR was 2.4 dB. PD retest CoR varied with glaucoma severity, ranging from 1.24 dB for PD ≤ 2.8 to 3.1 dB for PD > 5.7. The 10° test CoR for OD was 2.1 dB and PD retest CoR ranged from 1.58 for PD ≤ 2.8 to 2.4 for PD > 5.7.

Conclusions: In glaucoma patients, retest variance for some decibel values does not seem to increase with increasing eccentricity in the M700. OD values as graded by the M700 do not appear to correspond well with the amount of visual field loss, and are not directly comparable to Mean Deviation results reported by other perimeters. Pattern Defect values in the M700 appear to correlate well with the degree of field loss.

Introduction

Retest variability compromises the ability of standard automated perimetry (SAP) to detect true progression in glaucoma. Variability in successive tests may be due to patient factors such as small fixational eye movements during testing,[1] limited neural dynamic range,[2] or inherent problems in some SAP test strategies such as spatial under-sampling.[3, 4]

Different test strategies available in the same instrument (e.g. Swedish Interactive Threshold Algorithm (SITA) and Standard Full Threshold in the Humphrey Field Analyser (HFA)) may also report different visual field defect extents,[5] depths[5, 6] and differing levels of retest variability.[7, 8]

The HFA 30-2 and the Medmont M700 "*Central 30°-test*" (hereafter referred to as the 30°-test) differ substantially in the distribution and spacing of test points. The HFA has test points arranged in a square grid pattern with each point separated by 6°, with the central 10° of the 30-2 test only having 12 test points. The M700 30°-test points are arranged in radial patterns around fixation (Fig.1A).

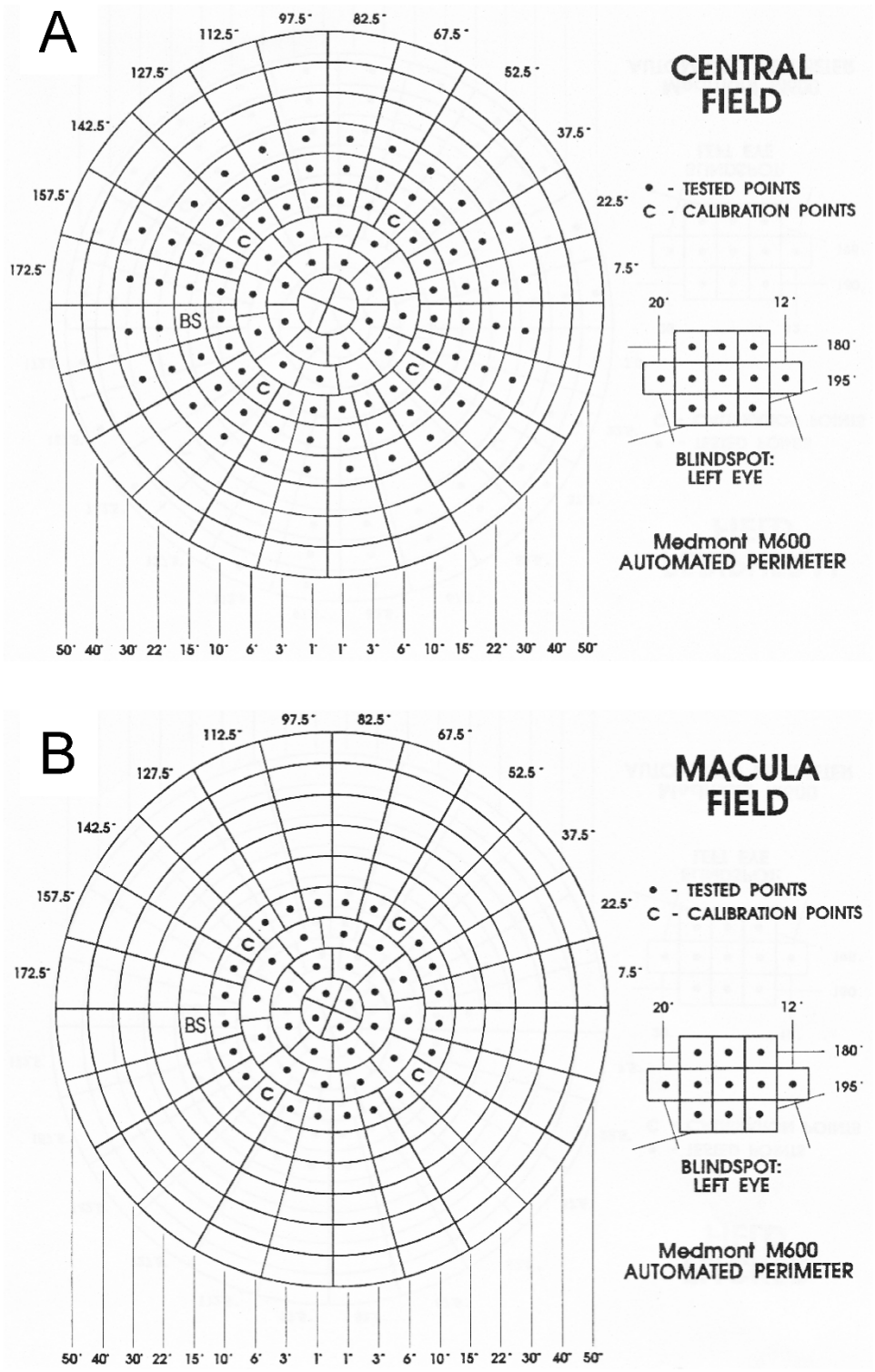


FIGURE 1.

Panel A illustrates the Medmont M600 and M700 Central (30°) field test pattern layout, left eye shown here. Stimulus locations are represented by a black dot, and only those areas containing a black dot, and the calibration points, (C), are tested.

Panel B illustrates the Medmont M600 and M700 Macula (10°) test pattern layout, left eye shown here. Note that the stimulus layout only differs from the central 10° of the 30°-test by the addition of the four central-most test points.

The M700 30°-test has a significantly increased central sampling rate compared to the HFA 24-2 and 30-2 protocols, which may result in better detection of central functional loss given that small but deep retinal nerve fibre layer defects may not be picked up as a scotoma using a 6° grid.[9, 10] The central 10° stimulus layout in the M700 30°-test is an easily accessible clinical tool for detecting central and peripheral field loss within a single wide field test.

The M700 "*Macula test*" (hereafter referred to as the 10°-test) stimulus layout only differs from the central 10° of the 30°-test by the addition of four test points located 1° from fixation (Fig.1B).

To facilitate detection of progression, we propose to establish event based progression criteria for the M700 visual field indices of Pattern Defect (PD) and Overall Defect (OD) for the 30°-test and the 10°-test protocols.

We intend to examine the retest variance of the central 10° of the 30°-test, and compare this to the retest variance of the 10°-test. Comparing the Test-Retest Variability (TRV) obtained from each strategy will determine whether any differences in repeatability exist between both strategies.

Retest variance of peripheral points in healthy individuals has been reported to be larger in mid-peripheral points than in points closer to fixation using the HFA.[11, 12] It may be that the higher density central sampling rate in the M700 may result in a different profile of variance to that of the peripheral field.[3, 4] We will therefore also compare the retest variance of the central 10° fields with the outer 20° of the 30°-test.

Methods

Twenty four glaucoma subjects were enrolled in this study. The research conformed to the tenets of the Declaration of Helsinki. Informed consent was obtained from the subjects after the nature and possible consequences of the test protocols were explained to the subjects. The appropriate consent forms were signed under ANU/ACT- Health protocol 7/07.667.

The patients had been diagnosed as having glaucoma by an ophthalmologist. Mean age was 73.6 years (range: 52.3 to 87.8), mean visual acuity was 6/8.1 (range: 6/5+ to 6/15-) and mean refractive error was +1.1/-0.9 D (range: sphere +3.50 to -1.50 D, cylinder: 0 to -1.75 D).

The data set consisted of 40 eyes of 21 patients (7 males) with 160 tests in total. Pearson's correlation coefficient (using PD) was 0.297, indicating a low level of correlation between eyes. That being said, no statistics where between - eye correlation would come into play were estimated.

No subjects were perimetrically naïve. Mean time between the subjects' most recent field test prior to the study, and the first field test undertaken in the study was 6.1 months (Standard Deviation (SD) 4.7 months). No training test was performed on any subjects.

A Medmont M600 bowl and Medmont M700 software (Version 3.9.7.) (Medmont P/L, Nunawading, Victoria, Australia) was used. The M600 bowl uses Goldmann size III (0.43°) rear projection light emitting diodes (peak wavelength 565 nm). The fixation target is a yellow light emitting diode with a peak wavelength of 583 nm. The maximum stimulus brightness is 318 cd/m² (0 dB)

and is attenuated in sixteen 3 ± 1 dB levels. The background illumination is 10 apostilb (3.2 cd/m²).

All tests used the *fast threshold* strategy, with automatic blind spot location. The fast threshold strategy employs the following test order of presentation points: The calibration points (“C” instead of a dot in Figures 1A and 1B) are fully thresholded to initialise the neighbourhood prediction function. This strategy estimates the projected threshold value for a point based on the completed points adjacent to it. An age and population based probability function is applied to each new test point based on the results from the surrounding exposed points. The true threshold level is then assessed using repeated exposures and statistical techniques. The fast threshold strategy does not employ any spatial smoothing, as individual points are not averaged with their neighbours. Since the calibration points for the 30°-test and 10°-test remain the same, there is no alteration to the fast threshold strategy for each test.

All tests were carried out in the same room, with the lights turned off, to ensure consistent background illumination, and all subjects were continuously monitored throughout their testing. The test order at each visit was as follows: 30°-test on the right eye, followed by a break of approximately four minutes, with the room lights off. A 30°-test was then carried out on the left eye, followed by a break of approximately fifteen minutes, with the room lights turned on. The 10°-test was then carried out on the right eye, followed by a break of approximately four minutes, with the room lights turned off. The 10°-test was then carried out on the left eye.

The lights in the room remained on while preparing and advising each patient before the commencement of any right eye test. The lights were then turned off

by the examiner, who sat down and commenced the test. Since the same examiner performed all the tests in exactly the same manner, there should be no differences in right eye sensitivities.

The two test sessions were carried out one week apart. The first test session was carried out at a randomly allocated time of day for each subject, but the second session was done at the same time of day to minimise circadian variations in test results. Prior to the first test, the height of the test lens holder and the canthus level indicator to the chinrest was measured for each subject. These figures were used to centre the test lens (48 mm diameter) and position the subject for all subsequent tests. These positions were not adjusted during testing. Pupils were not dilated and the standard M700 response time setting was used.

Some tests with reliability criteria outside M700 recommendations were included in the study. The reliability criteria suggested for the M700 are such that fixation losses higher than 20%, false positives higher than 20% and false negatives higher than 33% indicate low patient reliability. No subject had more than one reliability index exceed the recommended criteria in any test, although some subjects had a single unreliable index in more than one test used in the study. False negative results exceeding the recommended criteria all occurred in significantly damaged fields,[13] and only two false positive results exceeded the recommended criteria. Tests with fixation losses outside the recommended reliability criteria were included after due consideration was given to the examiner's assessment of patient reliability. There were 21 such tests, although that number reduced to 6 if reliable fixation loss criteria was increased to 33%.[14]

The M700 incorporates an asterisk based staging system on its visual field printouts (Table 1), indicating the degree of abnormality in OD or PD. The number of asterisks are not staged according to any percentile values. One asterisk (*), two asterisks (**) and three asterisks (***) indicate a mild, moderate or severe deviation from normal respectively.

TABLE 1.
M700 Overall Defect and Pattern Defect severity staging criteria (taken from the M700 manual). The severity of Overall Defect is staged according to age. The severity of the defect on the printout is indicated by the number of asterisks, and Pattern Defect has no age related severity staging.

	Age	*	**	***
Overall Defect	1-45	-2.6	-3.72	-4.92
	46 - 60	-2.8	-4.05	-6.10
	> 60	-3.2	-5.95	-8.91
Pattern Defect	Not applicable	≥ 2.8	≥ 5.7	≥ 8.6

The M700 users' manual defines OD as the mean difference between the age normal Hill of Vision (HoV) and the mean deviation (or patient based HoV). The patient HoV deviation shows the difference between the patients' test results and what the patients' HoV would look like without any localised test defects present. An algorithm is used to fit a HoV to the patients visual field, ignoring outliers and defect areas. Values for the patient HoV are only displayed if variation out to 15° is ± 4.5 dB, and beyond 15° ± 6.5 dB. A positive number indicates better than normal vision, and a negative number indicates a depressed field. The OD is calculated using a trimmed mean of the total

deviations, with the amount of trimming being influenced by the severity of disease present, the presence of diffuse losses and a high number of false positive results.[15]

The M700 uses an algorithm to calculate an estimated HFA Mean Deviation (MD) value. The HFA MD value is calculated using an older Medmont global index, the Average Defect (AD), from the M600. The AD index is not displayed on any M700 test report.

PD is a measure of the clustering and depth of defects. It is a scaled mean value of the product of a points' HoV deviation and that of its neighbours, qualifying the extent to which deviations are spatially correlated or clustered. Randomly distributed deviations from the patients HoV result in a small PD, whereas clusters of deviations cause the index to increase.

TRV may be reported using the Limits of Agreement (LoA), which are the mean difference between measurements ± 1.96 times the SD of the difference between measurements. The Coefficient of Repeatability (CoR) can be derived from a Bland-Altman plot by subtracting the mean difference from the upper 95% LoA. TRV in the current study will be reported as the CoR, a single figure which can be used to define the TRV for M700 test parameters.

Statistical analysis was carried out using Medcalc version 12.4 (Medcalc Software bvba, Ostend, Belgium) and Matlab (Matlab 6.1, The Mathworks Inc, Natick, MA, 2000). For comparison of retest variance, all measures from the 10°-test excluded the four central most points (Fig.1B) to ensure congruity with the central 10° of the 30°-test.

Results

Test retest variability

The 30°-test has 99 test points (excluding the blind spot). The overall average number of presentations was 206 presentations/99 test points (2.08 /point). Average test duration for the 30°-tests was 5 minutes and 44 seconds. Test duration was highly negatively correlated with MD ($r = -0.82$, $p < 0.0001$), indicating that increasing test time was associated with increasing levels of visual field damage.

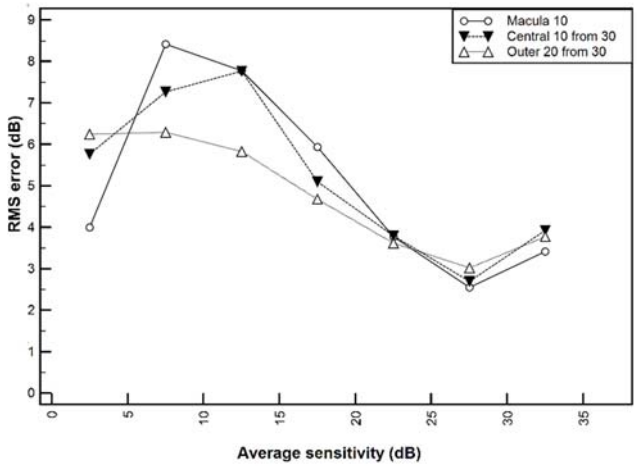


FIGURE 2.

Average test point RMS errors versus average sensitivity for the M700. The standard errors of the mean (SEM) were all less than 0.5 (median 0.287), with the exception of 5 results whose SEM were1.00, 0.800, 0.743, 0.720 and 0.663.
Explanation of the legend:
-- Macula 10: 10°-test (excluding the central four points)
-- Central 10 from 30: the central 10° results from the 30°-test
-- Outer 20 from 30: the outer 20° results from the 30°-test

TRV in the outer 20° of the 30°-test (Fig.2) was lower than that for corresponding sensitivities (from 5dB to 20 dB) in the central 10° of the 30°-test and the 10°-test. Reduced TRV at sensitivities close to 0 dB are due to limitations in the dynamic range of the instrument.[7]

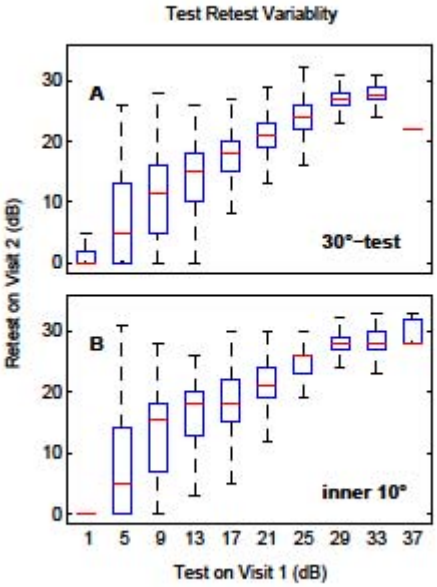


FIGURE 3.

Test retest variability results sensitivity results for the M700 30°-test and the inner 10° results from the 10° test and the central 10° of the 30° test. The blue orange line represents the median retest result, the blue box the inter-quartile range (25th and 75th percentiles) and the black bars the 5th and 95th percentiles.(dB = decibels)

The percentiles of the test-retest results for the 30°-test and the combined results of the central 10° of the 30°-test and the 10°-test are shown in Figure 3. TRV is lower for the 0 to 1 dB bins due to the dynamic range of the instrument. The anomalous results for 34 to 37 dB values are due to small sample sizes. TRV remains high for the 30°-test for initial test values below 17 dB, which is consistent with the findings from other studies.[2, 7]

To test for the effect of eccentricity on variance, TRV for points in the central 10° of the 30°-test were compared to TRV of points in the outer 20° of the same test. For each subject, the mean sensitivity (from the two tests) at each point was calculated and compared to the absolute value of the difference between visits at each point. The effect of age and gender was also assessed.

TABLE 2.
Multiple linear regression analysis of retest variability of points in the outer 20° of the 30°-test compared to that of the central 10° of the same test. Mean sensitivity = the mean sensitivity at each test point from the two test results, SE = the standard error of measurement, t = the t test result and p = probability ($p < 0.05$ indicates significant). ± 95 % CL = the 95% confidence limits for each coefficient. Age represents decades relative to the mean age of the subjects in the study. The results for each variable are in decibels, with the exceptions of age, which is in dB/decade from the mean age of the subjects in the study, and Mean Sensitivity, which is in dB (error)/ dB (sensitivity).

	Coefficient	SE	t	p	Multiplier	-95% CL	+95% CL
Constant	5.01	0.80	6.31	0	3.21	2.23	4.61
Eccentricity	0.03	0.02	1.64	0.10	1.01	1.00.	1.02
Age	0.77	0.18	4.26	0.01	1.19	1.10	1.30
Sex	-0.23	0.33	-0.68	0.50	0.95	0.82	1.10
Mean Sensitivity	0.08	0.03	-2.50	0.01	0.98	0.97	1.00

The proportion of variance (r^2) accounted for in the above model is 0.381. Whilst sex did not have any significant effect on TRV, age was found to have a significant effect (0.77 dB/decade from the mean age of the study subjects, $p < 0.01$). Eccentricity appeared to have little effect on retest error (0.03 dB ± 0.02, $p = 0.10$), whereas mean sensitivity had a more significant effect (0.08 dB (error)/dB (sensitivity), $p = 0.01$). These results indicate that in glaucoma subjects, using the M700, eccentricity does not have a significant effect on TRV.

Overall Defect and Pattern Defect Variability

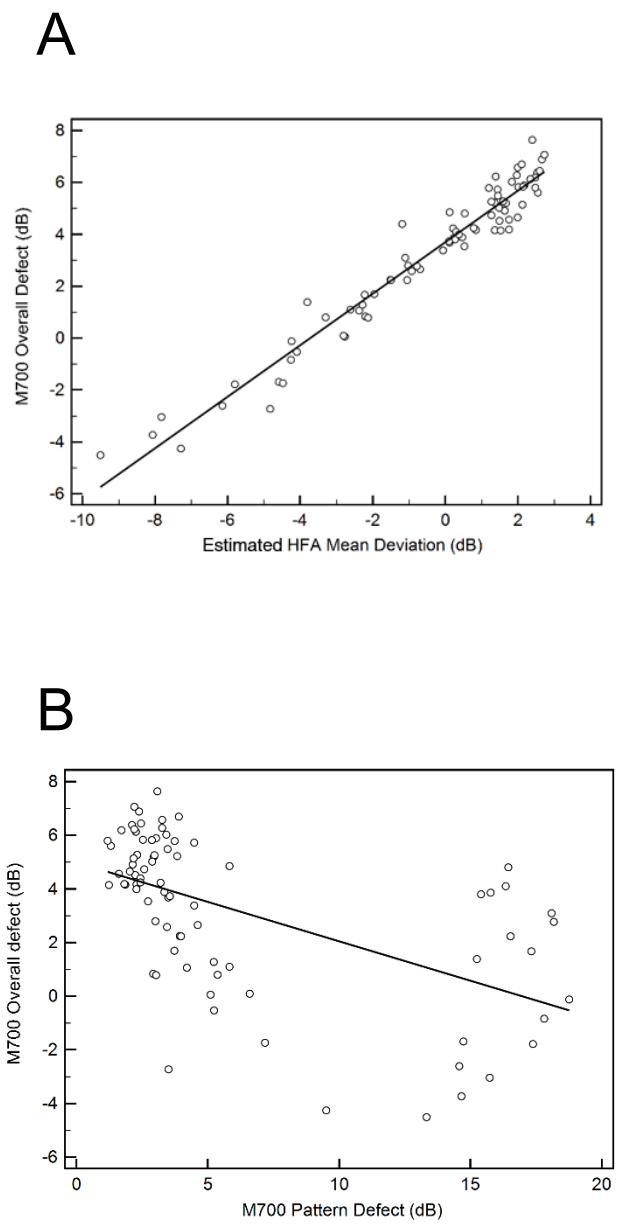


FIGURE 4.

Panel A shows the scatter plot and regression line of M700 Overall Defect versus estimated HFA Mean Deviation (as calculated by the M700).
Panel B shows the scatterplot and regression line of M700 PD against estimated HFA Mean Deviation (as calculated by the M700). The slope of the regression line is - 0.32. (dB = decibels).

OD and estimated HFA MD are highly correlated (r = 0.97) (Fig.4). Using Bland-

Altman analysis (not shown), the mean difference between the HFA MD and

M700 OD was found to be - 3.70 dB. The LoA obtained were - 3.70 ± 1.3 (- 2.40 to - 5.01) dB.

In Figure 4B, Pearson's correlation coefficient was r = - 0.61. Despite this moderately negative correlation, there is a noticeable spread of values above and below the regression line for all levels of PD.

TABLE 3.
Overall Defect Coefficient of Repeatability values for the 30°-test and 10°-test strategies. All values are in decibels.

M 700 Test	Average Overall Defect value	Range of results obtained	Coefficient of Repeatability
30° test	+ 2.88	-4.50 to 6.58	2.4
10° test	+ 2.94	-4.91 to 8.01	2.1

The average OD value of approximately + 2.9 dB (Table 3) is almost 6 dB higher than the M700 would indicate as being mildly abnormal (Table 1) for the average age of this glaucoma cohort. The reproducibility of the OD values was relatively good, with the CoR obtained being 2.1 to 2.4 dB (Table 3).

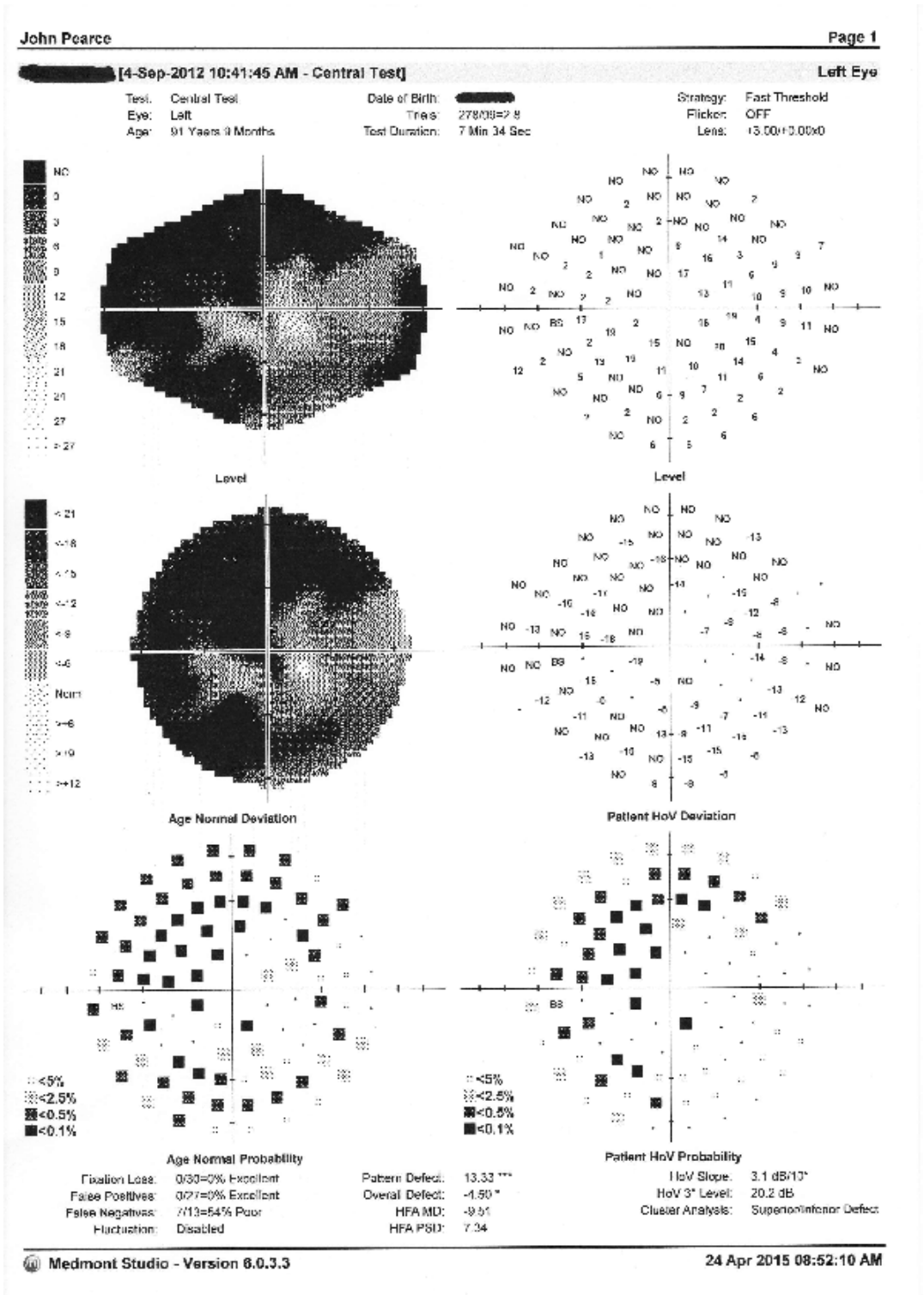


FIGURE 5.

M700 test printout (Version 6.0.3.3) showing a severely damaged field. OD - 4.50 indicates an M700 grading of a mildly abnormal field, whereas the HFA MD of – 9.51 indicates a moderately damaged field and is more reflective of the amount of visual field damage present.

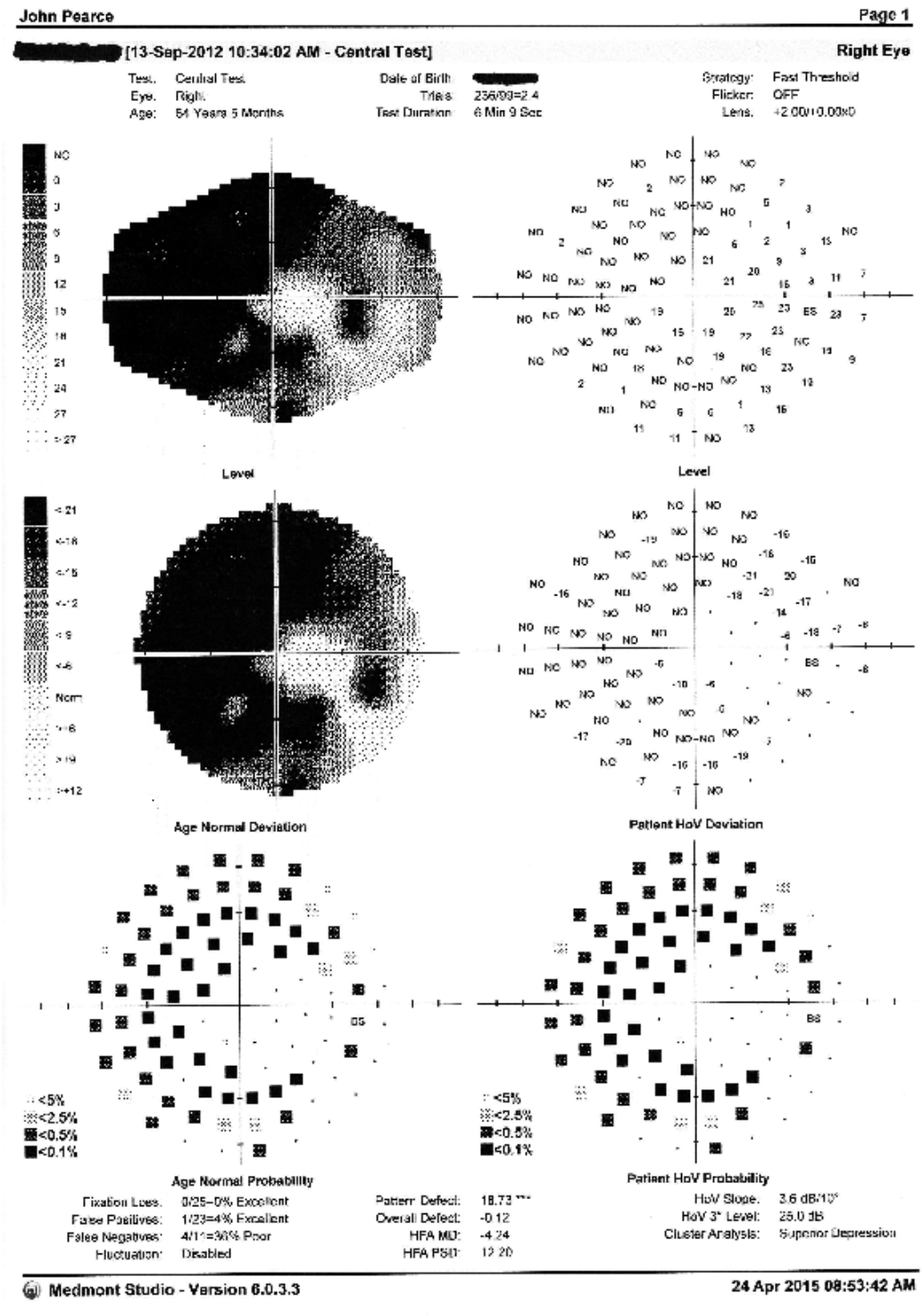


FIGURE 6.

M700 test printout (Version 6.0.3.3) showing a severely damaged field. OD – 0.12 indicates an M700 grading of normal, whereas the HFA MD of -4.24 indicates a mildly abnormal field which again is a more accurate reflection of the amount of field damage present.

Figures 5 and 6 demonstrate the disparity between OD and estimated HFA MD.

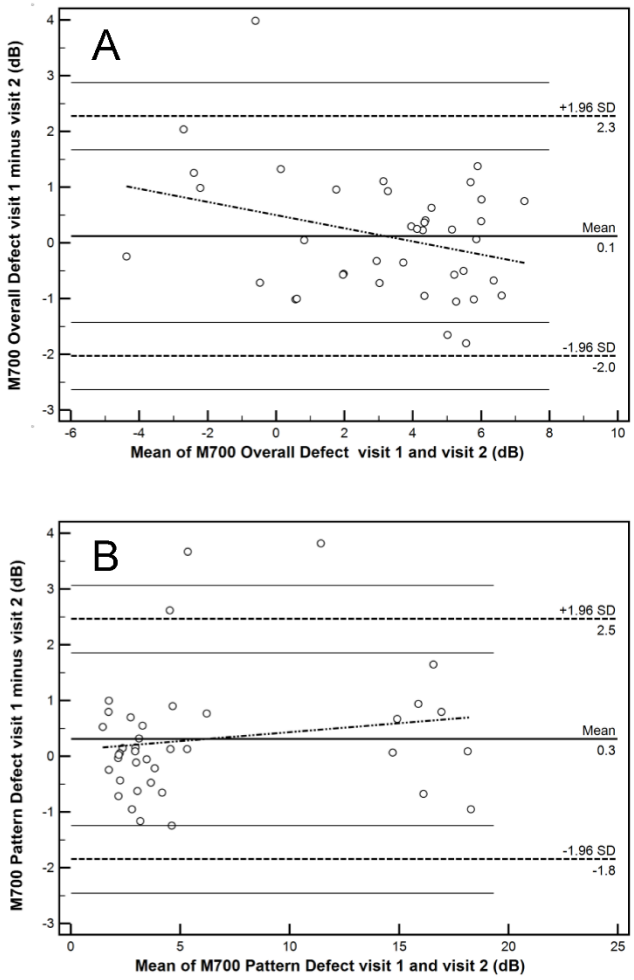


FIGURE 7.

Panel A shows the Bland-Altman plot for the M700 Overall Defect index. Panel B shows the Bland-Altman plot for the M700 Pattern Defect index (dB = decibels).

Figures 7A and 7B demonstrate the limits of agreement for M700 PD and OD. For PD, the LoA were 0.3 ± 2.2 (-1.9 to 2.5 dB). For OD, the LoA were 0.1 ± 2.2 (- 2.0 to 2.3 dB). A reviewer queried the normality of the distribution of the between visit differences, and so the validity of computing the LoA's as 1.96 × SD. A Kolmogorov-Smirnov test for normality found both the OD and PD differences to be normally distributed. A bootstrap method, based on 10,000 rounds of sampling with replacement, found good agreement with the 1.96 × SD method (e.g. the PD 5th and 95th percentiles were –1.04 ± 0.16 and 2.81 ± 0.91

(mean ± SE)). Given the SE, these do not appear very different to –1.9 and 2.5, and if anything, the 1.96 × SD method is conservative and may overestimate the error.

The mean and the CoR of the OD plot (Fig. 7A) is a reflection of the M700 method of calculation of OD. Points outside the age normal hill of vision are not used in the calculation of OD. As points of higher sensitivity demonstrate lower variability (Fig. 3), the variability of the OD descriptor may thus be reduced. The PD results (Fig. 7B) show the relationship of increasing variability with increasing PD (Table 4).

TABLE 4.

M700 Pattern Defect (PD) Coefficient of Repeatability (CoR) values. All results are in decibels. The CoR for the M700 severity classification of PD 5.71 to 8.6 (Table 1) was not calculated as there were only three subjects at this level. These results were therefore combined with the results for PD > 8.6 (9 subjects) to give a total of 12 results used in the calculation of the CoR for PD > 5.7

	30°-test CoR	10°-test CoR
Pattern Defect ≤ 2.8	1.24	1.58
Pattern Defect 2.81 - ≤ 5.7	1.13	1.9
Pattern Defect > 5.7	3.1	2.4
Range of values	1.18 – 18.75	0.29 – 18.59

Table 4 shows PD results for the cohort of subjects in the current study (cf. severity criterion, Table 1). The PD results are more reflective of the varying

levels of glaucomatous field loss than were the OD values. Severely abnormal individual PD results were obtained, as opposed to no moderately or severely abnormal OD results. The CoR was found to vary according to the level of PD.

Discussion

Retest variability

The form of the RMS error (Fig. 2) for the 30°-test is broadly similar to that reported for the HFA 30-2 test.[7] Since average dB results from the M700 have been reported to be approximately 5 dB lower than the HFA,[16] Figure 2 needs to be shifted to the right by 5 dB to compare with the HFA.[7] In Figure 2, the outer 20° of the 30°-test showed slightly lower retest variability than both results from the central 10° for decibel values from 5 dB to 20 dB, and quite similar values for all other decibel values. Points in the central 10° of the visual field reportedly have less variability than peripheral points in healthy subjects,[11] but this may not be the case for points of equal sensitivity for glaucoma patients using the M700.

Variability has been reported to increase rapidly as sensitivity decreases,[17] so the results of the current study may be biased if there is decreased sensitivity in the central 10°. Analysis of the 30°-tests shows that for the central 10° points, the average point sensitivity was 22.8 dB (1,757 points, excluding two 37 dB and one 38 dB outliers). The outer 20° of the same fields showed an average point sensitivity of 17.4 dB (2,118 points, which excluded one 37 dB outlier and the points directly above and below the blind spot). The lack of increase in peripheral TRV does not appear to be due to decreased sensitivity of the central points.

The two prevailing theories on sources of the high TRV in perimetry are those of Gardiner et al.[2] and Maddess.[3, 4] Gardiner et al. contend that noise in near-dead/dysfunctional retinal ganglion cells is the main cause. Under this proposal, TRV will be high even when spatial variability of the visual field is low. Maddess contends that spatial under-sampling of the jagged visual field profile, combined with known fixation eye jitter, is a strong contribution factor. Related ideas have been proposed by Wyatt et al.,[1] and the under-sampling effect can be large even when the damage level is lower.

Figures 2 and 3 of this paper are very similar to related plots in other papers.[7] It is apparent in those plots that significant TRV exists at even moderate damage levels. The finer sampling of the central field as used by the M700 might be thought to mitigate under-sampling, but the studies of Maddess[3, 4] indicate that even finer-grained sampling is needed. This issue will remain unresolved until very fine-grained spatial sampling from fixation to periphery is undertaken in a fairly large number of subjects, an onerous task that no-one has so far attempted.

TRV did not increase with eccentricity when comparing points of equal decibel values, and eccentricity was not found to have a significant effect on retest variance in the M700 *30°-test* (Table 2). Changes in peripheral point sensitivities may therefore have the same diagnostic importance as the same level of change in more centrally located points of equal sensitivity, and this finding needs to be incorporated into the interpretation of M700 field test results.

Overall Defect and Pattern defect variability

In this study, the OD disease staging (Table 1, Figs. 5 and 6) did not seem to correlate with the amount of field loss present. Whilst the OD index might be

assumed to be the equivalent of MD in the HFA, HFA MD equivalents give a mildly damaged field (Fig. 6) and a moderately damaged field (Fig. 5), whereas OD criteria has Figure 6 being normal and Figure 5 only mildly abnormal.

The most usable aspect of OD in the M700 may be the observation of change greater than the CoR of 2.4 dB in the 30°-test (Table 3). Although OD may decline monotonically due to its method of calculation (using only points on the patient's Hill of Vision considered to be within normal limits), this figure needs to be applied carefully in clinical practice

Pattern deviation measures in extremely damaged eyes may return a normal probability plot using the HFA.[18] Data from Landers et al.[19] suggests that PD continues to increase even in severely damaged fields with the M600 (an earlier version of the M700), as evidenced by the large PD's for the severely damaged fields in Figures 5 and 6. A study utilising the International Classification of Disease and Health related problems Glaucoma Staging Codes found that Pattern Standard Deviation (PSD) in the HFA was significantly higher in the severe stage group than in the moderate stage group.[20] The results of these studies[19, 20] seem to indicate that significantly damaged fields (up to the level encountered in the current study) may be able to use event based criteria as reported in the current study. M700 CoR values obtained reflect the importance of considering the amount of PD present when considering what constitutes statistically significant change, as found with MD in the HFA by Tattersall et al.[21]

Although trend analysis may be used to determine statistically significant change over time without the requirement for a reliable baseline field, a sufficient number of fields required for this analysis may be difficult to obtain in many clinical situations. Event based criteria may not always be the optimum

method of determining progressive field loss, but they are an accessible diagnostic tool for clinicians who may only have two fields per year with which to detect statistically significant progressive field loss.

Acknowledgements: This research was supported by the Australian Research Council through the ARC Centre of Excellence in Vision Science (CE0561903), and the Federal Government through the Nursing and Allied Health Scholarship Support Scheme. The views expressed in this paper do not necessarily represent those of the NAHSSS, its administrator, Services for Australian Rural and Remote Allied Health or the Australian Government Department of Health. Medmont figures and tables (©Medmont) have been used with the friendly permission of Medmont. None of the authors of this paper have any commercial associations that may result in any conflict of interest in connection with the submitted manuscript. Until recently, Maddess received royalties for sales of the FDT matrix perimeters until mid-2015, and has patents and equity in a possible new objective perimeter being developed by NuCoria Pty Ltd.

References

1. Wyatt HJ, Dul MW, Swanson WH. Variability of visual field measurements is correlated with the gradient of visual sensitivity. Vision Res 2007;47:925-936.
2. Gardiner SK, Swanson WH, Goren D, Mansberger SL, Demirel S. Assessment of the reliability of standard automated perimetry in regions of glaucomatous damage. Ophthalmology 2014;121:1359-1369.
3. Maddess T. The influence of sampling errors on test-retest variability in perimetry. Invest Ophthalmol Vis Sci 2011;52:1014-1022.
4. Maddess T. Modeling the relative influence of fixation and sampling errors on retest variability in perimetry. Graefes Arch Clin Exp Ophthalmol 2014;252:1611-1619.
5. Aoki Y, Takahashi G, Kitahara K. Comparison of Swedish interactive threshold algorithm and full threshold algorithm for glaucomatous visual field loss. Eur J Ophthalmol 2007;17:196-202.
6. Budenz DL, Rhee P, Feuer WJ, McSoley J, Johnson CA, Anderson DR. Comparison of glaucomatous visual field defects using standard full threshold and Swedish interactive threshold algorithms. Arch Ophthalmol 2002;120:1136-1141.
7. Artes PH, Iwase A. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. Invest Ophthalmol Vis Sci 2002;43:2654-2659.
8. Saunders LJ, Russell RA, Crabb DP. Measurement precision in a series of visual fields acquired by the standard and fast versions of the Swedish interactive thresholding algorithm: analysis of large-scale data from clinics. JAMA Ophthalmol 2015;133:74-80.
9. Asaoka R. Mapping glaucoma patients' 30-2 and 10-2 visual fields reveals clusters of test points damaged in the 10-2 grid that are not sampled in the sparse 30-2 grid. PLoS One 2014;9:e98525.
10. Burk RO, Tuulonen A, Airaksinen PJ. Laser scanning tomography of localised nerve fibre layer defects. Br J Ophthalmol 1998;82:1112-1117.
11. Heijl A, Lindgren G, Olsson J. Normal variability of static perimetric threshold values across the central visual field. Arch Ophthalmol 1987;105:1544-1549.
12. Asman P, Heijl A. Weighting according to location in computer-assisted glaucoma visual field analysis. Acta Ophthalmol (Copenh) 1992;70:671-678.
13. Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability? Invest Ophthalmol Vis Sci 2000;41:2201-2204.
14. Bickler-Bluth M, Trick GL, Kolker AE, Cooper DG. Assessing the utility of reliability indices for automated visual fields. Testing ocular hypertensives. Ophthalmology 1989;96:616-619.
15. Vingrys AJ, Zele AJ. Robust indices of clinical data: meaningless means. Invest Ophthalmol Vis Sci 2005;46:4353-4357.
16. Pye D, Herse P, Nguyen H, Vuong L, Pham Q. Conversion factor for comparison of data from Humphrey and Medmont automated perimeters. Clin Exp Optom 1999;82:11-14.
17. Russell RA, Crabb DP, Malik R, Garway-Heath DF. The relationship between variability and sensitivity in large-scale longitudinal visual field data. Invest Ophthalmol Vis Sci 2012;53:5985-5990.

18. Blumenthal EZ, Sapir-Pichhadze R. Misleading statistical calculations in far-advanced glaucomatous visual field loss. Ophthalmology 2003;110:196-200.
19. Landers J, Sharma A, Goldberg I, Graham S. A comparison of global indices between the Medmont Automated Perimeter and the Humphrey Field Analyzer. Br J Ophthalmol 2007;91:1285-1287.
20. Parekh AS, Tafreshi A, Dorairaj SK, Weinreb RN. Clinical applicability of the International Classification of Disease and Related Health Problems (ICD-9) glaucoma staging codes to predict disease severity in patients with open-angle glaucoma. J Glaucoma 2014;23:e18-22.
21. Tattersall CL, Vernon SA, Menon GJ. Mean deviation fluctuation in eyes with stable Humphrey 24-2 visual fields. Eye (Lond) 2007;21:362-366.

Chapter 2

Inter-visit test-retest variability of the Topcon 3D OCT-2000 in glaucoma

Abstract

Purpose: To determine the inter-visit test-retest variability (TRV) of a spectral domain optical coherence tomograph, the Topcon 3D OCT-2000, in the measurement of optic nerve head topography, peripapillary retinal nerve fibre layer (pRNFL) and macular ganglion cell complex (GCC) parameters in glaucoma patients. We also examine whether TRV with this instrument varies with the extent of glaucomatous damage.

Method: Twenty five subjects with varying degrees of glaucoma severity provided 41 eyes with usable results for the study. *3D Disc Retinal Nerve Fibre Layer Analysis* and *Macula V* (GCC) scans were repeated one week apart, at the same time of day, to determine the inter-visit TRV. TRV was determined using Bland-Altman Limits of Agreement (LoA) and the resulting Coefficients of Repeatability (CR).

Results: The overall horizontal and vertical cup/disc ratio CR's were 0.05 and 0.07 respectively. The GCC CR was 2.9 μm. In contrast, average pRNFL TRV expanded with increasing damage, with the LoA being well fitted by ± (34.67 - 0.294(*d*)), where *d* is the pRNFL thickness. A more complex model, with constant LoA of ± 5.61 μm at d > 82 μm, and linearly expanding TRV below that, achieved marginal significance ($p < 0.06$).

Conclusions: The repeatability of GCC measurements with this instrument was excellent. The determination of statistically significant change in average pRNFL should take into account average pRNFL thickness.

Keywords: Topcon 3D OCT-2000; inter-visit test-retest variability; glaucoma; macular ganglion cell complex; retinal nerve fibre layer

Spectral domain optical coherence tomography has become widely used to detect structural changes in glaucomatous eyes. Although spectral domain optical coherence tomography has been reported to be able to detect structural changes in eyes prior to functionally detectable changes,[1] an understanding of the test-retest variability inherent in any given instrument is necessary in order to differentiate true progressive change from test-retest variability in both glaucoma suspects and glaucoma patients.

The rapid introduction of new spectral domain optical coherence tomographs means that clinicians are often using instruments that have little test-retest variability data available. The Topcon 3D OCT-2000 has had relatively few studies conducted on it to provide test-retest variability information in healthy subjects, let alone glaucoma patients. We[2] have reported inter-visit test-retest variability values for the Topcon 3D OCT-2000 in healthy subjects. A study by Pierro et al.,[3] using the Topcon 3D OCT-2000, provided a figure for the standard deviation of the total data set rather than a value for test-retest variability, and in addition, only calculated this intra-visit value for healthy subjects. Other papers investigating the Topcon optical coherence tomograph have not generated test-retest variability data,[4] have used the older Topcon 3D OCT-1000[5] or have compared the diagnostic ability of different scan protocols in the Topcon[6] rather than generating test-retest variability values.

The Topcon 3D OCT-2000 provides retinal nerve fibre layer thickness trend analysis for the superior and inferior quadrants, as well as for the average

peripapillary retinal nerve fibre layer. At present, the Topcon does not incorporate a trend analysis function for optic nerve head parameters or macular ganglion cell complex measurements to assist in determining whether true progression has taken place.

Event based criteria for retinal nerve fibre layer, ganglion cell complex and optic nerve head parameters enable clinicians to differentiate progression from variability in a way that is easily managed in clinical practice, where patients' may attend irregularly, and can be used in conjunction with trend analysis. Trend based analyses are probably preferable, but often in the clinic, insufficient data is available, so event based criteria can be useful. The purpose of this study is to develop event based criteria for the Topcon 3D OCT-2000 peripapillary retinal nerve fibre layer, optic nerve head topography and total ganglion cell complex thickness parameters to assist clinicians using this instrument to differentiate test-retest variability from true progression in glaucoma patients.

Method

Twenty five glaucoma subjects were enrolled in the study. The research conformed to the tenets of the Declaration of Helsinki. Informed written consent was obtained after the nature and possible consequences of the test protocols were explained to the subjects. The appropriate information and consent forms were approved under ANU/ACT Health protocol 7/07.667.

The primary inclusion criteria for the study was that the subjects had been diagnosed with glaucoma by an ophthalmologist, and had either undergone, or

were undergoing, pressure lowering treatment by topical medications, laser trabeculoplasty procedures, surgical trabeculectomy, or a combination of treatments. All subjects were fair-skinned Caucasians of European origin, and native speakers of English. Patient demographic and clinical data is presented in Table 1.

Table 1.

Demographic and clinical data of glaucoma subject eyes used in the study. dB = decibels, M700 = Medmont M700 Automated perimeter, HFA = Humphrey Field Analyser

	Glaucoma subjects n = 24
Female (n)	15
Mean age (years)	74.9 ± 10.5
M700 Overall Defect (dB)	3.5 ± 3.0
M700 Pattern Defect (dB)	6.0 ± 5.6
Estimated HFA Mean Deviation (dB)	- 0.2 ± 2.6
Mean refractive error (Sphere) (Dioptres)	+ 1.1 ± 1.1
Mean refractive error (Cylinder) (Dioptres)	- 0.9 ± 0.5
Mean visual acuity	6/8.3 ± 2.5

Exclusion criteria were spherical refractive error outside ± 4 Dioptres, cylinder > 3D and best corrected visual acuity worse than 6/18. On that basis, three eyes were excluded from the study (due to a large disciform macular scar, best corrected visual acuity < 6/60 (one eye), branch retinal vein occlusion and macular oedema (one eye) and failed corneal graft with best corrected visual

acuity < 6/60 (one eye). One subject was excluded as they only attended one test session.

Of the 45 remaining eyes, four were excluded from 3D disc scan analysis due to Weiss rings in the vitreous creating artefacts in the image analysis. Five eyes were excluded from the ganglion cell complex scan analysis due to unacceptable image qualities (one eye), reanalysed scans (when reanalysing from Version 7.11 to Version 8.00) being analysed incorrectly (one eye) and significant structural macular abnormalities such as epiretinal membrane (one eye) and large sub-foveal drusen (two eyes).

As a result, 22 subjects (41 eyes) had a usable *3D Disc Retinal Nerve Fibre Layer analysis scan* (hereafter referred to as the 3D disc scan), and 22 subjects (40 eyes) had a usable *Macula V scan* (hereafter referred to as the ganglion cell complex scan).

Visual field tests were carried out using the Medmont M700 automated perimeter (Version 3.9.7) (Medmont P/L, Nunawading, Victoria, Australia) at the same visits as the Topcon scans. For a detailed description of the Medmont M700, please refer to Pearce and Maddess.[7]

The Topcon utilises a super luminescence diode (peak wavelength 840 nm) as a light source, and has an A-scan speed of 27,000 per second. The Topcon 3D disc scan is carried out over a 6.0 × 6.0 mm square scan area, and comprises 128 horizontal B-scans. Its A-scan rate of 512 A-scans per B-scan gives a total of 65,636 A-scans. The depth of these scans is 2.4 mm. The instrument algorithm then calculates the location of the optic disc centre, and the retinal

nerve fibre layer information is generated at the edge of a circle of 3.4 mm diameter. The disc topography information is obtained from the information in the original cube scan.

The ganglion cell complex scan measures the thickness of the retinal nerve fibre layer, ganglion cell layer and inner plexiform layer, which together constitute the ganglion cell complex. It consists of a scan area of 7.0 × 7.0 mm, in which 128 B-scans are taken. With 512 A-scans per B-scan, this gives a total of 65,636 A-scans. The instrument algorithm selects the best 6.0 × 6.0 mm position having the fovea in the centre, and the total number of B-scans is therefore 110, resulting in 56,320 A-scans. Information is extracted from the selected area, and the individual layer thicknesses are calculated for each point over this area.

The first cohort (12 subjects with usable scans), were scanned using software Version 7.11. The second cohort (10 subjects with usable scans), were scanned using Version 8.00, giving the total number of subjects used in the study of 22. Retest scans on each subject were all done with the same software version as the first scan. The Version 7.11 scans were then re-analysed using Version 8.00 software to ensure all statistical analyses were carried out using the same software version.

All scans were carried out with undilated pupils. We did not dilate the patients because that may not be routinely done in clinical practice, and our objective was to replicate real-world conditions. That being said, unlike older time-domain methods,[8, 9] it appears that dilation is less of an issue with spectral domain optical coherence tomography devices.[10]

All scans were carried out in the same room, under the same lighting conditions for each scan, and all scans were carried out by the same operator (JGP). Each subject attended two test sessions, the second session being conducted one week after the first. The scans on each subject were carried out at the same time of day at each session. This was done to minimise possible circadian variations in axial length,[11] ocular aberrations,[12] optic nerve head topography[13] or retinal nerve fibre layer thickness. No scans were manually altered, and all calculations of disc parameters, peripapillary retinal nerve fibre layer thickness and ganglion cell complex thickness were carried out by the instrument itself. Scans were not evaluated for segmentation errors, as this analysis would not be routinely carried out in clinical practice.

Any scans that registered poor image quality due to eye movement, or were found to be unable to be properly analysed, were deleted, and another scan taken. Image quality for the scans used in this study therefore ranged from 65 to 100 (mean 84, standard deviation 7.7) for the peripapillary retinal nerve fibre layer scans and 63 to 98 (mean 81, standard deviation 8.0) for the ganglion cell complex scans. These figures reflect the fact that a few scans were included which had sub-optimal image qualities, because in clinical practice it is not always possible to obtain high quality images.

Retest variability may be reported using the limits of agreement, which are defined as the mean difference between measurements ± 1.96 times the standard deviation of the difference between measurements. The coefficient of repeatability is another method which can be used to assess repeated measurements on the same instrument. The coefficient of repeatability can be

derived from a Bland-Altman plot by subtracting the mean difference from the upper limit of agreement.

Retest variability in the current study (for all parameters, with the exception of average peripapillary retinal nerve fibre layer), will be represented as the coefficient of repeatability, which should be interpreted with the foregoing definition in mind. The coefficient of repeatability will therefore consist of a single figure, which can be used to define the inter-visit test-retest variability of the Topcon for any given test parameter under consideration in the current study.

On reviewing the data for average peripapillary retinal nerve fibre layer, we also considered whether test-retest variability varied with the degree of glaucomatous damage. Accordingly, we applied two methods to quantify test-retest variability for average peripapillary retinal nerve fibre layer thickness, both based on the methods of Bland and Altman.[14] The first of these were standard Bland-Altman plots (Fig. 1 A, B), which are appropriate when it is clear that the differences between tests for each eye do not scale with the means of the two tests on each eye. The second is a regression based approach for the case where the test-retest variability scales with the mean differences in some way. We are grateful to a reviewer for suggesting this method.

Basically, this method uses regression of the data, *d*, to provide a fitted function, *f(d)*, and then regression on the absolute value of the residuals, *r*, to provide a fit to the error, *f(r)*, with the final 95% Limits of Agreement being $f(d) \pm 1.96kr(d)$, k being a correction for the width of the distribution of the absolute value of the residuals: $k=\sqrt{\pi}/2$.[14] In some cases, multiple regression models of the data and

residuals were performed that had 3 fitted parameters rather than the 2 fitted parameters of the linear regression, with F-change statistics being used to test if more 3-parameter models were statistically justified ($p < 0.05$) compared to the simpler models..

The Bland-Altman plots and other statistical analyses were carried out using Medcalc version 12.4 (Medcalc Software bvba, Ostend, Belgium) and Matlab (Matlab 6.1, The Mathworks Inc, Natick, MA 2000).

Results

Data from both eyes of eligible subjects were used in test-retest variability analysis. The between-eye correlation for average peripapillary retinal nerve fibre layer thickness was moderate ($r = 0.44$), while between-eye M700 Pattern Defect correlation was lower ($r = 0.30$).

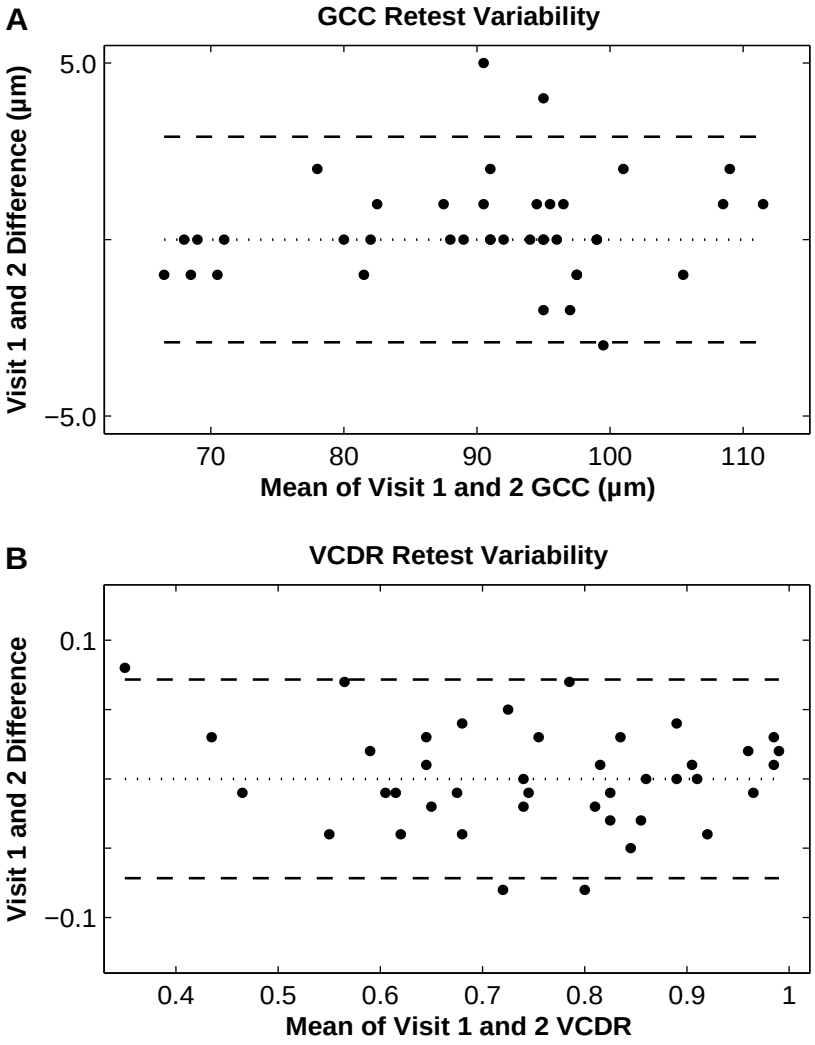


Figure 1 A, B.
Panel A illustrates the Bland-Altman difference plot for the macular ganglion cell complex (GCC) scan, showing test-retest variability (TRV) to be constant at all levels of GCC thicknesses encountered in this study. The Limits of Agreement (LoA) are therefore constant, and the coefficient of repeatability (CR) derived from them is ± 2.9 µm
Panel B illustrates the Bland-Altman difference plot for vertical cup/disc ratio (VCDR). TRV remains constant at all VCDR values, the LoA therefore remain constant and the derived CR is 0.07

The test-retest variability characteristics of the ganglion cell complex and vertical cup/disc ratio are illustrated in Fig. 1 A, B. Horizontal cup/disc ratio is not shown for brevity. The Bland-Altman plots (Fig. 1 A, B) show that the level of test-retest variability remains relatively constant for both parameters, irrespective of the mean measurement values. For this reason, standard Bland-Altman analysis was used to derive the coefficient of repeatability values quantified in Table 2.

Table 2.

Inter-visit test-retest variability of optic nerve head, peripapillary retinal nerve fibre layer and macular ganglion cell complex measurements. CR = Coefficient of Repeatability, SD = Standard Deviation, CV = Coefficient of Variation (%), HCDR = Horizontal cup/disc ratio, VCDR = Vertical cup disc ratio, µm = micrometres, mm^2 = square millimetres, GCC = macular ganglion cell complex, average pRNFL = average peripapillary reinal nerve fibre layer

	CR	Mean	SD	CV
Disc area (mm^2)	0.27	2.32	0.43	18.5
Rim area (mm^2)	0.24	1.10	0.38	34.5
HCDR	0.05	0.71	0.13	18.3
VCDR	0.07	0.75	0.16	21.3
GCC thickness (µm)	2.9	90.2	11.6	12.9
Average pRNFL (µm)	*	85.8	17.8	20.7

* no value for average pRNFL CR given, since test-retest variability varied depending on thickness, as discussed below (Fig. 2 and statistical analysis).

We also examined coefficients of variation (Table 2). Although the coefficient of variation is not a directly applicable measure of test-retest variability, it does enable a comparison of the spread in the data sets for different parameters on the same subjects. The higher the coefficient of variation, the greater the range

of values in the data set. For repeat measures, this indicates lower repeatability. The coefficient of variation ranged from 13% (ganglion cell complex thickness) to 35% (rim area). Although the mean thickness for average peripapillary retinal nerve fibre layer and ganglion cell complex did not differ greatly, the coefficient of variation for average peripapillary retinal nerve fibre layer was noticeably higher.

A reviewer pointed out that using raw data for the calculation of coefficient of variation ignores the floor effect present in some measurements, such as average peripapillary retinal nerve fibre layer. Strictly speaking, this approach is a more accurate appraisal of the coefficient of variation, but the non-neural components of average peripapillary retinal nerve fibre layer may vary with age[15] and subject. Furthermore, given that the non-neural components of average peripapillary retinal nerve fibre layer and optic nerve head parameters may also be contributing to test-retest variability, their exclusion may also introduce further errors into coefficient of variation calculations.

Our coefficient of variation values have therefore been calculated from the raw data, and this seems to be in line with similar calculations carried out in other studies.

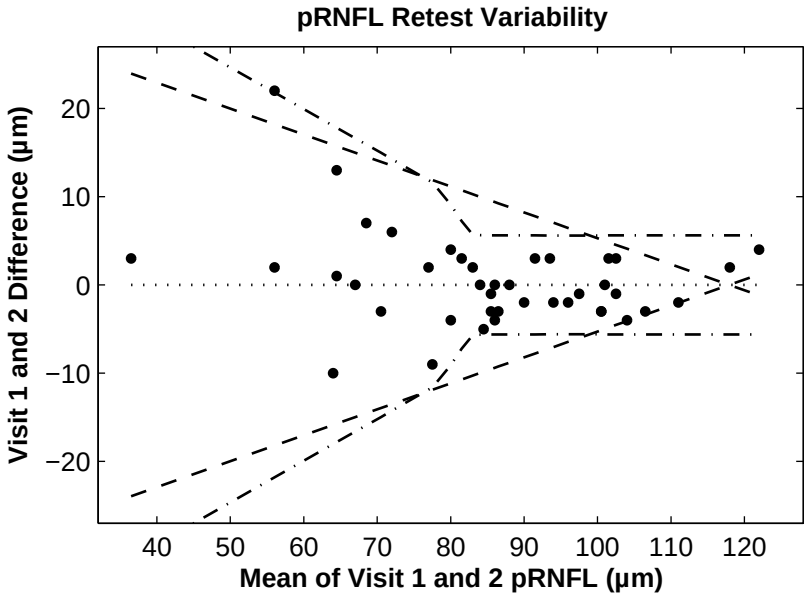


Figure 2.
Bland-Altman plot of average peripapillary retinal nerve fibre layer (pRNFL) results. The straight dashed lines represent the regression based limits of agreement of the absolute values of the differences, the lower dashed line being a reflection of the upper dashed line about zero. The dash-dot lines represent Monte-Carlo fits to the data and their residuals using a 3-parameter model, which were marginally significantly better than the straight line ($p < 0.06$). That set of fits produced a transition zone that on average occurred at 81.5 µm, with 95% confidence limits of 75.8 µm and 87.2 µm

Inspection of the peripapillary retinal nerve fibre layer data (Fig. 2) suggested that the test-retest variability might become larger with decreasing thickness. At the suggestion of a reviewer, we used a regression based method, originally described by Bland and Altman,[14] to quantify this. Rather than fit separate lines to the positive and negative test-retest variability data, we modified the method to fit the absolute values of the differences and then reflected the resulting limits of agreement line about zero. The results are shown as the pair of straight dashed lines in Fig. 2. F-change statistics indicated that the sloping line model

was significantly better ($p = 0.012$) than the constant limits of agreement as found in a standard Bland-Altman plot (e.g. Fig. 1 A, B). The formula for the linear regression based limits of agreement is $\pm (34.67 - 0.294(d))$, where d is the average peripapillary retinal nerve fibre layer thickness.

Although the model parsimoniously captures the trend of increasing test-retest variability with increasing damage, it suffers from being inaccurate at large (normal) peripapillary retinal nerve fibre layer thicknesses, where the lines intersect at around 117 μm. We therefore decided to examine if models with 3 parameters could produce a statistically significantly better fit than the 2-parameter (constant + slope) line models, again employing F-change statistics to see if the penalty of adding the third fitted parameter was justified.

Two types of 3-parameter models were considered. Both types included an average peripapillary retinal nerve fibre layer thickness breakpoint, where one model operated below the break point and a different model above it. For each break point, these were fitted as a single multiple regression model. The models were examined using a standard iterative Monte Carlo cross-validation, with 1000 repeats. On each repeat, the fits were to the original data plus normally distributed noise with standard deviation 1.0. Re-running the models with different random number seeds indicated that with 1000 repeats, the outcomes were stable to 2 decimal points. On each repeat, 27 break points were tested, beginning with a break at 72 μm through to 104 μm, providing a minimum of 10 points below the first break point for the regression to operate on. At each breakpoint, F-change statistics were computed between the multiple regression results and the best fitting function out of 27 was recorded for each repeat.

The first pair of multiple regression models compared a constant limits of agreement (one parameter) with a constant + slope below the breakpoints and a second constant above (three parameters), which we called the constant limits of agreement vs. constant + slope + constant model comparison. The second set of models compared the single constant + slope (2 parameters, i.e. the dashed lines of Fig. 2), and the previous case of the slope + constant below the break points, and a second constant above the break points, which we refer to as the constant + slope vs. constant + slope + constant comparisons.

For constant limits of agreement vs. constant + slope + constant, the more complex model was always selected at a mean ± standard deviation significance of $p = 0.010 \pm 0.008$. The mean breakpoint was 81.5 μm, with 95% confidence limits of 75.8 μm and 87.2 μm. For the second set of models, the constant + slope + constant models only had marginal significance compared to the constant + slope models ($p = 0.059 \pm 0.028$), and the mean break points were the same. The dash-dot lines of Fig 2 are based on the means of all the fits to the data and their residuals (Methods). Because the best break point varied slightly, there is a transition zone derived from the 1000 fits, each of which had a single sharp kink. Below this transition zone, the formula for the limits of agreements are $\pm (48.2 - 0.471(d))$, and above it the constant limits of agreement are ± 5.61 μm.

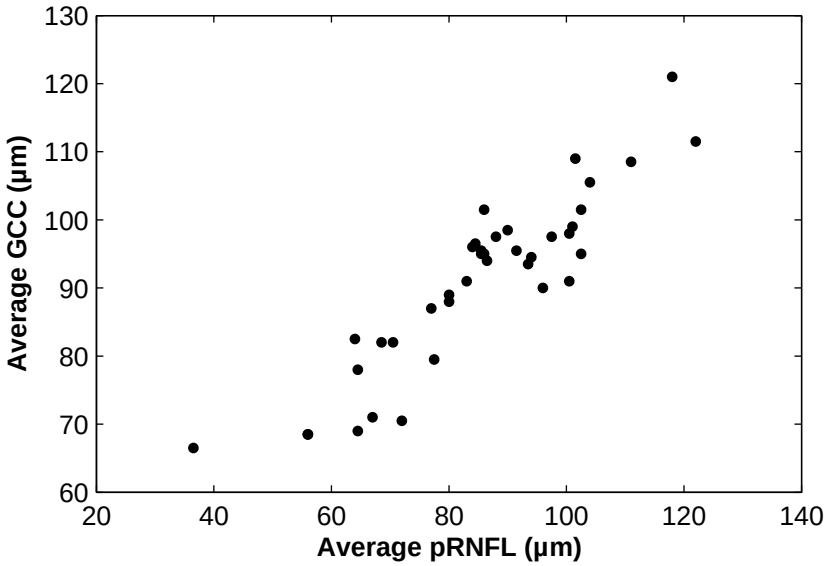


Figure 3.
Scatter plot for average peripapillary retinal nerve fibre layer (average pRNFL) versus macular ganglion cell complex (GCC) thickness. These two parameters were highly correlated (Pearson's correlation coefficient: r = 0.91)

The calculation of a correlation co-efficient in Fig. 3 is permissible, as correlation looks at the level of association rather than the level of agreement.[16] A regression line was not included in Fig. 3, since linear regression is not valid for comparison studies. If the axes are switched, the correlation does not change, but the slope and intercept of the regression line will.[16] Although this effect is lower when the correlation is high, as in Fig. 3, a regression line would still violate the principles of Bland-Altman analysis, and is therefore not included in Fig. 3.

The determination of measurement error is important, since it can affect statistical analysis and interpretation.[17] An alternative method determining the retest variability of the data would therefore have been to calculate Intra-class

Correlation Coefficients (ICC), where a high ICC value means that device noise (variability) is low in comparison to inter-subject variability.[18]

Here, we could have opted for the ICC2 method, which is used when a random sample of k raters (devices) rate the performance of multiple subjects, and can be reported as the ICC2k, where k reflects the means of the devices.

While these coefficients are useful for relative comparisons, they are not clinically applicable as they do not provide a figure for the disagreement between measurements.[19] We therefore did not follow this line of investigation, since the Bland-Altman analysis already provided a clinically applicable quantitative method that also corrects for regression to the mean.

The average peripapillary retinal nerve fibre layer and ganglion cell complex thickness were among the most highly correlated parameters in the Topcon (Fig. 3). This was despite the varying amounts of test-retest variability in relation to average peripapillary retinal nerve fibre layer thickness, which did not occur for ganglion cell complex measurements (Fig. 1A). Pearson's correlation coefficient, r, was 0.91 for these two parameters.

Conclusions

Disc topography

Horizontal cup/disc ratio and vertical cup/disc ratio test-retest variability values demonstrated a high degree of repeatability (Table 2). Horizontal cup/disc ratio showed slightly less variability than the vertical cup/disc ratio, and this was perhaps due to the decreased variability in horizontal disc diameter measurements compared to vertical disc diameter test-retest variability.

Subjective evaluation of progressive changes to the optic nerve head in glaucoma can be difficult. In one study, using glaucoma specialists to determine progressive disc change in glaucoma, in 40% of cases judged to have progressed, the “worse/progressed” disc photo was actually taken first, and the ability of glaucoma specialists (inter-observer agreement) to judge progressive disc change from stereo-photographs to be only slight to fair.[20] When viewing stereoscopic disc photographs, glaucoma experts were found to differ in cup/disc ratio estimates by up to 0.16.[21] The low test-retest variability in the Topcon cup/disc ratio measurements (Table 2) appears to provide an objective measurement which may provide an additional resource to circumvent the difficulties associated with subjective assessment of progressive changes to the optic nerve head in glaucoma patients.

Although the coefficient of repeatability for rim area appears quite small, at 0.25 mm² it is relatively large in relation to the mean rim area of the subjects in the study, 1.10 mm² (Table 2). Rim area measurements also recorded the highest coefficient of variability of any parameter (35%, Table 2). While the rim area coefficient of repeatability may not be as sensitive for progression detection as other disc topography parameters in the Topcon, it remains an important adjunctive measure for the assessment of progressive change to the optic nerve head.

Peripapillary retinal nerve fibre layer

The slope + constant model of Fig. 1B (dashed lines) was a significantly better model than the less complex constant limits of agreement model (e.g. as in Fig. 1A, B) at $p = 0.012$. Given its problems in describing the data at larger

peripapillary retinal nerve fibre layer thicknesses, we attempted two other types of models, each of which permitted a change of behaviour at some characteristic thickness. Compared to a constant limits of agreement, these models were always significantly better, even when their extra complexity was compensated for by use of F-change statistics. Compared to the single line model however, the most complex model (constant + slope + constant) was only marginally significantly better ($p = 0.059 \pm 0.028$) than the linear regression (constant + slope). On balance, they are worth considering.

The analysis suggests that above about 82 μm, the optical coherence tomograph data is dominated by a constant added noise, which might represent the measurement error of the device alone. Below that thickness, some (possibly biological) process seems to take over that has expanding error, which is presumably related to glaucoma.

More data would be needed to validate the breakpoint models, but the simpler slope + constant model certainly quantified the increasing error with increasing damage in glaucoma patients. Clinically speaking, that model is quite relevant because presumably average peripapillary retinal nerve fibre layer thicknesses > 110 μm should probably be regarded as normal, and only at substantially lower thicknesses, where the linearly expanding error is valid, does one want to consider whether change on a later patient visit represents true progression or not.

Increased test-retest variability with lower average peripapillary retinal nerve fibre layer thickness has been reported using time domain optical coherence tomography.[22, 23] Using the RTVue 100 spectral domain optical coherence

tomograph, Li et.al[24] also reported that within-subject standard deviation and within-subject coefficient of variation were higher, and precision lower, for average retinal nerve fibre layer in eyes with severe glaucoma as compared to eyes with mild and moderate glaucoma.

In eyes with markedly reduced peripapillary retinal nerve fibre layer thickness, the increasing percentage of the retinal nerve fibre layer composed of non-neural components, in addition to the existing non-neural components around the optic nerve head, such as blood vessels, may further affect the ability of the instruments segmentation algorithm to consistently delineate the peripapillary retinal nerve fibre layer.

In summary: in the Topcon, for average peripapillary retinal nerve fibre layer, the level of disease present needs to be taken into consideration when determining the level of test-retest variability in the average peripapillary retinal nerve fibre layer measurement.

Macular ganglion cell complex

Although the coefficient of variation for macular ganglion cell complex thickness was 0.15, in absolute terms this translates to a small figure in terms of micrometres, i.e. a coefficient of repeatability of 2.9 μm (Table 2). The lowest macular ganglion cell complex thickness in the current study was 66 μm. Given that the coefficient of repeatability for the macular ganglion cell complex in the current study for this subject was 2.9 μm (Table 2), this represents the possible detection of loss due to disease progression of only 4% in a glaucoma patient with existing significant structural loss. Moreover, test-retest variability for the

macular ganglion cell complex was not found to vary in relation to the mean macular ganglion cell complex thickness (Fig. 1A). Clinically, this allows the application of a single test-retest variability value for any macular ganglion cell complex measurement, which greatly facilitates progression detection.

The Coefficient of Repeatability of 2.9 μm for the macular ganglion cell complex is quite good, and more recent studies have reported that measurements of the ganglion cell inner plexiform layer was able to detect progressive change in eyes with severely advanced glaucoma.[25]

It should be borne in mind that over a sufficiently long period of time, physiological structural change may induce statistically significant change in macular ganglion cell complex thickness. Given that macular ganglion cell complex thicknesses are compared to an age matched normative database, the result is colour coded as to what percentile the subjects' thickness lies in.

For those within the 95th percentile (coded green) and those in the 95th to 99th percentile (coded yellow), a change in macular ganglion cell thickness that does not induce a change percentile ranking may thus be more likely to be interpreted as physiological change. Significant change with an associated change in percentile ranking may be more likely to represent pathological change.

For those patients already ranked outside the 99th percentile, the clinician would need to take into account the patients overall presentation to determine whether any significant change is physiological or pathological in origin.

Although macular ganglion cell complex thickness values did not, on average, differ greatly from average peripapillary retinal nerve fibre layer values (Table 2), the test-retest variability was lower for the macular ganglion cell complex measurements. Measured average peripapillary retinal nerve fibre layer thickness may vary depending on where the instrument algorithm locates the centre of the optic nerve head. Delineating a consistent point as the centre of the (relatively large and complicated) optic nerve head may be more difficult than localising the centre of the (relatively small and uncomplicated) foveal pit consistently. Structural elements at the optic nerve head, such as blood vessels, which are smaller and less obtrusive in the macular scans, may also play a part in average peripapllary retinal nerve fibre layer test-retest variability. This additional complexity can be visualised in the retinal nerve fibre layer circular tomogram, which shows a quite uneven retinal nerve fibre layer surface and a quite uneven demarcation between the retinal nerve fibre layer and the underlying layers.

In contrast, examination of the macular ganglion cell complex scans shows a much smoother presentation of the retinal nerve fibre layer. This presumably facilitates more easily and consistently defined boundaries for the instrument algorithms. The consistency of macular ganglion cell complex scans may depend on the structural integrity of the foveal pit, and macular abnormalities, such as epiretinal membranes, may cause the automated scan centration algorithm to analyse scan data well away from the actual fovea.

Further investigations into the reasons for the discrepancy between average peripapillary retinal nerve fibre layer and macular ganglion cell complex test-retest variability may be warranted in the future

Acknowledgements: This research was supported by the Australian Research Council through the ARC Centre of Excellence in Vision Science (CE0561903), and the Australian Federal Government through the Nursing and Allied Health Scholarship Support Scheme. The views expressed in this paper do not necessarily represent those of the NAHSSS, its administrator, Services for Australian Rural and Remote Allied Health or the Australian Government Department of Health. The Australian Federal Government National Eye Health Initiative provided a grant which enabled the purchase of the Topcon 3D-OCT 2000 used in this study.

References

1. Kotera Y, Hangai M, Hirose F, Mori S, Yoshimura N. Three-dimensional imaging of macular inner structures in glaucoma by using spectral-domain optical coherence tomography. Invest Ophthalmol Vis Sci 2011;52:1412-1421.
2. Pearce JG, Pearce ME, Maddess T. Variability of peripapillary retinal nerve fibre layer measurements in four Spectral Domain Optical Coherence Tomographs. Clin Exp Ophthal 2013 2013;41:131.
3. Pierro L, Gagliardi M, Iuliano L, Ambrosi A, Bandello F. Retinal nerve fiber layer thickness reproducibility using seven different OCT instruments. Invest Ophthalmol Vis Sci 2012;53:5912-5920.
4. Fanihagh F, Kremmer S, Anastassiou G, Schallenberg M. Optical coherence tomography, scanning laser polarimetry and confocal scanning laser ophthalmoscopy in retinal nerve fiber layer measurements of glaucoma patients. Open Ophthalmol J 2015;9:41-48.
5. Hirasawa H, Araie M, Tomidokoro A, et al. Reproducibility of thickness measurements of macular inner retinal layers using SD-OCT with or without correction of ocular rotation. Invest Ophthalmol Vis Sci 2013;54:2562-2570.
6. Kim YJ, Kang MH, Cho HY, Lim HW, Seong M. Comparative study of macular ganglion cell complex thickness measured by spectral-domain optical coherence tomography in healthy eyes, eyes with preperimetric glaucoma, and eyes with early glaucoma. Jpn J Ophthalmol 2014;58:244-251.

7. Pearce JG, Maddess T. Retest Variability in the Medmont M700 Automated Perimeter. Optom Vis Sci 2016;93:272-280.
8. Paunescu LA, Schuman JS, Price LL, et al. Reproducibility of nerve fiber thickness, macular thickness, and optic nerve head measurements using StratusOCT. Invest Ophthalmol Vis Sci 2004;45:1716-1724.
9. Kamppeter BA, Schubert KV, Budde WM, Degenring RF, Jonas JB. Optical coherence tomography of the optic nerve head: interindividual reproducibility. J Glaucoma 2006;15:248-254.
10. Garas A, Vargha P, Hollo G. Reproducibility of retinal nerve fiber layer and macular thickness measurement with the RTVue-100 optical coherence tomograph. Ophthalmology 2010;117:738-746.
11. Chakraborty R, Read SA, Collins MJ. Diurnal variations in axial length, choroidal thickness, intraocular pressure, and ocular biometrics. Invest Ophthalmol Vis Sci 2011;52:5121-5129.
12. Chakraborty R, Read SA, Collins MJ. Diurnal variations in ocular aberrations of human eyes. Curr Eye Res 2014;39:271-281.
13. Sehi M, Flanagan JG, Zeng L, Cook RJ, Trope GE. The association between diurnal variation of optic nerve head topography and intraocular pressure and ocular perfusion pressure in untreated primary open-angle glaucoma. J Glaucoma 2011;20:44-50.
14. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135-160.
15. Harwerth RS, Wheat JL. Modeling the effects of aging on retinal ganglion cell density and nerve fiber layer thickness. Graefes Arch Clin Exp Ophthalmol 2008;246:305-314.
16. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. Ultrasound Obstet Gynecol 2003;22:85-93.
17. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420-428.
18. Ionan AC, Polley MY, McShane LM, Dobbin KK. Comparison of confidence interval methods for an intra-class correlation coefficient (ICC). BMC Med Res Methodol 2014;14:121.
19. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehabil 1998;12:187-199.
20. Jampel HD, Friedman D, Quigley H, et al. Agreement among glaucoma specialists in assessing progressive disc changes from photographs in open-angle glaucoma patients. Am J Ophthalmol 2009;147:39-44 e31.
21. Varma R, Steinmann WC, Scott IU. Expert agreement in evaluating the optic disc for glaucoma. Ophthalmology 1992;99:215-221.
22. Wu Z, Vazeen M, Varma R, et al. Factors associated with variability in retinal nerve fiber layer thickness measurements obtained by optical coherence tomography. Ophthalmology 2007;114:1505-1512.
23. Youm DJ, Kim H, Shim SH, et al. The effect of various factors on variability of retinal nerve fiber layer thickness measurements using optical coherence tomography. Korean J Ophthalmol 2012;26:104-110.
24. Li JP, Wang XZ, Fu J, Li SN, Wang NL. Reproducibility of RTVue retinal nerve fiber layer thickness and optic nerve head measurements in normal and glaucoma eyes. Chin Med J (Engl) 2010;123:1898-1903.
25. Belghith A, Medeiros FA, Bowd C, et al. Structural Change Can Be Detected in Advanced-Glaucoma Eyes. Invest Ophthalmol Vis Sci 2016;57:OCT511-518.

Chapter 3

Synchronous diurnal variation in test-retest variability of standard automated perimetry and spectral domain optical coherence tomography

Abstract

Purpose To investigate diurnal variation of summary measures, and test- retest variability (TRV) of those measures, arising from repeated spectral domain optical coherence tomography (SD-OCT) and standard automated perimetry (SAP) tests.

Methods Healthy subjects had repeated OCT and SAP tests of both eyes at three times: 8:00, 13:30 and 17:30, spanning normal office hours. The fixed test times for the repeated testing of each of the eyes of the seven subjects were designed to minimise sources of variance other than those under investigation. Those data were compared with a finer grained temporal analysis, in which one subject completed 4 repeated SAP tests at 8 intervals 1.5 hours apart. TRV was estimated as the 95% confidence intervals (CI_{95}) in the summary measures as obtained from Bland-Altman plots.

Results SD-OCT and SAP summary measures calculated at 8:00, 13:30 and 17:30 showed no diurnal variation. The TRV (CI_{95}) for horizontal and vertical cup/disc ratios (HCDR and VCDR) were significantly larger at 13:30, compared to 8:00. The TRV for peri-papillary RNFL mean thickness was significantly larger at 13:30 compared to either 8:00 or 17:30, as was the TRV for most SAP measures. The fine-grained time analysis agreed, showing the largest TRV at about 14:00 for three of four SAP summary measures.

Conclusions TRV for SD-OCT and SAP summary measures may vary similarly over the course of the day, even when the mean measures themselves show no diurnal change. The presence of subtle variations in visual field indices, as

reported in Appendix 1, indicates that avoiding testing in the early afternoon may reduce the TRV of SD-OCT and SAP testing.

Introduction

Fluctuating sensitivity to light during the day has been documented in some animals,[1] and in humans, circadian variations in visual discrimination thresholds and detection thresholds have been noted.[2] Circadian rhythms have been noted in retinal electrophysiological testing and psychophysical measures in healthy subjects,[3] and changes to the b-wave implicit time in the human electro-retinogram have been found to exhibit significant diurnal variation.[4]

Patients with post-retinal lesions have been found to have diurnal variations in functional visual field size,[5] but this variation only appeared to occur in patients who demonstrated a gradual change in increment threshold between the intact field and the scotoma. Glaucomatous damage extends to the brain[6] and retino-geniculo-cortical pathway.[7] Post-retinal damage might thus contribute to diurnal variation in glaucomatous fields.

Significant changes in several visual field test indices during the course of a normal working day have been reported for glaucoma patients.[8] Visual field sensitivity has also been found to be influenced by seasonal factors, with mean deviation being reported to be significantly higher in winter than summer.[9]

Most of the above studies used the Humphrey Field Analyser (HFA). Here we examine the Medmont M700, which has a very different test pattern to the HFA. Since standard automated perimetry (SAP) remains the gold standard for the detection of functional losses due to glaucomatous optic neuropathy,[10] it is important to identify any factors which might influence Test-Retest Variability

(TRV) To investigate this in a clinically meaningful way, we complete up to four retests per subject at three times of day that encompass normal office hours.

Spectral domain optical coherence tomography (SD-OCT) measures structural parameters of the optic nerve head and retina. Retinal thickness of healthy subjects has not been reported to show diurnal variation when measured,[11] although diurnal variation in the amount of macular oedema of patients with central retinal vein occlusion has been documented.[12] Diurnal variations in optic disc topography have been noted in cases of extremely high variations in IOP,[13] and have also been found in healthy subjects.[14] We therefore decided to acquire SD-OCT data with the same schedule as for the SAP.

For each data type, we provide data on the diurnal variation in key summary measures and then look for diurnal variation in the TRV of those parameters. Given that SAP and SD-OCT data were obtained at the same times, we also examined correlations between various parameters at the different test times.

Methods

A total of seven subjects were recruited for this study. The research conformed to the tenets of the Declaration of Helsinki, with informed written consent obtained as per ANU/ACT Health protocol 7/07.667. At the time of testing, no subject had any ocular pathology, and in particular, their optic nerve heads were assessed as having no glaucomatous changes.

All visual field tests were carried out on the Medmont M700 (Version 3.9.7, Medmont P/L, Nunawading, Victoria, Australia) automated perimeter utilising the *Central 30° test* (hereafter referred to as the 30°-test) using the fast

threshold test strategy and automated blind spot detection. The test pattern is depicted in Figure 1.

As some of the subjects were perimetrically naive, or had not undergone a visual field test for some time, a non-recorded field test was carried out on each eye to familiarise them with the test procedure. This familiarisation test was carried out prior to the commencement of the first day of the study, with the mean time between the familiarisation test and the first study test being 2.7 months (Standard Deviation (SD) 2 months).

For the six healthy subjects (mean age 29.5 years, SD 15.5 , range 15 to 50, refractive error range -0.75 to +1.25D), all visual field tests were carried out by the same experienced operator. When being set up for their demonstration tests, the canthus height from the chinrest was measured, recorded, and used to set up each patient in all subsequent tests. If a test lens was required (two subjects), the correct position was determined at the familiarisation test.

Due to the length of time taken for the number of tests involved, no strict pre-test protocols were undertaken by any subjects (i.e. there were no abstinence from caffeine, and food was eaten at variable times before testing, etc.). None of the subjects were smokers.

M700 Version 3.9.7 used here has been described in detail.[15] It uses rear projection light emitting diodes (peak wavelength 565 nm) which are Goldmann size 3. The maximum stimulus brightness is 318 cd/m² (0 dB), and is attenuated in sixteen 3 ± 1 dB levels. The background illumination is 10 apostilb (3.2 cd/m²) and uses automatic level control, with the fixation target being a yellow light emitting diode (peak wavelength 583 nm).

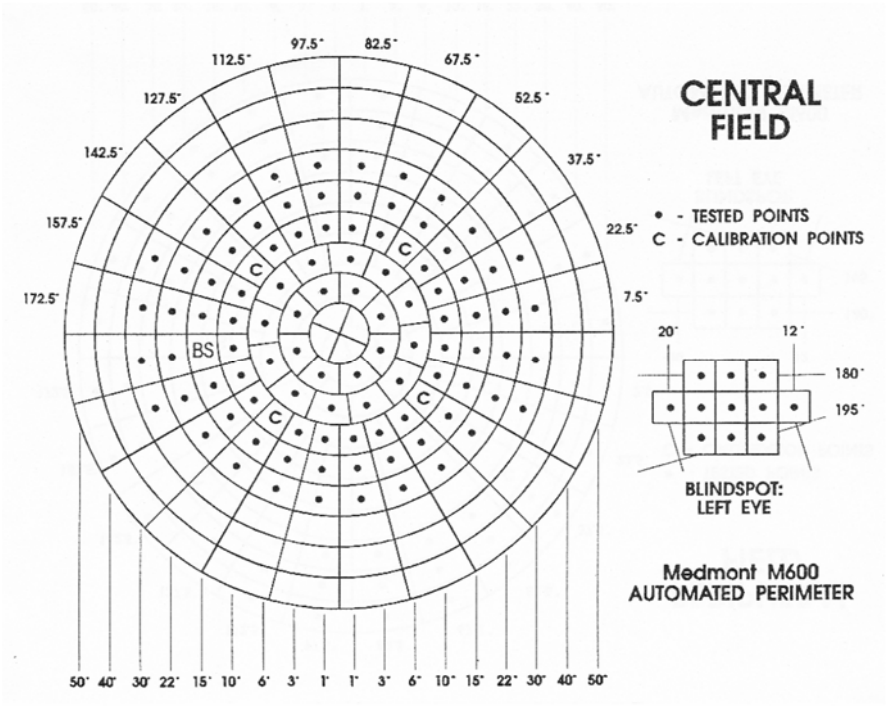


FIGURE 1.
M600 and M700 Central 30° test point layout, left eye shown here. The Goldmann size 3 stimuli are represented as black dots, and only those regions with a black dot (and the calibration points, C) are tested with the Central 30° test protocol. The stimuli are basically in a log polar layout, and are thus approximately scaled for retinal magnification. The stimulus layout provides much denser central sampling than other perimeters

The summary indices under consideration in the M700 were Overall Defect (OD), Pattern defect (PD), Hill of Vision 3° Level (HoV3) and Hill of Vision Slope (HoVS). OD is calculated as the mean difference between the age normal Hill of Vision (HoV) and the patient based HoV. A positive value for OD means the patient has a HoV that is higher (or more sensitive) than expected for their age, while a negative value indicates a lower sensitivity than expected for their age. A decrease in OD is therefore a measure of the decrease in the overall field sensitivity.

PD is a measure of the clustering and depth of defects, with clustered deviations from the patients' HoV resulting in an increase in PD while randomly distributed deviations do not contribute as much to PD, resulting in a lower value for PD. To calculate HoVS and HoV3°, the patients' HoV is estimated to fit a linear slope in decibels (dB/10°) for HoVS, and provide a dB level at 3° of eccentricity for HoV3°.

None of the six subjects had more than one test on each eye on any day, with tests carried out at close to 8.00 am, 1.30 pm and 5.30 pm (Table 1). The mean number of days for each subject to complete all the tests was 18 (SD 6 days).

In auxiliary testing, one single healthy subject, a 52 year old male (JGP), completed a total of four 30°-tests at eight different times of day (32 tests in total). Commencing at 8.00 am, tests were undertaken at intervals of 1.5 hours over a period of 28 days. Testing was from 4th June to 2nd July, as opposed to the other subjects, which had their tests carried out in the (southern hemisphere) summer time.

TABLE 1.
Test time statistics for the seven subjects for morning, midday and afternoon testing. Times were calculated using the time recorded on the first field test printout (the right eye in all cases).

Time referred to in text	Mean commencement time	Standard deviation (minutes)
8.00 am	8.07 am	14
1.30 pm	1.24 pm	36
5.30 pm	5.22 pm	16

For all subjects, the right eye was always tested first, and there was a three minute break between tests on the right eye and the subsequent test on the left eye. During this period, the subjects remained in the testing room, with the lights out, while waiting to undertake the next test. All tests used in this study were reliable according to the test reliability indices.

After finishing the last SAP test, the six subjects then underwent a single 3D disc scan on each eye using the Topcon 3D OCT-2000 SD-OCT (Topcon Corporation, Tokyo, Japan). Subjects who had not had an OCT scan prior to the commencement of this study underwent a non-recorded 3D disc scan on each eye to familiarise themselves with the procedure.

The optic nerve head scan protocol used in this study (3D Disc scan) covers an area of 6.0 × 6.0 mm, comprising of 128 horizontal B-scans and a total of 65,636 A-scans. The disc centre is calculated and the RNFL thickness measurements are calculated at the edge of a circle 3.4 mm in diameter.

Statistical analysis was carried out using Medcalc version 12.4 (Medcalc Software bvba, Ostend, Belgium) and Matlab (Matlab 6.1, The Mathworks Inc., Natick, MA, 2000).

Results

Visual field data

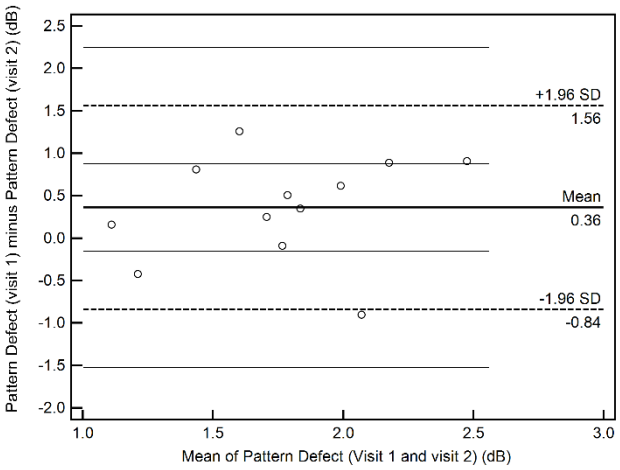


FIGURE 2.
Bland-Altman plot for M700 Pattern Defect for visual field tests carried out on the six subjects at 1.30 pm. The mean difference between scans is 0.36 dB. The Coefficient of Repeatability is 1.2 dB and the Limits of Agreement are –0.84 to + 1.56 dB. The 95% confidence limits (± 1.96 × SD) are taken around the mean difference, and, in this example, are ± 1.2 dB.

Bland-Altman plots (Fig. 2) are a graphical method often used to examine the limits of agreement between repeated measurements using the same instrument.[16] The presentation of a Bland-Altman plot enables a visual inspection of the data to be carried out, with any irregularities or significantly abnormal values easily identified and examined. The method is unaffected by regression to the mean. The standard deviation (SD) of individual differences is used as a measure of repeatability,[17] with the 95% confidence limits being calculated as the mean difference ± 1.96 SD.

We have chosen this method to calculate event based criteria which can be easily understood and applied clinically to differentiate instrument TRV from true progressive change.

TABLE 2.
Mean values for visual field test indices, right and left eyes combined, at each time point for the six healthy subjects for the M700 30°-test. All results are in decibels, with the exception of Hill of Vision Slope, which is in dB/10°.

	Overall Defect	Pattern Defect	Hill of Vision slope	Hill of Vision 3° level
8.00 am	5.50	1.77	1.63	30.5
1.30 pm	5.33	1.77	1.55	30.2
5.30 pm	5.38	1.70	1.58	30.2

Table 2 shows that average visual field index results did not alter significantly between any test times. Hill of Vision 3° levels were largely unchanged at any time point, despite this index having the largest confidence interval of 2.7 dB at 8.00 am.

TABLE 3.
Possible diurnal variation as measured by Bland-Altman 95% confidence intervals for M700 30°-test visual field indices at three different times of day for the six healthy subjects. The figures in brackets below are the bootstrap (n = 100,000; sampling with replacement) median 95% confidence intervals ± the standard error, allowing significance testing. T test p values for significance between midday versus morning and afternoon tests were all $p < 0.03$, with the exception of 8.00 am Hill of Vision 3° level ($p = 0.49$). All measurements are in decibels (dB), with the exception of Hill of Vision slope, which is in dB/10°.

	Overall defect	Pattern defect	Hill of vision slope	Hill of vision 3° level
8.00 am	1.5 (1.42 ± 0.22)	1.0 (0.94 ± 0.18)	1.6 (1.54 ± 0 23)	2.7 (2.54 ± 0.45)
1.30 pm	1.8 (1.75 ± 0.20)	1.2 (1.16 ± 0.24)	2.0 (2.00 ± 0.32)	2.7 (2.55 ± 0.66)
5.30 pm	1.4 (1.32 ± 0.22)	0.9 (0.89 ± 0.23)	1.3 (1.23 ± 0.23)	1.2 (1.16 ± 0.22)

For the six healthy subjects, CI_{95} were all significantly larger at 1.30pm for all visual field parameters, with the exception of Hill of Vision 3° (HoV 3°) level at 8.00 am. Note that the bootstrap values agree closely with the original CI_{95} for all parameters, and the CI_{95} values derived from the study are therefore slightly conservative.

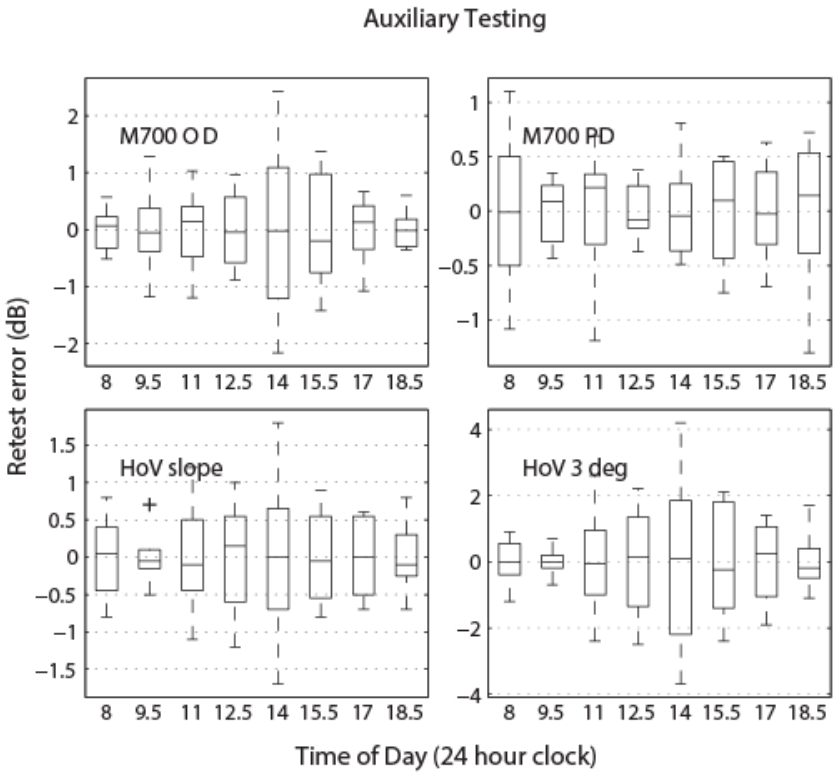


Figure 3.
Boxplots for M700 visual field indices for the auxiliary subject at all test times throughout the day. Mean retest error is represented by the orange bar, with the interquartile and 5th and 95th percentiles represented by the box and whiskers respectively.

For the auxiliary subject, Overall Defect (OD) did not show any statistically significant differences, while morning and afternoon testing showing reduced TRV ($p < 0.02$) for all other parameters except HoV slope at 5.30 pm ($p = 0.1$) (Fig.3).

Boxplots for the auxiliary subject at all time points tested during the day (Fig. 3) show that TRV was largest at about 2.00 pm for all parameters except PD, and

least for HoVS and HoV 3° at 9.30 am. These results appear to reinforce the finding that TRV for visual field parameters is higher in the middle of the day.

SD-OCT data

TABLE 4.

Mean values of the Topcon 3D OCT-2000 average pRNFL, horizontal cup/disc ratio (HCDR) and vertical cup/disc ratio (VCDR) at each time point for the six healthy subjects. Average pRNFL is in micrometres.

Time of day	Average pRNFL (μm)	HCDR	VCDR
8.00 am	113.6	0.26	0.26
1.30 pm	113.2	0.27	0.28
5.30 pm	113.8	0.27	0.28

TABLE 5.

Possible diurnal variation in the TRV measured by Bland-Altman 95% confidence intervals for average pRNFL, horizontal cup/disc ratio (HCDR) and vertical cup/disc ratio (VCDR) at three different times of day for the six healthy subjects. The figures in brackets below are the bootstrap (n = 100,000; sampling with replacement) median 95% confidence intervals ± the standard error, allowing significance testing. Statistically significant t test probabilities are shown where applicable.

Time of day	Average pRNFL (μm)	HCDR	VCDR
8.00 am	5.2 (5.0 ± 1.20) P < 0.001	0.02 (0.02 ± 0.004) P = 0.001	0.07 (0.063 ± 0.017) P = 0.06
1.30 pm	9.9 (9.6 ± 1.59)	0.07 (0.067 ± 0.023)	0.09 (0.080 ± 0.023)
5.30 pm	3.8 (3.8 ± 1.03) P < 0.001,	0.06 (0.062 ± 0.022)	0.06 (0.08 ± 0.023)

As with the visual field indices (Table 2), average values for SD-OCT parameters (Table 4) did not vary significantly according to time of day. From Table 5 however, average pRNFL TRV was significantly higher at 1.30 pm than in the morning and afternoon and both CDR parameters has statistically significant reductions in TRV in the morning as compared to 1.30 pm.

The magnitude of this difference appears to be clinically significant for pRNFL not only in percentage terms, but also in terms of micrometres (9.9 μm versus 3.8 μm), which could have a clinically significant effect on the analysis of serial average pRNFL test results with this instrument.

Correlations

Table 6.

Pearson's correlation coefficients for M700 visual field indices and Topcon SD-OCT parameters at each time point for the six healthy subjects. OD = Overall Defect, PD = Pattern Defect, HoV 3° = Hill of Vision 3° level, HoV S = Hill of Vision Slope, pRNFL = average peripapillary retinal nerve fibre layer, HCDR = Horizontal cup/disc ratio, VCDR = Vertical cup disc ratio. Values highlighted in **bold** had $p \leq 0.05$.

8.00 am	OD	PD	HoV 3°	HoV S	pRNFL	HCDR
PD	**0.68**					
HoV 3	**0.40**	**0.46**				
HoVS	- 0.26	- 0.18	**0.57**			
pRNFL	- 0.35	0.02	0.10	- 0.16		
HCDR	**0.69**	**0.40**	- 0.20	**- 0.51**	**- 0.50**	
VCDR	**0.67**	0.37	- 0.22	**- 0.50**	**- 0.53**	**0.99**
1.30 pm						
PD	**0.81**					
HoV 3	**0.68**	**0.56**				
HoVS	- 0.25	- 0.29	0.31			
pRNFL	- 0.25	- 0.28	0.2	0.11		
HCDR	0.21	0.09	- 0.2	0.04	- 0.39	
VCDR	0.21	0.07	- 0.2	0.06	**- 0.40**	**0.99**
5.30 pm						
PD	**0.50**					
HoV 3	**0.56**	**0.44**				
HoVS	- 0.34	- 0.36	0.16			
pRNFL	- 0.02	0.25	**0.46**	- 0.15		
HCDR	0.24	- 0.10	- 0.30	0.08	- 0.38	
VCDR	0.24	- 0.07	- 0.31	0.10	- 0.39	**0.99**

Table 6 shows that correlation coefficients between some parameters varied greatly at different times of day. The greatest difference in values was in the correlation between Hill of Vision Slope (HoV S) and horizontal cup/disc ratio

(HCDR) and vertical cup/disc ratio (VCDR) at 8.00 am compared to both later times of day.

There were 17 correlations which did not show consistently high correlation coefficients throughout the day (shaded areas, Table 6). At 8.00 am, 8 of these correlations were significant, compared to only one value out of the 17 at each of the other two time points. Overall defect (OD), Pattern defect (PD) and HoV S were all much more significantly correlated with HCDR and VCDR at 8.00 am than at any other time points.

The correlation between HoV S and the Hill of Vision 3° level (HoV 3°) was significant at 8.00 am, but declined during the day and was not significant at any other time points. This decline in correlation during the day was also evident for PD versus HCDR and VCDR.

Discussion

Although it has been suggested that visual field testing should be carried out at consistent times of day,[18] the optimal time of day has not been specifically addressed. In addition, most clinical practice protocols do not incorporate SD-OCT testing at specific or consistent times of day. The results from the current study seem to indicate that there are clinically significant alterations in TRV in SAP and SD-OCT dependent on the time of day at which the test is undertaken, and that there may be certain times of day when TRV is lower.

A reduction in SAP TRV of 20% may allow progression to be detected one visit sooner.[19] This level of reduction appears to be obtainable in the M700 by the simple expedient of not carrying out visual field tests in the early afternoon. Early morning testing may also significantly reduce the TRV of M700 tests. The

caveat to this is that the results in the current study were obtained from healthy patients, and may not be applicable to glaucoma patients.

Whilst the results of the current study have not established any significant alteration to SAP index values, average pRNFL thickness or CDR's in healthy patients over the course of a normal day, it has demonstrated that TRV may vary according to the time of day. Of particular interest is the fact that average pRNFL thickness remained virtually unchanged at different times of day, despite the significantly different TRV at 1.30 pm compared to 8.00 am and 5.30 pm (Table 2).

Changes to TRV according to the time of day may have a considerable impact on clinical decision making using both trend based and event based progression criteria. Whilst the underlying mechanisms for these changes remain unresolved at present, there are several possible explanations for the changes to perimetric TRV. For example, it has been reported that eating lunch may induce a decrease in the ability to maintain attention[20] and high starch and high sugar meals have been reported to slow the reaction time to visual stimuli presented in the periphery.[21]

Explanations for the alterations to structural TRV are a little less clear, although significant diurnal changes to optic nerve head topography has been reported in both glaucoma subjects and healthy individuals[14] which may explain changes to ONH TRV. Foveal retinal thickness has been reported to show no significant diurnal thickness variation.[22] A more recent study[23] reported that although no significant diurnal changes in average pRNFL thickness were detected, the highest value was recorded at 9.00 am and the lowest value was recorded at

6.00 pm. Interestingly, this study also reported the standard deviation in average pRNFL to be the smallest at 12 pm.

The possibility that average pRNFL is undergoing some sort of structural change between early morning and late afternoon is a possible explanation for our finding in regard to average pRNFL TRV.

Correlations between some parameters did vary depending on the time of day testing was undertaken (Table 6). Given the apparent lack of change in mean visual field index and SD-OCT parameters, the basis for any alterations to correlations between these parameters may lie in the changes to TRV due to time of day (Tables 3 and 5).

Although further investigations would be required to confirm the results from the current study, these results raise the possibility that structure-function relationships may not be consistent at different times of day, which may have implications for progression detection methods which combine structural and functional analysis protocols. The results from the current study seem to indicate that for structure–function analysis, consistent early morning testing may result in generally higher correlations.

Acknowledgements: This research was supported by the Australian Research Council through the ARC Centre of Excellence in Vision Science (CE0561903), and the Australian Federal Government through the Nursing and Allied Health Scholarship Support Scheme (NAHSSS) and the National Eye Health Initiative, which enabled the purchase of the Topcon 3D-OCT 2000 used in this study. The views expressed in this paper do not necessarily represent those of the NAHSSS, its administrator, Services for Australian Rural and Remote Allied

Health of the Australian Government Department of Health. The Medmont figure (©Medmont) has been used with the friendly permission of Medmont.

References

1. Tassi P, Pins D. Diurnal rhythmicity for visual sensitivity in humans? Chronobiol Int 1997;14:35-48.
2. Tassi P, Pellerin N, Moessinger M, Hoeft A, Muzet A. Visual resolution in humans fluctuates over the 24h period. Chronobiol Int 2000;17:187-195.
3. Tuunainen A, Kripke DF, Cress AC, Youngstedt SD. Retinal circadian rhythms in humans. Chronobiol Int 2001;18:957-971.
4. Hankins MW, Jones RJ, Ruddock KH. Diurnal variation in the b-wave implicit time of the human electroretinogram. Vis Neurosci 1998;15:55-67.
5. Zihl J, Poppel E, von Cramon D. Diurnal variation of visual field size in patients with postretinal lesions. Exp Brain Res 1977;27:245-249.
6. Gupta N, Yucel YH. What changes can we expect in the brain of glaucoma patients? Surv Ophthalmol 2007;52 Suppl 2:S122-126.
7. Yucel Y. Central nervous system changes in glaucoma. J Glaucoma 2013;22 Suppl 5:S24-25.
8. Junoy Montolio FG, Wesselink C, Gordijn M, Jansonius NM. Factors that influence standard automated perimetry test results in glaucoma: test reliability, technician experience, time of day, and season. Invest Ophthalmol Vis Sci 2012;53:7010-7017.
9. Gardiner SK, Demirel S, Gordon MO, Kass MA. Seasonal changes in visual field sensitivity and intraocular pressure in the ocular hypertension treatment study. Ophthalmology 2013;120:724-730.
10. Erb C, Gobel K. [Functional glaucoma diagnosis]. Ophthalmologe 2009;106:375-385; quiz 386.
11. Jo YJ, Heo DW, Shin YI, Kim JY. Diurnal variation of retina thickness measured with time domain and spectral domain optical coherence tomography in healthy subjects. Invest Ophthalmol Vis Sci 2011;52:6497-6500.
12. Gupta B, Grewal J, Adewoyin T, Pelosini L, Williamson TH. Diurnal variation of macular oedema in CRVO: prospective study. Graefes Arch Clin Exp Ophthalmol 2009;247:593-596.
13. Lee BL, Zangwill L, Weinreb RN. Change in optic disc topography associated with diurnal variation in intraocular pressure. J Glaucoma 1999;8:221-223.
14. Sehi M, Flanagan JG, Zeng L, Cook RJ, Trope GE. The association between diurnal variation of optic nerve head topography and intraocular pressure and ocular perfusion pressure in untreated primary open-angle glaucoma. J Glaucoma 2011;20:44-50.
15. Pearce JG, Maddess T. Retest Variability in the Medmont M700 Automated Perimeter. Optom Vis Sci 2016;93:272-280.
16. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-310.
17. Hanneman SK. Design, analysis, and interpretation of method-comparison studies. AACN Adv Crit Care 2008;19:223-234.
18. Fingeret M. Factors that influence automated perimetry. Invest Ophthalmol Vis Sci 2012;53:7018.

19. Turpin A, McKendrick AM. What reduction in standard automated perimetry variability would improve the detection of visual field progression? Invest Ophthalmol Vis Sci 2011;52:3237-3245.
20. Smith AP, Rusted JM, Eaton-Williams P, Savory M, Leathwood P. Effects of caffeine given before and after lunch on sustained attention. Neuropsychobiology 1990;23:160-163.
21. Smith A, Leekam S, Ralph A, McNeill G. The influence of meal composition on post-lunch changes in performance efficiency and mood. Appetite 1988;10:195-203.
22. Read SA, Collins MJ, Alonso-Caneiro D. Diurnal variation of retinal thickness with spectral domain OCT. Optom Vis Sci 2012;89:611-619.
23. Sharifipour F, Farrahi F, Moghaddasi A, Idani A, Yaseri M. Diurnal Variations in Intraocular Pressure, Central Corneal Thickness, and Macular and Retinal Nerve Fiber Layer Thickness in Diabetics and Normal Individuals. J Ophthalmic Vis Res 2016;11:42-47.

Preface to Appendices

In the published chapters of the thesis (Chapters 1 and 2), test-retest variability (TRV) was presented as the Coefficient of Repeatability (CoR) at the request of the reviewers. The CoR is calculated from Bland-Altman plots as the difference between the upper Limit of Agreement (LoA) and the mean difference between measurements.

In the Appendices and Chapter 3, TRV is presented as the 95% confidence interval (CI_{95}), which is represented as ± the difference between the upper and lower LoA from the Bland-Altman plot divided by two. When carrying out repeat measurements on the same instrument in each Bland-Altman plot, the mean difference is close to zero, and the CI_{95} is usually equal to the CoR.

When comparing instruments, the differences between measurements is reported as ± LoA.

Appendix 1

Circadian variation in visual field test indices during a 24 hour period

Abstract

Background Variability in visual field test indices has been documented due to testing being carried out at different times of day. At present, it is unknown whether there is any circadian variations in visual field test-retest variability (TRV) or visual field indices over a 24 hour period. The purpose of this study is to determine whether any clinically significant circadian variations exist in visual field test results over a 24 hour period.

Method A very well characterised subject was recruited for the study. Testing was carried out on the Medmont M700 automated perimeter, commencing at 8.00 am and being repeated on both eyes every one and a half hours over a 24 hour period, providing 64 test results. The study was conducted in two sessions, with each session six months apart near the equinoxes to ensure no variability was induced by the length of day.

Results No significant alterations to visual field sensitivities, indices or TRV were detected when comparing daytime test results with those during the night. Although some alterations to visual field indices appeared consistent, they were not statistically significant. Factor analysis showed that two underlying factors accounted for approximately 80% of the variability in the data.

Conclusions The results of this study indicate that subtle variations in visual field indices may be present over a 24 hour period. In Chapter 3, significant alterations to the test-retest variability in visual field indices was reported. As such, *c*linicians should give some consideration to scheduling visual field test at the same time of day in order to avoid any possible variability being induced by testing at different times of day.

Introduction

There is evidence that there are daily fluctuations in visual function, with the visual discrimination threshold being low in the morning and increasing progressively over the day.[1] Electrophysiological testing and psychophysical measures have elicited circadian rhythms in in healthy subjects,[2] although statistically significant circadian changes in retinal may be too small to have any clinically significant impact when assessing visual function.[3] Human cone responses have also been found to be highest at about 20.00 h and lowest at 06.00 h, suggesting an endogenous rhythm of cone responsiveness to light.[4]

Intrinsically photosensitive retinal ganglion cells (ipRGCs) are the main relay to the brain for rod and cone signals responsible for circadian photoentrainment,[5] and the ipRGCs controlled post illumination pupil response has been found to demonstrate a circadian rhythm independent of external light cues.[6] Moderate to severe glaucoma has been found to affect the ipRGC mediated post-illumination pupil response,[7] with implications for circadian photoentrainment.

Although patients would not be expected to routinely undertake visual field tests outside normal office hours, the presence of detectable circadian variations in visual field sensitivity could have implications regarding the initiation of glaucomatous change and progression in the human eye.

In clinical practice, there are few instances where patients would be required to undertake more than one standard automated perimetry (SAP) test in one visit. However, SAP testing is generally carried out on older patients, who may experience variable levels of general health and wellbeing.

Repeated testing with Frequency Doubling Technology (FDT) over six tests in two sessions has shown a decrease in sensitivity of approximately 2 dB between the first and sixth tests,[8] although inter-eye fatigue has not been found to influence mean defect or test reliability indices using the Humphrey Field Analyser (HFA) Swedish Interactive Threshold Algorithm (SITA).[9]

Fluctuations in general health in an older population may have some effect on SAP testing via some attenuation of cognitive function and attentional abilities, which could contribute to SAP variability and alterations to summary indices. If the results of SAP testing can be affected by changes in aspects other than progression in eye disease, the patients' general health and demeanour should perhaps be taken into account when assessing SAP test results.

When investigating circadian influences and the test-retest variability (TRV) of an instrument, we need to consider other factors which may be contributing to the total variance. Some examples of factors other than machine variance which may contribute to the total variance include subject variance (the number of subjects) and the number of operators. The total variance (σ^2_{total}) can thus be represented as

1) $\sigma^2_{total} = \sigma^2_{retest} + \sigma^2_{subject\ (n)} + \sigma^2_{\ operator\ (m)} + \sigma^2_{\ circadian} + \sigma^2_{annual} + \sigma^2_{error}$.

In this study, we have used a single subject and a single operator in order to minimise the variance from these two components, and focus on the retest and circadian contributions by taking many repeats over several days. In addition to isolating machine variance, our study protocol also allows us to examine any within-subject differences between eyes.

Using this strategy, we thus intend to determine whether any changes to visual field test sensitivity and M700 summary indices occur over a 24 hour period, and whether any detectable circadian influences on these parameters exist. We also intend to investigate the effect of fatigue and attentional change via repeated field testing over the 24 hour period. Although being tested at 2.00 am, after being woken from sleep, may not exactly replicate the effects of old age, diminished cognitive skills and/or poor general health, the possible effects of repeated testing and some sleep deprivation will be assessed.

Methods

One subject, a 51 year old male, JP, was enrolled in this study and the appropriate consent forms were signed under ANU/ACT Health protocol 7/07.667. At the time of testing, the subject had no ocular pathology and had his optic nerve head had been assessed as having no glaucomatous changes. Intra-ocular pressures for the subject were within normal limits, and central corneal thickness (CCT) was included in this assessment. Recent visual field tests on the subject have demonstrated no visual field defect, and previous 3D disc scans and ganglion cell complex scans using the Topcon 3D-OCT 2000 spectral domain optical coherence tomograph have revealed no abnormalities of the optic nerve head or the peripapillary retinal nerve fibre layer.

The standard automated perimeter used for the visual field testing in this study is the Medmont M700 Software Version 3.9.7. (Medmont P/L, Nunawading, Victoria, Australia) using an M600 bowl. The M600 bowl is a part hemispherical bowl perimeter, with a radius of 350 mm and a test distance of 400 mm. It uses rear projection light emitting diodes (pale green in colour, peak wavelength 565 nm) which are Goldmann size III, and the stimulus size is not adjustable. The

maximum stimulus brightness is 318 cd/m^2 (0 dB) and is attenuated in sixteen 3 ± 1 dB levels. The background illumination is 10 apostilb (3.2 candela/m^2) using automatic level control, with the fixation target being a yellow light emitting diode of peak wavelength 583 nm.

The subject has had numerous field tests done with the M700. The M700 Central 30° test (hereafter referred to as the *30°-test*) was used, with the fast threshold strategy. The subject was corrected for the test distance of 400 mm using a corrective lens of +3.00 D, with a test lens diameter of 48 mm. The canthus height alignment was recorded for the first test by an assistant, and remained unchanged for all subsequent tests. All tests were carried out in the same darkened room by the same operator. The right eye was tested first in all test sessions. There was a three minute break between tests on each eye, as this would be considered a common, if not maximal, time for patients to rest between tests in clinical practice. The subject remained in the darkened test room during this three minute break. After the visual field tests were completed, the subject underwent 3D disc scans using the Topcon 3D-OCT 2000 spectral domain optical coherence tomograph.

The summary indices under consideration in the M700 were Overall Defect (OD), Pattern defect (PD), Hill of Vision 3° Level (HoV3) and Hill of Vision Slope (HoVS). OD is calculated as the mean difference between the age normal Hill of Vision (HoV) and the patient based HoV. A positive value for OD means the patient has a HoV that is higher (or more sensitive) than expected for their age, while a negative value indicates a lower sensitivity than expected for their age. A decrease in OD is therefore a measure of the decrease in the overall field sensitivity.

PD is a measure of the clustering and depth of defects, with clustered deviations from the patients' HoV resulting in an increase in PD while randomly distributed deviations do not contribute as much to PD, resulting in a lower value for PD. To calculate HoVS and HoV3°, the patients' HoV is estimated to fit a linear slope in decibels (dB/10°) for HoVS, and provide a dB level at 3° of eccentricity for HoV3°.

Each test session commenced at 8.00 am, and testing was carried out every one and a half hours for a total period of 24 consecutive hours (i.e. 16 tests per eye in total). The subject slept (or attempted to sleep) between tests from 9.30 pm onwards, and during the day carried out all normal activities such as shopping, etc. The subject took all his medications at the same time he normally does (at 6.30 am) on the day the testing commenced. The prescribed medications were Lipitor (Atorvastatin) 40 mg, Perindo (Perindopril erbumine) 4 mg. The subject also took his daily non-prescribed multi-vitamin (Cenovis Multivitamins and Minerals).

The test sessions were carried out on two separate occasions, the first in late March, the second in late September. Although those months represent different seasons, the length of day was consistent for each test session to avoid any effect on the results being attributable to different lengths of day.

Two statistical methods were used to analyse the data. Bland-Altman difference plots were used to determine the 95% confidence intervals when assessing TRV. Bland-Altman plots are a graphical method often used to examine the limits of agreement between repeated measurements using the same instrument.[10] The presentation of a Bland-Altman plot enables a visual

inspection of the data to be carried out, with any irregularities or significantly abnormal values easily identified and examined. The method is unaffected by regression to the mean. The standard deviation (SD) of individual differences is used as a measure of repeatability,[11] with the 95% confidence limits being calculated as the mean difference ± 1.96 SD of the differences. Bland-Altman analysis was carried out using MedCalc v.16.4.3 (MedCalc Software, bvba Ostend, Belgium).

Bland Altman plots have been described and presented in Chapters 1, 2 and 3 of this thesis, and for the sake of avoiding repetition, we do not present a figure here.

Factor analysis was carried out in order to provide an assumption free examination of the major sources of variance over 24 hours, and the extent to which they might display any circadian rhythm(s). Our main method involved computing the principal components of the measured variables, sometimes called principle components analysis (PCA). PCA is a method of data reduction which finds orthogonal (un-correlated) factors, each of which is a linear combination (weighted average) of the original variables. These regression weights are often called *factor loadings*.

Factors accounting for a large proportion of the data represent putative under-lying mechanisms in the data that drive the pattern of observed correlations between the measured variables.[12] In the present case, we also examine how the discovered factors vary according to the time of day, the so-called *factor scores*. A secondary objective is to examine how well a small number of the larger factors (larger in terms of proportion of variance accounted for in the

data) give a balanced account of the data (measured by what are called the *communalities*).

Factor analysis is thus an exploratory type of statically analysis that might point to previously unsuspected drivers of variance in the data, which in turn might be the subject for further study. We also examined an alternative method for identifying underlying factors using a maximum likelihood method. This can identify non-orthogonal factors, but summary statistics indicated that this method was not better than the simpler PCA method, and so we do not report on it here. The factor analyses were done in MATLAB version 2014a (The MathWorks Inc., Natick, Massachusetts, United States).

Results

Table 1.

Bland-Altman 95% confidence intervals derived from tests taken during both 24 hour test periods. The results have been divided into daylight/office hours (8.00 am to 6.30 pm) and night (8.00 pm to 6.30 am), with each period consisting of eight field tests per eye (16 tests in total).

	8.00 am to 6.30 pm	8.00 pm to 6.30 am
Pattern Defect (dB)	± 0.8	± 0.9
Overall Defect (dB)	± 2.2	± 2.2
Hill of Vision slope (dB/10°)	± 2.0	± 1.9
Hill of Vision 3° level (dB)	± 4.0	± 3.5

From Table 1, it can be seen that there was no obvious change in confidence intervals for any visual field test parameter associated with tests carried out during the day and tests carried out at night. It should be remembered that these confidence intervals were derived from two sets of results six months apart. The effect of fatigue during the nocturnal tests appeared to be inconsequential.

The consistency of the raw values obtained at both sessions is demonstrated in Tables 2 and 3, which list the average value for each parameter for each eye during the day and night sessions respectively.

Table 2.

Mean daytime and night-time parameter values of each eye. There were eight tests per eye during the day and night subdivisions.

	Right eye		Left eye	
	Day	Night	Day	Night
Pattern defect (dB)	2.3	2.0	2.5	1.9
Overall defect (dB)	7.6	6.9	6.9	6.9
Hill of Vision Slope (dB/10°)	2.4	2.1	2.0	2.1
Hill of Vision 3° level (dB)	32.5	31.5	31.4	31.3

Most parameters demonstrated no consistent changes in index values at night as compared to during the day, once again indicating no effect from fatigue during nocturnal testing. The exception to this was PD, which did show a consistently lower value for each eye at night, although a paired sample t-test did not show any significance in the observed difference.

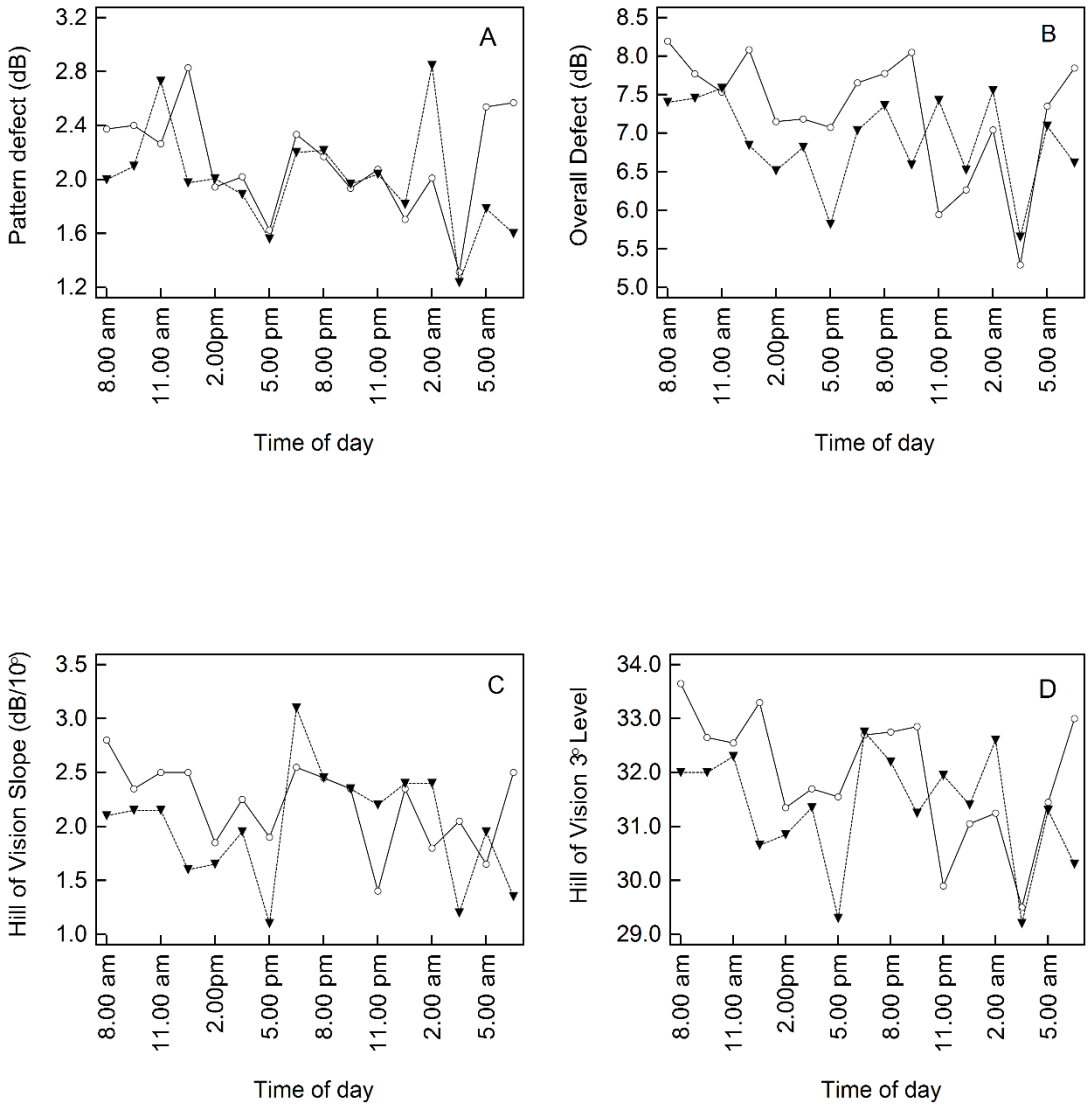


Figure 1.

Panels A – D show the results at each time point for each eye for the summary indices of the M700. The results for each eye were averaged over the two test sessions.

Legend: ○…….○ Right eye
▼……▼ Left eye

Figures 1A to 1D show each summary index at each time point, results from both sessions combined. Within-eye correlations were mostly significant, or near significant, suggesting the averaging of results from each eye from both sessions. It can be seen from these graphs that the results varied in

consistency between eyes. This inconsistency is not unexpected, given that variability is one of the biggest problems with SAP, and there were no results that fell outside the confidence intervals for each parameter.

The only markedly different observation involving both eyes was a decrease in visual field sensitivity (OD and HoV3) and PD at 3.30 am. Although appearing to be considerably different to observations at other time points, this difference still fell within the 95% confidence intervals from the mean values from each parameter (Tables 1 and 2).

Factor analysis can be used to identify any underlying, but unknown, variables that may be influencing the variance for different parameters at different times of day in our data. These unknown variables are called factors, and there are the same number of factors as there are variables, with each factor explaining a proportion of the overall variance in the variables under examination. The factors are assigned a number according to the proportion of variance in the data that is explained by the factor. Factor 1 is thus responsible for the largest proportion of variance, and so on. We therefore used factor analysis to determine whether there was any underlying commonality which may be contributing to the variance in our results. Factor analysis was carried out on the four summary indices of the M700, those being Overall Defect, Pattern Defect, Hill of Vision Slope and Hill of Vision 3º level.

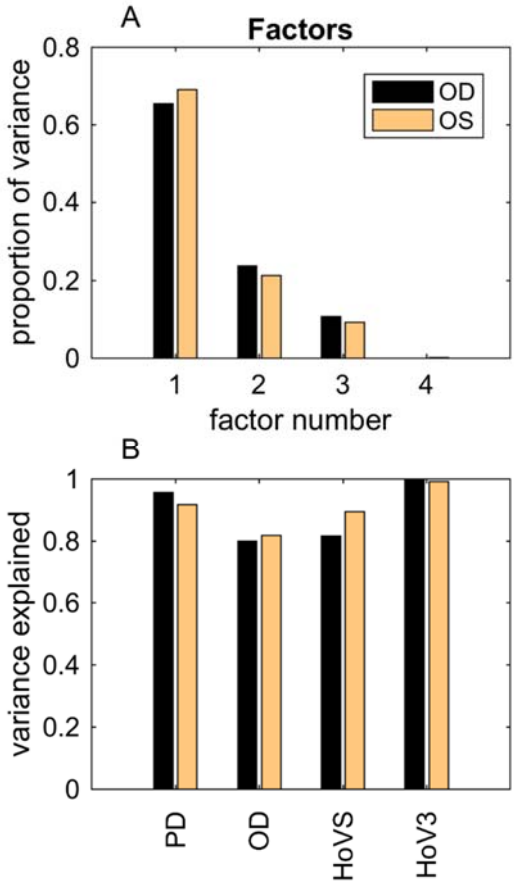


Figure 2
Panel A shows the proportion of variance explained by the four main factors derived from factor analysis. Approximately 80% of the total variance is explained by factors 1 and 2, and all factors show a high degree of consistency between eyes.

Panel B shows the proportion of variance explained (using the two factor model) for each M700 index under consideration. Between-eye consistency remains high, as does the proportion of variance explained by factors 1 and 2.
PD = Pattern Defect, OD = Overall defect, HoVS = Hill of Vision Slope, HoV3 = Hill of Vision 3° level.

Fig. 2A shows that over 80% of the variance is explained by two factors (Factor 1 and Factor 2) while Fig. 2B shows that the proportion of variance explained by the two factors was consistently high for each M700 index. Fig. 2 A and B also demonstrate the consistency of the results for each eye. This consistency is quite interesting in view of the time between test sessions (6 months) and the TRV inherent in SAP testing generally.

Factor loadings are an extension of factor analysis, where the loadings represent how much a factor explains each variable examined in factor analysis. The loadings are assigned a value between ± 1. The closer the factor loading value to either + or – 1, the stronger the effect of the factor on the variable under consideration. A value of zero indicates that the factor does not explain any variance in the variable, while a negative value indicates an inverse relationship between the factor and the variable.

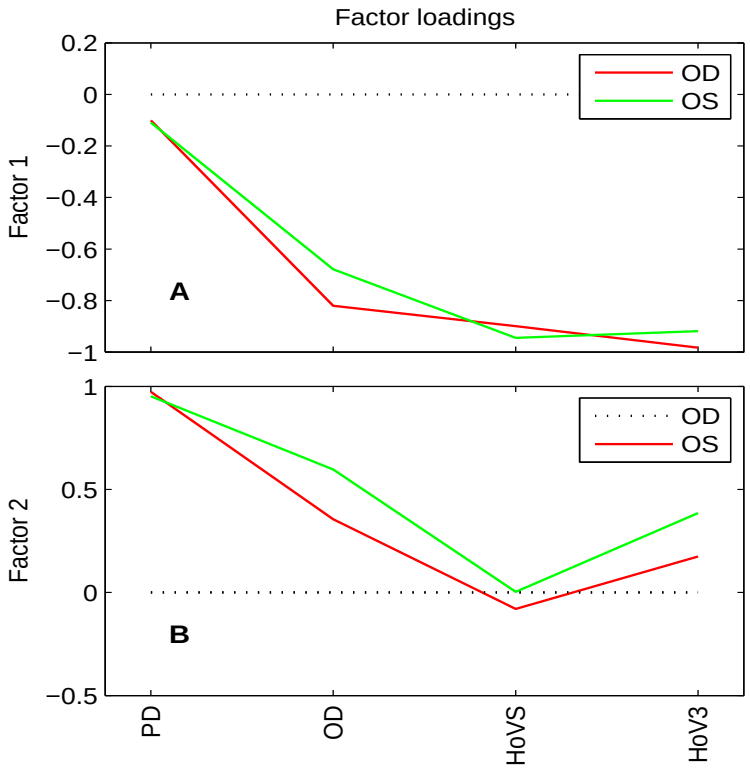


Figure 3 A, B

Panel A shows the factor loadings for each eye for each M700 index for Factor 1. Pattern Defect (PD) had the least weighting while the other indices had quite high factor loadings and were reasonably similar. **Panel B** shows the factor loadings for each eye for each M700 index for Factor 2. In contrast to Factor 1, PD had the highest weighting for Factor 2, and HoVS had a noticeably lower weighting than OD and HoV3°.

PD = Pattern Defect, OD = Overall defect, HoVS = Hill of Vision Slope, HoV3 = Hill of Vision 3° level.

Figure 3 A, B shows the factor loadings for each eye and each parameter for Factors 1 and 2. Between-eye consistency remained high for both factors and all parameters. Index weightings varied between factors, with PD having the least weighting for Factor 1 and the highest weighting for Factor 2. Thus the factors split neatly between those related to the mean amount of damage (Overall defect) (Factor 1), and the Pattern Defect (Factor 2). Mathematically, this corresponds roughly to the factors being related to the first and second moments of the distributions of the visual field damage. Given that these two factors account for such a large proportion of the variance, the consistency between eyes and M700 indices for each factor is quite encouraging.

Factor scores can be derived from factor analysis, and were used to indicate how much each factor varied at different time points during the day. We have therefore determined the factor scores for Factors 1 and 2 for each time point used in the current study (Fig. 4 A, B). These figures show the contribution of each factor to the variance encountered at different time points.

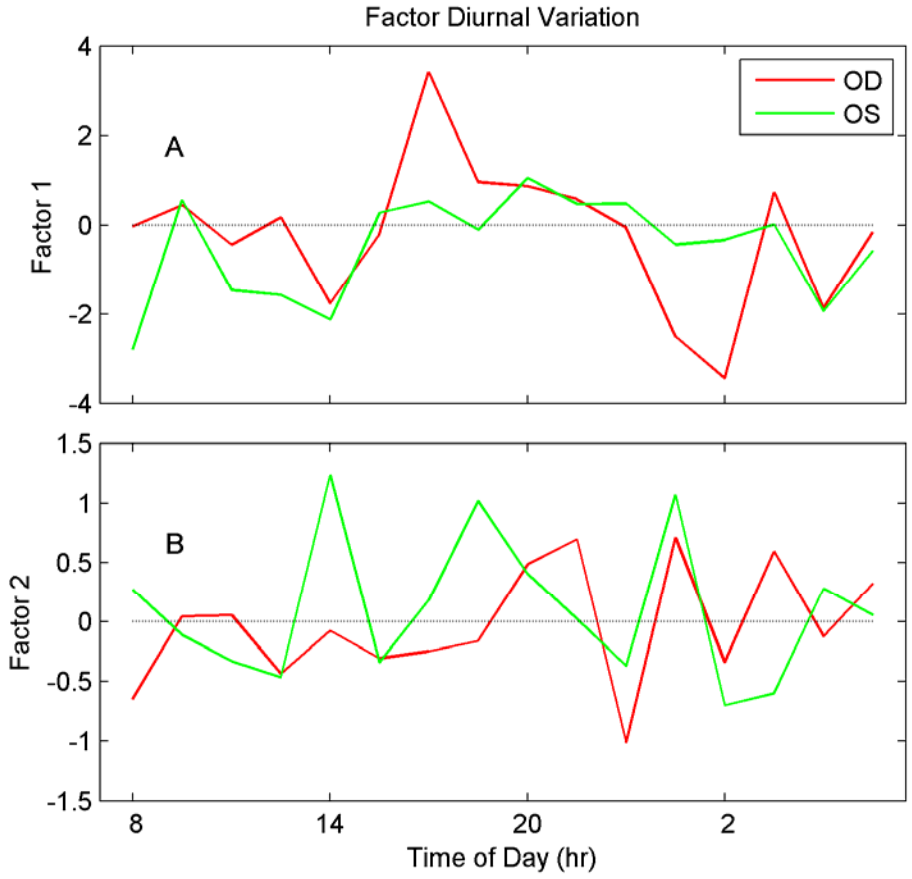


Figure 4

Panel A shows the factor scores for each eye for Factor 1
Panel B shows the factor scores for each eye for Factor 2
OD = Right eye, OS = Left eye

The factor scores were inconsistent between eyes, although Factor 1, which accounts for approximately 65% of the variance in the data, does show some large differences at different time points in the right eye.

Discussion

The effect of sleep interruption to the subject's visual field test performance during the night is difficult to quantify. In the first instance, there have been no comparative studies where visual field tests have been carried out over a 24 hour period, or between the hours of 8.00 pm to 6.30 am. Similarly, the effect on

task performance due to the subject of being awoken every 1.5 hours during the night has not been explicitly investigated in this regard.

There is some evidence to suggest that sleep deprivation may result in the degradation of attention and visual tracking abilities[13] and that circadian variations for various attentional aspects have reported a decrease in accuracy and an increase in reaction time between 04:00 hours and 07:00 hours.[14] A central discrimination and peripheral detection task was reported to show an increase in overall task performance in the morning which remained stable in the afternoon,[15] although this was a diurnal study only.

As none of our visual field test indices varied outside the 95% confidence intervals for each visual field summary index, it is hard to say whether the time of day and /or fatigue had any effect on either visual field sensitivity or subject performance. Had we been able to detect any significant alterations in our visual field summary indices, then the influence of fatigue and/or time of day would then have had to have been addressed and quantified.

The results from the current study did not demonstrate any statistically significant variations in summary index values or TRV when comparing values at different time points. In addition, no changes were observed in parameters derived from tests undertaken during the night (Tables 1 and 2), when some effects of fatigue may have been expected to contribute to TRV and/or changes to visual field index values. Any apparent changes in index values are overshadowed by a lack of consistency, and by having the changes in parameters falling within the mean ± 95% confidence intervals established for

the M700 in this study and, for OD and PD, the confidence intervals established using a cohort of glaucoma patients.[16]

Factor analysis indicated that a large proportion of the variance, approximately 80%, could be accounted for by two underlying factors (Fig. 2 A, B). Factor loadings (Fig. 3 A, B) showed excellent between-eye correlations for these factors in relation to each M700 index. Although factor scores (Fig. 4 A, B) were a little inconsistent, the results of the current study seem to provide some direction towards the investigation of what may constitute some of these underlying factors.

An important aspect of factor analysis was the consistency of both factor analysis and factor loadings for each summary index. The consistency of the two major underlying factors over the time period between test sessions (six months) is a positive indication that there may be a consistent, and as yet unresolved, component of SAP TRV that requires further investigation. Given that subject variance was reduced to one in the current study, it also remains to be seen whether inter-individual differences in any underlying factors exists.

Despite our inability to establish any statistically significant differences, the possibility of circadian variations occurring in visual field test results remains, as for example shown by Factor 1 OD (Fig. 4A). These changes may impact on SAP TRV, and in the current absence of any other information regarding circadian variation in visual field test parameters, it may be better for clinicians to review their visual field test appointment times and ensure consistent test times.

As described in the methods, a decrease in OD represents a decrease in a patients overall visual field sensitivity. Although the possibility of a change in visual field sensitivity (OD and HoV3) and PD at 3.30 am exists (Fig. 1 A, B, D), this is not clinically significant from a practical patient management point of view. It may be significant when considering the pathogenesis of glaucoma, as the peak circadian IOP in healthy eyes has been reported to occur at approximately 4.00 am[17] or 5.30 am.[18] In addition, optic nerve head perfusion has been reported to be significantly reduced at night in normal subjects,[19] which may be related to our findings regarding visual field sensitivity at 3.30 am.

The consistency of results for both eyes at 3.30 am raises the possibility of circadian changes in visual field parameters that should be assessed using a larger cohort of subjects or over a larger number of test sessions. It is suggested that assessment initially be carried out during normal expected appointment times (i.e. 8.00 am to 6.30 pm), in order to facilitate subject enrolment and develop clinically meaningful values for TRV at different times of day. A study encompassing nocturnal SAP testing could be carried out either concurrently or separately, depending on the availability of subjects for such a study.

References

1. Tassi P, Pellerin N, Moessinger M, Hoeft A, Muzet A. Visual resolution in humans fluctuates over the 24h period. Chronobiol Int 2000;17:187-195.
2. Tuunainen A, Kripke DF, Cress AC, Youngstedt SD. Retinal circadian rhythms in humans. Chronobiol Int 2001;18:957-971.
3. Lavoie J, Gagne AM, Lavoie MP, Sasseville A, Charron MC, Hebert M. Circadian variation in the electroretinogram and the presence of central melatonin. Doc Ophthalmol 2010;120:265-272.
4. Danilenko KV, Plisov IL, Cooper HM, Wirz-Justice A, Hebert M. Human cone light sensitivity and melatonin rhythms following 24-hour continuous illumination. Chronobiol Int 2011;28:407-414.
5. Do MT, Yau KW. Intrinsically photosensitive retinal ganglion cells. Physiol Rev 2010;90:1547-1581.
6. Zele AJ, Feigl B, Smith SS, Markwell EL. The circadian response of intrinsically photosensitive retinal ganglion cells. PLoS One 2011;6:e17860.
7. Feigl B, Mattes D, Thomas R, Zele AJ. Intrinsically photosensitive (melanopsin) retinal ganglion cell function in glaucoma. Invest Ophthalmol Vis Sci 2011;52:4362-4367.
8. Artes PH, Nicolela MT, McCormick TA, LeBlanc RP, Chauhan BC. Effects of blur and repeated testing on sensitivity estimates with frequency doubling perimetry. Invest Ophthalmol Vis Sci 2003;44:646-652.
9. Barkana Y, Gerber Y, Mora R, Liebmann JM, Ritch R. Effect of eye testing order on automated perimetry results using the Swedish Interactive Threshold Algorithm standard 24-2. Arch Ophthalmol 2006;124:781-784.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-310.
11. Hanneman SK. Design, analysis, and interpretation of method-comparison studies. AACN Adv Crit Care 2008;19:223-234.
12. Reyment RJ, K. Applied factor analysis in the natural sciences. Cambridge University Press 1996.
13. Heaton KJ, Maule AL, Maruta J, Kryskow EM, Ghajar J. Attention and visual tracking degradation during acute sleep deprivation in a military sample. Aviat Space Environ Med 2014;85:497-503.
14. Valdez PRCG, Aida; Talamantes Javier; Armijo, Pablo & Borrani, Jorge. Circadian rhythms in components of attention. Biological Rhythm Research 2005;36:57-65.
15. Fimm B, Brand T, Spijkers W. Time-of-day variation of visuo-spatial attention. Br J Psychol 2016;107:299-321.
16. Pearce JG, Maddess T. Retest Variability in the Medmont M700 Automated Perimeter. Optom Vis Sci 2016;93:272-280.
17. Agnifili L, Mastropasqua R, Frezzotti P, et al. Circadian intraocular pressure patterns in healthy subjects, primary open angle and normal tension glaucoma patients with a contact lens sensor. Acta Ophthalmol 2015;93:e14-21.
18. Liu JH, Kripke DF, Hoffman RE, et al. Nocturnal elevation of intraocular pressure in young adults. Invest Ophthalmol Vis Sci 1998;39:2707-2712.
19. Osusky R, Rohr P, Schotzau A, Flammer J. Nocturnal dip in the optic nerve head perfusion. Jpn J Ophthalmol 2000;44:128-131.

Appendix 2

Circadian variation in optic nerve head topography and peripapillary retinal nerve fibre layer thickness measurements using spectral domain optical coherence tomography

Abstract

Background Clinical assessment of the optic nerve head and retinal nerve fibre layer now involves the use of a range of objective measures, such as spectral domain optical coherence tomography (SD-OCT). At present, it is not known whether measurement variability might be influenced by intrinsic circadian rhythms present in the eye itself. The object of this study is to determine whether any variability in the measurement of ocular parameters exists over a twenty four hour period, and whether the amount of change is statistically significant.

Methods One subject was recruited for this study, a 52 year old male. 3D disc scans using the Topcon 3D-OCT 2000 SD-OCT were carried out on both eyes every one and a half hours over a twenty four hour period (16 scans/eye/day). The testing was repeated six months later to obtain two sets of results for comparative purposes, comprising 64 repeat measurements.

Results Circadian variation was consistently found in some pRNFL parameters in one eye, but not in the other eye, while circadian disc topography changes were noted in both eyes. Factor analysis was used to determine that three (as yet undefined) factors were responsible for 71.8% of the variability found in this study.

Conclusions Circadian variation in optic nerve head and pRNFL parameters may exist, but the amount of RNFL variation was inconsistent between eyes and was not found to have any significant influence on results obtained during normal office hours. The circadian variability found in some parameters may have implications for the initiation and progression of glaucomatous optic

neuropathy, and further studies with glaucoma patients may reveal greater levels of circadian variation in those patients.

Introduction

The repeatability of optic disc topographical measurements and peripapillary retinal nerve fibre layer (pRNFL) measurements with spectral domain optical coherence tomography (SD-OCT) has generally been assessed during normal clinic hours in most studies. At present, it is not known whether there is a circadian component to the variability encountered with SD-OCT measurements at other times of day outside normal office hours. The presence of any circadian variations to optic nerve head topography and the pRNFL may have implications in the onset and progression of glaucomatous optic neuropathy.

Diurnal fluctuations have been found to occur in ocular parameters such as intra-ocular pressure (IOP).[1] Although a more recent study reported that healthy subjects do not produce a sustained and repeatable diurnal IOP pattern in the short term,[2] this result was not confirmed by later studies which found that healthy subjects did exhibit significant diurnal variation in IOP.[3, 4] Diurnal variations occur in corneal thickness,[5] and although changes to corneal biomechanical properties such as corneal hysteresis have not been found,[6] there have been reports of refractive instability associated with diurnal variation in patients who have undergone radial keratotomy.[7]

Retinal thickness has not been found to have any diurnal variation in healthy patients using macular scan protocols with SD-OCT,[8, 9] although diurnal variation in the amount of macular oedema present in patients with central retinal vein occlusion has been documented using SD-OCT,[10] as has significant

24 hour variation in macular thickness and volume measurements in patients with diabetic macular oedema.[11] Choroidal thickness has been found to demonstrate significant diurnal variation in thickness using SD-OCT in healthy patients, and was correlated with systolic blood pressure.[12] Diurnal variations in optic disc topography have been noted in cases of extremely high variations in IOP,[13] and have also been found in healthy patients,[14] which may be related to diurnal fluctuations in optic nerve head perfusion[15] and ocular blood flow parameters.[16] The influence of diurnal variations in the eyes' vascular system has also been documented regarding the incidence of branch retinal vein occlusion, with the most common period being between 6.00 am and noon.[17]

At present, no studies have been carried out on circadian variation in pRNFL thickness, and studies on diurnal macular thickness variation have only been carried out over shorter periods such as ten hours,[9] or only having patients tested at two time points such as 8.00 am and 6.00 pm.[8] Similarly, the investigation of variations in optic nerve head topography has only been carried out over a maximum of 15 hours.[14] Although these studies do encompass normal office hours testing times, they have not addressed the possibility of an underlying circadian cycle in other SD-OCT parameters which might occur over a full 24 hour period.

As described in Appendix 1, in order to isolate machine and circadian variance from other extraneous sources of variance, only one subject and one operator were used during both test sessions. By eliminating between subject variance, we will thus be able to assess the effect of any circadian influences occurring in both eyes, and also identify any between-eye differences.

The purpose of this study is to investigate whether any circadian cycles might influence SD-OCT measurements and contribute to *test-retest variability* (TRV) As a complement to an earlier study,[18] where good baseline measurements of RNFL and optic nerve head topography variability for healthy patients were obtained for the Topcon 3D- OCT 2000, these results will be used to determine whether any significant circadian variations may occur with SD-OCT measurements in clinical practice.

Methods

One subject, (JP, a 51 year old male) was enrolled in the study with the appropriate consent form signed under ANU/ACT Health protocol 7/07/.667. At the time of testing, the subject had no ocular pathology, including no glaucomatous changes to his optic nerve head. Intra-ocular pressures were R 12 mmHg and L 13 mmHg, with central corneal thicknesses of R 525.5 μm and L 524 μm. Standard automated perimetry results were within normal limits, and pRNFL and GCC scans using the Topcon 3D-OCT 2000 prior to this study have shown no abnormal results.

JP had been a subject in a previous study,[18] where a total of eight SD-OCT scans were performed on each eye (two scans per eye at each visit, each visit was one week apart and all scans performed at approximately the same time each day) using the Topcon 3D-OCT 2000. The TRV results from this study were then compared to the 24 hour TRV results from the current study. The same SD-OCT operator was used in both of these studies, to eliminate the possibility of any inter-operator error.

The instrument used was the Topcon 3D-OCT 2000, which has a light source consisting of a super luminescent diode (wavelength 840 nm). The A-scan speed is 27,000 A-scans per second, with the Topcon 3D disc scan being carried out over a 6.0 x 6.0 mm square scan area. The 3D disc scan comprises 128 horizontal B-scans, with 512 A-scans per B-scan, with a scan depth of 2.4 mm. The disc centre is located, and RNFL thickness information is obtained from a 3.4 mm diameter circle created around the disc centre. The disc topography information is extracted from the information in the original cube scan.

The scans were carried out every one and a half hours, commencing at 8.00 am in the morning, and continuing throughout the night until the last scan at 6.30 am, providing 16 scans/eye/day. During the day, normal activities such as eating, shopping and any other tasks were carried out as usual. After the 9.30 pm scan, the subject attempted to (or did) sleep between scans whilst lying down.

The subjects' general health and medications did not change between sessions. The subject took all his medications at the same time he normally does (at 6.30 am) on the day the testing commenced. The prescribed medications were Lipitor (Atorvastatin) 40 mg, Perindo (Perindopril erbumine) 4 mg, and the subject also took his daily non-prescribed multi-vitamin (Cenovis Multivitamins and Minerals). The medications were not taken again until after the testing had finished at 6.30 am the next day.

The testing was done on two separate occasions, the first session being done on 23/24th March using version 7.11 software on the Topcon 3D-OCT 2000 and the second session was done on the following 23/24th October using Version

8.00 software. The Version 7.11 results were re-analysed using Version 8.00. The times of year were consistent for length of daylight hours, to reduce the influence of any circadian rhythms associated with the length of day.

As in the previous Appendix (Appendix 1: Circadian variation in visual field test indices during a 24 hour period) two statistical methods were used to analyse the data. Bland-Altman difference plots, as described in Appendix 1 and presented in Chapters 1, 2, and 3, were used to determine the 95% confidence intervals when assessing TRV. Bland-Altman analysis was carried out using MedCalc v.16.4.3 (MedCalc Software, bvba Ostend, Belgium).

Factor analysis, as described in Appendix 1, was also carried out on all the repeated data in order to provide an assumption free examination of the major sources of variance in the data, and the extent to which they might display some circadian rhythm(s). The factor analyses were done in MATLAB version 2014a (The MathWorks Inc., Natick, Massachusetts, United States).

By carrying out measurements at 16 time points per day for each OCT variable, we were able to obtain information about rhythms in the data for up to 8 cycles per day. Therefore, rather than just looking at the weighted average of the variables, we also examined fits of sinusoidal circadian rhythms to each of the individual OCT measures. To do this, we fitted multiple regression models of the form

1) $V(t) = C + S(t) + \cos(2\pi f(t)) + \sin(2\pi f(t)) + \cos(2\times 2\pi f(t)) + \sin(2\times \pi f(t)) + \cos(3\times 2\pi f(t)) + \sin(3\times 2\pi f(t))$

where C is a constant (the mean level of a variable), S(t) is a constant slope, and n×2πf(t) are harmonic frequencies at n = 1, 2, or 3 cycles/day. The cosine and sine pairs for each frequency allow arbitrary phases (delays) to be fitted.

For each variable, we used more and less complex models, with the less complex models involving subsets of the elements of Eq. 1, and compared the models using F-change statistics. In this way, we found the most parsimonious models that explained the temporal variation of individual variables over 24 hours. Generally, these models contained a constant and terms for 1 and 2 cycles per day. Of the OCT variables showing significant rhythmical effects, most were small in amplitude compared to their mean value over the day, so we compared the rhythmical effects for different OCT variables in terms of their percentage change over the day. The multiple regression analyses were done in Matlab ver 2014a (The MathWorks Inc., Natick, Massachusetts, United States).

Results

Table 1.

Bland-Altman 95% confidence intervals (CI_{95}) obtained from Bland-Altman plots for Topcon optic nerve head and retinal nerve fibre layer parameters. The 24 hour CI_{95} were obtained from the two 24 hour sessions (64 scans from both 24 hour sessions) in the current study. Data from a previous study[18] are presented for comparison. Those Topcon CI_{95} values were obtained from 32 scans (two subjects, two scans on each eye at each visit, on four occasions, one week apart at the same time of day). HCDR = horizontal cup/disc ratio, VCDR = vertical cup disc ratio, pRNFL = *peripapillary retinal nerve fibre layer*. RNFL values are in micrometres.

	Topcon CI_{95} n = 32 (Pearce and Maddess)[18]	24 Hour overall CI_{95} n = 64	24 Hour day CI_{95} n = 32	24 Hour night CI_{95} n = 32	Paired samples (day cf. night) t-test p values
Disc area (mm^2)	0.36	0.49	0.42	0.55	0.47
Rim area (mm^2)	0.24	0.39	0.31	0.46	0.71
HCDR	0.07	0.05	0.04	0.05	0.27
VCDR	0.08	0.07	0.07	0.07	0.28
RNFL Quadrants					
Superior	8.3	10.8	10.6	11.0	0.04
Nasal	15.3	19.1	21.8	16.4	0.16
Inferior	11.1	9.8	9.7	8.7	0.02
Temporal	9.8	10.9	10.4	11.7	0.85
Average pRNFL	4.9	6.6	6.5	6.8	0.01

Table 1 shows that overall TRV for RNFL and ONH parameters were similar in the current and previous study.[18] From the current study, daytime CI_{95} values did not differ greatly from the night-time values. Paired sample t-tests showed a significant difference in the mean values for daytime versus night-time results for average pRNFL and superior and inferior quadrants. In all three instances, the mean night-time thickness was lower than that of the daytime thickness by approximately 2 μm.

Given that multiple parameters were assessed in Table 1, a full Bonferroni correction was applied to the p values. Assuming substantial correlation between eyes, this correction left only average pRNFL as significant.

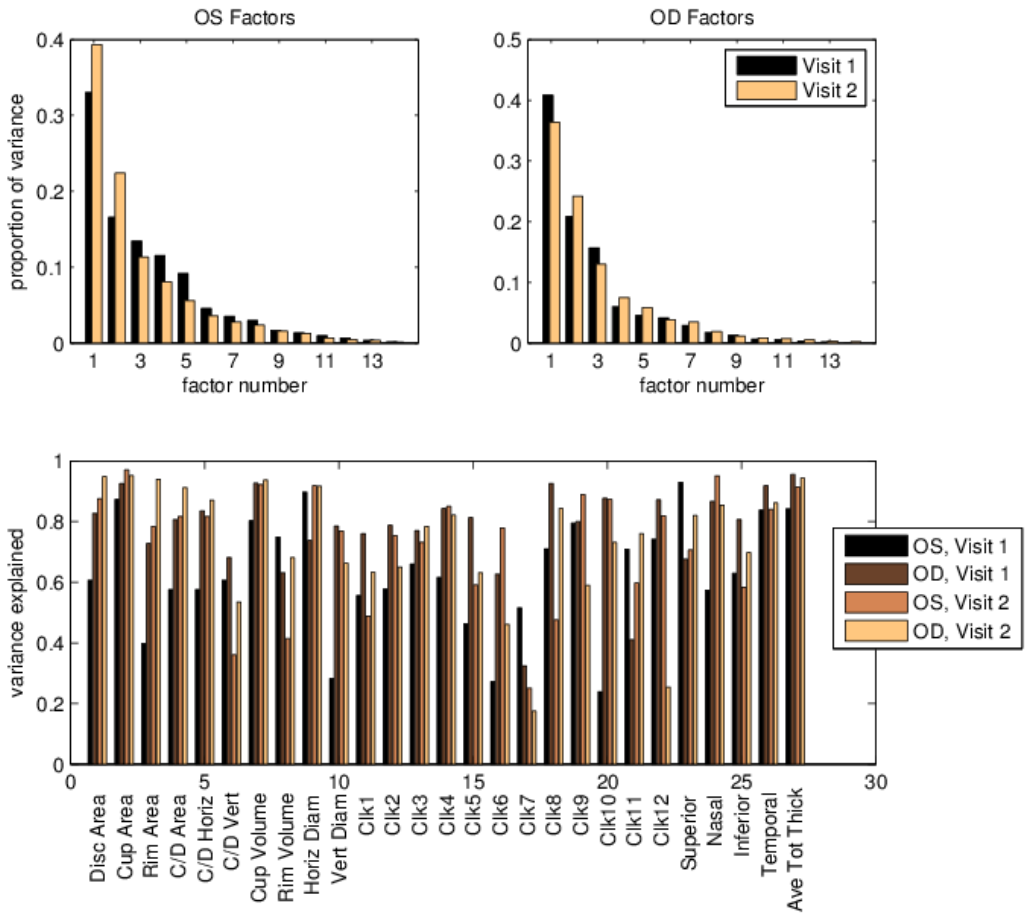


Figure 1.

Scree plots showing the proportion of variance due to each factor number and the proportion of variance explained for each variable (the communalities) using a three factor model.

Factor analysis is a statistical method used to explain the variability of the observed variables in terms of underlying unknown variables, which are known as factors. Here, the objective is discover if there is a small number of factors explaining most of the data, and then to examine the time evolution of those factors. This provides an assumption free assessment of diurnal variation.

Communalities represent the amount of explained variance in each variable under consideration. Using all factors, each variable has a communality of 1, with a reduction in factors used resulting in a value less than 1. The closer the

communalities are to 1 using a reduced number of factors is an indication of how well the retained factors explain the original data.[19] Also, when using a small number of factors, it is desirable that the communalities are uniformly high, indicating that the factors represent all the measured variables well.

From Figure 1, it can be seen that three (as yet undefined) factors are sufficient to explain 71.8% of the variability in the results obtained, and the median proportion of each variable explained by the three combined factors, the communalities, is 77.0%, with only clock hour 7 not consistently explained.

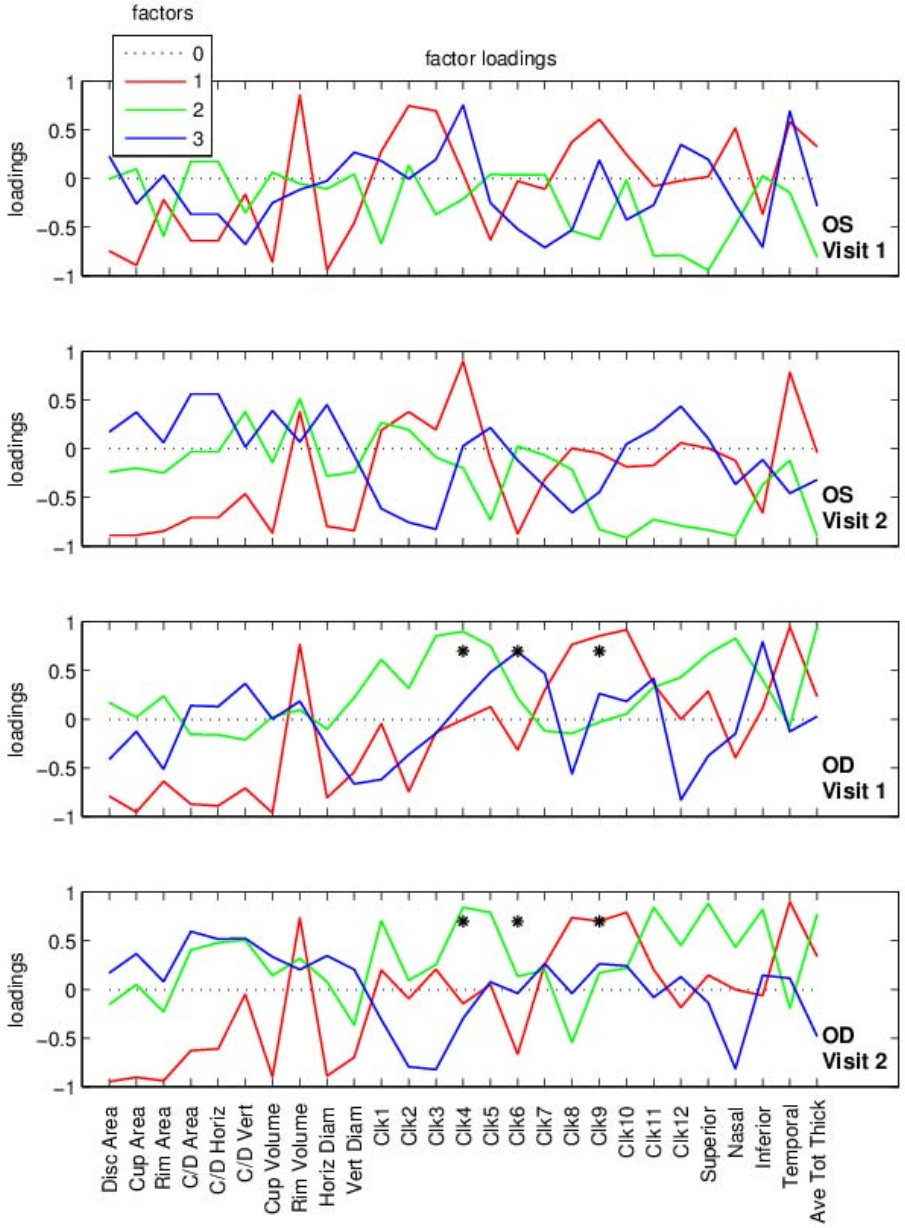


Figure 2.

Factor loading graphs for each eye, showing the weights between the variables and the three factors at visit one and visit two. The asterisks on the OD graphs represent clock hours 4, 6 and 9 and highlight the consistent loading for factors 1, 2 and 3 for those parameters. The first 10 loadings are very consistent across visits and eyes, especially for Factor 1, which accounts for the single largest proportion of variance (Fig. 1). The dotted line in each panel (0 in the legend) corresponds to a weight of 0.

Factor loadings are the regression coefficients between each factor and the original data, and represent how much a given factor explains each variable involved in the original factor analysis. Factor loadings are given a relative weight between ± 1. The closer the factor loading to either + or – 1, the stronger the influence of the factor on the variable. A negative factor loading indicates an inverse relationship between the factor and the variable, while a value of zero indicates no influence of the factor on the variable.

Disc parameters (Fig. 2) show some highly consistent loadings in both eyes at both visits (e.g. rim volume) with Factor 1, which accounts for around 40% of the variance in both eyes at both visits (Fig 1, top panels). Other disc parameters related to rim volume, such as rim area, cup volume, horizontal and vertical disc diameter and vertical C/D ratio (Fig. 2) also show very consistent loading for Factor 1 across eyes and visits.

RNFL parameters demonstrated much less consistency across factor loadings, with only the temporal quadrant showing a consistent loading in both eyes at both visits. Most other parameters showed inconsistency not only between eyes, but also between visits for each eye. Clock hours 4, 6 and 9 showed a high level of consistency for factors 1 and 2 between visits for the right eye (* Fig. 2), but this finding was not replicated in the left eye.

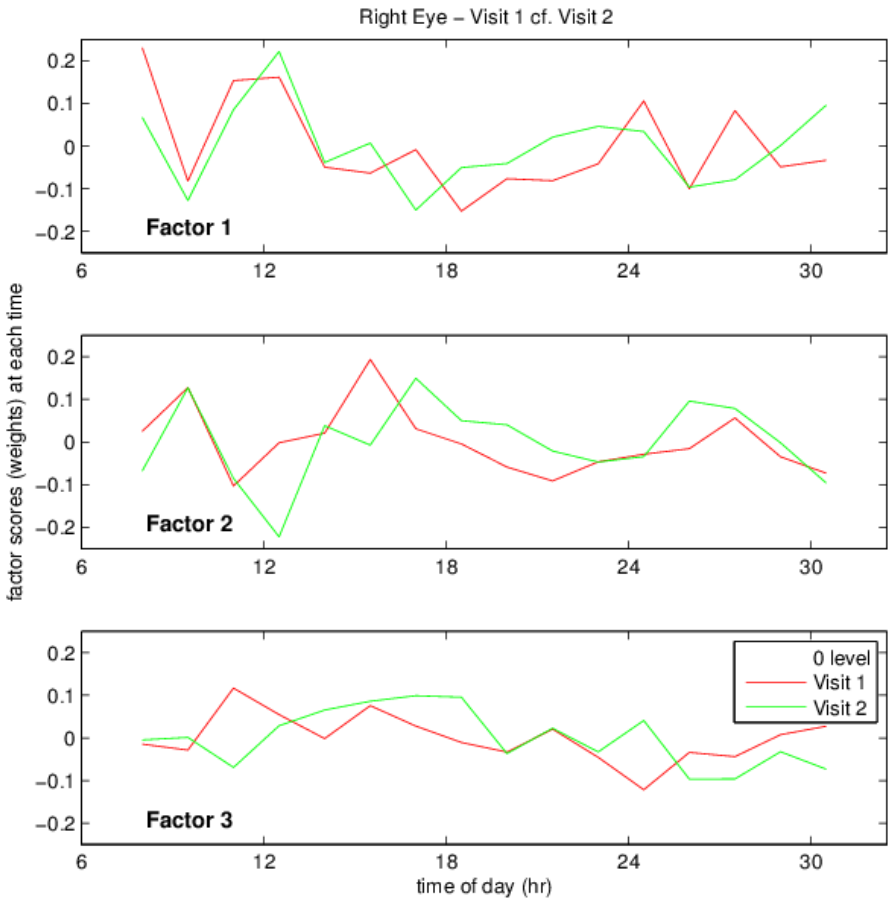


Figure 3.

Factor scores for factors 1, 2 and 3 for the right eye, visit 1 versus visit 2. The graphs for factor scores for the left eye have not been included as none of the factor score correlations were significant for the left eye. Factor scores 1 and 2 appear to have the greatest variation overnight.

In the present example the factor scores represent the influence of a given factor over time. The scores thus represent an assumption free model of the natural temporal evolution of the main components of variance in the OCT data. In the right eye, factor scores were reasonably consistent for factors 1, 2 and 3 (Fig. 3). Factor scores for the left eye (not shown) were not, emphasising the variability inherent in the investigation of physiological processes in living organisms.

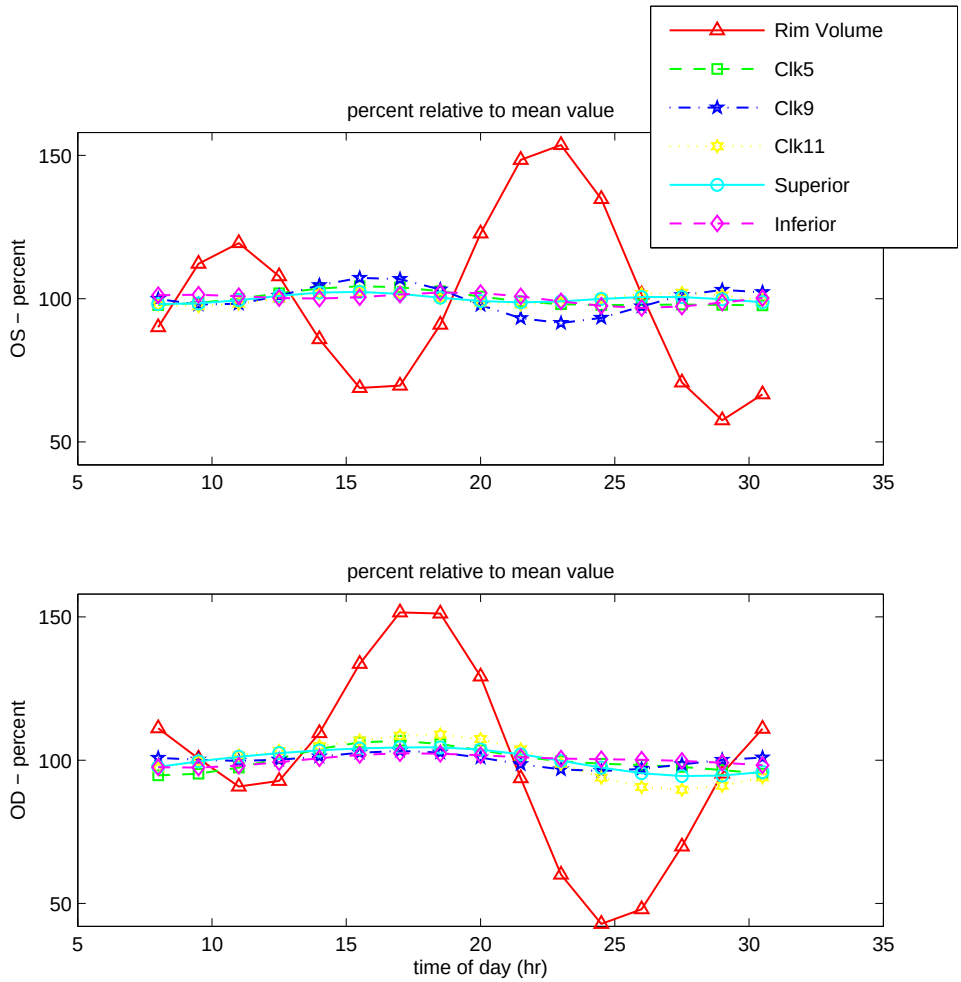


Figure 4.

Data fitted with 1 or 2 cycles/24 hour period, showing the change in percentage terms (relative to the mean) for those parameters showing significant ($p < 0.05$) variation in percentage terms relative to the mean.

An advantage of having 16 time points per day meant that in principle, we could fit sinusoidal variations up to rates of 8 cycles/day. We examined models with up to 4 cycles/day, but only found significant fits for a limited number of variables, and then only using 1 and 2 cycles/day. Fitting combinations of 1 and 2 cycles/day to the data (Fig. 4) demonstrated a large cyclical variation for rim volume ($p < 0.001$), although the fitted waveform for each eye was almost perfectly out of phase with the other. Since this asymmetry is unlikely to be a

physiological phenomenon, it appears that the majority of the effect is due to test-test variability. Other parameters were found to have significant cyclical variations ($p < 0.05$), but in percentage terms, these rhythms were relatively small.

Discussion

Due to the nature of the study protocol, we were unable to measure IOP's at the same time that visual fields and OCT scans were carried out. Whether or not this measurement would have improved or confused our findings is debateable, given the disagreement as to whether detectable circadian rhythms in IOP actually exist.

Evidence for the existence of consistent circadian rhythms in intra-ocular pressure (IOP) seems conflicting at best. Kida et al.[1] reported IOP to be higher during the nocturnal period in healthy subjects, and Mosaed et al.[20] reported the same finding for both younger (21.6 ± 2.0) and older (58.1 ± 7.6) healthy patients as well as untreated glaucoma patients (60.2 ± 12.0). However, a later study by Realini et al.[2] reported that the eyes of healthy subjects did not produce a sustained and repeatable diurnal IOP pattern.

In glaucoma, Romanet et al.[21] reported the 24 hour IOP rhythm to be reversed in glaucoma subjects, with IOP values higher in the day than in the night. Renard et al. [22] reported a similar finding, noting that although a nyctohemeral rhythm existed in most normal tension glaucoma subjects, the acrophase occurred diurnally in 54.5% of subjects, nocturnally in 36.4% and 9.1% had no nyctohemeral rhythm. Again, in contrast, a later study by Realini et al.[23] reported that, like the healthy subjects in their previous study,[2] treated primary

open angle glaucoma patients did not produce a sustained and repeatable diurnal IOP pattern.

In view of the foregoing, quantifying any changes in structure or function in relation to IOP may be problematic.

Retinal nerve fibre layer

Comparison of daytime and night-time mean RNFL thicknesses revealed significant differences in the superior and inferior quadrants and average pRNFL (Table 1). The lower nocturnal thicknesses may indicate an alteration in some component, or components, of the RNFL, the clinical implications of which remains unclear at present. The detection of circadian changes in pRNFL parameters may also have practical applications in assessing SD-OCT TRV.

Several RNFL parameters demonstrated significant ($p< 0.05$) cyclical variation, including superior and inferior quadrant pRNFL (Fig. 4). Although the percentage change in the cycles for these two parameters was relatively small, there is at least some consistency in circadian variation for these parameters in mean thickness (Table 1) and cyclical variation (Fig. 4).

Factor analysis revealed that three as yet undefined underlying factors accounted for a large proportion of the variance (71.8%) in the data (Fig.1), with most parameters under consideration consistently explained. Whilst Factor 1 has not yet been identified, the presence of a single factor which contributes so much to disc topography and pRNFL circadian variation may be an important driver in the onset and progression of glaucomatous optic neuropathy.

In a previous study,[18] in three SD-OCT's (Optovue RTVue-100, Nidek RS-3000 and Zeiss Cirrus HD-OCT), the temporal pRNFL quadrant was found to have

the lowest TRV out of all the quadrants for three instruments, and was the second least variable quadrant for the Topcon when analysing scans carried out at the same time of day. The repeatability of temporal quadrant measurements with SD-OCT strengthens the findings of a consistent effect of Factor 1 in this parameter (Fig. 2).

The right eye had much more consistent factor loadings for Factors 1 and 2 in RNFL results, particularly at clock hours 4, 6 and 9 as highlighted by the asterisks in Figure 2. Clock hour 4 was shown to be highly affected by factor 3, while clock hour 9 was strongly affected by factor 1. Clock hour six was affected by factor 1, but to a lesser extent than clock hour 9. Each eye had differing amounts of consistency between factor loadings at both visits (Fig. 2), and factor scores (Fig. 3) (the importance of each factor at each time point) were also more consistent for the right eye, but not the left.

The repeatable findings of cyclical variation in factor loadings for some disc and RNFL parameters indicate that there may be changes to the ONH and RNFL associated with the time of day that testing is undertaken. The inconsistency between each eye in this regard is harder to explain. Instrument variability may play a part, but that variability exists equally for both eyes. The strategy of reducing subject variance in the current study has provided the advantage of exposing the lack of congruency between eyes.

Disc parameters

Optic nerve head topographical features did not show any obvious alterations to daytime TRV compared to night-time TRV, and in contrast to RNFL thickness,

no optic disc parameters demonstrated any significant circadian changes to their mean values (Table 1).

Disc parameters demonstrated reasonably consistent factor loadings (Fig. 2), both between eyes and between visits. These loadings were also quite often at high levels, indicating a strong underlying factorial influence on disc topography. Given the dynamic nature of the optic nerve head due to its extensive vasculature, changes to optic nerve head perfusion may play some role in these underlying factors.

Rim volume demonstrated the largest cyclical variation in percentage terms (Fig. 4). Although each eye appears to be out of phase with the other in Fig. 4, factor loadings (Fig. 2) showed Factor 1 was able to consistently strongly explain the variance for rim volume in both eyes at both visits. The apparent discrepancy between the opposing cycles in Fig. 4 and the consistent factor loadings in Fig. 2 defies explanation at this point in time, but further studies may shed some light on the circadian influences on this parameter.

The current study has been able to demonstrate the presence of circadian variations in several pRNFL thickness parameters. Factor analysis has also been able to show that circadian rhythms may be present in some pRNFL and ONH parameters. The clinical importance of these findings remains to be seen, but their initial detection is an important starting point into investigating the possible role of circadian rhythms in glaucoma onset and progression, as well as any implications for TRV using SD-OCT.

References

1. Kida T, Liu JH, Weinreb RN. Effect of 24-hour corneal biomechanical changes on intraocular pressure measurement. Invest Ophthalmol Vis Sci 2006;47:4422-4426.
2. Realini T, Weinreb RN, Wisniewski SR. Diurnal intraocular pressure patterns are not repeatable in the short term in healthy individuals. Ophthalmology 2010;117:1700-1704.
3. Read SA, Collins MJ, Iskander DR. Diurnal variation of axial length, intraocular pressure, and anterior eye biometrics. Invest Ophthalmol Vis Sci 2008;49:2911-2918.
4. Chakraborty R, Read SA, Collins MJ. Diurnal variations in axial length, choroidal thickness, intraocular pressure, and ocular biometrics. Invest Ophthalmol Vis Sci 2011;52:5121-5129.
5. Giraldez-Fernandez MJ, Diaz-Rey A, Garcia-Resua C, Yebra-Pimentel-Vilar E. [Diurnal variations of central and paracentral corneal thickness and curvature]. Arch Soc Esp Oftalmol 2008;83:183-191.
6. Shen M, Wang J, Qu J, et al. Diurnal variation of ocular hysteresis, corneal thickness, and intraocular pressure. Optom Vis Sci 2008;85:1185-1192.
7. Koppen C, Gobin L, Tassignon MJ. Intacs to stabilize diurnal variation in refraction after radial keratotomy. J Cataract Refract Surg 2007;33:2138-2141.
8. Jo YJ, Heo DW, Shin YI, Kim JY. Diurnal variation of retina thickness measured with time domain and spectral domain optical coherence tomography in healthy subjects. Invest Ophthalmol Vis Sci 2011;52:6497-6500.
9. Read SA, Collins MJ, Alonso-Caneiro D. Diurnal variation of retinal thickness with spectral domain OCT. Optom Vis Sci 2012;89:611-619.
10. Gupta B, Grewal J, Adewoyin T, Pelosini L, Williamson TH. Diurnal variation of macular oedema in CRVO: prospective study. Graefes Arch Clin Exp Ophthalmol 2009;247:593-596.
11. Kotsidis ST, Lake SS, Alexandridis AD, Ziakas NG, Ekonomidis PK. 24-Hour variation of optical coherence tomography-measured retinal thickness in diabetic macular edema. Eur J Ophthalmol 2012;22:785-791.
12. Tan CS, Ouyang Y, Ruiz H, Sadda SR. Diurnal variation of choroidal thickness in normal, healthy subjects measured by spectral domain optical coherence tomography. Invest Ophthalmol Vis Sci 2012;53:261-266.
13. Lee BL, Zangwill L, Weinreb RN. Change in optic disc topography associated with diurnal variation in intraocular pressure. J Glaucoma 1999;8:221-223.
14. Sehi M, Flanagan JG, Zeng L, Cook RJ, Trope GE. The association between diurnal variation of optic nerve head topography and intraocular pressure and ocular perfusion pressure in untreated primary open-angle glaucoma. J Glaucoma 2011;20:44-50.
15. Osusky R, Rohr P, Schotzau A, Flammer J. Nocturnal dip in the optic nerve head perfusion. Jpn J Ophthalmol 2000;44:128-131.
16. Pemp B, Georgopoulos M, Vass C, et al. Diurnal fluctuation of ocular blood flow parameters in patients with primary open-angle glaucoma and healthy subjects. Br J Ophthalmol 2009;93:486-491.
17. Oh J, Oh IK, Huh K. Diurnal variation of the incidence of symptomatic branch retinal vein occlusion. Ophthalmologica 2007;221:251-254.

18. Pearce JG, Maddess T. Comparing the peripapillary retinal nerve fibre layer inter-visit test-retest variability of four different spectral domain optical coherence tomographs Chapter 6, Part 1 2016;PhD Thesis, Austarlian National University
19. Field A, Miles J, Field Z. Discovering statistics using R. Sage Publications Ltd 2012.
20. Mosaed S, Liu JH, Weinreb RN. Correlation between office and peak nocturnal intraocular pressures in healthy subjects and glaucoma patients. Am J Ophthalmol 2005;139:320-324.
21. Romanet JP, Maurent-Palombi K, Noel C, et al. [Nyctohemeral variations in intraocular pressure]. J Fr Ophtalmol 2004;27 Spec No 2:2S19-12S26.
22. Renard E, Palombi K, Gronfier C, et al. Twenty-four hour (Nyctohemeral) rhythm of intraocular pressure and ocular perfusion pressure in normal-tension glaucoma. Invest Ophthalmol Vis Sci 2010;51:882-889.
23. Realini T, Weinreb RN, Wisniewski S. Short-term repeatability of diurnal intraocular pressure patterns in glaucomatous individuals. Ophthalmology 2011;118:47-51.

Appendix 3: Part 1 (3.1)

Comparing peripapillary retinal nerve fibre layer inter-visit test-retest variability of four spectral domain optical coherence tomographs

Abstract

Background To investigate the inter-visit test-retest variability (TRV) of four different spectral domain optical coherence tomographs (SD-OCT's) while isolating machine variance from inter-subject, circadian and other contributions to TRV. We develop event based criteria for each instrument to enable the differentiation of TRV from true progression.

Methods We investigated the Topcon 3D OCT- 2000, Nidek RS-3000, Optovue RTVue-100 and Zeiss Cirrus HD-OCT. Contributions from circadian and inter-subject variance were minimised by performing repeated tests at the same time of day. There were four test sessions, one week apart over four consecutive weeks, using two well characterised normal subjects. Two scans on each eye of each subject at each session resulted in a total of 32 scans per SD-OCT.

Results: Bland-Altman analysis of average pRNFL thickness yielded CI_{95} of 4.4 μm for the Cirrus, 4.9 μm for the Topcon, 5.4 μm for the Optovue, and 9.5 μm for the Nidek. Differences in average pRNFL thicknesses between instruments were similar to those obtained from the normative databases and/or the manufacturers FDA 510(k) data. Most OCT measurements were significantly correlated across the instruments. Inter-ocular asymmetry results differed between individuals and between instruments.

Conclusions: The TRV for each instrument was less than 10% of measured average pRNFL. Between-instrument confidence intervals and mean differences between instruments are required to facilitate the interpretation of measurements obtained from patients tested on different instruments. Inter-ocular asymmetry diagnostic criteria applicable to one instrument may not be transferable to other instruments.

Introduction

When using the older time domain Stratus OCT (Carl Zeiss, Meditec Inc), the single best parameter selected from either retinal nerve fibre layer (RNFL) or optic nerve head (ONH) measurements both showed greater diagnostic power than subjective assessment of the ONH by general ophthalmologists.[1] More recently, potentially more accurate spectral domain optical coherence tomography (SD-OCT) has become an integral procedure in the detection and monitoring of glaucomatous optic neuropathy.

The differentiation of true progression from test-retest variability (TRV) may be compromised in newer instruments, because there may be little data available as to what constitutes statistically significant change in the peripapillary retinal nerve fibre layer (pRNFL). It is important to establish the level of machine variance present when carrying out repeat measurements in order to be able determine whether true progression has occurred.

The total variance recorded form any instrument may incorporate many independent factors, such as individual subject variance and diurnal and seasonal factors. These factors can be summarised as:

$$(1)\ \sigma^2_{total} = \sigma^2_{retest} + \sigma^2_{subject\ (n)} + \sigma^2_{operator\ (m)} + \sigma^2_{diurnal} + \sigma^2_{annual} + \sigma^2_{error}.$$

The terms $\sigma^2_{subject\ (n)}$ and $\sigma^2_{operator\ (m)}$ are the sums of variance from the *n* subjects studied and the *m* operators used. The current study has been designed to isolate machine variance from other possible causes of variance by eliminating diurnal and seasonal factors, as well as reducing subject variance by limiting the sample size to two subjects. The two subjects were very well characterised before testing. The total variance therefore can now be expressed as:

(2) $\sigma^2_{total} = \sigma^2_{retest} + \sigma^2_{subject\ (n = 2)} + \sigma^2_{operator\ (n = 1\ or\ 2)} + \sigma^2_{error}$

To generate clinically meaningful values for TRV, we undertook a relatively large number of scans (32 per OCT) over a four week period, at the same time of day, to create inter-visit values, which are more clinically relevant than intra-visit values. The alternative would be to test a very large number of subjects at many times, operators and seasons, and then fit a multivariate model to attempt to disentangle the sources of variance. The TRV data here were compared with the instruments normative databases and/or the manufacturers FDA 510(k) data, and a study examining the same four OCTs in diverse patients at various times and seasons.[2]

We determine the inter-visit levels of pRNFL measurement variability in four SD-OCT devices (Topcon 3D OCT-2000, Optovue RTVue-100, Nidek RS-3000 and Zeiss Cirrus HD-OCT), and present these values as 95% confidence intervals derived from Bland-Altman plots. These event based criteria are intended to be easily applied in clinical practice to determine statistically significant change from baseline pRNFL thickness.

As pRNFL values have been reported to be not interchangeable between instruments,[3] it may also be the case that diagnostic criteria established for one instrument may not be applicable to other instruments. We therefore also examine the inter-ocular asymmetry results from these instruments to determine the level of agreement in this parameter between instruments.

Methods

In order to isolate instrument variability from subject variability, the data set was restricted to two healthy subjects. A 51 year old male (JGP) and a 15 year old female (MEP) were recruited for the study, with the appropriate information and consent forms signed under ANU/ACT Health protocol 7/07.667. Refractive errors ranged between plano and +1.25 D of hyperopia, with up to 0.25 D of astigmatism. They had no glaucomatous changes to their optic nerve heads, visual field tests for both patients using the Medmont M700 (Version 3.9.7) automated perimeter (Medmont P/L, Nunawading, Victoria, Australia) demonstrated no visual field defects, and neither patient had any active ocular pathology.

JGP was recently found to be very similar to six age normal subjects, both in terms of mean results and TRV, in visual field tests repeated four times at 8.30 am, 1.30 pm and 5.30 pm.[4] JGP has also recently completed extensive 24 hour testing in SAP and SD-OCT,[5] indicating his diurnal variance was small. MEP was included here to span the effects of age, and was also found to have similar mean results and TRV in SAP and SD-OCT as six age normal subjects in a recent study.[4]

Each subject was tested on each OCT at approximately the same of day. The mean test times were 7.02 am (Standard Deviation (SD) 13 minutes) with the Topcon 3D-OCT 2000, 10.31 am (SD 18 minutes) with the Nidek RS-3000, 9.25 am (SD 43 minutes) with the Cirrus HD-OCT and 11.47 am (SD 53 minutes) for the Optovue RTVue-100.

It is important to note that each instrument was located in four different locations, necessitating travel time, and occasionally waiting time, between having scans carried out on different instruments on the same day. This test protocol was intended to replicate variations in patient presentations to clinical practice.

The protocol for each instrument was that the right eye and left eye for each subject were scanned once, and then scanned again on the same subject immediately afterwards, meaning that the subjects did not move away from the instruments in between any scans. Each session was carried out at intervals of one week for four consecutive weeks, giving a total of eight scans on each eye with each instrument.

All tests on the Optovue RTVue-100, Cirrus HD-OCT and Nidek RS-3000 were carried out by the same operator (that is, a different operator was used for each machine, but it was the same operator for each individual machine). With the Topcon 3D-OCT 2000, all scans done on MEP carried out by JGP, and the scans done on JGP were carried out by MEP.

None of the scans used in this study were manually modified, and all scans had an image/scan quality greater than 60 out of 100. The scan protocols and technical information for each OCT are summarised in Table 1.

The number of scans per point in the Nidek is adjustable, with the default setting being 512 A-scans per B-scan, with the number of B-scans being 128. The instrument in this study used 256 A-scans per B-scan, with 256-B scans, which is covered by the Nidek normative database.

Table 1.

Summary of the various scan protocols and technical information for each SD-OCT used in the current study.

	Topcon 3D OCT-2000	Nidek RS-3000	RTVue-100	Cirrus HD-OCT
Software	Version 7.11	Navis Ex 1.3.0.3	A5,1,0,90.	4.5.1.11
Scan name	3D Disc Scan	Disc Map Scan	RNFL 3.45	Optic Disc Cube Scan
Scan area	6.0 × 6.0 mm	6.0 × 6.0 mm	6.0 × 6.0 mm	6.0 × 6.0 mm
Number of B scans	128	256	101	400
Number of A scans	65,636	65,536	51,813	80,000
Number of operators	2	1	1	1

The RTVue-100 generates two pRNFL thickness measurements: the optic nerve head map protocol (NHM4), and the RNFL 3.45 which is used to generate the RNFL glaucoma report. The RNFL 3.45 uses direct scanning, while the NHM4 uses re-sampling from datasets, although there is no significant difference between the data obtained from both protocols.[6] The pRNFL data used in this study for the RTVue-100 were taken from the RNFL 3.45 results.

Statistical analysis in this study was carried out using Medcalc version 12.4 (Medcalc Software bvba, Ostend, Belgium).

Results

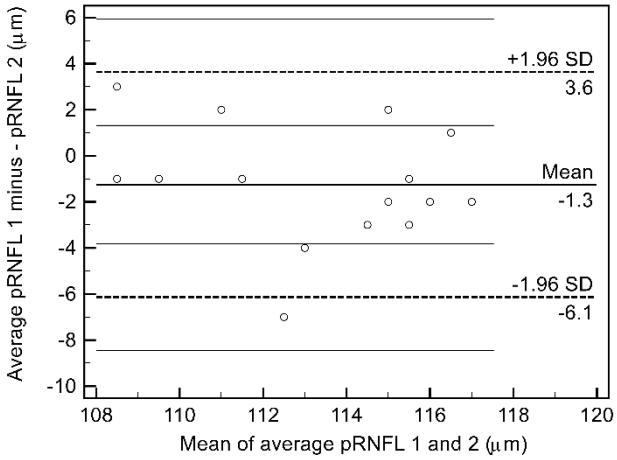


Figure 1.

Bland-Altman difference plot of average pRNFL values. The limits of agreement are -6.1 to +3.6 µm, the mean difference between visits is -1.3 µm and the 95% confidence interval is ± 4.9 µm.

Figure 1 provides an example of a Bland-Altman difference plot, which we use to determine the levels of TRV present in each instrument parameter. In the above example, the difference between each pair of average pRNFL measurements is plotted against the mean of the two measurements. The Limits of Agreement (LoA) are calculated as 1.96 × the Standard Deviation (SD) of the differences between visits, centred around the mean difference between visits. Since 95% of the differences between visits lie within the LoA, we therefore use the LoA to determine the 95% confidence intervals (CI_{95}) which we use to quantify TRV.

Table 2.

Bland-Altman inter-visit 95% confidence intervals (μm) for quadrant and average pRNFL thickness for each instrument ($n_{(scans)} = 32$)

	Superior Quadrant	Nasal Quadrant	Inferior Quadrant	Temporal Quadrant	Average pRNFL
Topcon	8.3	15.3	11.1	9.8	4.9
Nidek	19.2	36.1	30.6	15.3	9.5
RTVue	15.9	19.3	20.2	14.8	5.4
Cirrus	19.1	16.4	13.9	7.5	4.4

The Bland-Altman confidence intervals in Table 2 were derived by pairing the first right eye result and second right eye result at each visit, with the same being done for the left eye. These confidence intervals are therefore derived from a series of intra-visit results over the duration of the study to create an inter-visit value.

Although average pRNFL confidence intervals for three of the instruments were quite similar, the Nidek had a much larger confidence interval than the other three instruments. Quadrant intervals varied from 7.5 μm (Cirrus, temporal quadrant) to 36.1 μm (Nidek, nasal quadrant), with the temporal quadrant providing the least variability amongst three of the instruments.

Table 3.

Bland-Altman 95% confidence intervals (CI_{95}, in µm) and correlation coefficients (r) for average pRNFL and quadrant measurements. Av = average pRNFL, S = superior, N = nasal, I = inferior, T = temporal quadrants.

	Topcon					Nidek					RTVue				
	Av	S	N	I	T	Av	S	N	I	T	Av	S	N	I	T
Nidek															
CI_{95}	11.5	19.7	25.7	26.9	20.3										
r	0.66	**0.31**	0.59	0.75	0.77										
RTVue															
CI_{95}	11.3	16.5	21.4	18.2	18.4	10.1	19.9	31.6	32.2	19.9					
r	0.79	0.57	0.75	0.80	0.84	0.77	0.45	0.51	0.59	0.61					
Cirrus															
CI_{95}	7.9	14.3	15.6	21.5	12.5	8.8	18.6	30.2	29.1	18.8	7.8	18.5	24.1	18.3	14.6
r	0.74	0.75	0.74	0.79	0.97	0.81	0.57	**0.35**	0.69	0.71	0.87	0.56	0.68	0.84	0.82

Table 3 shows the LoA for each instrument when comparing measurements between instruments. Almost all correlations (not shown) were highly significant at the $p < 0.01$ value, with the exceptions of the Topcon – Nidek in the superior quadrant and the Cirrus – Nidek in the nasal quadrant.

The problem with applying LoA to measurements from different instruments is that they are dependent on the mean differences between instruments. The mean difference between instruments may vary greatly, depending on the information source. This variability in mean differences is outlined in Tables 4 and 5, where the normative databases and FDA 510(k) data have been evaluated to illustrate the variation in mean differences between instruments that have been reported.

It can be seen that the results from the current study (Table 5) more closely reflect the mean difference between instruments for both the normative database and FDA 510(k) data than the study by Pierro et al.[2]

The determination of between instrument differences in pRNFL thickness should incorporate a large number of scans per subject and be independent of diurnal, seasonal and inter-subject variability (cf. Equation 2). These protocols may result in a more accurate measure of between instrument differences in pRNFL thickness measurements.

Table 4.

Average pRNFL thicknesses (μm) in the 4 OCT's obtained from various sources. FDA 510(k) refers to the pre-market submissions made for each instrument to the United States Food and Drug Administration to demonstrate that the device is as at least as safe and as effective (i.e. substantially equivalent to) a currently legally marketed device that is not subject to pre-market approval. The values in Table 4 are those supplied by the manufacturers to meet the FDA 510(k) requirement.

	Topcon	Nidek	RTVue	Cirrus
Current study	113.44	115.34	116.81	107.69
Pierro et al.[3]	106.51	102.43	103.9	90.08
Normative databases	97.87	105.48 (Caucasian)	102.14 (Caucasian)	
FDA 510(k)		101.0	100.82	93.0

Table 5.

Mean difference between average pRNFL thicknesses (μm) for the 4 OCT's using the values from Table 4. The instrument in the top row has the mean difference value in the column below added to (or subtracted from) it to obtain the equivalent average pRNFL thickness in the other instrument. N/A = not available

		Topcon	Nidek	RTVue
Nidek	Current study	+ 1.9		
	Pierro et. al.	- 4.08		
	Normative database	+ 3.13		
	FDA 510(k)	N/A		
RTVue	Current study	+ 3.37	+ 1.47	
	Pierro et. al.	- 2.61	+ 1.47	
	Normative Database	+ 4.27	+ 1.14	
	FDA 510(k)	N/A	- 0.18	
Cirrus	Current study	- 5.75	- 7.65	- 9.12
	Pierro et. al.	- 16.43	- 12.35	- 13.82
	Normative database	N/A	N/A	N/A
	FDA 510(k)	- 4.87	- 8.0	- 9.14

Table 6.

Maximum intra- and inter-visit average PRNFL inter-ocular asymmetry values (μm).

Instrument	Maximum Intra-visit inter-ocular asymmetry values (mean value in brackets)		Maximum Inter-visit inter-ocular asymmetry value	
Subject	JP	MP	JP	MP
Topcon	9 (4.4)	3 (1.4)	9	4
Nidek	13 (9.6)	18 (7.5)	19	18
RTVue	13 (10.8)	9.4 (3.1)	14.43	10.66
Cirrus	7 (4.8)	9 (5.5)	8	9

Table 6 illustrates not only the differences in the range of inter-ocular asymmetry values for different instruments, but also the naturally occurring variability between individual subjects. This finding seems to vindicate the protocol used in this study, which reduced inter-subject variance by limiting the sample size and having a large number of scans taken over a significant time frame. The largest average intra-visit difference between subjects was generated by the RTVue 100 (7.7 μm), while the largest inter-visit inter-ocular asymmetry values were found using the Nidek.

Discussion

Limits of Agreement

In terms of pRNFL measurements, all instruments were strongly positively correlated with each other, at statistically significant levels. Thus, simple algorithms can be used calculate a theoretical measurement value in one instrument using an actual measurement from another instrument. The inter-instrument LoA generated in this study (Table 3) are a more accessible method of detecting significant change in measurements recorded on different instruments. The mean differences in the current study, around which these confidence intervals are based, seem to correlate well with data from other sources (normative databases and FDA 510(k), Table 5), suggesting that they may be an accurate reflection of the true difference between instruments.

The inter-instrument CI_{95} values in the current study are particularly important for those clinicians involved the co-management of glaucoma suspects, whereby different instruments may be used by individual practitioners. The LoA derived from the current study (Table 3) are an easily accessible resource for clinicians wishing to compare tests between different instruments.

Peripapillary retinal nerve fibre layer

Average pRNFL retest variability in the current study for the Cirrus was 4.4 μm. This compares closely with intra-visit studies involving healthy subjects such as Hong et al.[7] and Tan et al.,[8] which reported test retest variability of 4.5 μm and 4.89 μm respectively. The CI_{95} for average pRNFL with the RTVue obtained in the current study was 5.4 μm, which is higher than that found in other studies using healthy patients, with Gonzalez-Garcia et al.[9] finding intra-visit minimum

detectable change (MDC) of 3.1 μm[10] and Garas et al.[11] reporting inter-session test retest variability to be 4.25 μm.

Inter-visit retest variability in average pRNFL with the Topcon in the current study was 4.9 μm. Although higher than the intra-visit MDC of 4.2 μm[10] obtained with healthy subjects using the Topcon 3D-OCT 1000 by Menke et al.,[12] this is perhaps a reflection of the inter-visit nature of the current study. The Nidek was reported by Pierro et al.[2] as having an intra-visit standard deviation (SD) of 6.54 μm for average pRNFL. The SD of the Nidek in the current study was 7.42 μm, which agrees quite closely with the present value of 6.54 μm, and any difference may be a reflection of the fact that the current study generated inter-visit values.

Quadrant TRV was greater than average pRNFL in all instruments (Table 2), and reflects the increasing variability associated with smaller segment measurements. The figures supplied in Table 2 should be useful in detecting statistically significant change in healthy patients when assessing inter-visit changes to quadrant thicknesses.

Although scans in this study were not evaluated for segmentation errors, automated segmentation has been reported to decrease average peripapillary retinal nerve fibre layer thickness by only 1.6 μm in comparison to manual refinement.[13] Although seemingly small, this figure was found to be statistically significant, and it was sufficient to change the glaucoma staging of 8.5% of scans overall, with 23.7% of borderline scans being reclassified as normal.

The fact that manual refinement and segmentation error ratios were not carried out in our study may be considered a limitation. However, the purpose of our

study was to quantify TRV in each instrument under the conditions that would be most likely to be encountered in clinical practice. To this end, we have analysed the data as it would be in routine clinical practice, and as recommended by the manufacturer, in order for our results to be applicable to clinical practice.

Analysis of the segmentation error ratio for each instrument may have shed light on the reasons for differences in TRV between instruments, but would not have added to the clinically applicable intent of the study.

One of the limitations of the current study became apparent after the completion of a later study (Chapter 3), where we reported that TRV for several SD-OCT scan parameters was found to be significantly higher during the middle of the day using the Topcon. On that basis, the instruments used more towards the middle of the day in this study (Nidek (10.31 am ± 18 minutes) and the RTVue 100 (11.47 am ± 53 minutes)) may have had their reported TRV impacted by the time of day they were used.

Given that TRV for the Nidek and RTVue was higher than that of the Topcon, further studies would need to be carried out using those instruments to determine whether significant diurnal alterations are detectable with those instruments, and that difference would have to be then quantified.

The results from the current study, where time of day testing was controlled for, still provides important TRV characteristics for the instruments under investigation as we did control for consistent time of day testing, an aspect that has not been carried out in previous studies. The current study provides

comparative data for use in future studies, where the randomization of time of day is carried out, or where all instruments are tested at the same times of day.

Inter-ocular asymmetry

The results from the current study indicate that measurement from different instruments may provide different values for inter-ocular asymmetry (Table 6). An inter-ocular difference exceeding 9 μm when measured with Cirrus is considered statistically significant asymmetry, which is suggested to be indicative of early glaucomatous damage.[14] Using the diagnostic criteria reported for the Cirrus,[14] subject JGP has significant asymmetry which may indicate early glaucomatous change (RTVue, Nidek), or no significant asymmetry and no indication of early glaucomatous change (Cirrus, Topcon).

Therefore, not only do different instruments give different measurement results when the same parameters are measured, but diagnostic criteria pertaining to one instrument may not be transferable to other instruments. These results indicate that the use of inter-ocular asymmetry as a definitive diagnostic feature is somewhat doubtful, and this view is in agreement with Anton et al.,[15] who found no difference between glaucomatous, ocular hypertensive and normal eyes in parameters describing RNFL asymmetry between eyes.

The use of a small number of subjects in the current study has enabled a greater number of scans to be taken over a longer period of time, reducing the contribution of subject variance to the total variance. The time frame encompassed by the current study has created a more realistic appraisal of instrument variability than other studies carried out in single sessions or over much shorter periods of time. The mean differences between instruments for average pRNFL in the current study agree quite closely with the differences

obtained from the normative databases and FDA 510(k) data. The results from the current study should therefore be regarded as being statistically robust and clinically applicable.

Acknowledgements: This research was supported by the Australian Research Council through the ARC Centre of Excellence in Vision Science (CE0561903), and the Australian Federal Government through the Nursing and Allied Health Scholarship Support Scheme (NAHSSS) and the National Eye Health Initiative, which enabled the purchase of the Topcon 3D-OCT 2000 used in this study. The views expressed in this paper do not necessarily represent those of the NAHSSS, its administrator, Services for Australian Rural and Remote Allied Health or the Australian Government Department of Health. I would like to thank Max Astri for the use of the Cirrus HD-OCT, Tony Burgun and Mick Brennan for the use of the RTVue 100 and Sallyanne Morrison for the use of the Nidek RS-3000.

References

1. Vessani RM, Moritz R, Batis L, Zagui RB, Bernardoni S, Susanna R. Comparison of quantitative imaging devices and subjective optic nerve head assessment by general ophthalmologists to differentiate normal from glaucomatous eyes. J Glaucoma 2009;18:253-261.
2. Pierro L, Gagliardi M, Iuliano L, Ambrosi A, Bandello F. Retinal nerve fiber layer thickness reproducibility using seven different OCT instruments. Invest Ophthalmol Vis Sci 2012;53:5912-5920.
3. Leite MT, Zangwill LM, Weinreb RN, et al. Effect of disease severity on the performance of Cirrus spectral-domain OCT for glaucoma diagnosis. Invest Ophthalmol Vis Sci 2010;51:4104-4109.
4. Pearce JG, Maddess T. Diurnal variation in summary measures and test-retest variability of standard automated perimetry and spectral domain optical coherence tomography To be submitted 2015.
5. Pearce JG, Maddess T. Daily variations in spectral domain OCT measurements. Abstract presentation, Australasian Ophthalmic Vision Science Meeting 2012.
6. Shin CJ, Sung KR, Um TW, et al. Comparison of retinal nerve fibre layer thickness measurements calculated by the optic nerve head map (NHM4) and RNFL3.45 modes of spectral-domain optical coherence tomography (RTVue-100). Br J Ophthalmol 2010;94:763-767.
7. Hong S, Kim CY, Lee WS, Seong GJ. Reproducibility of peripapillary retinal nerve fiber layer thickness with spectral domain cirrus high-definition optical coherence tomography in normal eyes. Jpn J Ophthalmol 2010;54:43-47.
8. Tan BB, Natividad M, Chua KC, Yip LW. Comparison of retinal nerve fiber layer measurement between 2 spectral domain OCT instruments. J Glaucoma 2012;21:266-273.
9. Gonzalez-Garcia AO, Vizzeri G, Bowd C, Medeiros FA, Zangwill LM, Weinreb RN. Reproducibility of RTVue retinal nerve fiber layer thickness and optic disc measurements and agreement with Stratus optical coherence tomography measurements. Am J Ophthalmol 2009;147:1067-1074, 1074 e1061.
10. Araie M. Test-retest variability in structural parameters measured with glaucoma imaging devices. Jpn J Ophthalmol 2013;57:1-24.
11. Garas A, Vargha P, Hollo G. Reproducibility of retinal nerve fiber layer and macular thickness measurement with the RTVue-100 optical coherence tomograph. Ophthalmology 2010;117:738-746.
12. Menke MN, Knecht P, Sturm V, Dabov S, Funk J. Reproducibility of nerve fiber layer thickness measurements using 3D fourier-domain OCT. Invest Ophthalmol Vis Sci 2008;49:5386-5391.
13. Mansberger SL, Menda S, Fortune B, Gardiner SK, Demirel S. Automated segmentation errors when using optical coherence tomography to measure retinal nerve fiber layer thickness in glaucoma. Am J Ophthalmol 2016.
14. Mwanza JC, Durbin MK, Budenz DL. Interocular symmetry in peripapillary retinal nerve fiber layer thickness measured with the Cirrus HD-OCT in healthy eyes. Am J Ophthalmol 2011;151:514-521 e511.
15. Anton A, Moreno-Montanes J, Blazquez F, Alvarez A, Martin B, Molina B. Usefulness of optical coherence tomography parameters of the optic disc and the retinal nerve fiber layer to differentiate glaucomatous, ocular hypertensive, and normal eyes. J Glaucoma 2007;16:1-8.

Appendix 3: Part 2 (3.2)

Ganglion cell complex measurement comparison between two spectral domain optical coherence tomographs

Abstract

Background To interpret what changes in repeated measurements of ganglion cell complex (GCC) thickness represent, the reproducibility of instrument measurements needs to be known. In this study, we compare two spectral domain optical coherence tomographs (SD-OCT), the Topcon 3D-OCT 2000 and Optovue RTVue-100. We intend to determine the level of test retest variability (TRV) in their measurement of ganglion cell complex parameters and the measurement agreement between instruments.

Methods Two healthy subjects underwent 8 scans on each eye with each instrument to concentrate on TRV rather than between subject variance. To avoid any circadian variation, each subject was tested on each SD-OCT at approximately the same time in the morning (SD < 60 minutes). There were four sessions in total. Two tests were done on each eye of each patient in each session, and the sessions were carried out at intervals one week apart for four consecutive weeks.

Results The maximum inter-visit variation in any individual measurement with the Topcon 3D-OCT 2000 was 3 µm, and the maximum inter-visit variation in total GCC measurement was 2 µm. The maximum inter-visit variation for any individual parameter with the Optovue RTVue-100 was 5.9 µm, and the maximum inter-visit variation in total GCC was 5.9 µm. For total GCC thickness, Bland-Altman analysis gave 95% confidence intervals of 1.3 µm for the Topcon 3D-OCT 2000 and 4.7 µm for the Optovue RTVue-100. The average total GCC thicknesses obtained with each instrument were significantly different, being on average 12.4 µm greater with the Topcon 3D-OCT 2000.

Conclusions The average ganglion cell complex thicknesses obtained with these two instruments are significantly different and are not interchangeable. The levels of TRV in both instruments needs to be taken into account when reviewing serial GCC scans on patients.

Introduction

Newer technologies, such as spectral domain optical coherence tomography (SD-OCT), have become an integral component in the management of glaucoma patients and glaucoma suspects. The introduction of new instruments from different manufacturers means that many instruments are available that have no published information on their retest variability, their measurement agreement with other instruments for the same parameters, or what their event based criteria of change might be. Modern instruments incorporate test protocols that are able to assess ganglion cell complex (GCC) parameters for glaucomatous change. The GCC is comprised of the inner retinal layers, those being the retinal nerve fibre layer, the ganglion cell layer and inner plexiform layer.

Although GCC thickness is an important diagnostic parameter, some studies have found that it has no greater diagnostic ability than peripapillary retinal nerve fibre layer (pRNFL) analysis in detecting early, moderate or severe glaucoma.[1] Sensitivity, specificity and diagnostic accuracy have been found to be low when using GCC analysis in patients with pre-perimetric glaucoma.[2] Additionally, pRNFL thickness measurement has been reported to be the better individual OCT parameter for detecting perimetric glaucoma using the Optovue RTVue (Model RT 100).[3]

In contrast, recent studies have shown that the combined inner retinal layer thickness (nerve fibre layer, ganglion cell layer and inner plexiform layer) is significantly less in patients with suspected glaucoma and pre-perimetric glaucoma[4] and that the thickness of the GCC is significantly less in the normal hemifield of glaucomatous eyes (with visual defects restricted to one hemifield) compared to normal control eyes.[5] A recent study using the RTVue[6] found that 38% of glaucoma suspects had an abnormal GCC thickness while only 13% had an abnormal pRNFL thickness.

Given the potential importance of GCC measurements in glaucoma detection and management, the purpose of this study is to generate some event based criteria for progression detection in the Topcon 3D-OCT 2000 and RTvue-100. These can then be applied in clinical practice when rate of change information is not available.

It is also our intention to determine the measurement agreement between instruments for the GCC parameter, and establish confidence intervals to enable the detection of progressive GCC thinning using measurements taken on both instruments. These criteria are intended to enable clinicians to easily determine whether a change in GCC thickness is representative of instrument variability or genuine structural change, and to enable the comparison of measurements obtained from patients on these instruments.

Methods

Two patients were recruited for the study, which comprised 8 scans per SD-OCT per eye: a 51 year old male (JP) and a 15 year old female (MP). The research conformed to the tenets of the Declaration of Helsinki, with informed written consent obtained under ANU/ACT Health protocol 7/07.667. Their refractive errors ranged between plano and +1.25 D of hyperopia, with up to 0.25 D of astigmatism. They had both been assessed as having no glaucomatous changes to their optic nerve heads by experienced optometrists. Intra-ocular pressures for JP were RE 12 mmHg and LE 13 mmHg with central corneal thicknesses of R 525.5 µm and L 524 µm. Intra-ocular pressures for MP were R 18 mmHg and L 17 mmHg with central corneal thicknesses of R 555 µm and L 524 µm. Visual field tests for both patients using the Medmont M700 automated perimeter (Version 3.9.7) demonstrated no visual field defect, and neither patient had any active ocular pathology.

To minimise the possibility of any circadian variations in results, each subject was tested on each machine at approximately the same time on a Thursday morning. The mean test time with the Topcon 3D OCT 2000 was 7.02 am (Standard Deviation (SD) 13 minutes) and the Optovue RTVue-100 mean test time was 11.47 am (SD 53 minutes). Two tests were done on each eye of each patient within each session, with the protocol for each instrument being that a right eye and left eye for each patient were scanned once, and then the right eye and left eye were scanned again on the same patient immediately afterwards. This means that the subjects did not move from the chinrest of each instrument between scans. Each session was carried out at intervals one week

apart for four consecutive weeks, giving each subject a total of eight scans on each eye with each instrument.

All tests on the RTVue-100 were carried out by the same experienced operator. They were all practice staff who had been trained to carry out testing on the relevant SD-OCT's. The other operator of the Topcon (JP) is an optometrist. The scans done on JP were carried out by MP.

None of the GCC scans used in this study were modified after the scans were taken. All scans had an image quality greater than 60 out of 100. The Topcon mean image quality was 93 (SD 3.1) and the RTVue mean image quality was 82 (SD 3.5).

The Topcon 3D-OCT 2000 (Software Version 7.11, Topcon Corporation, Tokyo, Japan) utilises a super luminescence diode (peak wavelength 840 nm) as a light source, and has an A-scan speed of 27,000 per second. The Macula V scan (which is used to measure the thickness of the nerve fibre layer (NFL), ganglion cell layer (GCL) and inner plexiform layer (IPL), which together constitute the ganglion cell complex (GCC)), consists of a scan area of 7.0 × 7.0 mm. 128 B-scans are taken, with 512 A-scans per B-scan, giving a total of 65,636 A-scans. The instrument algorithm selects the best 6.0 × 6.0 mm position having the fovea in the centre, and the information is extracted from the selected 6.0 × 6.0 mm area. The individual layer thicknesses are calculated for each point over this area.

The Optovue RTVue-100 (Algorithm Version A5,1,0,90. Software Version #5,1,0,90, Optovue Inc, Fremont, California) uses a light source with a peak wavelength of 840 ± 10 nm and has an A-scan rate of 26,000 A-scans per second. The GCC scan pattern consists of 15 vertical lines covering a 7.0 × 7.0

mm square, centred 1 mm temporal to the fovea centre. It includes a horizontal line scan for vertical scan registration. The scan captures 14,928 data points, and a 6.0 mm diameter circle is used to generate the GCC information.

Statistical analysis was carried out using Medcalc version 12.4 (Medcalc Software bvba, Ostend, Belgium).

Results

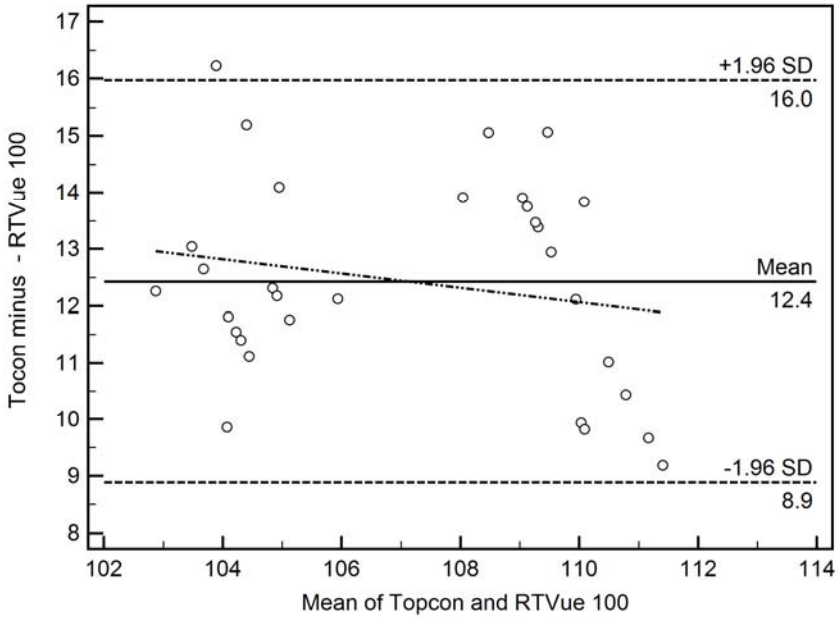


Figure 1.

Bland-Altman difference plot of Topcon GCC thickness measurements compared to RTVue 100 GCC thickness measurements (in micrometres, µm). The mean difference between the instrument measurements is 12.4 µm, and the Limits of Agreement are 8.9 to 16.0 µm.

Figure 1 shows the relationship between the Topcon and RTVue total GCC thickness measurements. The Limits of Agreement (LoA) are 8.9 to 16.0 µm, with a mean difference of 12.4 µm (Topcon greater than RTVue). When applying the 95% confidence interval of ± 3.6, the mean difference between instruments of 12.4 µm needs to be applied also. The instrument measurements were highly correlated at r = 0.84 (Pearsons' correlation coefficient).

Table 1.

Total GCC thickness results for the Topcon 3D-OCT 2000. All measurements are in micrometres, SD = Standard deviation. N = 8 per eye, except there were only 7 scans used for JP (R eye) due to one scan being analysed incorrectly.

Subject	Eye	Range	Mean ± SD
JP	R	111 – 112	111.5 ± 0.6
	L	109 – 110	109.8 ± 0.5
MP	R	115 - 117	115.9 ± 0.6
	L	115 - 117	116.0 ± 0.5

Table 2.

Total GCC thickness results for the Optovue RTVue-100. All measurements are in micrometres, SD = Standard Deviation.

Subject	Eye	Range	Mean ± SD
JP	R	95.8 – 100.6	98.5 ± 1.6
	L	96.7 – 99.1	98.0 ± 0.9
MP	R	100.9 – 106.8	104.5 ± 2.1
	L	101.1 – 105.0	102.7 ± 1.1

From Tables 1 and 2, it can be seen that the average GCC thickness for both subjects is 12.4 µm greater with the Topcon (Fig. 1). This difference was consistent between subjects, with the average value for both eyes in each subject being 12.4 µm greater in the Topcon. Inter-subject average thickness differences were consistent between instruments, with subject MP having an average (for both eyes) GCC thickness 5.3 µm greater than JP in the Topcon, and 5.3 µm greater in the RT-Vue 100.

Table 3.

Bland-Altman 95% confidence intervals for the Topcon 3D-OCT 2000 and the Optovue RTVue-100 total GCC, superior GCC and inferior GCC thickness. Measurements are in micrometres.

Instrument	Total GCC	Superior GCC	Inferior GCC
Topcon 3D-OCT 2000	1.3	0.9	1.7
Optovue RTVue-100	4.7	5.1	4.4

Table 3 shows that the TRV of the Topcon was lower for total, superior and inferior GCC thickness measurements than that of the RTVue.

Discussion

The 95% confidence limits (CI_{95}) established for total GCC thickness in healthy subjects using the Topcon 3D-OCT 2000 was 1.3 µm. This is a significant improvement in previously reported repeatability obtained with the Topcon 3D-OCT 1000, where a difference of 6 µm between visits in pre-presbyopic individuals was considered to be structural change rather than true variability.[7]

The CI_{95} for total GCC thickness in the current study was 4.70 µm for the RTVue-100. This result indicates that the RTVue-100 has less precision than the Topcon when measuring GCC thickness, which is consistent with the findings regarding retinal nerve fibre layer measurement by Buchser et al.[8] In that study,[8] the RTVue-100 demonstrated higher measurement variability than both the Cirrus HD OCT and the Topcon 3D-OCT 1000, and that this imprecision was higher for healthy eyes, as used in the current study.

The variability in hemifield thickness measurements obtained from the RTVue 100 were higher than those obtained by Garas et al.,[9] where inter-visit test-

retest variability was reported as 4.51 µm and 3.93 µm for superior and inferior GCC thicknesses (respectively) in non-glaucomatous subjects.

The difference in average total GCC thickness measurement between these instruments was relatively large, being on average 12.4 µm. This finding is consistent with findings from other studies[10, 11] that measurements that are stated in units of micrometres from different instruments cannot be directly compared.

The GCC measurements from each instrument were highly correlated. For clinicians wanting to compare total GCC thickness measurements on the same patient from the different instruments the used in this study, a CI_{95} of ± 3.6 µm was established for measurements between instruments. This CI_{95} can only be applied after adding or subtracting the mean difference of 12.4 µm (as required) to account for the mean difference in total GCC thickness measurements between instruments. *This information is critical for clinicians attempting to differentiate variability from true progression in patients who have had GCC measurements taken on one instrument which need to be compared to measurements taken on another.*

Glaucoma detection and management requires the ability to differentiate progression from variability in SD-OCT scans. Highly repeatable results can provide useful data for detecting early retinal changes associated with ganglion cell loss and provide another reliable objective measure for the assessment of glaucomatous progression. The results obtained in this study indicate that GCC measurements show a high level of repeatability in the instruments tested, and may therefore be a useful parameter for monitoring progressive changes in GCC parameters. The variability found in both instruments in this study should

be taken into consideration by clinicians when determining glaucomatous change their patients.

The use of healthy subjects to determine these confidence intervals means that these values may not be transferable to a glaucoma population. For that reason, the confidence intervals would be most applicable in the detection of early progression in a glaucoma suspect/early disease stage patient population.

The small number of subjects in the current study was intended to isolate instrument variability from subject variability. Specifically, the undertaking of a larger number of scans per eye over a significantly longer time period than would be encountered in most studies has been designed to more effectively determine instrument variability. This alternative approach has generated higher test-retest variability than that obtained in other studies with a larger number of subjects,[9] making these findings clinically relevant and applicable.

Acknowledgements: This research was supported by the Australian Research Council through the ARC Centre of Excellence in Vision Science (CE0561903), and the Australian Federal Government through the Nursing and Allied Health Scholarship Support Scheme. The views expressed in this paper do not necessarily represent those of the NAHSSS, its administrator, Services for Australian Rural and Remote Allied Health or the Australian Government Department of Health. The Australian Federal Government National Eye Health Initiative provided a grant which enabled the purchase of the Topcon 3D-OCT 2000 used in this study. I would also like to thank Tony Burgun and Mick Brennan (Burgun and Brennan Optometrists) for the use of their RTVue 100 and support staff in this study.

References

1. Kim NR, Lee ES, Seong GJ, Kim JH, An HG, Kim CY. Structure-function relationship and diagnostic value of macular ganglion cell complex measurement using Fourier-domain OCT in glaucoma. Invest Ophthalmol Vis Sci 2010;51:4646-51.
2. Rolle T, Briamonte C, Curto D, Grignolo FM. Ganglion cell complex and retinal nerve fiber layer measured by fourier-domain optical coherence tomography for early detection of structural damage in patients with preperimetric glaucoma. Clin Ophthalmol 2011;5:961-9.
3. Huang JY, Pekmezci M, Mesiwala N, Kao A, Lin S. Diagnostic power of optic disc morphology, peripapillary retinal nerve fiber layer thickness, and macular inner retinal layer thickness in glaucoma diagnosis with fourier-domain optical coherence tomography. J Glaucoma 2011;20:87-94.
4. Kotera Y, Hangai M, Hirose F, Mori S, Yoshimura N. Three-dimensional imaging of macular inner structures in glaucoma by using spectral-domain optical coherence tomography. Invest Ophthalmol Vis Sci 2011;52:1412-21.
5. Takagi ST, Kita Y, Yagi F, Tomita G. Macular retinal ganglion cell complex damage in the apparently normal visual field of glaucomatous eyes with hemifield defects. J Glaucoma 2012;21:318-25.
6. Ganekal S. Ganglion cell complex scan in the early prediction of glaucoma. Nepal J Ophthalmol 2012;4:236-41.
7. Bruce A, Pacey IE, Dharni P, Scally AJ, Barrett BT. Repeatability and reproducibility of macular thickness measurements using fourier domain optical coherence tomography. Open Ophthalmol J 2009;3:10-4.
8. Buchser NM, Wollstein G, Ishikawa H, Bilonick RA, Ling Y, Folio LS, Kagemann L, Noecker RJ, Albeiruti E, Schuman JS. Comparison of retinal nerve fiber layer thickness measurement bias and imprecision across three spectral-domain optical coherence tomography devices. Invest Ophthalmol Vis Sci 2012;53:3742-7.
9. Garas A, Vargha P, Hollo G. Reproducibility of retinal nerve fiber layer and macular thickness measurement with the RTVue-100 optical coherence tomograph. Ophthalmology 2010;117:738-46.
10. Huang J, Liu X, Wu Z, Guo X, Xu H, Dustin L, Sadda S. Macular and retinal nerve fiber layer thickness measurements in normal eyes with the Stratus OCT, the Cirrus HD-OCT, and the Topcon 3D OCT-1000. J Glaucoma 2011;20:118-25.
11. Giani A, Cigada M, Choudhry N, Deiro AP, Oldani M, Pellegrini M, Invernizzi A, Duca P, Miller JW, Staurenghi G. Reproducibility of retinal thickness measurements on normal and pathologic eyes by different optical coherence tomography instruments. Am J Ophthalmol 2010;150:815-24.

Appendix 3: Part 3 (3.3)

Comparison of optic nerve head topography measurements between three different spectral domain optical coherence tomographs

Abstract

Purpose To establish event based criteria for detecting structural change in optic nerve head topographical features using three different spectral domain optical coherence tomographs (SD-OCT's).

Method Two healthy patients were recruited for the study, and testing was carried out using the Topcon 3D OCT-2000, Nidek RS-3000 and the Optovue RTVue-100. There were four test sessions, carried out at one week intervals over four consecutive weeks. Scans were taken at approximately the same time of day on each instrument, and each session consisted of two scans being taken on each eye of each subject. This gave a total of eight scans on each subject eye on each instrument.

Results Disc area, cup area and cup to disc ratio measurements were not interchangeable between instruments.SD-OCT portrayal of cup to disc ratio differs from subjective clinical examination of the optic nerve head. Horizontal cup/disc ratio 95% confidence intervals (CI_{95}) were 0.07, 0.23 and 0.18 in the Topcon, Nidek and RTVue respectively. Vertical cup/disc ratio CI_{95} were 0.08, 0.18 and 0.12 in the same instrument order.

Discussion Highly repeatable SD-OCT cup/disc ratio measurements may be able to detect glaucomatous changes to cup/disc ratios earlier than clinical subjective observation. Instrument retest variability should be taken into account when defining progressive change to optic nerve head morphology. This study provides some event based guidelines for detecting progressive change to cup/disc ratios in healthy patients in the instruments under consideration.

Introduction

Historically, direct observation of the optic nerve head (ONH) has been one of the mainstays of glaucoma detection and monitoring, and remains so even in the presence of modern visual field assessment techniques and imaging devices currently available. Changes in appearance, or the presence of suspicious features noted during direct examination of the ONH, are the precursor to further investigations with other methods. Unfortunately, early changes to ONH morphology can be difficult to detect even with direct observation. Using optic nerve head photography, over 80% of eyes may have conversion to visual field loss before discernible disc changes are noted.[1]

The advent of spectral domain optical coherence tomography (SD-OCT) ONH imaging has greatly enhanced our ability to detect glaucomatous change at a much earlier stage. It has been reported that cup to disc ratio (CDR) using SD-OCT has the best diagnostic ability to discriminate between normal and glaucomatous patients' out of all available SD-OCT parameters, including ganglion cell complex (GCC) analysis and peripapillary retinal nerve fibre layer (pRNFL) thickness measurements.[2]

It is important to recognise that whilst SD-OCT ONH parameters have been found to have the same level of discriminatory ability to differentiate normal eyes from those with mild glaucoma to the same extent as RNFL thickness parameters,[3] other studies have found that optic disc parameters not only had lower diagnostic accuracy than RNFL and GCC parameters, they also had a significantly lower positive likelihood ratio than the other parameters.[4]

The assessment of disc size is an important clinical aspect of glaucoma diagnosis, and in conjunction with vertical cup /disc ratio, can be used as part of

the calculation to determine retinal ganglion cell numbers.[5] Cup to disc ratio has always been perhaps the most important disc feature routinely recorded by clinicians for many years, and vertical CDR has been found to be the best individual optic nerve head parameter to differentiate normal eyes from early glaucoma[6] and glaucoma suspects from perimetric glaucoma.[7]

Detecting progressive change in any given parameter requires some insight as to what constitutes variability in any given instrument with respect to the structure being measured. The purpose of this study is to determine the test-retest variability (TRV) present in three commonly used SD-OCT devices. The Topcon 3D OCT-2000 and Nidek RS-3000 are relatively new and have little information available regarding their repeatability. The RTVue-100 has been used extensively in other studies, which will provide a benchmark for comparative purposes.

The intention of the current study is to establish event based criteria for ONH topographical features which are clinically important in the detection and ongoing management of glaucoma. These confidence intervals are intended to be an easily accessed and applied method for clinicians to determine statistically significant change in these parameters in their clinical populations.

Methods

There were two patients were recruited for the study, a 51 year old male (JGP) and a 15 year old female (MEP), with the appropriate consent forms signed under the ANU/ACT Health protocol 7/07.667. Refractive error in these patients ranged from plano to +1.25 D of hyperopia, with a maximum astigmatic correction of 0.25 D. They were both assessed as having no glaucomatous changes to their optic nerve heads and intra-ocular pressures (IOP) were within

normal limits for both patients. Standard automated perimetry was performed on both patients using the Medmont M700 automated perimeter (Central 30° test), which demonstrated no visual field defects, and neither patient had any active ocular pathology.

To avoid any diurnal effects on scan results, each subject was tested on each instrument at approximately the same time of day at one week intervals for four consecutive weeks. The mean test times were: Topcon 3D-OCT 2000 7.02 am (Standard Deviation (SD) 13 minutes), Nidek RS-3000 10.31 am (SD 18 minutes) and the Optovue RTVue-100 11.47 am (SD 53 minutes). At each session, two scans were done on each eye of each patient. The right eye and left eye for each patient was scanned once, and the right eye and left eye were then scanned again on the same patient immediately afterwards. The patient did not move from the chinrest of each instrument in between scans. The sessions were carried out at intervals one week apart for four consecutive weeks, giving a total of eight scans on each eye with each instrument.

To avoid the possibility of inter-operator factors influencing the results, all scans on the RTVue-100 and Nidek were carried out by the same operator (that is, a different operator was used for each machine, but it was the same operator for each individual machine). The exception to this was with the Topcon, with the scans done on MEP carried out by JGP, and the scans on MEP carried out by JGP. None of the disc scans used in this study were manually modified after the scans were taken. All scans had an image/scan quality greater than 60 out of 100.

The Topcon 3D- OCT 2000 (Software Version 7.11, Topcon Corporation, Tokyo, Japan) 3D disc scan is carried out over a 6.0 × 6.0 mm square scan

area and is composed of 128 horizontal B-scans which have 512 A-scans per B-scan resulting in a total of 65,636 A-scans with A-scan depth of 2.4 mm. From these scans, the disc centre is located and RNFL thickness values are measured at the edge of a circle 3.4 mm in diameter. The disc topography information is obtained from the information in the original cube scan.

The Nidek RS 3000 (Software version Navis Ex 1.3.0.3, Nidek Corporation Limited, Aichi, Japan) Disc Map scan consists of a 6.0 × 6.0 mm square scan area. The number of scans per point is adjustable, with the default setting being 512 A-scans per B-scan, with the number of B-scans being 128. The instrument used in this study used 256 A-scans per B-scan, with 256 B-scans, and this scan protocol is still covered by the Nidek RS-3000 normative database. This scan protocol results in 65,536 A-scan points with a depth of 2.1 mm. RNFL and disc parameters are calculated from the scans taken in the cube scan and RNFL parameters are generated at the edge of circle with a radius of 1.725 mm.

The Optovue RTVue-100 (Algorithm Version A5,1,0,90. Software Version #5,1,0,90, Optovue Inc, Fremont, California) optic nerve head scan protocol uses 12 radial B-scans 3.4 mm in length to determine disc margin and ONH topography measurements and 13 circular B-scans, with diameters ranging from 1.3 to 4.9 mm to create the pRNFL map. The disc scan covers a region of 6.0 × 6.0 mm containing 101 horizontal lines, with each line scan comprising 513 A-scans to give a total of 51,813 A-scans.

Statistical analysis in this study was carried out using Medcalc version 12.4 (Medcalc Software bvba, Ostend, Belgium).

Results

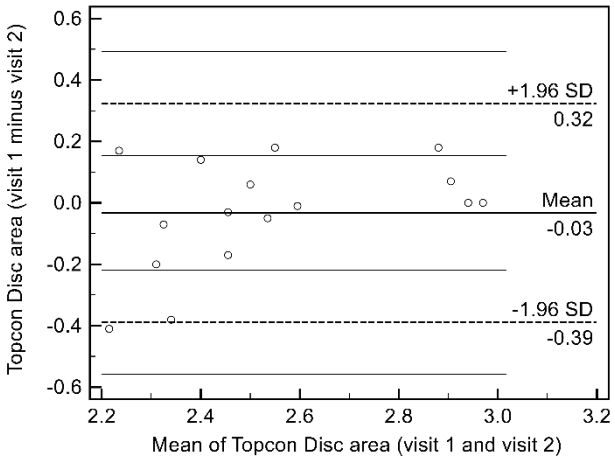


Figure 1.
Bland-Altman difference plot showing the data for the Topcon disc area measurements. The Limits of Agreement (LoA) are ± 1.96 × the Standard deviation (SD) of the differences between visits. In this example, the LoA are from -0.39 to +0.32. The 95% confidence limits are ± 0.36 from the mean difference between visits. Measurements are in mm^2

Retest variability may be reported using the Limits of Agreement (LoA), which are defined as the mean difference between measurements ±1.96 times the standard deviation (SD) of the difference between measurements. This value (± 1.96 SD) is the repeatability coefficient, or 95% confidence interval, outside of which will fall only 5% of pairs of measurements on the same subject. Differences between pairs of measurements that fall within the 95% confidence limits are considered to be TRV, while those that lie outside the confidence limits are considered to be true change in the parameter under consideration.

Table 1.
Inter-visit Bland-Altman 95% confidence intervals for each instrument for disc area, cup area, rim area, linear cup/disc ratio and vertical cup disc ratio using results from both subjects in this study. For patient MEP, the Topcon 3D-OCT 2000 rim area figures were calculated by subtracting cup area from disc area, because when there was no cup area measurement, for some reason the Topcon 3D-OCT 2000 indicated a rim area of zero. The rim area figures for both patients on the Nidek RS-3000 were calculated as above (i.e. by subtracting cup area from disc area), as it is not a parameter included on their printout. (mm^2 = square millimetres, HCDR = horizontal cup/disc ratio, VCDR = vertical cup/disc ratio)

	Disc area (mm^2)	Cup area (mm^2)	Rim area (mm^2)	HCDR	VCDR
Topcon	0.36	0.19	0.24	0.07	0.08
Nidek	1.22	0.25	1.23	0.23	0.18
Optovue	0.03	0.21	0.21	0.18	0.12

The 95% confidence intervals (CI_{95}) were calculated by pairing the right and left eye results at each visit and then combining the eight sets of pairs for each eye to generate the inter-visit values. Although disc area confidence intervals varied substantially between instruments (Table 1), variability in this fundamental parameter did not necessarily translate into significantly different confidence intervals in other disc parameters. For example, while the Nidek and RTVue-100 had the highest and lowest disc area confidence intervals respectively, cup area confidence intervals with the two instruments did not differ greatly. Similarly, although the Topcon disc area confidence interval was greater than

the RTVue-100, linear and vertical C/D ratios in the Topcon were lower than in the RTVue 100.

Table 2.
Average results for cup area, rim area, linear cup/disc ratio and vertical cup/ disc ratio for patient JGP (who had discernible and measurable disc cupping). Subject MEP was excluded from the cup area, rim area and vertical and linear C/D ratio results in this table due to having essentially flat discs, although she was included in the disc area results. The rim area figures for subject JGP on the Nidek were calculated by subtracting cup area from disc area, as it is not a parameter included on the Nidek printout. (HCDR = horizontal cup/disc ratio, VCDR = vertical cup/disc ratio, mm^2 = square millimetres)

	Topcon	Nidek	RTVue-100
Cup area (mm^2)			
RE	0.46	0.39	0.62
LE	0.56	0.43	0.64
Rim area (mm^2)			
RE	2.00	2.30	1.26
LE	1.72	2.27	1.34
HCDR			
RE	0.43	0.39	0.65
LE	0.49	0.38	0.66
VCDR			
RE	0.44	0.38	0.61
LE	0.57	0.42	0.62
Disc area (mm^2)			
JGP RE	2.46	2.68	1.88
JGP LE	2.27	2.70	1.99
MEP RE	2.50	3.49	1.91
MEP LE	2.92	2.76	1.88

Table 2 shows the average figures derived by each instrument for the listed disc parameters. Disc area results gave distinctly different values for each instrument, with the RTVue-100 having consistently smaller values than the other instruments. The effect of having this smaller initial disc area value,

combined with a slightly larger cup area figure than the other instruments, resulted in significantly lower rim area values for the RTVue-100. Whilst no instrument can be said to provide a definitive value for disc area, smaller initial rim area values for the RTVue-100 might influence the ability to detect progression via loss of rim area in this instrument.

Vertical and horizontal CDR's were also consistently larger in the RTVue 100 than the other two instruments. It is a distinct possibility that the RTvue 100 does not delineate cupping of the ONH any differently than other instruments, but the smaller disc area measurement means that the resultant CDR's may be larger (in ratio terms) than in reality.

Table 3.
Disc area Pearson correlation coefficients (r) and Bland-Altman 95% confidence intervals (CI_{95}) and statistical significance (p) values for each instrument. This table includes values for both subjects in the study. (mm^2 = square millimetres)

Instrument	Topcon 3D-OCT 2000	Nidek RS-3000
Nidek RS-3000		
CI_{95}	1.55 mm^2	
r	0.020	
p	0.916	
Optovue RTVue-100		
CI_{95}	0.57 mm^2	1.08 mm^2
r	-0.590	-0.051
p	0.0004	0.7811

The Nidek did not correlate well with either of the other instruments when considering disc area (Table 3). This may be a reflection of greater variability in disc area calculation (Table 1), a generally higher value for disc area than the other instruments (Table 2), or a combination of both.

Conclusions

These results show that disc parameter measurements are not inter-changeable between these instruments. There were large differences in average disc area measurements (Table 2), with the RTVue-100 evaluating disc areas for both patients as being significantly smaller than the other two instruments. This finding is consistent with the fact that the RTVue-100 has been shown to give disc areas 0.4 mm² smaller, on average, than Stratus time domain OCT.[8] With the average difference between the Topcon and the RTVue-100 in this study being 0.62 mm², it appears that the RTVue 100 also gives smaller disc area measurements when compared to other SD-OCT's.

It seems that disc area cannot be conclusively measured by any individual SD-OCT instrument, and that like pRNFL measurements, any measurement only has relevance to itself for determining change over time in a particular instrument rather than supplying an anatomically correct figure.

Variability in horizontal and vertical cup/disc ratios was significantly different between instruments. The variability in results obtained for normal patients using the RTVue 100 in this study were approximately double that obtained using the RTVue 100 by Gonzalez-Garcia.[8] This may be a reflection of differences in test protocols, with the subjects in the Gonzalez-Garcia study scanned three times in one session versus the eight scans over four sessions one week apart in the current study. The scan protocols in the current study would seem to be more typical of clinical reality, with the result that although only one subject was used, it appears that a greater number of scans over extended time frames can generate greater instrument retest variability. If the figures for variability generated by Gonzalez-Garcia et al.[8] were applied to subject JGP, he would be demonstrating glaucomatous progression over a

period of four weeks. As this is clearly not the case, the larger confidence intervals created in this study may thus be more clinically relevant.

The figures for cup area (Table 2) also revealed considerable differences between instruments, which again highlights the difficulty of using this important measurement in glaucoma diagnosis and management when comparing values from different instruments. Although vertical and horizontal cup/disc ratios are listed separately, cup area is vitally important as it determines the area of neuro-retinal rim remaining on any given disc.

The rim area figure is dependent on the previous two results, and as would be expected from the foregoing, the rim area figures from each instrument are significantly different to each other (Table 2). Rim area has been found to be one of the least reliable optic nerve head measurements,[9] although the results for the Topcon and RTVue-100 from this study would seem to indicate that they may be sensitive enough (in percentage terms) to detect structural changes in glaucoma suspects and early glaucoma prior to the detection of functional progression.

The usefulness of rim area to detect progression in more advanced glaucoma decreases to the point where the percentage of remaining rim area lost in order to detect true progression becomes quite substantial. The smaller initial rim area values for patients with the RTVue-100 may also impact on the ability of rim area in this instrument to differentiate early progression from retest variability. Perhaps for this reason, rim area has been found to have the best diagnostic accuracy among ONH parameters in differentiating between normal eyes and glaucoma suspect eyes,[10] where rim area would be expected to be larger than in moderate to advanced cases.

An important finding in the current study is the difference between stereoscopic slit lamp estimation of CDR's and the OCT results obtained. For patient JGP, CDR's were measured as 0.3 (horizontally and vertically) right and left by an experienced optometrist, and all SD-OCT results were above the clinical estimation. This finding may be attributed to the fact that the disc margin defined by clinical observation is probably not the innermost edge of Bruch's membrane used by SD-OCT to delineate the disc margin,[11] and that cup margins on fundus photographs only correspond 73.6% of the time with the hyporeflective regions of the interior border obtained with SD-OCT images.[12]

The CDR's found in this study replicated the results of a study that found that the RTVue-100 gave higher vertical and horizontal CDR's than a time domain OCT (Stratus OCT).[13] The results from the current study would seem to indicate that this difference also applies when comparing the RTVue 100 to the Topcon and Nidek SD-OCT's used in this study.

Subjective assessment of changes to CDR can be subject to problems in the detection of actual change. One study found that expert readers examining optic disc stereo-slides can demonstrate excellent reproducibility in evaluating glaucomatous optic disc change,[14] while another found the ability of glaucoma specialists (inter-observer agreement) to judge progressive disc change from stereo-photographs to be only slight to fair.[15] When viewing stereoscopic disc photographs, glaucoma experts were found to differ in CDR estimates by up to 0.16.[16] In another study, using glaucoma specialists to determine progressive disc change in glaucoma, in 40% of cases judged to have progressed, the "worse/progressed" disc photo was actually taken first.[15] Inter-observer experience has also been found to influence inter-observer agreement in the

assessment of cup/disc ratio,[17] and inter-observer agreement in assessing stereo-slides for glaucomatous change between non-expert ophthalmologists was significantly lower than that of glaucoma specialists, whose own inter-observer agreement was only moderate.[18] Knowing the chronological order stereophotographs were taken in has also been found to provide considerably different interpretations of glaucomatous change when compared to observations made without this knowledge.[19]

Objective assessment of disc parameters with SD-OCT eliminates the subjective aspects of variability as outlined above. The confidence interval for HCDR was 0.07 and VCDR 0.08 for the Topcon for patient JGP (Table 1). The amount of detectable change in CDR with the Topcon SD-OCT appears to be smaller than that which would be expected to be detected subjectively by binocular slit lamp fundoscopy or serial fundus photography, and also eliminates the effect of subjective differences in interpretation as noted above. Despite the fact that even small changes in CDR may represent the loss of a significant number of ganglion cells, especially in discs with large CDR's,[5] confidence intervals of such small magnitude as found in this study may assist in the detection of structural progressive change prior to the detection of functional progression using SAP, before the loss of significant numbers of RGC's and earlier than subjective clinical observation.

Although one criticism of the current study may be the small sample size, this deficiency has been overcome to some extent by the number of scans carried out on each subject and the time frame over which they were taken. In the absence of any other criteria for these parameters in these instruments, a clinically useful starting point has been established. Perhaps more importantly,

the design of this study enabled the generation of larger confidence intervals for CDR in the RTVue 100 than those of a much larger study,[8] providing some validation of the results obtained in the current study and the study design itself.

Acknowledgements: This research was supported by the Australian Research Council through the ARC Centre of Excellence in Vision Science (CE0561903), and the Australian Federal Government through the Nursing and Allied Health Scholarship Support Scheme (NAHSSS) and the National Eye Health Initiative, which enabled the purchase of the Topcon 3D-OCT 2000 used in this study. The views expressed in this paper do not necessarily represent those of the NAHSSS, its administrator, Services for Australian Rural and Remote Allied Health of the Australian Government Department of Health. For the other instruments used in this study, I would like to thank Tony Burgun and Mick Brennan (Burgun and Brennan Optometrists) for the use of the RTVue 100 and Sallyanne Morrison (Morrisons Family Eye Care Centre) for the use of the Nidek RS-3000.

References

1. Quigley HA, Katz J, Derick RJ, Gilbert D, Sommer A. An evaluation of optic disc and nerve fiber layer examinations in monitoring progression of early glaucoma damage. Ophthalmology 1992;99:19-28.
2. Schulze A, Lamparter J, Pfeiffer N, Berisha F, Schmidtmann I, Hoffmann EM. Diagnostic ability of retinal ganglion cell complex, retinal nerve fiber layer, and optic nerve head measurements by Fourier-domain optical coherence tomography. Graefes Arch Clin Exp Ophthalmol 2011;249:1039-45.
3. Mwanza JC, Oakley JD, Budenz DL, Anderson DR. Ability of cirrus HD-OCT optic nerve head parameters to discriminate normal from glaucomatous eyes. Ophthalmology 2011;118:241-8 e1.
4. Garas A, Vargha P, Hollo G. Diagnostic accuracy of nerve fibre layer, macular thickness and optic disc measurements made with the RTVue-100 optical coherence tomograph to detect glaucoma. Eye (Lond) 2011;25:57-65.
5. Tatham AJ, Weinreb RN, Zangwill LM, Liebmann JM, Girkin CA, Medeiros FA. The relationship between cup-to-disc ratio and estimated number of retinal ganglion cells. Invest Ophthalmol Vis Sci 2013;54:3205-14.
6. Li S, Wang X, Wu G, Wang N. Evaluation of optic nerve head and retinal nerve fiber layer in early and advance glaucoma using frequency-domain optical coherence tomography. Graefes Arch Clin Exp Ophthalmol 2010;248:429-34.
7. Huang JY, Pekmezci M, Mesiwala N, Kao A, Lin S. Diagnostic power of optic disc morphology, peripapillary retinal nerve fiber layer thickness, and macular inner retinal layer thickness in glaucoma diagnosis with fourier-domain optical coherence tomography. J Glaucoma 2011;20:87-94.
8. Gonzalez-Garcia AO, Vizzeri G, Bowd C, Medeiros FA, Zangwill LM, Weinreb RN. Reproducibility of RTVue retinal nerve fiber layer thickness and optic disc measurements and agreement with Stratus optical coherence tomography measurements. Am J Ophthalmol 2009;147:1067-74, 74 e1.
9. Olmedo M, Cadarso-Suarez C, Gomez-Ulla F, Val C, Fernandez I. Reproducibility of optic nerve head measurements obtained by optical coherence tomography. Eur J Ophthalmol 2005;15:486-92.
10. Pomorska M, Krzyzanowska-Berkowska P, Misiuk-Hojlo M, Zajac-Pytrus H, Grzybowski A. Application of optical coherence tomography in glaucoma suspect eyes. Clin Exp Optom 2012;95:78-88.
11. Reis AS, Sharpe GP, Yang H, Nicolela MT, Burgoyne CF, Chauhan BC. Optic disc margin anatomy in patients with glaucoma and normal controls with spectral domain optical coherence tomography. Ophthalmology 2012;119:738-47.
12. Kotera Y, Yasuno Y, Hangai M, Inoue R, Makita S, Nakanishi H, Yamanari M, Yoshimura N. Comparison of spectral domain optical coherence tomography and color photographic imaging of the optic nerve head in management of glaucoma. Ophthalmic Surg Lasers Imaging 2009;40:255-63.
13. Kim NR, Kim JH, Kim CY, Jun I, Je Seong G, Lee ES. Comparison of the Optic Nerve Imaging by Time-Domain Optical Coherence Tomography and Fourier-Domain Optical Coherence Tomography in Distinguishing Normal Eyes From Those With Glaucoma. J Glaucoma 2011.
14. Zeyen T, Miglior S, Pfeiffer N, Cunha-Vaz J, Adamsons I. Reproducibility of evaluation of optic disc change for glaucoma with stereo optic disc photographs. Ophthalmology 2003;110:340-4.

15. Jampel HD, Friedman D, Quigley H, Vitale S, Miller R, Knezevich F, Ding Y. Agreement among glaucoma specialists in assessing progressive disc changes from photographs in open-angle glaucoma patients. Am J Ophthalmol 2009;147:39-44 e1.
16. Varma R, Steinmann WC, Scott IU. Expert agreement in evaluating the optic disc for glaucoma. Ophthalmology 1992;99:215-21.
17. Hanson S, Krishnan SK, Phillips J. Observer experience and Cup:Disc ratio assessment. Optom Vis Sci 2001;78:701-5.
18. Breusegem C, Fieuws S, Stalmans I, Zeyen T. Agreement and accuracy of non-expert ophthalmologists in assessing glaucomatous changes in serial stereo optic disc photographs. Ophthalmology 2011;118:742-6.
19. Altangerel U, Bayer A, Henderer JD, Katz LJ, Steinmann WC, Spaeth GL. Knowledge of chronology of optic disc stereophotographs influences the determination of glaucomatous change. Ophthalmology 2005;112:40-3.

Thesis summary

Summary

Glaucoma is an enigmatic disease, with various underlying pathophysiological processes contributing to the loss of retinal ganglion cells and the resultant structural and functional changes which characterise the condition. The key to effective glaucoma treatment protocols remains the early detection of the disease itself, and the earliest detection of any progression subsequent to the instigation of topical and/or surgical therapies.

Current clinical practice for the detection of glaucoma and glaucomatous progression is also multifactorial. Intra-ocular pressure measurement, central corneal thickness measurement, direct observation of the optic nerve head, retinal reflectivity, structural assessment using *spectral domain optical coherence tomography* (SD-OCT), and functional investigations using *standard automated perimetry* (SAP) provide a variety of inputs to assist in detection and treatment.

SD-OCT and SAP are pivotal strategies involved in detecting structural and functional change in glaucoma, but both demonstrate an inherent amount of *test-retest variability* (TRV). The main purpose of our study has been to determine the levels of TRV present in two commonly used SD-OCT and SAP instruments (Table 1). We have also set out to investigate whether any external factors might influence TRV, and whether any of these factors could be ameliorated in clinical practice settings.

In Chapter 1 of the current study, we investigated the TRV characteristics of a cohort of glaucoma subjects using the Medmont M700 automated perimeter. As a result of this study, we were able to develop easily applicable event based criteria to differentiate TRV from progression in the summary indices of the M700 (e.g. Table 1 below). We anticipate that these criteria will assist in the detection and management of glaucoma in patients all over the world.

Our investigations of TRV of the various M700 test protocols also revealed some hitherto unreported findings in regard to eccentricity and TRV. We were able to demonstrate, for the first time, that TRV using SAP did not increase with eccentricity for points of equal sensitivity (Fig. 1). This finding may have significant implications for automated progression detection algorithms currently utilised by some automated perimeters.

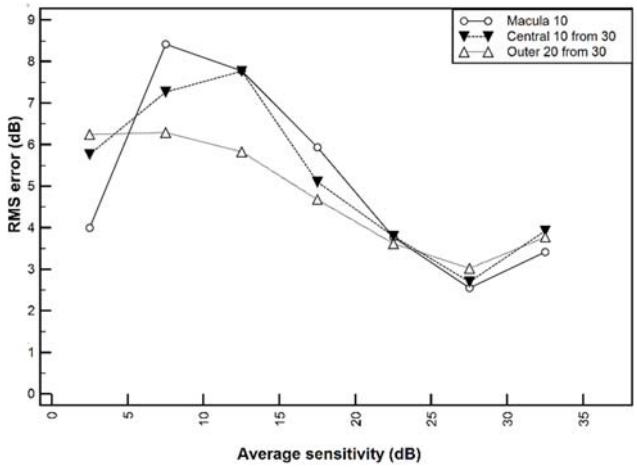


Figure 1.

Average test point RMS errors versus average sensitivity for the M700. The standard errors of the mean (SEM) were all less than 0.5 (median 0.287), with the exception of 5 results whose SEM were1.00, 0.800, 0.743, 0.720 and 0.663. Explanation of the legend:

-- Macula 10: 10°-test (excluding the central four points)
-- Central 10 from 30: the central 10° results from the 30°-test
-- Outer 20 from 30: the outer 20° results from the 30°-test

In Chapter 2, we investigated the TRV characteristics of a relatively new SD-OCT instrument, the Topcon 3D OCT-2000. Using a cohort of glaucoma subjects, we developed event based criteria for the main summary indices of this instrument (Table 1), facilitating the differentiation of TRV from true progressive change with this instrument. Perhaps the most critical finding was that TRV for the macular ganglion cell complex (mGCC) scan was lower than that of average peripapillary retinal nerve fibre layer (pRNFL) thickness. We anticipate that this finding may significantly enhance progression detection with this instrument by providing clinicians with TRV criteria for the two main summary indices.

Table 1.

Summary of event based progression criteria for the Medmont M700 and the Topcon 3D OCT-2000 derived from the current study

		Event based progression criteria
Chapter 1		
Medmont M700 (Central 30° test) (glaucoma patients)		
Overall Defect (dB)		2.4
Pattern Defect (dB)	≤ 2.8	1.24
	2.81 – ≤ 5.7	1.13
	> 5.7	3.1
Chapter 2		
Topcon 3D OCT-2000 (glaucoma patients)		
Disc area (mm^2)		0.27
Rim area (mm^2)		0.24
Horizontal cup/disc ratio		0.05
Vertical cup/disc ratio		0.07
Macula ganglion cell complex (μm)		2.9
Average retinal nerve fibre layer (μm)		± (26.67 μm – 0.294 (d)) (d = measured average retinal nerve fibre layer thickness)
Appendix 3		
Topcon 3D OCT-2000 (healthy patients)		
Part 1		
Average retinal nerve fibre layer (μm)		4.9
Part 2		
Macula ganglion cell complex (μm)		1.3
Part 3		
Disc area (mm^2)		0.36
Rim area (mm^2)		0.24
Horizontal cup/disc ratio		0.07
Vertical cup/disc ratio		0.08

Although the possibility of any diurnal influence on TRV was not suspected at the time, our studies into TRV in the Topcon and the M700 were designed to minimise the influence of any diurnal variations in TRV. We achieved this by conducting the tests on each subject at the same time of day at each session to minimise any possible diurnal influences. In Chapter 3, we conducted a study to

determine whether any clinically significant alterations in TRV might be present in SAP or SD-OCT tests conducted at different times during normal office hours.

The results from this study appeared to indicate that clinically significant alterations in TRV did in fact exist in relation to the time of day testing was undertaken, being largest in early afternoon for both methods. For the first time, we were able to quantify the effect of circadian variations in the various summary measures. Moreover, we were then able to develop clinically applicable guidelines for practitioners to introduce to their clinical protocols. Specifically, we were able to show that conducting tests in the early morning and late afternoon resulted in lower levels of TRV in both SAP and SD-OCT. Reducing TRV in both test protocols should result in better progression detection, and thus improved patient treatment outcomes.

In addition, we were also able to report that correlations between structure and function also appeared to vary throughout the day. This previously unreported aspect of the relationship between structure and function may have considerable implications for progression detection algorithms that utilise the integration of structural and functional parameters.

Following on from our findings in Chapter 3, we also investigated whether any circadian cycles or influences might be present in SAP and SD-OCT. In Appendices 1 and 2, we therefore carried out testing with both instruments every hour and a half for 24 hours, on two separate occasions. Whilst no significant circadian alterations to visual field indices were found in SAP (Appendix 1), factor analysis was able to demonstrate that two underlying factors were responsible for approximately 80% of the variance in the data. Factor loadings were quite consistent between eyes, and we were able to show that Factors 1 and 2 had quite strong effects on certain parameters under investigation. Given the time between test sessions of six months, this consistency was quite remarkable. The factor scores indicated that Factor 1 in particular showed large variation overnight but little during the day. The detection of these previously unreported factors, which may be contributing to SAP TRV, is both important and encouraging.

Some results seemed to be intriguing enough to warrant more extensive investigation, particularly the consistency of changes to visual field sensitivity at

3.30 am as reported in our 24 hour SAP investigation (Appendix 1). We believe that our pioneering investigation, with the first reported SAP testing over a 24 hour period, may lead to other studies investigating circadian aspects of structure, function, TRV and the investigation of underlying factors contributing to TRV as detected in Appendix 1.

Our findings with SD-OCT (Appendix 2) were also enlightening. We were able to detect statistically significant circadian variations in several average peripapillary retinal nerve fibre layer parameters. Factor analysis also showed that the first (as yet unexplained) factor (Factor 1) accounted for approximately 35% of the variance present in the 27 SD-OCT parameters under consideration. Factor loadings were able to demonstrate some highly consistent loadings with Factor 1 for optic disc parameters, although RNFL parameters demonstrated less consistency when factor loadings were analysed.

Whilst our findings in regard to the Topcon (Chapter 2) were important in their own right, we also decided to compare the TRV of this instrument to several other SD-OCT instruments. We felt it would be instructive to know whether the level of TRV in the Topcon was comparable to other SD-OCT's. In order to accomplish this, we therefore compared the Topcon with three other SD-OCT's: the Cirrus HD-OCT, the Optovue RTVue 100 and the Nidek RS-3000. To isolate machine variance form other sources of variance, our study protocol was designed to reduce other sources of variance by conducting many repeats on a few subjects at fixed times.

Table 2.

Summary of event based progression criteria for the other SD-OCT instruments investigated in the thesis (Nidek RS 3000, Optovue RTVue 100, Cirrus HD-OCT)

	Event based progression criteria
Nidek RS-3000	
Average retinal nerve fibre layer (μm)	9.5
Disc area (mm^2)	1.22
Rim area (mm^2)	1.23
Horizontal cup/disc ratio	0.23
Vertical cup/disc ratio	0.18
RTVue 100	
Average retinal nerve fibre layer (μm)	5.4
Disc area (mm^2)	0.03
Rim area (mm^2)	0.21
Horizontal cup/disc ratio	0.18
Vertical cup/disc ratio	0.12
Macula ganglion cell complex (μm)	4.7
Cirrus HD-OCT	
Average retinal nerve fibre layer (μm)	4.4

In Appendix 3 (Part 1), we compared average pRNFL TRV between the Topcon, the Nidek, the RTVue and the Cirrus. Our findings (Tables 1 and 2) showed that the Topcon demonstrated the same levels of repeatability of the two well studied instruments, the RTVue and the Cirrus. Whilst attempting to develop TRV criteria to assist clinicians comparing scans from different instruments, we realised developing such criteria was dependent on the mean difference between the measurements on different instruments. It became apparent that many different studies, and the instruments normative databases themselves, often gave different values for average pRNFL population thicknesses, and thus many different mean differences could be derived. The inconsistency of these values made it difficult to apply confidence intervals between machine measurements, given that no single definitive machine average pRNFL measurement existed.

As we had reduced subject variance in our test protocol, the results from our study would appear to be the first to quantify the true difference between instrument measurements for the parameter of average pRNFL. The creation of a true mean difference between measurements then enabled the application of the Limits of Agreement between instruments we developed, enabling clinicians, for the first time, to be able to accurately determine whether true progression had taken place when comparing pRNFL scans taken on different instruments.

In Appendix 3 (Part 2) we found that the TRV in the Topcon for mGCC scans (Table 1) was lower than that of the RTVue 100 (Table 2). Critically, for clinicians attempting to compare results obtained on one instrument with those obtained on the other, we found that there was a significant difference in the measurement of mGCC thickness in these instruments (12.4 μm). Taking this mean difference into account, we were also able to develop TRV criteria for measurements obtained on these instruments to further assist clinicians comparing test results to be able to more accurately determine the presence, or otherwise, of true progressive change.

In Appendix 3 (Part 3), we examined the TRV of the Topcon when assessing optic nerve head topographical features, and compared this to the TRV of the RTVue 100 and the Nidek for the same parameters. Once again, the Topcon demonstrated comparable TRV to the other two instruments (Tables 1 and 2). The repeatability of the Topcon appears to make the objective determination of early structural changes in disc topography an important adjunctive measure to subjective clinical observation. Disc topography assessment with the Topcon is therefore a useful adjunctive measure with which to assess and manage glaucoma suspects and glaucoma patients.

Future directions

The results from these studies into TRV in SAP and SD-OCT have provided the basis for continuing research into other aspects of TRV and glaucoma progression protocols. Perhaps the most critical aspect for future studies would be to determine whether TRV varies in relation to the time of day in glaucoma subjects (as well as in normal controls as shown here). Any findings in this regard may lead to a refinement of glaucoma progression detection criteria,

both in SAP and SD-OCT, and assist in the earlier determination of true progression. Any investigation of this nature would require a large cohort of glaucoma subjects (to enable a reasonable number of subjects with differing degrees of disease severity at each time point).

Concurrently, the identification of any underlying factors which may be influencing TRV may also be pivotal in providing further understanding of the pathophysiology of glaucoma. We envisage a comprehensive study which would encompass a large range of ocular parameters in order to determine which factors may be positively associated with diurnal variations in TRV.

In view of our findings that TRV for points of equal sensitivity, using SAP, does not increase with eccentricity, a review of current progression detection criteria incorporated in automated perimeters may be enlightening. It is a distinct possibility that incorporating our findings into future instrument progression detection algorithms may facilitate earlier and/or more accurate progression detection in these instruments.

Conclusions

Whilst individual studies can rarely provide seismic changes to glaucoma detection and management protocols, some of the findings derived from the current study appear to have the potential to appreciably alter some aspects of glaucoma diagnostic and management protocols. Perhaps more importantly, these findings may stimulate continued research into aspects of TRV which had not been previously recognised. It is our hope that the results from our research has positively contributed to the detection, diagnosis, management and treatment of this potentially blinding disease.