University of Bradford eThesis

EN FACE OCT IMAGING FOR THE ASSESSMENT OF GLAUCOMA

R. CHELONI

PhD

2021

En face OCT imaging for the assessment of glaucoma

Riccardo CHELONI

Submitted for the Degree of

Doctor of Philosophy

School of Optometry and Vision Science

Faculty of Life Sciences

University of Bradford

2021

Abstract

Riccardo Cheloni

En face OCT imaging for the assessment of glaucoma

Keywords: glaucoma, en face imaging, optical coherence tomography, retinal nerve fibre layer.

Glaucoma is a leading cause of irreversible vision loss globally, and demands early and accurate diagnosis. OCT has become a key investigative technique in glaucoma, and, although it provides invaluable clinical support, detection of early glaucoma remains imperfect. Recent OCT developments enabled direct assessment of retinal nerve fibre bundle (RNFB) reflectance in en face OCT images. The technique has considerable potential in the assessment of glaucoma, yet it has limited clinical usability due to an incomplete understanding of RNFB features in healthy and glaucoma eyes and the lack of accepted methods to identify reflectance defects. This thesis aimed to better understand characteristics of RNFB reflectance in en face OCT imaging and to develop objective methods to extract defects in this domain.

Structural and functional measures of glaucoma changes were collected in eyes with established glaucoma and age-similar controls. Results showed that the healthy configuration of RNFB varies across the retina and between different eyes. We developed a method for automated and objective examination of reflectivity changes in en face images. This method considers individual anatomy and varying RNFB configuration, and found more abnormalities than previous approaches. Measures of en face reflectance and conventional retinal nerve fibre layer thickness were strongly related. The agreement between changes of reflectance and visual function was moderate-to-good, and both testing domains presented concordant abnormalities in all tested eyes.

Following further minimisation of artefacts in en face images, direct use of reflectance analysis or its combination with perimetry appear viable and with significant potential for clinical examination of glaucoma.

Acknowledgements

I want to take the opportunity to express my gratitude to the many people who knowingly or unknowingly enabled me to undertake and complete this PhD.

I would like to thank all participants who gave their time to take part in our studies and made this project possible. A huge thanks goes to my supervisor Dr Jonathan Denniss, for having considered my application for this post and granting me the opportunity to work on this project. I have been privileged to undertake my PhD under his supervision and to get to know and work alongside a fantastic person, before than a great supervisor. He never failed to give me the right amount of challenge and also to provide full support, being this in the form of educational or professional development or directed to address any private matter (including any national park's walking itineraries). I wish him to remain as he is in his years to come, and to never lose the capability of putting himself in others' positions.

Thanks to all my fellow postgrads, with whom I have shared fantastic moments (and coffees) at the university. It was a privilege to work alongside all of you. A particular mention to my fabulous colleague Dr Habiba Bham, for her invaluable help at all stages of this project. I express my sincere thanks to my second supervisor Prof Marina Bloj, and to Mr Simon Dewsbery (Leeds Teaching Hospital) for helping with the data collection. I also wish to thank Prof Andrew Turpin (University of Melbourne) for assistance with OCT data extraction. I am grateful to all BSOVS staff members for having created a friendly and stimulating working environment. In particular, thanks to Brendan (Prof Barrett), Ian (Dr Pacey), James (Dr Heron) and Chris (Dr Davey) for all binocular vision chats, bike rides, wine tastings and pub peer discussions which made my years in Bradford a truly unforgettable time. I am also grateful to Dr Annette Parkinson, for her support and contribution to my further optometric training.

Thanks to my parents, my brothers, all other family members and friends in Italy, the UK, and elsewhere. Throughout these years of academic studies, they have always supported me, even when a bit neglected at times. The

Table of contents

List of figures

List of tables

List of abbreviations

ACG   Angle Closure Glaucoma

AllVis   All visible RNFBs slab

AUC   Area Under (receiver operating characteristic) Curve

BestVis   Best Visibility slab

BMO   Bruch's Membrane Opening

BMO-MRW Bruch's Membrane Opening – Minimum Rim Width

CCT   Central Corneal Thickness

CI   Confidence Interval

cpRNFL   circumpapillary Retinal Nerve Fibre Layer

GAT   Goldmann Applanation Tonometry

GCC   Ganglion Cell Complex

GCIPL   Ganglion Cell Inner Plexiform Layer

GCL   Ganglion Cell Layer

GHT   Glaucoma Hemifield test

ICC   Intraclass Correlation Coefficient

ILM   Inner Limiting Membrane

IOP   Intraocular Pressure

IPL   Inner Plexiform Layer

MD   Mean Deviation

OAG   Open Angle Glaucoma

OCT   Optical Coherence Tomography

ONH   Optic Nerve Head

pAUC   partial Area Under (receiver operating characteristic) Curve

POAG   Primary Open Angle Glaucoma

PSD   Pattern Standard Deviation

RGC   Retinal Ganglion Cell

RNFB   Retinal Nerve Fibre Bundle

RNFL   Retinal Nerve Fibre Layer

ROC   Receiver Operating Characteristic

SAP   Standard Automated Perimetry

SD   Standard Deviation

SD-OCT   Spectral Domain OCT

SITA   Swedish Interactive Threshold Algorithm

SLO   Scanning Laser Ophthalmoscope

SMAS   Summary of Multiple Anatomically-adjusted Slabs

VF   Visual Field

1. Introduction

1.1 Glaucoma

1.1a Types of glaucoma

The term Glaucoma encompasses a heterogeneous group of progressive optic neuropathies, all associated with characteristic structural changes affecting the optic nerve head (ONH) and visual sensitivity (Foster et al., 2002, Jonas et al., 2017). While in the past, clinicians and researchers had the tendency to define glaucoma according to the presence of high intraocular pressure (IOP) (Cedrone et al., 2008), it is now clearly established that a multitude of factors in addition to raised IOP determine the disease (Blumberg et al., 2015).

An initial differentiation in glaucoma subgroups can be made with regard to the morphology of the anterior chamber angle (Weinreb et al., 2016), which is defined by the relationship taken between the anterior surface of the iris and the peripheral cornea at the limbus (Weinreb et al., 2014). The angle itself is directly connected to Schlemm's Canal through the trabecular meshwork and represents the access to the main source of aqueous humour drainage. This, combined with the uveo-scleral outflow route allows retention of the IOP within normal limits (Jonas et al., 2017). According to the width of the angle, glaucoma can be classified as Open Angle (OAG) or Angle Closure Glaucoma (ACG) (Jonas et al., 2017, Weinreb et al., 2016). The former describes a condition of free access of the aqueous humour to the trabecular meshwork. In OAG, the fault to the drainage is thought to be caused by an increased internal resistance to the outflow which can be either idiopathic, as in the case of Primary Open Angle Glaucoma (POAG), or induced by a detectable cause such as the accumulation of pigment or exfoliative material (i.e. Secondary Open Angle Glaucoma) (Weinreb et al., 2014). On the contrary, when the aqueous humour outflow is blocked at the angle, the patient may present ACG. With respect to the causative mechanism of angle closure, ACG can be either primary or secondary (Jonas et al., 2017). Again, in secondary ACGs there is an additional condition (e.g. new blood vessels at the iris) which is causing the

occlusion of the angle. Figure 1.1 shows different configurations of the angle in healthy eyes, primary OAG and ACG.
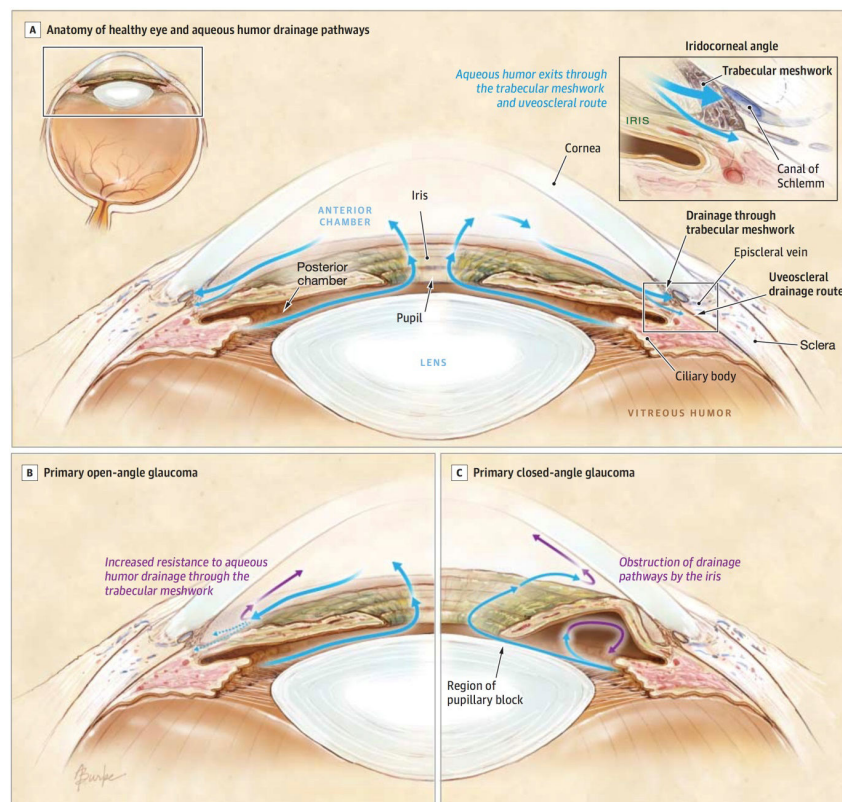
Figure 1.1 Anatomy of the anterior and posterior chamber in healthy eyes (a) and in different types of primary glaucoma: open angle (b) and closed angle (c). Reproduced from (Weinreb et al., 2014).

Both OAG and ACG are typically acquired disorders, usually affecting patients in adulthood (Weinreb et al., 2014). Nonetheless, congenital glaucoma has also been described. In these cases the outflow route through the trabecular meshwork and the Schlemm's canal is often only partially developed at birth, resulting in IOP increase (Ko et al., 2015).

As anticipated above, the role of IOP in OAG has now been reconsidered to be only a major risk factor and not the primary causative factor (Blumberg et al., 2015). Indeed, it is now recognised that typical OAG lesions, such as neuro-retinal thinning and disc haemorrhages, can be documented with normal IOP values, defining the so-called normal tension glaucoma (Drance et al., 2001). Normal tension glaucoma can be considered as an open-angle glaucomatous neuropathy with IOP within 21mmHg, i.e. within 'normal' values (Killer and Pircher, 2018). Even though several mechanisms are likely to contribute to the disease, the condition is far from understood and further

research is required (Killer and Pircher, 2018). Besides normal tension glaucoma, patients also present with anomalously high IOP values, yet without any detectable glaucomatous damage (Weinreb et al., 2016). These patients are encompassed in the ocular hypertension group (Schulzer et al., 1998, National Institute of Health and Care Excellence, 2017).

Among the many glaucoma subtypes, this thesis concerns open angle glaucoma, of which POAG is the most common form (Yamamoto et al., 2014).

1.1b POAG epidemiology

Glaucoma is among the main causes of irreversible blindness (Bourne et al., 2018, Flaxman et al., 2017, Steinmetz et al., 2021), leading to considerable interest in understanding the epidemiology of this disorder from the early decades of the previous century (Cedrone et al., 2008). However, data from studies conducted between 1920 and 1970 are likely relevant to ocular hypertension rather than glaucoma, since in those days the disease was regarded as synonymous of raised IOP (Cedrone et al., 2008). The first study performing visual field (VF) testing on every single participant enrolled dates back to 1980 (Bengtsson, 1981). Nonetheless, performing perimetry on all participants studied had still been a rare approach in the following years, with most of the less recent studies not assessing VF on all the included sample (Cedrone et al., 2008). In 2014 it was estimated that around 3.5% of the 40-80 aged global population was affected by any type of glaucoma (Tham et al., 2014). Of these cases, more than 80% were ascribable to POAG, whereas only 15% were reported as ACG (Tham et al., 2014).

POAG prevalence

Among the latest study reporting POAG prevalence from a global perspective (Figure 1.2), Kapetanakis and colleagues reported in 2015 that an estimated 57 million people suffered from POAG globally, corresponding to a 2.2% prevalence (Kapetanakis et al., 2016). POAG's prevalence in Europe was estimated to be around 2.0% in 2015, affecting almost 8 million people (Kapetanakis et al., 2016). According to the study group, Europe's figures are expected to remain stable in the following years, whereas Latin America, Asia

and Africa will face an increase in POAG cases due to a steadier aging of the population in these countries (Kapetanakis et al., 2016).

Figure 1.2 Global prevalence of POAG in adults over 40 years of age, color-coded according to different regions. Reproduced from (Weinreb et al., 2016), drawing data from (Kapetanakis et al., 2016).

Inherent characteristics of POAG, namely the lack of symptoms at the earlier stages may result in late detection of the disease, unless opportunistic screening of asymptomatic patients is performed (Hohn et al., 2018, Chan et al., 2017). It should be underlined that also in high income countries POAG remains largely undetected. Data from a recent systematic review considering only population based studies found 66% of people with glaucoma to be undetected in Europe (Soh et al., 2021). Most up-to-date data in these settings are from a population based study conducted in Northern Ireland, where 67% of POAG cases did not self-report to be affected by this disease (McCann et al., 2020). Undetected glaucoma has considerable clinical implications, since it will result in late presentation, ultimately leading to considerable vision loss at diagnosis. For instance, data from British studies indicates that up to 20% of glaucoma patients might present advanced VF loss (mean deviation (MD) of -12dB or worse) in the worst eye at their first hospital visit (Jones et al., 2020, Boodhna and Crabb, 2015).

POAG risk factors

Although some factors are still debated, epidemiologic studies performed so far have identified a number of risk factors for the development of POAG,

thoroughly summed in a number of published reviews (McMonnies, 2017, Zhao et al., 2015, Jain et al., 2017, Blumberg et al., 2015, Kapetanakis et al., 2016, Leske, 2007, Hollands et al., 2013, Coleman and Miglior, 2008). Table 1.1 at the end of this section summarises the major risk factors for POAG as reported in recently published guidelines and reviews.

- **Age**

  Increasing age is a well-established risk factor for glaucoma (Leske, 2007, Chan et al., 2017, McMonnies, 2017, Hollands et al., 2013, Coleman and Miglior, 2008).

- **Ethnicity**

  Varying rates of POAG prevalence have been consistently reported in population-based studies according to different ethnicities, and those with African ancestry are up to 3 times more at risk of developing POAG (Tham et al., 2014).

- **Family-History**

  A positive family history is a well-established risk factor of POAG (Leske, 2007, McMonnies, 2017, Hollands et al., 2013, Khawaja and Viswanathan, 2018). Patients could be up to 110% more likely to develop OAG in case of 1st degree relative affected by glaucoma (McMonnies, 2017).

- **IOP**

  High IOP represents the major modifiable risk factor for POAG (Chan et al., 2017, Leske, 2007, Blumberg et al., 2015, Coleman and Miglior, 2008). The relationship between OAG and IOP is clearly depicted in incidence studies, where the relative risk of developing the condition can increase more than tenfold in eyes with baseline IOP>25mmHg compared to 17-19mmHg ranges (Leske, 2007).

- **Myopia**

  Short-sightedness is frequently described as a considerable risk factor for the development of OAG (Leske, 2007), with a positive correlation between the degree of myopia and POAG risk itself (McMonnies, 2017, Hollands et al., 2013). Yet, it should be noted that this relationship was mainly detected in cross-sectional studies while many longitudinal

investigations have failed to confirm myopia to be a risk factor for glaucoma or its progression (Gordon et al., 2002, Koh et al., 2021, Leske et al., 2004, Founti et al., 2020). Accordingly, more research is needed to fully understand the nature of this relationship.

- **Corneal-Biomechanical-Properties**

  A number of studies have reported a lower central corneal thickness (CCT) and higher corneal stiffness as risk factor for the development of POAG (Belovay and Goldberg, 2018, Susanna et al., 2019, Qassim et al., 2020).

Additional risk factors are frequently reported in the literature, including ocular perfusion pressure, systemic vascular risk factors such as diabetes and systemic hypotension and hypertension (systolic blood pressure greater than 160 mm Hg), migraine, and gender (Grzybowski et al., 2020, Kim et al., 2020). The relationships between these factors and glaucoma onset and/or progression remain less clear and still debated.

Table 1.1
Major and recognised risk factors for POAG as reported from recent guidelines and reviews. The green circle (O) represents a risk factor which has been reported in the cited document, whereas the red cross (X) indicates that the factor was not reported.

Risk Factor	(Jonas et al., 2017)	(Weinreb et al., 2016)	(Prum et al., 2016)	BMJ Best Practice (Amerasinghe, 2018)	(The College of Optometrists, 2018)
Older Age	O	O	O	O	O
Raised IOP	O	O	O	O	O
Black/African descent	O	O	O	O	O
Positive family history	O	O	O	O	O
Medium/high myopia (>3 - 6 SD)	O	O	O	O	O
Thin CCT (<555µm)	O	O	O	O	O
Male Gender	X	O	X	X	X
Diabetes	X	X	O	X	O
Low Ocular Perfusion	X	X	O	X	X

1.2 POAG in the clinic

POAG primarily affects the axons of retinal ganglion cells (RGCs) at the optic disc and is currently assessed through the examination of the ONH integrity and its related measures (e.g. retinal nerve fibre layer (RNFL) integrity and/or thickness). Additionally, the measurement of visual function through VF testing is a mainstay in glaucoma care. The aforementioned testing strategies can be described as the structure and function domains of glaucoma, respectively.

1.2a POAG clinic

Case presentation: signs of damage

Timely diagnosis of POAG based on patient presentation is often missed because of two main characteristics of the glaucomatous damage, which is i) painless, and ii) affects the visual function to a symptomatic level only at moderate to advanced stages of the disease (Jonas et al., 2017, Weinreb et al., 2014). As a consequence, patients usually seek eye care when the disease is progressed (Jonas et al., 2017, Crabb et al., 2017). Diagnosis of POAG is made when the typical RNFL and ONH changes are detected (Weinreb et al., 2014). A corresponding VF defect is often associated, albeit is not necessarily required for the diagnosis (Weinreb et al., 2016, Prum et al., 2016). Identification of POAG relies on a combination of information derived from multiple examinations, which are reported below.

Examination of the Optic Nerve Head and the Retinal Nerve Fibre Layer

Examination by mean of the slit lamp is reported as the standard technique for the assessment of the ONH and the surrounding ocular structures (Weinreb et al., 2016), and a number of signs have been described according to the appearance of those structures in POAG (Jonas et al., 2017, Weinreb et al., 2016, National Institute of Health and Care Excellence, 2017, Prum et al., 2016).

- Loss, narrowing and notching of the neuroretinal rim.
- Optic cup enlargement and a consequent increase of the cup-disc ratio, with especially the vertical ratio reported to be affected (Tuulonen and Airaksinen, 1991, Jonas et al., 1999). Observing an increment of the

baseline cup-disc ratio represents a more meaningful sign of glaucoma than a large value itself (Amerasinghe, 2018).

- Deepening of optic cup which determines a growing excavation of the ONH.
- Thinning of the RNFL.
- Development or enlargement of beta zone peripapillary atrophy.
- Optic disc haemorrhages, typically small and flame-shaped.
- Nasalisation of central ONH vessels.
- Baring of circumlinear vessels.
- Increase of neuroretinal rim pallor.

The examination of the ONH and the RNFL is a corner stone in glaucoma investigation, providing solid and cogent information about any structural damage (Prum et al., 2016). Although slit lamp examination is described as the gold standard technique, considerable drawbacks exist when compared with newly developed and objective technologies. As will be observed in greater detail in section 1.3, a major limitation is the high between-observer variability, particularly in the early stage of the condition (Jonas et al., 1999, Abrams et al., 1994, Hadwin et al., 2013). This leads to considerable disagreement between graders in the detection and grading of ONH defects, even when considering glaucoma specialists (Jampel et al., 2009). Major challenges in the accurate detection of ONH damage by clinical observation arise from the considerable anatomical variations of the dimension and aspect of the ONH in healthy subjects (Weinreb et al., 2016, Weinreb et al., 2014, Jonas et al., 1999).

A substantial support to clinicians has derived from the development of more objective ONH and RNFL imaging techniques which allow for an accurate and reproducible quantitative approach in the assessment of glaucoma (Weinreb et al., 2016, Weinreb et al., 2014). Although several instruments have been proposed, the scan speed of optical coherence tomography (OCT), combined with its enhanced image resolution has prompted the majority of glaucoma clinicians to the use of this instrument to quantify structural lesions (Weinreb et al., 2016, Prum et al., 2016).

Visual field testing: Perimetry

It is frequently reported that structural defects need to be progressed, and a considerable number of RGCs be loss or damaged, before a glaucomatous defect could be detectable in standard perimetry (Kerrigan-Baumrind et al., 2000, Harwerth et al., 1999, Quigley et al., 1989). For instance, some reports suggest that OCT-measurable thinning of the RNFL could anticipate VF defects by up to 8 years (Kuang et al., 2015). Although these claims are still criticised and might suffer some overestimates (Hood, 2019), the presence of reduced visual sensitivity measured by perimetry is not always a required criteria for glaucoma diagnosis, i.e. pre-perimetric glaucoma (Weinreb et al., 2016). Notwithstanding, perimetry allows a psychophysical quantification of glaucomatous defects and albeit presenting imperfections, is an irreplaceable mean of documenting and monitoring any change of visual function in this disease (Jonas et al., 2017, Weinreb et al., 2016, National Institute of Health and Care Excellence, 2017, Prum et al., 2016). As it will be discussed in detail in section 1.4, the reference standard approach to VF examination is standard automated perimetry (SAP), typically measuring sensitivities in the central 24-30°, on each side of fixation (Weinreb et al., 2016, Prum et al., 2016).

Tonometry

Even though raised IOP is not a requirement in the diagnosis (Weinreb et al., 2016), and is likely to be excluded in present and future definitions of glaucoma (Quigley, 2018), IOP is a central risk factor for the development and progression of the disease (Jonas et al., 2017). More importantly, its reduction is the only proven treatment of glaucoma (Jonas et al., 2017) and, therefore, IOP assessment is a mainstay in glaucoma clinic (Jonas et al., 2017, National Institute of Health and Care Excellence, 2017, Prum et al., 2016). The most widely accepted instrument for the measurement of IOP is Goldmann Applanation Tonometry (GAT), and its usage is recommended in patients with glaucoma (Prum et al., 2016). Although GAT is largely accepted, the obtained values are influenced by biomechanical corneal properties (Weinreb et al., 2016). More precisely, softer or thinner corneas tend to yield underestimated

IOP values. On the other hand, thicker or stiffer corneas lead to GAT readings that overestimate the actual IOP value (Belovay and Goldberg, 2018).

Central Corneal Thickness (CCT)

This measurement has largely become recommended in the care of glaucoma patients (Prum et al., 2016, National Institute of Health and Care Excellence, 2017). Determination of CCT allows for both a better interpretation of the IOP value and the assessment of a risk factor that is suggested to be potentially independent to IOP (National Institute of Health and Care Excellence, 2017). The value of 555µm was identified as lower limit of normality, below which there is an increased risk of POAG, as reported in studies predominantly assessing Caucasian eyes (Prum et al., 2016).

Gonioscopy

It is only after the assessment of the peripheral anterior chamber configuration, by gonioscopy, that a diagnosis of POAG rather than other forms of glaucomas can be made (Jonas et al., 2017, National Institute of Health and Care Excellence, 2017, Prum et al., 2016). In fact, secondary mechanisms of reduced aqueous drainage (e.g. exfoliative material or pigment) might be detected on examination (Prum et al., 2016). Alternatively, the angle itself could be closed.

1.2b Treatment

The overall goal of glaucoma care is the retention of the individual's quality of life through the preservation of vision (Prum et al., 2016). To date, this is achieved by impeding or minimising further glaucoma progression via the administration of treatments aimed at lowering the IOP (Jonas et al., 2017, Weinreb et al., 2016, Conlon et al., 2017). IOP reduction is thus far the only proven treatment to minimise further glaucomatous damage (Garway-Heath et al., 2015), with evidence from randomised controlled trials indicating the incidence of progressive POAG to be significantly reduced (Prum et al., 2016, Garway-Heath et al., 2015). Furthermore, reduction of glaucoma progression has been demonstrated across all stages and subgroups of POAG (Weinreb et al., 2014). Patients with ocular hypertension, early-moderate or advanced

POAG, as well as those with normal tension glaucoma, all showed a reduced progression rate following IOP lowering treatments (Weinreb et al., 2016). Once glaucoma is detected, clinicians managing the patients will set a certain amount of IOP reduction, which if achieved over time aims for a stable structural and functional status (Prum et al., 2016).

Several elements will be considered when a decision on whether or not to treat a specific patient is made, including stage of the disease, the rate and the risk of progression, and lastly the patient's life expectancy (Weinreb et al., 2016). Once treatment is advocated, the required magnitude of IOP reduction is selected by the clinician on the basis of individual specific factors (Jonas et al., 2017, Prum et al., 2016, Conlon et al., 2017). This evaluation allows to define a target IOP, at which the rate of progression is expected to be significantly reduced (Weinreb et al., 2014), and further VF loss should be unlikely to result in reduced quality of life over the remaining lifetime (Prum et al., 2016). Once the treatment has commenced, follow-up visits aim to monitor IOP and reassess the target value in light of signs of progression at the ONH, RNFL and VF (Jonas et al., 2017, Prum et al., 2016).

A number of IOP-lowering treatments are currently available, usually classified into medical (largely administered using eye-drops), laser-based and surgical. Eye drops, and especially prostaglandin analogues, are reported as the most effective medical treatment for lowering IOP, and have been the first-line treatment for many years in POAG (Conlon et al., 2017, Qureshi et al., 2021). However, this conventional paradigm could be reorganised in the close future. Indeed, recent evidence showed that laser treatment of the trabecular meshwork (i.e. Selective Laser Trabeculoplasty) may be an effective first line treatment, with comparable cost effectiveness to prostaglandin analogous (Gazzard et al., 2019). The treatment aims at an increased permeability of the trabecular meshwork, and would offer many advantages over daily administration of drops, including reduction of compliance issues and local and general side effects (Gazzard et al., 2019, Qureshi et al., 2021).

1.2c Diagnosis of glaucoma

Glaucoma case identification encompasses a number of signs currently used to determine the presence of the condition. The following sections will focus on the performance of different testing strategies in glaucoma diagnosis, resulting in a large use of statistical indices linked to categorisation performance. A brief review of these indices is reported below.

Classifiers and diagnostic accuracy measures

Differentiating between heathy and diseased patients is among the most relevant tasks in health care (Kumar and Indrayan, 2011, Hajian-Tilaki, 2013), and the necessity of comparing different diagnostic tests in this context is a common circumstance. For instance, multiple tests might become available for the detection of a disease. Alternatively, a new procedure could be proposed for the identification of a certain pathology, at reduced costs and invasiveness for the patient. In both eventualities, statistical procedures are required to perform a quantitative comparison of the performance of the available procedures (Saunders et al., 2015).

A classification system, or classifier, is a test (i.e. index test), used to predict the actual status of a group of subjects (Fawcett, 2006, Linnet et al., 2012). According to different testing procedures, either a discrete outcome (e.g. positive/negative) or a continuous measure can be the output. In the latter case a threshold value must be applied to differentiate between a positive and a negative prediction (Fawcett, 2006, Hajian-Tilaki, 2013). Following the comparison between the actual status of a patient and the prediction of the index test, a contingency table (Figure 1.3) can be drawn for a primary visualisation of diagnostic performance (Tharwat, 2018, Fawcett, 2006, Mallett et al., 2012). The true status is typically assessed by a reference standard test, with this term being increasingly preferred to the term 'gold standard' due to potential uncertainty around validity (Trevethan, 2017). The reference standard could be a more accurate but also more expensive and/or invasive test, a combination of tests, or evidence available from longitudinal clinical examination or a panel of experts (Kumar and Indrayan, 2011, Hajian-Tilaki, 2013). When a person with the disease is correctly identified by the index test,

this is defined as a true positive. On the other hand, if a negative test result was obtained it would have been classified as false negative. While considering healthy persons, true negatives are identified in case of a correct prediction, whereas patients classified positive by the test result in a false positive (Saunders et al., 2015, Kumar and Indrayan, 2011). Consideration must be paid to the two different types of error which could be performed when examining a patient with any index test. A positive result in a healthy patient (i.e. false positive) could be considered as a false alarm (Saunders et al., 2015). In glaucoma, this may produce an incorrect referral, adversely affecting both the patient, in terms of anxiety and concerns, and the healthcare system because of the waste of resources. Furthermore, a false positive could result in overtreatment of healthy patients, with obvious burden associated (Saunders et al., 2015, Karvonen et al., 2018). On the other hand, missing the identification of patients truly affected by the disease (i.e. false negative) carries consequences which might be highly detrimental according to different diseases. Generally, a patient could be incorrectly reassured, and timely treatment might not be administered at earlier and more favourable stages (Saunders et al., 2015).
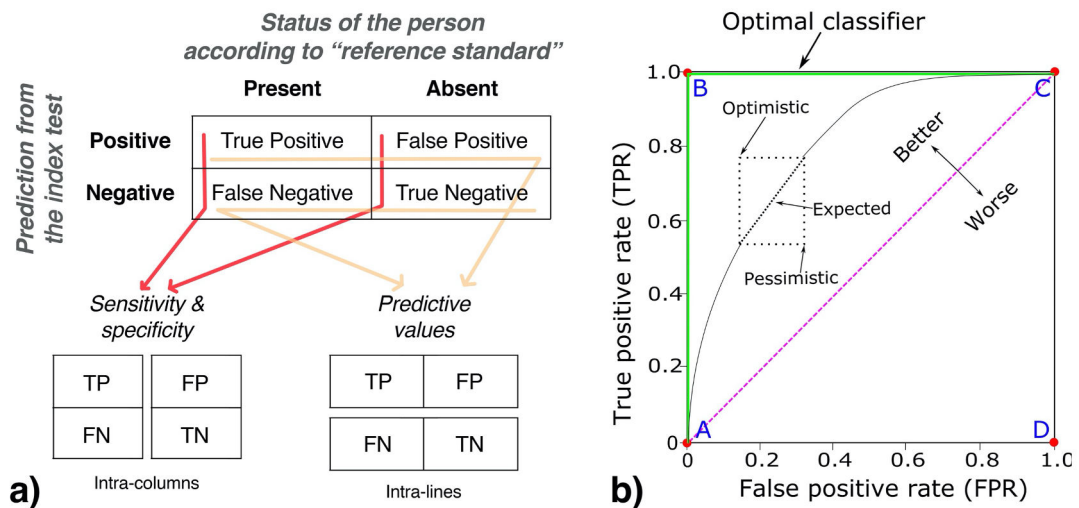
Figure 1.3 Main diagnostic accuracy statistical measures. In (a), a 2x2 contingency table shows the different possible classifications. Indices such as specificity and sensitivity can be derived by calculation intra-columns, whereas predictive values are derived from calculation intra-lines. In (b), ROC curve showing (in green) the ideal classifier line and in purple the results of mere chance classification. Adapted from (Trevethan, 2017, Tharwat, 2018). TP = True Positive, FN = False Negative, FP = False Positive, TN = True Negative.

The algebraic combination of all the subclasses identified in 2x2 tables determines several indices which are used to describe the diagnostic accuracy of the test of interest. Among the others, sensitivity and specificity are reported as the most widely used measures (Hajian-Tilaki, 2013). Sensitivity defines the proportion of subjects with the disease who have been correctly classified with the test. Conversely, specificity, also called true negative rate, is equal to the ratio between true negative and all subjects without the disease (Tharwat, 2018, Saunders et al., 2015, Trevethan, 2017). The former allows us to measure the accuracy of the test among people with the disease, whereas the latter gives indication among healthy people (Tharwat, 2018). Measures strictly related to sensitivity and specificity are the false positive rate and the false negative rate, respectively defined by 1-specificity and 1-sensitivity.

A thorough understanding of the diagnostic performance of a test, which also considers different cut-off points for continuous outcomes, can be achieved with receiver operating characteristic (ROC) analyses (Fawcett, 2006, Mallett et al., 2012). As shown in Figure 1.3b, this is performed by generating plots of sensitivity (true positive rate) against the false positive rate (i.e. 1-specificity), on the y and x axis respectively (Tharwat, 2018, Fawcett, 2006, Hajian-Tilaki, 2013, Linnet et al., 2012). ROC curves enable simultaneous overview of sensitivity and specificity of a given test and the diagnostic capability can be observed at a glance from the localisation of the curve in the ROC space. Overall, better performance is shown by points located close to the top-left region of the ROC space (Hajian-Tilaki, 2013). The top left corner (maximum sensitivity and specificity) indicates the circumstance where all subjects with and without the condition are correctly classified by the index test. This point is associated to the ideal classifier, described by the green line in Figure 1.3b. On the opposite extreme, points lying on the diagonal line (purple in Figure 1.3b) pertain to a test performing similar to chance classification, where only ~50% of subjects with or without the condition are correctly classified.

While a detailed picture of diagnostic performance can be achieved by visualising ROC curves, practicality in the comparison of different tests often requires single quantitative measures (Fawcett, 2006). The majority of diagnostic accuracy studies in glaucoma employed the area under (receiver

operating characteristic) curve (AUC) to this end (Chen et al., 2018). AUC is the area underneath a specific ROC curve (Hanley and McNeil, 1982), and it ranges between 0.5 and 1 (Tharwat, 2018, Fawcett, 2006), with 1 representing the perfect classifier with neither false positives nor false negatives. As a rule of thumb, a test can be considered excellent when presenting an AUC between 0.9 and 1 and good when AUC is between 0.8 and 0.9. Tests with AUC between 0.7-0.8 and 0.6-0.7, are considered to provide fair and a poor diagnostic capability, respectively (Chen et al., 2018, Fawcett, 2006).

Limitations of diagnostic tests

Ideally, only tests with 100% sensitivity and specificity would be available in clinics. However, this is seldom the case in reality where classifiers are at their best only able to minimise the overlap between the distributions of healthy and diseased patients. As a consequence, suboptimal levels of sensitivity and specificity are found and a trade-off between the two must often be selected. One test could perform similar to another, with slightly better specificity and worse sensitivity or vice versa (Saunders et al., 2015).

While evaluating the most valuable test, it is crucial to consider the clinical context in which the test will be applied (Mallett et al., 2012, Linnet et al., 2012), as misclassification of healthy and diseased patients is unlikely to be equally important (Mallett et al., 2012). Additionally, the clinical and epidemiological characteristics of the disease must also be considered (Mallett et al., 2012). For instance, serious pathology, such as life threatening ocular tumours, demand highly sensitive tests, since a false negative error in these cases would have dramatic consequences (Saunders et al., 2015). Similarly, potentially severe diseases with lack of symptoms at their onset, disorders with effective treatment at their early stages, as well as conditions quickly progressive are better assessed with this type of test (Trevethan, 2017). On the other hand, conditions with low prevalence in the population, relatively unserious, with slow progression rates, and associated to either high cost or considerable invasiveness of the verification procedures and follow-up, demand tests with a low rate of false positives and high specificity (Saunders et al., 2015, Trevethan, 2017). This should allow healthcare services to

minimise unnecessary costs, avoid over treatment and the overall inconveniences caused to patients (Trevethan, 2017).

Translation of these concepts to glaucoma requires several considerations. As discussed above, early diagnosis is highly recommended in diseases such as glaucoma, capable of producing irreversible vision loss (Fry et al., 2018, Porciatti and Ventura, 2012). Nonetheless, the condition has relatively low prevalence in the overall population and is generally slowly progressing. This, coupled with the substantial economic and patient-related burden of diagnosis, make highly specific tests preferable for the assessment of glaucoma (Mwanza et al., 2018a, Fallon et al., 2017). This is reflected in the frequent evaluation of diagnostic tests at high levels of specificity, between 95% and 99% (Michelessi et al., 2015).

1.2d Glaucoma challenges in the clinic

As described above, the early diagnosis of POAG is desirable to prevent any visual impairment (Fry et al., 2018, Porciatti and Ventura, 2012), and clinicians face the challenge of detecting of the earliest modifications. In fact, when glaucoma approaches its moderate-severe stages, both structural and functional changes are usually readily identifiable, making diagnosis straightforward. In similar circumstances, a reliable diagnosis is likely performed by eye-care practitioners without the need of additional investigations, such as advanced structural examination. Accordingly, one of the real challenges confronted by glaucoma clinicians is the early diagnosis of the disease. When this requirement is translated into more practical terms, early detection corresponds to the differentiation between glaucoma suspects and individuals presenting early glaucomatous modifications (Tatham et al., 2015). As discussed in detail in next sections (see sections 1.3 and 1.4), glaucoma detection at its earliest stages based on single cross-sectional structural examinations has been shown to be inaccurate, with the detection of a progressive defect indicated as a more reliable endpoint (Medeiros et al., 2009). Once a diagnosis of OAG has been confirmed by clinicians, a second relevant challenge is then introduced, namely the prompt and accurate detection of disease progression. This represents a crucial task in practice, enabling timely changes to the treatment regime to prevent further vision loss.

While aiming to achieve the two goals mentioned above, glaucoma clinicians make use of information from both the structural and the functional domains. Although it has been suggested that structural modifications may often precede VF defects as detected by SAP, the temporal relationship between structural and functional modifications in clinical settings is not fully understood and is heavily dependent on measurement techniques and criteria adopted (Tatham et al., 2015, Hood and De Moraes, 2018a). In the following sections (1.3, 1.4, 1.5), the role of both domains in glaucoma diagnosis as well as their relationship as represented in the recent literature will be reviewed.

1.3 The structural domain

The examination of structural changes in glaucoma is becoming increasingly central in clinics, as abnormalities affecting this domain may be detectable early in the disease course (Tatham et al., 2015). This is reflected in some more recently released guidelines in which, despite perimetry remaining essential for monitoring and staging glaucoma, VF defects are not necessarily required for diagnosis (World Glaucoma Association, 2016, Prum et al., 2016). Several devices have been proposed for the assessment of ocular structures in glaucoma patients, including scanning laser polarimetry, confocal scanning laser ophthalmoscopy, and OCT. However, the high reproducibility and resolution of OCT measurements resulted in substantial improvements (Tatham et al., 2015), and OCT is currently considered at the core of structural assessment in glaucoma clinics (Fallon et al., 2017, Lavinsky et al., 2017, Hood, 2017).

1.3a Optical Coherence Tomography

OCT enables acquisition of high axial resolution cross-sectional images of the retina and other ocular structures with high reproducibility (Fortune, 2018). Ocular imaging is accomplished in a non-invasive and non-contact fashion (Chen et al., 2018), and anatomic characteristics of the retina are depicted as a result of the varying reflectivity associated with different retinal layers (Chen et al., 2018, Huang et al., 1991). Since the introduction of the time-domain OCT three decades ago (Huang et al., 1991), the main limitations of this generation of instruments have been improved upon dramatically. These

included reduced resolution, slow acquisition speeds and the two-dimensionality of analyses. The more recent spectral-domain OCT (SD-OCT) mainly differs from the time-domain OCT for its light processing mechanisms, where a spectrometer has replaced the photodetector. In addition, the information retrieved is analysed according to a Fourier transformation. These modifications enabled a significant increase in performance of SD-OCTs when compared with previous generation instruments encompassing several aspects, including: i) ability to detect weaker signal (Chen et al., 2005), ii) ability to perform scans at a faster rate (Chen et al., 2018, Tatham et al., 2015, Chen et al., 2005, White et al., 2003), and iii) improvement of the mean axial resolution from 10µm to 5-7µm (Chen et al., 2005, White et al., 2003).

Current application of OCT in Glaucoma Clinics

The implementation of SD-OCT in glaucoma clinics has focused on the examination of three main retinal structures, namely i) the RNFL thickness along a circle around the optic disc, i.e. the circumpapillary RNFL (cpRNFL), ii) the macula; and iii) structures related to the ONH itself (Chen et al., 2018). A brief overview of the main protocols enabling these analyses follows.

cpRNFL thickness

Among the several approaches for OCT examination of glaucomatous changes, the analysis of cpRNFL thickness is at the core of the current assessment of patients with or at risk of glaucoma (Hood and De Moraes, 2018a, Kim and Park, 2018, Kansal et al., 2018, Chen et al., 2018). In a recent systematic review of the literature (Chen et al., 2018), among the 59 eligible studies using SD-OCT in glaucoma, 56 included analysis of the cpRNFL thickness, whereas macular and ONH parameters where collected in 36 and 23 studies, respectively. Capturing the thickness of the RNFL in proximity to the ONH enables the sampling of RGCs across the entire retina, since bundles of axons of all RGCs enter the optic nerve before leaving the eye. This critical benefit, combined with the relative ease of acquisition, are likely to account for the wide spread use of cpRNFL thickness analysis in research and clinical practice (Kansal et al., 2018, Bussel et al., 2014).

According to different manufacturers, cpRNFL analysis can be performed following either a series of concentric circular scans around the ONH (2.5mm to 4.9mm diameter), radial scans across the ONH or a volumetric scan (6x6mm or 7x7mm) of the same area (Chen et al., 2018). Raw OCT scans are analysed by in-built software, enabling automated detection of anterior and posterior limits of the RNFL (i.e. segmentation, Figure 1.4a). RNFL thickness maps are then generated at varying distances from the optic disc. In a study comparing different circle sizes of cpRNFL scans, a greater diagnostic capability was achieved by using the narrowest circle scan, among the ones available (e.g. 12º or 3.5mm from the Spectralis OCT, Figure 1.4d) (Gmeiner et al., 2016). A variety of indices can be determined following the measurement, simplistically divided into global (e.g. average cpRNFL thickness) and regional. In this latter case, the cpRNFL thickness values are averaged across a number of sectors around the ONH (usually from 4 to 12 sectors).
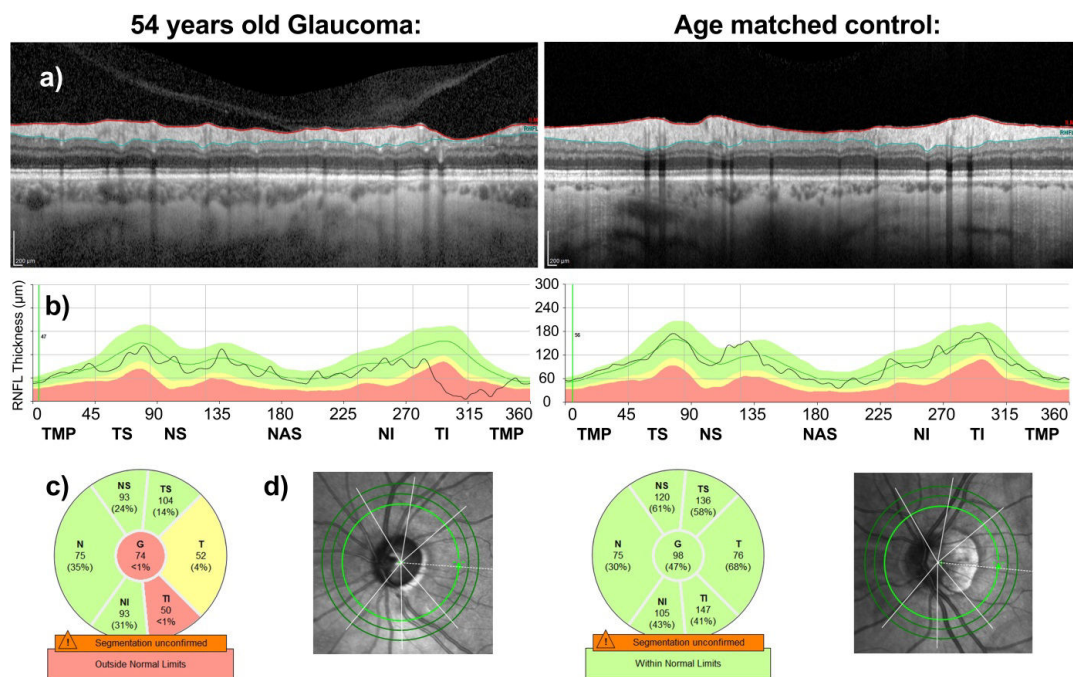
Figure 1.4 cpRNFL analysis of the left eyes of a 54 year old glaucoma patient (left panel) and an age matched healthy control (right panel). The B-scan enabling cpRNFL thickness measurement is shown in (a). Panels (b) and (c) show cpRNFL thickness profile and the ONH sectorial analysis, color-coded according probability levels from comparison with reference normative data: green (within normal limits), yellow (<5% of population norms), red (<1%). Panel (d) shows the location on the retina where the circle scan is acquired. In the left panel a focal thinning of the temporal-inferior sector, consistent with glaucoma, can be observed (a, b, c).

Among the several cpRNFL thickness metrics, researchers have underlined the importance of looking for localised damage, particularly when either the superior-temporal or the inferior-temporal cpRNFL sectors are affected (Tatham et al., 2015, Hood and De Moraes, 2018a). Indeed it is reported that these portions of the cpRNFL may be affected earlier and more frequently in glaucoma (Tatham et al., 2015).

Macula

Analysis of the macula is becoming ever more common in glaucoma, either to complement or as alternative to cpRNFL analyses (Oddone et al., 2016). The importance of examining the macula in glaucoma is supported by the growing evidence suggesting its early involvement in the disease (Tatham et al., 2015, Hood et al., 2013, Kim and Park, 2018). Additionally, there is a considerable anatomic rationale making the macula a suitable area for the identification of glaucomatous structural modifications (Kim and Park, 2018). In fact, approximately 50% of the RGCs are found in the macular region, in a multi-layered arrangement (Curcio and Allen, 1990). In addition, RGC bodies in this area of the retina present an increased size and are reported to be up to 20 times larger than their axons (Curcio and Allen, 1990). It has, therefore, been suggested that in disorders affecting RGCs, lesions could be promptly identified in the macula (Tatham et al., 2015, Hood and De Moraes, 2018a, Oddone et al., 2016). The refinement of macular layer segmentation within SD-OCT platforms enabled development of suitable indices for macular thickness analysis (Figure 1.5).
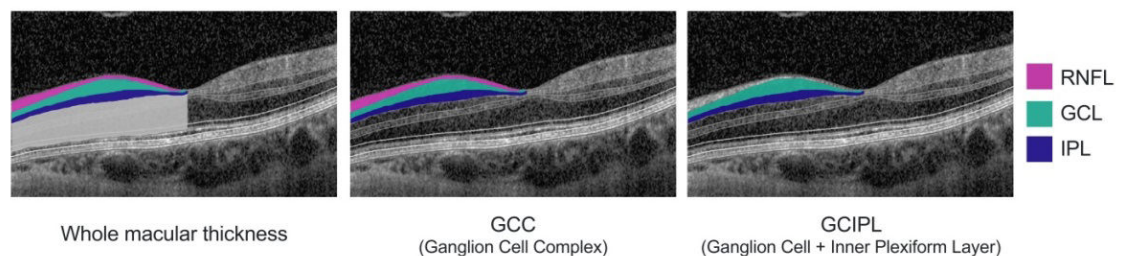
Figure 1.5 Conventional OCT segmentation of retinal layers at the macula. Different platforms may select different approaches in term of layers considered in the analysis. For instance, algorithms may consider either the whole macular thickness (left) or a combination of the RNFL, Ganglion cell layer (GCL) and inner plexiform layer (IPL).

Despite differences across OCT manufacturers, assessment of the macula in glaucoma typically focusses on the inner retinal layers, i.e. the RNFL, the ganglion cell layer (GCL) and the inner plexiform layer (IPL), or their combination (Kim and Park, 2018). For instance, the Ganglion Cells Analysis performed by the Cirrus SD-OCT (Carl Zeiss Meditec, Dublin, CA) represents a specific analysis of a macular cube scans. Here, as in many other commercially available devices, examination can focus either on the combination of the ganglion cell and inner plexiform layers (GCIPL) or on the so-called ganglion cell complex (GCC), computed as the combination of the GCIPL and the macular RNFL (Kim and Park, 2018). In the latter case the axons, bodies and dendrites of RGCs are all included in the analysis, since the RNFL, GCL and IPL (the retinal layers which are housing these underlined RGCs related structures) are respectively segmented and computed together (Chen et al., 2018). Interpretation of macular analyses is not different from that of cpRNFL. Indeed, glaucomatous damage is visualised as thinning of the retinal layers considered in the analysis (e.g. GCC or GCIPL), usually in an arcuate shape, predominantly located in the temporal macula and often respecting the horizontal midline (Kim and Park, 2018).

One alternative analysis of the macula in glaucoma was proposed by the Spectralis OCT. Until recently, this platform did not include segmentation of different macular layers, in favour of an analysis of the asymmetry of the posterior-pole conducted by considering whole macular thickness (Figure 1.6). Macular scans could indeed enable a more direct analysis of hemisphere asymmetry, which might prove particularly valuable to differentiate suspects from patients with early glaucoma changes (Figures 1.6, 1.7). In fact, glaucomatous defects often respect the horizontal midline, and asymmetrical thinning of the macula as opposite to comparison of thickness to normative values may be a more sensitive measure of early signs of glaucoma (Ha et al., 2020).
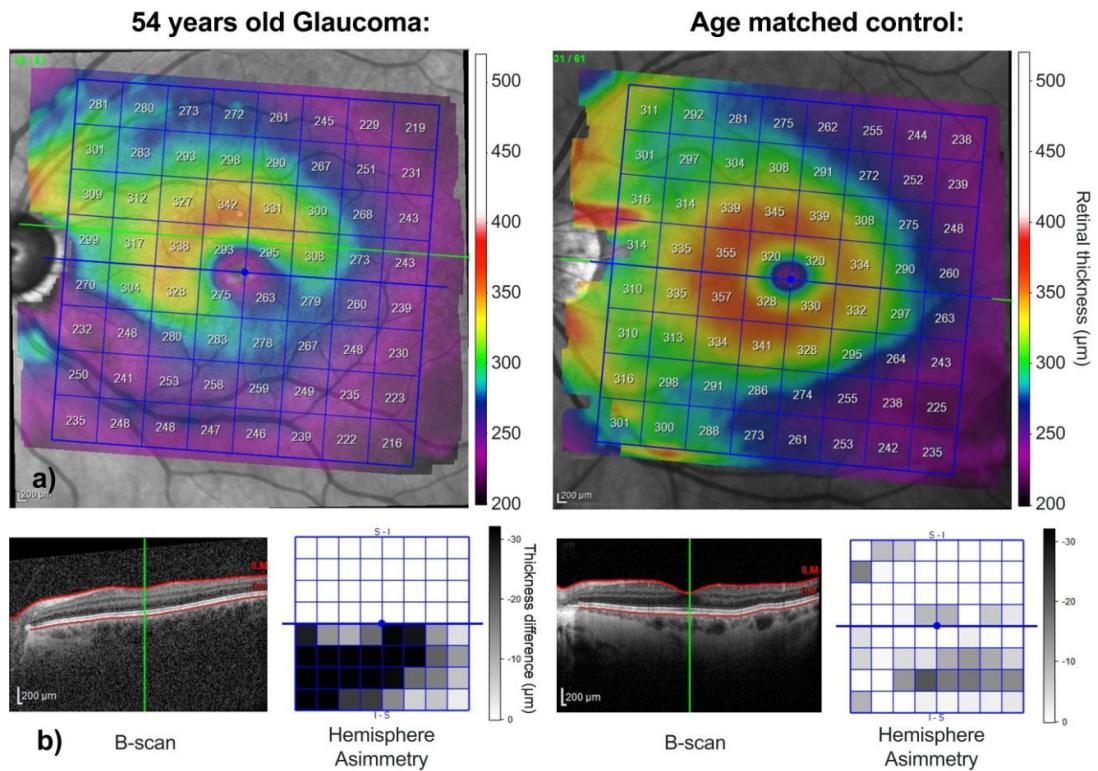
Figure 1.6 Macular posterior pole analysis as performed by the Spectralis OCT. The figure shows findings from the same glaucoma patient (left) and the healthy observer (right) in Figure 1.4. Panel (a) shows the colour-coded macular thickness analysis, and the grid is located according to the disc-fovea angle. A thinning affecting the inferior temporal region can be observed, consistently with panel (b), which is reporting a single B-scan and the asymmetry analysis between the superior and inferior hemispheres. Thickness difference is color-coded accordingly.

Nonetheless, since glaucoma typically affects RGCs, the outer layers of the retina should be relatively spared, potentially hampering the diagnostic accuracy of parameters considering whole retinal thickness (Tatham et al., 2015). Indeed, consistently with other devices the same manufacturer has recently introduced an additional analysis of single macular layers. Within the new analysis, loss of thickness of the macular RNFL, GCL and IPL can be evaluated as raw value or compared to normative data from age-matched healthy eyes (Figure 1.7).
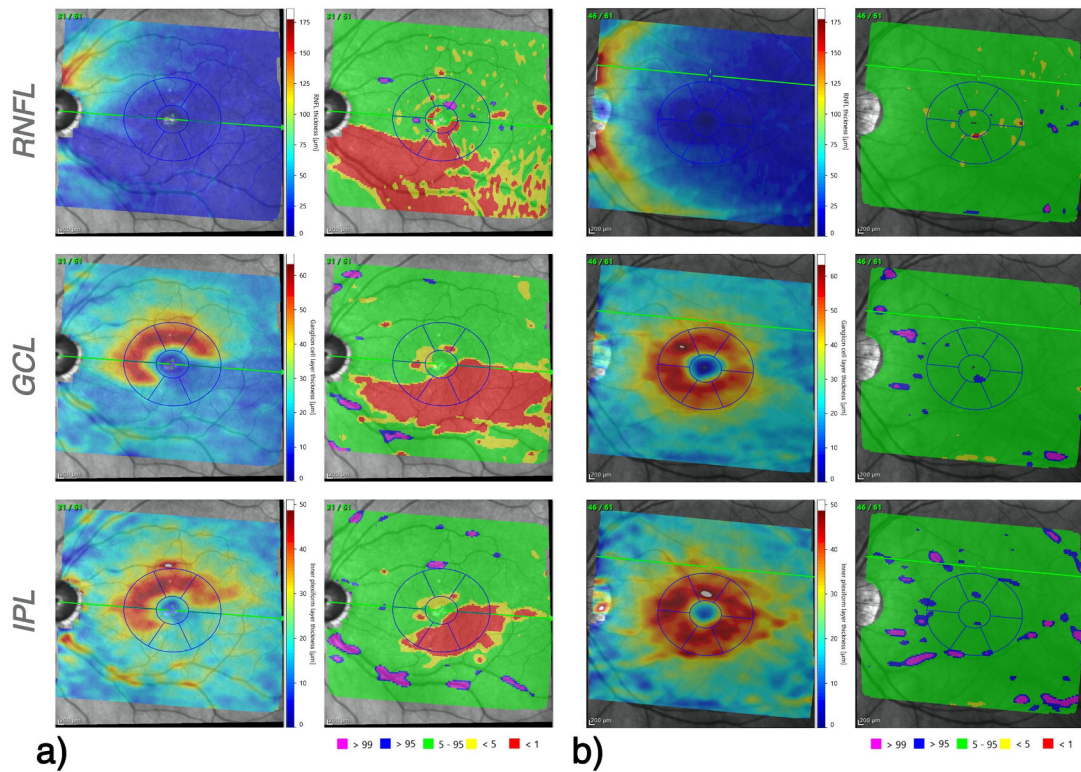
Figure 1.7 Macular single layer thickness analysis as performed by the Spectralis OCT, including the macular RNFL (top panels), Ganglion Cell Layer (GCL, medium panels) and the Inner Plexiform Layer (IPL, lower panels). Figure shows images from the same patients and healthy eye in Figures 1.4 and 1.6. The left panel in (a) shows the colour-coded thickness analysis for the three isolated layers. A reduction of thickness of the inferior or inferior-temporal regions is obvious in all macular sublayers. The right most panel in (a) shows the deviation maps; retinal areas are flagged in red when thickness falls below 1% of the normative population. (b) shows the same analysis for a healthy eye.

Overall, a strong correlation and topographic correspondence between cpRNFL and parameters from macular analyses has been suggested, allowing for them to be considered together (Kim and Park, 2018).

Optic Nerve Head

Following the introduction of ONH volumetric scans the measurement of disc area and pertaining volumetric parameters have become largely available (Chen et al., 2018). Notably, all these parameters are typically determined after having set a reference plane, at variable distance from the retinal pigment epithelium (Chen et al., 2018). Diagnostic performance of these indices was found to be poorer compared to cpRNFL and macular measurements, and therefore after initial wide adoption of ONH volumetric measures, cpRNFL and macular are typically preferred (Tatham et al., 2015).

A relatively new approach to ONH assessment with OCT is described by a group of measurements available in the Spectralis, which relies on automatic detection of Bruch's membrane opening (BMO) (Chen et al., 2018). After the BMO has been identified, the axis connecting the centre of this reference point, which corresponds to the centre of the ONH, with the fovea is also defined. A set of radial scans is then acquired in order to calculate the Bruch's membrane opening minimum rim width (BMO-MRW, Figure 1.8) (Chen et al., 2018). It has been suggested that BMO-MRW relies on more solid geometrical principles than similar biomarkers, since the rim thickness is measured along the minimum distance from the scleral canal to the nearest point of the optic disc surface (i.e. the vitreous-internal limiting membrane interface). The underlined method is supposed to identify the point in the optic cup where all the RGC axons run perpendicular to the measurement performed (Gmeiner et al., 2016, Tatham et al., 2015). As a consequence of this sound geometrical approach there could be a minimisation of any overestimation of rim thickness which could arise from tissue orientation problems when more conventional methods are considered (Tatham et al., 2015). Overall, this could result in an increased reliability of the measurement and greatest diagnostic accuracy (Fortune, 2018).
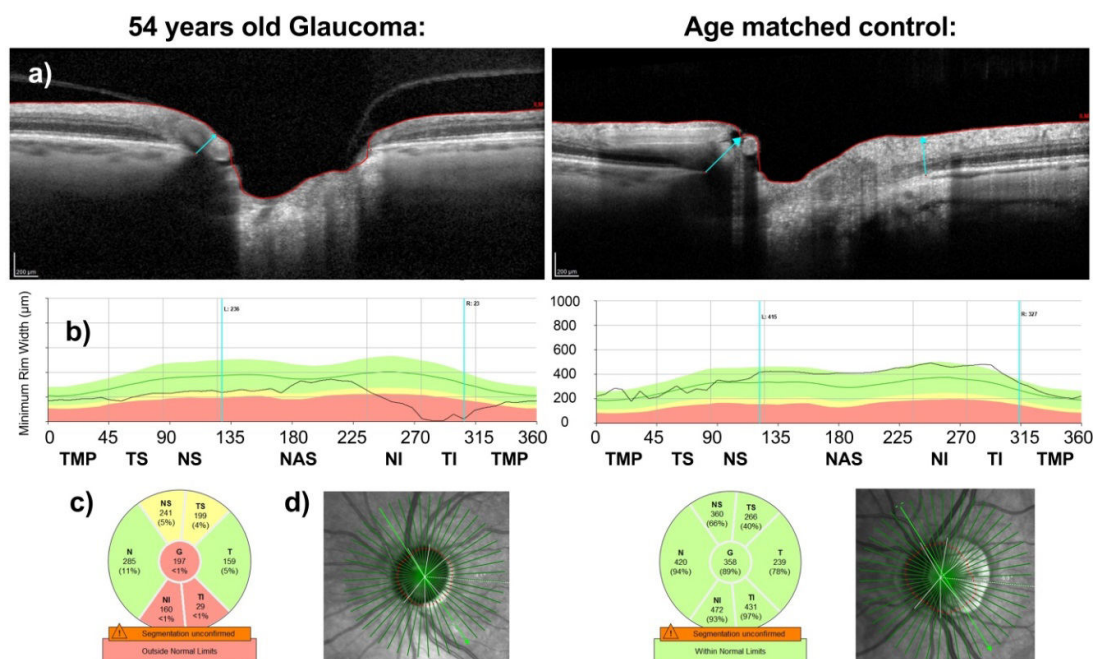
Figure 1.8 BMO-MRW analysis from the same glaucoma patient and healthy eye represented in previous Figures (1.4, 1.6, 1.7). Sub-sections of the figure (a, b, c, d) are consistent with those represented in Figure 1.4.

A key difference between these newer three-dimensional parameters and customary volumetric and/or superficial indices (e.g. cup volume and cup-disc ratio) is the independence of the measurement from a reference plane (Chen et al., 2018), which has been identified as a considerable limitation of those approaches adopting it (Fortune, 2018).

The few studies comparing BMO-MRW with conventional cpRNFL thickness analysis have showed mixed results, with earlier reports suggesting a better diagnostic accuracy, whereas others indicating more similar performance (Stagg and Medeiros, 2020). In one such study, La Bruna and colleagues showed good agreement between defects identified with the two methods (~75%), with borderline levels of significance, splitting of defects due to ONH sectorization and blood vessel interference being the main reasons for disagreement in the remaining 25% of eyes where defects did not agree (La Bruna et al., 2020). Consistent with findings above, a study assessing clinicians' evaluation of glaucoma eyes with cpRNFL and BMO-MRW reports, first in isolation and then combined, found similar diagnostic accuracy (Wu et al., 2020b). Agreement between-clinicians was also good, when judgment was driven by either one modality in isolation or the two combined together (Wu et al., 2020b).

1.3b Diagnostic ability of OCT

Diagnostic accuracy of OCT-related biomarkers has received considerable attention in recent years. Recent systematic reviews in this area agree that OCT provides considerable support to clinicians in glaucoma diagnosis, with cpRNFL-related parameters performing equally or slightly better than macular ones, at least in the general population (Fallon et al., 2017, Kansal et al., 2018, Oddone et al., 2016, Chen et al., 2018, Michelessi et al., 2020).

A meta-analysis of 150 studies aimed to clarify diagnostic accuracy of different OCT biomarkers at different stages of glaucoma (Kansal et al., 2018). Results showed pooled AUCs of 0.897 (95% Confidence Interval (CI): 0.887-0.906), 0.858 (95%CI, 0.835-0.880), and 0.885 (95%CI, 0.869-0.901) for the global cpRNFL thickness, the macular GCIPL and macular GCC, respectively. Although different indices performed similarly, a slight superiority of cpRNFL

was reported, with a marked trend toward a reduced accuracy in glaucoma diagnosis at earlier stages of the disease (Figure 1.9). Furthermore, the analysis confirmed how whole macular thickness performed worse than indices fed by individual sublayers (Kansal et al., 2018). A methodological concern in this study was its wide acceptance of reference standard tests adopted by the eligible studies, possibly causing heterogeneity and thus biased results. A slightly older meta-analysis (Oddone et al., 2016) had the objective to assess differences between cpRNFL thickness and macular OCT parameters (namely GCC and GCIPL) in the diagnosis of manifest glaucoma. The majority of included studies diagnosed glaucoma when a VF defect on SAP was associated to ONH defects. Overall, the global cpRNFL showed a similar diagnostic performance to the inferior cpRNFL sectors, which were the best among isolated sectors. Macular parameters had similar or only slightly poorer diagnostic capability compared to cpRNFL, with sensitivity values ranging between 0.65 and 0.75 at fixed level of specificity (0.90 to 0.95). An overall modest superiority of the cpRNFL was identified which, according to the authors' conclusion, made this largely adopted parameter still preferable when a single OCT protocol has to be performed (Oddone et al., 2016).

The meta-analysis by Kansal et al. presented above is noteworthy since it included a subgroup analysis for different stages of the disease (Figure 1.9). While considering perimetric glaucoma, findings largely mirrored what was observed in the overall population, with AUCs approaching 0.9 for some of the cpRNFL indices but not for macular ones (Kansal et al., 2018). Nevertheless, when the focus was directed on the pre-perimetric group (i.e. glaucoma patients with no VF defect), a drop in the AUC values was observed (Kansal et al., 2018). This has been subsequently confirmed in further studies, with a clear trend toward greater OCT diagnostic accuracy at more advanced stages of the disease (Chen et al., 2018).
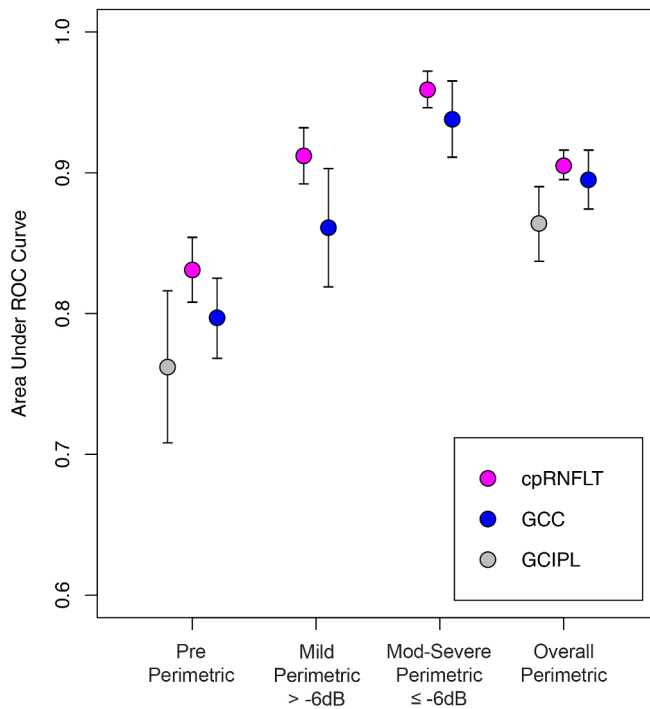
Figure 1.9 Area under receiver operating characteristic curve data, from a meta-analysis of OCT diagnostic accuracy at different stages of glaucoma. Diagnosis of pre-perimetric glaucoma was made according only to disc appearance, with no evidence of VF defects. Grouping between mild perimetric and moderate-severe perimetric glaucoma was made according to SAP mean deviation values better or worse than -6dB. Error bars show the 95% confidence interval. Replotted from (Kansal et al., 2018). cpRNFLT = cpRNFL thickness.

According to the findings above, it remains difficult to perform an accurate diagnosis of early glaucoma from a single cross-sectional examination of ONH and retinal structure in the absence of concomitant VF defect (Tatham et al., 2015, Mwanza et al., 2018a, Mwanza et al., 2018b). In such cases, clinicians can only diagnose pre-perimetric POAG with reasonable certainty when a progressive structural damage is observed during follow-up examinations compared to baseline (Tatham et al., 2015, Tatham and Medeiros, 2017). Indeed, the likelihood of future functional loss for an evolving defect on longitudinal observations is significantly increased compared to that for a single cross sectional observation of apparently abnormal findings (Miki et al., 2014, Mwanza and Budenz, 2018, Yu et al., 2016).

It is worth mentioning how findings from studies assessing diagnostic performance of clinical tests in general are often difficult to be transferred to the clinical context. The observed outcomes only apply to the sample studied, which is likely to differ to a varying extent from the clinical population in which the test will be used and where it will probably perform worse than reported (Ransohoff and Feinstein, 1978). Additionally, selection criteria vary significantly across different studies, resulting in target populations which might also differ significantly. More specifically to glaucoma, it could be thought

that OCTs and similar advanced instruments would be more useful for the assessment of difficult cases (e.g. suspicious ONH and normal VF), where other clinical tests are equivocal and the diagnosis is therefore uncertain (Michelessi et al., 2015). Notably, such cases typically do not meet inclusion criteria of diagnostic performance studies, leaving only readily identifiable cases of glaucoma and unequivocally healthy controls (Oddone et al., 2016, Rao et al., 2012). Yet, diagnostic studies preserve considerable insights to clinical practice, since they do enable a direct comparison of the relative performance of different metrics within a study (e.g. cpRNFL vs BMO-MRW), giving an indication of which measure is likely to perform best.

1.3c OCT to detect progression

As per the assessment of structural changes for glaucoma diagnosis, evaluation of progression within the structural domain has conventionally used subjective clinical evaluation of fundus photos, not without caveats (Tatham and Medeiros, 2017). Indeed, a poor-to-fair agreement between clinicians combined with the lack of a quantitative measure of progression rate were recognised limitations of this approach (Jampel et al., 2009). These shortcomings led to the introduction of more objective instruments in progression analysis (Tatham and Medeiros, 2017), with OCT now playing a key role.

The short term repeatability of OCT has been generally reported to be excellent for both cpRNFL and macular measurements, determining a great potential for monitoring progression with this instrument (Tatham and Medeiros, 2017). When assessing global metrics such as the average cpRNFL or macular GCIPL, test-retest repeatability of SD-OCT devices is in the order of 4-5µm (Leung et al., 2009, Mwanza et al., 2010, Kim et al., 2015). Repeatability in localised sectors is worse, in the order of 8-12 µm, though this is not consistent across instruments (Pierro et al., 2012, Reis et al., 2017).

Before attributing progressive thinning noted on SD-OCT to glaucoma progression, changes due to the normal ageing process should also be considered (Leung et al., 2012, Wu et al., 2017, Hammel et al., 2017). Even in healthy eyes, thinning of retinal layers for both macula and cpRNFL OCT

indices is observed over time and attributed to physiological effects of normal ageing (Miki et al., 2014). Unfortunately, OCT data from large-scale longitudinal studies are still scarce (Tatham et al., 2015). Among the few studies reporting physiological thinning of OCT measurements as a function of age, Leung and colleagues followed a small group of healthy individuals for 30 months and found cpRNFL to decrease by 0.52μm per year (95% CI, 0.17-0.86μm), significantly faster than for the macular GCIPL (-0.32μm/year) (Leung et al., 2012). Another study found a thinning rate in healthy subjects of -0.48μm/year and -0.14μm/year for cpRNFL and GCIPL respectively, after an average follow-up of 1.7 years (Hammel et al., 2017). One additional report followed 45 healthy participants for an average of 3.2 years, finding cpRNFL changes of -0.54±0.23μm/year (Wu et al., 2017). Overall, the data presented above were collected in limited samples, with variable follow-up durations and using varying devices and protocols. Although thresholds of 4-5μm have been proposed in the literature for determining modifications of average cpRNFL thickness and macular indices (Mwanza et al., 2010, Leung et al., 2009, Kim et al., 2015), consensus is lacking on a precise rate of progression that could confirm diagnosis of glaucoma.

When evaluating the ability of different OCT metrics to detect progression, attention should also be directed to the dynamic range of the underlined measures. For instance, the average cpRNFL ranges between a maximum of 100μm in healthy controls and an inferior boundary of 50μm (Tatham and Medeiros, 2017, Nouri-Mahdavi and Caprioli, 2015). Considering some of the suggested rates of minimum significant variation (5μm) of cpRNFL thickness, this would represent 10% of the whole dynamic range, possibly hampering the value of this measurement (Tatham and Medeiros, 2017).

Even though several prospective studies have explored the value of different OCT indices to assess the progression of POAG, the superiority of one metric over another is not clearly established, also because of the absence of a solid reference standard for progression (Tatham and Medeiros, 2017). Further, rates of change as identified, for example, at macular level are hardly comparable with those reported in the cpRNFL, with different baseline

thickness and dynamic ranges that would make a direct comparison inappropriate (Tatham and Medeiros, 2017).

1.3d Future perspectives of OCT examination and en face imaging

Although the introduction of SD-OCT made an invaluable contribution to structural assessment in glaucoma, shortcomings in the detection and monitoring of the disease preserve a solid rationale for further research (Mwanza and Budenz, 2018). Several areas of development of structural examination are currently explored in glaucoma, including use of different OCT technologies, such as OCT angiography (WuDunn et al., 2021), use of artificial intelligence for the analysis of OCT scans (Ting et al., 2019), and the examination of wider areas of the retina in wide-field imaging (Hood et al., 2016, Kim et al., 2021).

Novel areas of development also include the evaluation of RNFL reflectance with OCT en face images. Retinal nerve fibre bundles (RNFBs) are highly reflective compared to other retinal structures because of the ordered cytoskeletal structure of RGC axons (Huang et al., 2006). Accordingly, analysis of reflectivity might provide an additional source of information on RNFL status in glaucoma; for the same amount of thickness, RNFL could show different levels of reflectivity, according to integrity and density of RNFBs (Fortune, 2015, Vermeer et al., 2012).

Recent SD-OCT devices allow us to qualitatively explore reflectance in en face images of the RNFL (Ashimatey et al., 2018a, Hood et al., 2015, Sakamoto et al., 2019). En face analysis is often accomplished with dense volumetric scans of the retinal area of interest, by averaging the intensity of each A-scan over a certain range of depths into a two-dimensional image (Ashimatey et al., 2018a, Hood et al., 2015, Sakamoto et al., 2019, Mavrommatis et al., 2019). The result is a single transverse retinal section, or slab, frequently derived from a fixed retinal thickness (e.g. 50μm) beginning anteriorly from the inner limiting membrane (ILM) - vitreous interface (Hood et al., 2015, Sakamoto et al., 2019). In en face images, glaucomatous defects appear as regions of reduced reflectance, often in typical arcuate or wedge shapes as well as more generalised loss. Defects are likely to be induced by a combination of primary

loss of reflectivity, caused by cytoskeletal disruption (Huang et al., 2017, Huang et al., 2006), and thinning of the RNFL which leads to the inclusion of deeper hypo-reflective retinal layers in the slab (Hood et al., 2015). It should be noted that the assessment of RNFL reflectance has been performed in glaucoma clinics long before the introduction of OCT, either by ophthalmoscopy or fundus photography (Quigley et al., 1992, Tuulonen and Airaksinen, 1991). Yet, compared to these examination techniques, en face OCT has introduced appreciable advantages, including better visualisation of narrow defects and preserved bundles, the ability to examine below the superficial RNFL, and being less affected by lens opacities and light fundus pigmentation (Jung et al., 2018, Ji et al., 2020, Lim et al., 2020).

The growing interest in en face OCT imaging is at least in part explained by the potential for an earlier detection of glaucomatous changes, linked to the assessment of reflectivity (Mwanza and Budenz, 2018, Liu et al., 2014). In fact, animal models of experimental glaucoma suggest that RNFL thickness may show a measurable thinning only after 10-15% of RGC axons have been lost (Cull et al., 2012). On the contrary, results from studies conducted in similar models indicate that reflectivity of RNFBs may deteriorate earlier than a measurable thinning of the RNFL (Huang et al., 2011, Fortune et al., 2013, Dwelle et al., 2012), following cytoskeletal disruption of ganglion cell axons (Huang et al., 2017, Huang et al., 2006). Although the supporting evidence is not yet compelling and consistent replication in humans is lacking (Fortune, 2015), assessing reflectance of RNFBs could provide additional and potential earlier markers of glaucomatous changes (Hood et al., 2015). Further, as will be discussed in more detail in section 1.5, clinicians are frequently encouraged to confirm a suspected structural lesion in the functional domain, to increase confidence in diagnosis (Malik et al., 2012, Shigueoka et al., 2018, Raza et al., 2014, Hood et al., 2019). En face images may offer the opportunity to directly relate structure and function without use of dedicated mapping system between VF and ONH, thought to represent an additional source of noise in this relationship (Denniss et al., 2018). Similarly, newer VF strategies that aim to incorporate structural information for greater efficiency (Denniss et al., 2013, Montesano et al., 2018, Ganeshrao et al., 2015), or to assess in detail specific

regions of interest (Ballae Ganeshrao et al., 2018, Alluwimi et al., 2018a, Alluwimi et al., 2018b), may be facilitated by en face imaging.

The few studies conducted so far on en face OCT imaging in glaucoma showed a good correlation between reflectance changes and cpRNFL defects (Ashimatey et al., 2018a), and good agreement between residual RNFBs and preserved visual function in the macula (Sakamoto et al., 2019, Iikawa et al., 2020). One such study acquired wide-field OCT scans in 33 eyes with glaucoma to assess reflectivity defects of the RNFL – see Figure 1.10 for an example of the slab images and the analysis performed (Ashimatey et al., 2018a). The authors found a strong correlation with conventionally-detected cpRNFL defects, and the study confirmed how these measurements carry clinically-valuable information for the diagnosis and monitoring of glaucoma.
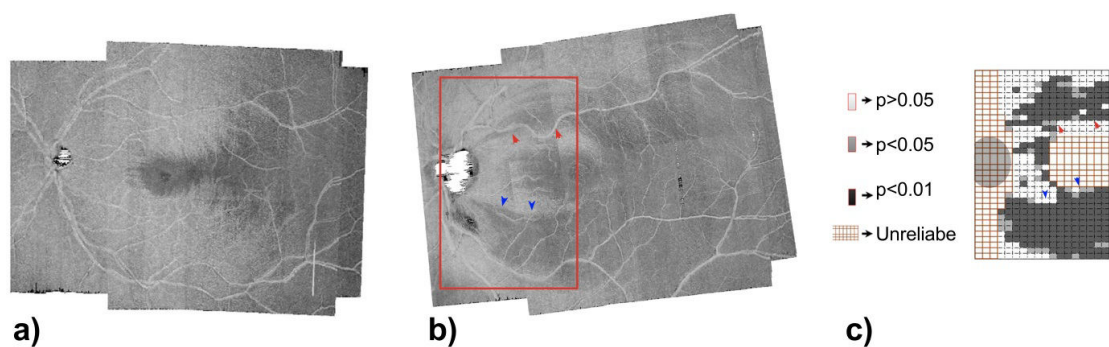
Fig.1.10 The Figure, reproduced from (Ashimatey et al., 2018a), shows the wide-field OCT montage of the en face RNFL of a healthy participant (a) and a glaucoma patient (b) presenting an inferior and a superior arcuate defects which can be observed as reflectivity loss. In (c) the analysis shows a possible objective approach to classify each pixel as normal or abnormal as a function of its reflectance.

1.4 The functional domain

1.4a The Visual Field

The visual field is the entire perception of the single eye when fixation is stable on a certain target (Schiefer et al., 2005). VF sensitivity is typically measured as differential light sensitivity and varies according to the distance from fixation, with a progressive reduction from the centre to the periphery (Traquair, 1939). Kinetic perimetry, as standardised by Goldmann, was among the first methods enabling systematic assessment of the VF. However, the considerable between-subject variability combined with suboptimal standardisation of test procedures lead to the development and widespread adoption of more repeatable examinations from the early 1960s. This

happened in the form of semi-automated static perimetry, followed by standard automated perimetry (SAP), only one decade later (Camp and Weinreb, 2017, Phu et al., 2017). SAP enabled the administration of standardised tests with a relative lack of operator-dependency, which overall yielded less variable results compared to kinetic modalities. The increased consistency of measurements partially accounted for the capacity to detect earlier glaucomatous defects, and made SAP the dominant methodology for VF assessment in glaucoma (Camp and Weinreb, 2017).

Loss of light sensitivity is one of the hallmarks of glaucoma disease, and its evaluation and quantification remains central in both the clinic and research (Jampel et al., 2011). Characteristics of VF defects in glaucoma often reflect the damage affecting the RGC axons at the ONH, and is determined by the bundle trajectory through the retina. This translates to the typical arcuate defects, nasal steps and more rarely temporal wedges, with the overall characteristic of VF defects typically respecting the horizontal midline at onset and to progress to the whole field at later stages (Anderson and Patella, 1992). However, beside the typically localised and vertically asymmetric clustered defects, a wide spread loss (Anderson and Patella, 1992) or a paracentral scotoma (Hood et al., 2013) might be the first functional defect.

1.4b Standard Automated Perimetry (SAP)

The term SAP refers to a general definition of any perimetric test adopting static white stimuli of constant size and varying intensity, presented over a grid of fixed locations of the VF upon an achromatic background (Jampel et al., 2011, Delgado et al., 2002, Phu et al., 2017). SAP is used to estimate the minimum luminance contrast required for a stimulus to be recognised from its background 50% of the time (differential threshold), across the area assessed. Despite its limitations, SAP still represents a fundamental clinical procedure in the evaluation and management of glaucoma, being the reference standard for the functional assessment of this condition (Camp and Weinreb, 2017, Wu and Medeiros, 2018, World Glaucoma Association, 2016). It is worth considering that structural measurements suffer limitations in advanced glaucoma, where patients face the greatest risk of severe visual impairment (de Moraes et al., 2016). Indeed, at these stages, assessment of progression

is routinely performed with SAP (Camp and Weinreb, 2017). Further, as indicated by recent guidelines (Prum et al., 2016, National Institute of Health and Care Excellence, 2017) the main goal in the management of glaucoma is the preservation of patient quality of life. Visual function is a key determinant of quality of life, and SAP measurements of functional loss are highly correlated with quality of life indices (Medeiros et al., 2015).

SAP can be administered through different platforms among which the Humphrey Field Analyzer (Carl Zeiss Meditec, Dublin, CA) and the Octopus perimeter (Haag-Streit, Koeniz, Switzerland) represent the two most widely adopted. In these instruments, light stimuli are presented at varying intensities according to a logarithmic scale (decibels, dB). The adopted dB scale does not directly describe the light intensity, instead it is a measure of the light attenuation achieved by neutral density filters of increasing absorption (Phu et al., 2017). For instance, 0dB represents the brightest stimulus in the Humphrey VF, with no attenuation applied, whereas 51dB is the maximum attenuation possible, being 5.1 log units less intense than brightest stimulus (Anderson and Patella, 1992). A logarithmic scale was selected against different measurement systems since luminous sensation follows a factorial nature. Accordingly, to produce constant differences between contiguous stimuli across the whole spectrum of presentations, light intensity must be increased according to multiplicative factor (see Figure 1.11). This allows a constant increase of likelihood of generating a visual sensation (Anderson and Patella, 1992).
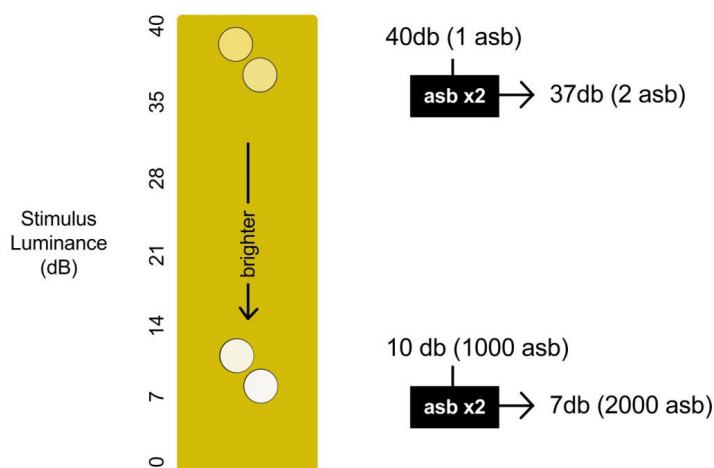
Figure 1.11 Logarithmic scaling across the whole SAP luminosity range, redrawn from (Anderson and Patella, 1992). The difference between stimulus intensity is constant in terms of dB, resulting in progressively larger increase in linear Apostilb (asb) units at lower dB values.

Stimuli of Goldmann size III, corresponding to a diameter of 0.43º, presented for typically 100-200 milliseconds on a background luminance of 10cd/m^2 at approximately 30cm, over the central 24º (24-2) represents the most implemented combination of settings in both clinical and research environment (Camp and Weinreb, 2017, Phu et al., 2017). In the Humphrey VF, the 24-2 area is assessed at 54 different locations, spaced 3° from the vertical and horizontal midline and 6° apart from each other (Anderson and Patella, 1992). Besides the largely adopted 24-2 grid in glaucoma, the 10-2 is a strategy targeting central vision which presents stimuli at 68 locations spaced 2° throughout the central 10° of the retina (Anderson and Patella, 1992). Example of 24-2 and 10-2 SAP grids are shown later in the section (Figure 1.13).

An overview of SAP results is facilitated by several statistical summary measures reported in VF printouts (Camp and Weinreb, 2017). Among the key analyses there is a comparison of measured VF sensitivities with age matched values from healthy subjects. The total deviation plot reports the point-wise differences between measured thresholds and the age-matched normative values at the same locations. These age normalised measurements of loss are summarised in the Mean Deviation (MD), computed as a weighted mean of the total deviation values, where greater weight is given to locations with lower variability in healthy eyes (Vianna and Chauhan, 2015). Pattern deviation plots are also reported and aim to enhance the visualisation of defects which could be hidden by either a generalised depression, as in the case of media opacities, or anomalously elevated hill of vision (Camp and Weinreb, 2017). This analysis allows us to identify the spatial configuration of a VF defect by comparing the sensitivity at each location with the mean sensitivity of the patient. Diffuse loss can, therefore, be compensated and more localised defects become more evident (Turalba and Grosskreutz, 2010). A global measurement of uniformity of VF defects is reported in the Pattern Standard Deviation (PSD). Lastly, another measure of VF status is the Glaucoma Hemifield Test (GHT). The GHT assesses sensitivity of vertically symmetric clusters of VF locations located at opposite areas in relation to the horizontal midline (Susanna et al., 1994). Asymmetric loss of either the superior or the inferior hemifield may be suggestive of glaucoma, since one

hemisphere is likely to be more affected then the other at early and moderate stages of the disease (Turalba and Grosskreutz, 2010).

SAP algorithms

SAP originally adopted staircase procedures characterised by presentations of stimuli with changes of intensity in small steps, usually 4 and 2dB, and fixed number of reversals (2 in case of the Humphrey VF Full Threshold procedure). Accordingly, the intensity of stimuli presented at each location is increased or decreased by 4dB until it goes from 'seen' to 'missed' or vice versa, termed a 'reversal'. Smaller stimulus changes (2dB) are then used to identify a 2nd reversal, which is used as the threshold estimate. This approach results in lengthy test times, being clinically unfriendly because of a considerable fatigue effect and ultimately contributing to compromised reliability (Delgado et al., 2002, McKendrick, 2005). Probably the most widely adopted SAP algorithms, the Swedish Interactive Threshold Algorithm (SITA) family were introduced in the late 1990s and have largely superseded staircase procedures (Bengtsson et al., 1998). The SITA algorithms allowed, through the reduction of presentations and other analytic implementations, to reduce SAP testing time by up to 50-70%, with considerably reduced fatigue and consequent better patient acceptability compared to Full Threshold (Jampel et al., 2011, Camp and Weinreb, 2017, McKendrick, 2005). Evidence has accumulated showing that SITA Standard produces interchangeable results with Full Threshold, in a substantially reduced amount of time and with less variability (Artes et al., 2002, Jampel et al., 2011, Delgado et al., 2002). The latest development of the SITA algorithms is SITA Faster, enabling further ~30% or ~50% shorter test duration compared to SITA Fast and Standard, respectively. Evaluation of this procedure is ongoing to fully understand the reliability and interchangeability with other conventional procedures in clinical settings (Heijl et al., 2019).

Although considerable reduction of test time was achieved by the SITA algorithms, the test time for such procedures is still considerable, taking up to 5-8 and 7-10 minutes per eye in healthy and glaucoma eyes, respectively (Wild et al., 1999, Shirato et al., 1999). One proposed solution to tackle the problem

of test time is *supra-threshold* perimetry, where stimuli are deliberately presented above the expected threshold sensitivity of the patient and locations are considered undamaged when these stimuli are seen (Katz et al., 1993).

Several parameters can be combined in supra-threshold perimetry aiming to achieve the optimal combination of sensitivity and specificity, still minimising test time (Turpin et al., 2016, Henson and Artes, 2002). For instance, a too bright stimulus – lower value in dB – is likely to determine a low sensitivity. Conversely, stimuli with too low intensity will have low specificity with a considerable number of false positives (Henson and Artes, 2002, Henson et al., 1999). According to Henson and colleagues, selecting the sensitivity to test with regards to normal values in age-matched healthy observers was no less accurate than methods estimating the height of the hill of vision of the subject and present accordingly (Henson et al., 1999). The number of trials at each location and the pass/fail criteria selected also demand consideration. Certain procedures use a single presentation, classifying either as normal or as defect when the stimulus is seen or missed, respectively. However, to reduce false positive and negative rates, different criteria have been adopted. For instance the presentation might be repeated if the first one is missed, i.e. a 1-2 pass fail criteria, where the observer has to respond to at least 1 of 2 presentations (Henson and Artes, 2002). According to recent report, a 1-3 criteria showed the greatest specificity among the tested criteria, yet the 1-2 was able to reach comparable levels of specificity with fewer presentations (Turpin et al., 2016).

1.4c SAP variability

Although being the most largely adopted method of assessing visual function in glaucoma, SAP is inherently dependent on the subject's response (McKendrick, 2005). This results in well-known limitations such as a considerable learning effect (Wild et al., 1991), the fatigue effect – especially for algorithms with longer test-times (Hudson et al., 1994), and the patient's sense of frustration that could arise from repeatedly missing dim lights (Jampel et al., 2011, Camp and Weinreb, 2017). Some of these drawbacks may contribute to the considerable between-test variability of SAP (Delgado et al., 2002, Camp and Weinreb, 2017, McKendrick, 2005, Aref and Budenz, 2017). Variability is physiologically encountered in healthy subjects and to an

increasing extent in glaucoma eyes. Consequently, the ability of either defining a defect or interpreting real modifications of damaged areas during follow-up is hampered (Jampel et al., 2011, Delgado et al., 2002, McKendrick, 2005, Turalba and Grosskreutz, 2010, Aref and Budenz, 2017). Notably, at advanced glaucoma stages, fluctuation of results might be so substantial to make the value of SAP in damaged locations questionable (Junoy Montolio et al., 2012). Indeed the magnitude of changes required to produce a statistically significant modification may overtake the dynamic range of the instrument itself (Delgado et al., 2002). As such, additional testing is required to ascertain disease diagnosis and progression, leading to potential delay in detecting rapid progression and accumulation of irreversible vision loss (Gardiner et al., 2014, Artes et al., 2005, de Moraes et al., 2016).

Variability in perimetry, and the pejorative effects due to increased disease severity have been known for a long time (Artes et al., 2005, Artes et al., 2002). Earlier findings were reinforced by Gardiner et al., in work suggesting that measures of sensitivity in areas with moderate-advanced damage do not provide useful information in the assessment of progression (Gardiner et al., 2014). Indeed, in this sample as the sensitivity approached 15-19dB, test-retest variability showed lower limits of 95% retest interval reaching 0dB (Figure 1.12). This would indicate that sensitivity of damaged locations on test repetition could assume any value between 0dB and ~20dB (Gardiner et al., 2014). Changes within this range might, therefore, not be indicative of actual progression but only an intrinsic characteristic of the measurements. A pragmatic implementation of these results into practice might be to reconsider the actual SAP dynamic range (i.e. the level of disease severities where reliable measurement can be obtained) as follows (Gardiner et al., 2014):

- Sensitivity greater or equal to 19dB: test-retest variability is reasonably low and results are reliable.
- Sensitivity between 19dB and 0dB: data are substantially variable with the only information that can be inferred being a reduced sensitivity, between 0 and 19dB.

- Sensitivity <0dB: measured at locations unable to see stimuli at their maximum contrast. This finding differs from 0dB, where the patient answers to brightest stimuli (0dB) less than 50% of the time.
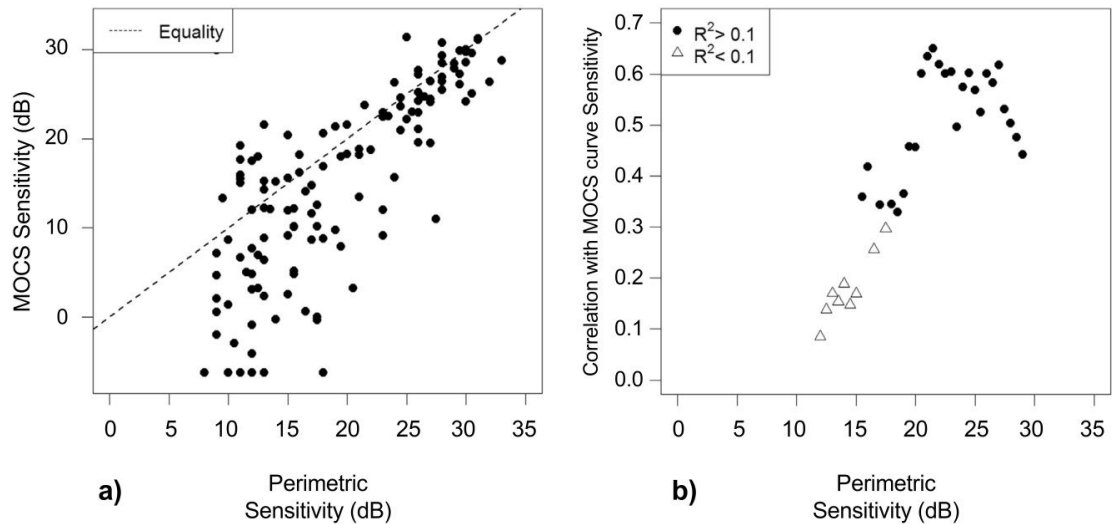
Figure 1.12 In (a), the scatter plot presents the relationship between SAP-measured sensitivities and corresponding values achieved with an alternative method (method of constant stimuli). (b) reproduces strength of correlation (coefficients of determination, R^2) between the measurements performed with the two methods as a function of sensitivity. Notably R^2 drops below 0.2 for sensitivities below 15dB. Reproduced from (Gardiner et al., 2014).

Studies overall indicate that caution is required when dealing with VF sensitivities approaching 20dB or worse (Gardiner et al., 2014, Artes et al., 2005, Artes et al., 2002). The floor measurement in SAP could be reconsidered to be around 15-19dB instead of the typical 0dB, since this lower limit is suggested to be the last value whose variability does not encroach 0dB (Wall et al., 2010).

1.4d SAP diagnostic accuracy

When considering diagnostic accuracy of any perimetric device, it must be remarked how the scientific literature on this topic is crowded by different criteria proposed either to define a VF defect or to diagnose glaucoma (Wu et al., 2020a). Within the definition of a VF defect, criteria presented by Hodapp et al, are often considered a landmark. Accordingly, a VF abnormality can be considered when all the following three conditions have been met: i) cluster of a minimum of three points with $p<0.05$ on the pattern deviation plot, including one point at $p<0.01$; ii) PSD less than 5%; and iii) GHT outside normal limits (Hodapp et al., 1993). Following this earlier definition, additional work has

presented newer and varied criteria to define glaucoma (Nouri-Mahdavi et al., 2011). It must be stressed that the adoption of different definitions will determine varying diagnostic capabilities of glaucoma (Wu et al., 2020a). Whereas some strategies might be more common and more widespread than others, this does not necessarily imply that they are supported by stronger evidence, and no consensus exists upon the best criterion to define a VF defect (World Glaucoma Association, 2016, Wu et al., 2020a, Karvonen et al., 2021).

Considering the lack of an established reference standard for assessing the diagnostic accuracy of SAP in the early stages of glaucoma, longitudinal ONH changes as observed in fundus photographs have been often adopted for this purpose (Tatham et al., 2015). In one such study, PSD and MD showed sensitivity of 48% and 55% respectively at a 90% fixed level of specificity, in a cohort of glaucoma participants with an average MD of -3.19dB (Sample et al., 2006). A significant report in this area is the study from Liu and colleagues, comparing several perimetry strategies, including 24-2 SITA Standard, with a structure-based reference standard (Liu et al., 2011). It must be noted that although the diagnosis of glaucoma was based upon OCT examination, allowing identification of early cases, glaucoma eyes had moderate-severe defect (average MD: -8.13dB), which might not be ideal to assess the ability to detect disease at its onset. Considering MD values, which performed similarly to PSD, SAP showed a sensitivity of 82% (95%CI, 70.5–92.6) at 90% specificity, which dropped down to 69.6% (95%CI, 51.8-87.5) when moderate to advanced defects were excluded from the analysis (Liu et al., 2011).

These results have helped to consolidate suggestions of relative inability of SAP to identify earlier glaucoma cases. Among the reasons that could explain this deficiency is the considerable overlap of RGC receptive fields with a consequent redundancy of coverage of a given retinal location (Delgado et al., 2002, Tatham et al., 2015, Harwerth and Quigley, 2006). Accordingly, the missed response of a damaged RGC when projecting at a certain retinal location might be overlooked when the same stimulus engaged the receptive field of an undamaged RGC sampling the same location. Furthermore, the use of a logarithmic scale to assess retinal sensitivity may compress

measurements, especially with earlier loss (Tatham et al., 2015, Nouri-Mahdavi et al., 2011). Finally, current criteria for VF defects are stringent, prioritising high specificity and low false positive rate in the diagnosis. This choice might also play a role in the reduced sensitivity to earlier stages of glaucoma (Phu et al., 2017).

1.4e New perspectives in perimetry

Among areas of further development in perimetry, there is the assessment of strategies to tackle SAP variability. One modality pursued to this end is the alteration of current algorithms, targeting strategies which make use of individual-specific information. Methods proposed along this line of research, aimed either to improve accuracy of estimation in defective areas or to reduce test time while maintaining same accuracy of measurements (McKendrick and Turpin, 2005). Nonetheless, it has been estimated that current SAP algorithms would require a 20-40% reduction of variability, for the impact of newer strategies to be clinically significant, i.e., allowing detection of progression 1 visit or 1 year earlier (Turpin and McKendrick, 2011). The same authors also concluded that similar levels of improvement were unlikely to be achieved by simply altering the main determinants of accuracy within measurement algorithms. In other words, increasing the number of presentations per location would be unlikely to achieve such reduction of variability. In fact, a typical algorithm would need at least 20 presentations per location to reduce variability by 20-40%, but fatigue effects and reduced attention from longer tests would be likely to disguise any potential improvement. As such, reductions of variability would rather derive from a revision of either the test stimuli and pattern or paradigms adopted (Turpin and McKendrick, 2011).

While some of the above approaches were oriented to more extensively test damaged areas to minimise variability, according to these latter indications extra presentations and time spent in damaged areas might be improperly invested. Indeed, the intrinsic variability in these locations might not allow the thresholds to be measured with greater accuracy and censoring them, by allocating a cut-off sensitivity (e.g. 19dB), might lead to similar or better performance (Pathak et al., 2017). Hence, different paradigms seem to be

advisable to gain additional information to support a more precise assessment of the disease.

Denser grids: Greater use of spatial information

One alternative in such direction might be a greater use of spatial information by implementing grids with greater resolution (Wu and Medeiros, 2018, Wu et al., 2019). In fact, commonly spaced grids (e.g. SAP 24-2, Figure 1.13) might identify defective areas, as well as their changes, with less precision compared to denser grids (Phu et al., 2017, Schiefer et al., 2003). The employment of patterns sampling the VF every 6° within the central 24-30º was likely chosen to reflect the time constraints associated with thresholding algorithms instead of the ideal spatial resolution (Numata et al., 2017). As a consequence, certain areas of the field are significantly under-sampled when assessed by current strategies. For instance, the central VF, corresponding with the maximum RGC density, is not assessed with adequate detail compared to the mid-periphery (Nouri-Mahdavi, 2014, Turpin et al., 2016). There is, in fact, an entire line of research building on this rationale and evaluating the benefit earlier employment of 10-2 strategies in glaucoma (Wu and Medeiros, 2018, Chakravarti et al., 2021, Montesano et al., 2021a).
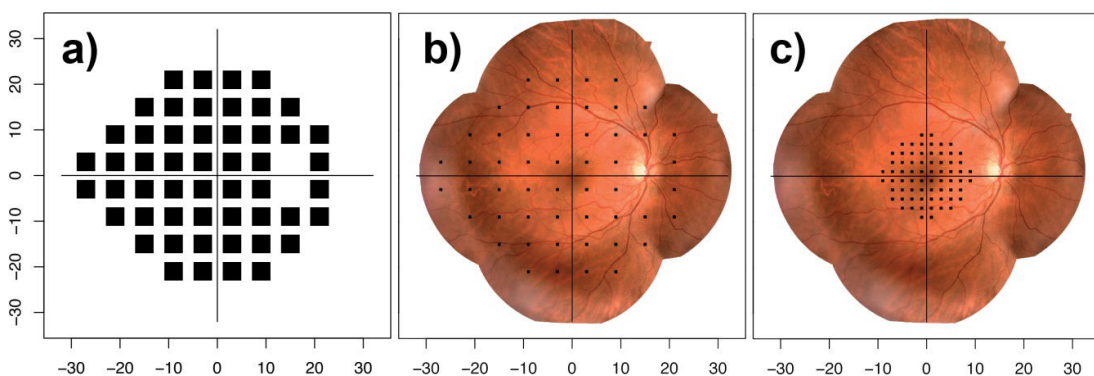
Figure 1.13 Different VF grids currently adopted in SAP strategies. (a) shows the 24-2 grid used in the Humphrey Field Analyzer. The sparse sampling of the VF is not immediately appreciable as in (b), where a 1:1 reproduction of the 24-2 grid and its stimuli are plotted on a wide-field fundus photograph (reproduced from https://www.centervue.com/). In (c) the same for the 10-2 grid.

Enlargement and/or deepening of existing defects are the dominant patterns of progression in glaucoma (Boden et al., 2004, Schiefer et al., 2003). This makes the assessment of scotoma borders critical for identification of glaucoma progression, yet, possibly limited by variability issues at damaged

locations (Gardiner et al., 2014). Nevertheless, detection of progression in previously normal locations, by means of a denser grid, may be a promising research direction.

Studies have shown the potential benefits of varying the resolution of the testing grid compared to current tests (Schiefer et al., 2003, Nevalainen et al., 2009). A significantly more accurate detection of glaucomatous defects was found in a study combining fundus-oriented perimetry with locally condensed grids (Schiefer et al., 2003). Similar results were achieved from a longitudinal report showing a greater sensitivity to identify progression when denser grids around defects were combined to typical 6° spaced locations (Nevalainen et al., 2009). Building on this promising evidence, another study tested a 0.5° dense perimetry pattern to identify the optimal resolution to maximise the detectability of glaucomatous abnormalities (Numata et al., 2017). Findings indicated that, especially within the central 10°, a density below 1.5° was optimum for identification of VF defects at a significantly greater accuracy (Numata et al., 2017, Numata et al., 2016).

An example of the implementation of individualised denser test patterns is GOANNA (Chong et al., 2014), an algorithm that automatically adds presentations at scotoma borders without increasing the testing time compared to conventional VF procedures. This strategy accomplished a more accurate characterisation of VF defects by presenting locally denser grids (3°) around scotomata (Chong et al., 2014). Further validation in a clinical scenario is required to confirm the promising results shown by computer simulation (Chong et al., 2015). In a similar approach, another algorithm was developed to add 10 extra locations to the standard 54 of the 24-2, around scotoma edges (Aoyama et al., 2014). In agreement with previous work (Chong et al., 2014), results showed that increasing the resolution of the grid might be beneficial to better define glaucomatous defects, particularly in areas of transition from damaged to normal (Aoyama et al., 2014).

Turpin and colleagues have recently introduced ARREST, a newer testing algorithm that firstly adopts a threshold strategy to create a baseline VF result (Turpin et al., 2018, Muthusamy et al., 2020). Subsequent examinations save

presentations by not performing complete threshold estimates at locations with reduced sensitivity (<17dB), deemed intrinsically unreliable. The time saved is rather spent to assess the areas surrounding defects in greater spatial detail. The algorithm presents stimuli at extra locations, along a 2º-dense grid, connected to regions marked as defective (<17dB). Although still requiring optimisation, the innovations introduced with ARREST are considerable. Firstly, the authors reinforced the rationale of censoring highly variable locations. In addition, greater accuracy in detection of glaucoma progression is sought through a paradigm change, which relies on spatial enlargement of defects rather than sensitivity changes from poorly reliable locations (Turpin et al., 2018, Muthusamy et al., 2020).

Overall, denser grids might represent a significant adjunct in the analysis of visual function in glaucoma with potential reduction in variability and a more precise definition of the spatial features of a scotoma, likely to yield to greater accuracy in diagnosis and progression detection (Wu et al., 2019). Potential concerns needing to be addressed before wider adoption of such approaches in clinics mainly apply to the test duration and the selection of additional locations which ideally should follow predefined and reproducible criteria.

1.5 The structure-function relationship in glaucoma

Limitations of current procedures used to evaluate structural and functional changes in glaucoma prevent the use of any single test for optimal diagnosis and monitoring. Consequently, clinicians perform an array of examinations to evaluate patients at risk of glaucoma, and in case of conflicting results and uncertain diagnosis, longitudinal signs of glaucoma are sought to achieve a more solid identification of the disease (Tatham et al., 2014). Combining structural and functional test results in glaucoma has evoked considerable attention through the years, due to the possibility of enabling a more confident diagnosis in case of corresponding defects across domains. Similarly, progression might be ascertained more consistently when information from structure and function are adequately integrated (Denniss et al., 2018, Malik et al., 2012, Tatham et al., 2014).

1.5a Early work and current status

Pioneering structure-function work dates back to the mid 1970's, when the first cross-sectional studies looked at the correspondence between enlarged cup-disc ratio and VF loss (Read and Spaeth, 1974, Douglas et al., 1974). However, it is the report from Quigley and colleagues, published more than 30 years ago, that is considered a first milestone within this research area (Quigley et al., 1989). The authors examined 6 post-mortem human eyes for which SAP results collected within 17 months of death were available, and showed that up to 20-40% of RGCs needed to be lost before a 5-10dB SAP loss would be measurable. Yet, results were later criticised for the small sample, and apparently for including eyes with RGC density significantly lower than the population average (Malik et al., 2012). Subsequent research in this area led to the suggestion that different structure-function models should be developed for different retinal eccentricities (Garway-Heath et al., 2000a): The greater RGC density of the central retina requires a greater number of cells to be damaged for a similar sensitivity defect compared to the periphery (Garway-Heath et al., 2000a). Further, discrepancy between measurement scales used for function (logarithmic, dB) and structure (linear, e.g. area in mm^2 and more recently thickness in µm) was also reported as a factor able to conceal the structure-function relationship (Garway-Heath et al., 2000a). These suggestions were in agreement with later evidence from animal models (Harwerth et al., 2004). Nonetheless, the linear relationship found in such structure-function models has been sought in human observers without similar success. Indeed, studies conducted and the models produced have achieved only a far from perfect structure-function relationship with considerable scatter (Yohannan and Boland, 2017).

Within current settings, clinicians are likely to assess the structure-function relationship to re-evaluate their confidence in a diagnosis of glaucoma once all examinations have been performed, with sizeable impact on clinical decisions (Lisboa et al., 2013, Malik et al., 2012). Despite the paucity of data describing how structural and functional data are related in everyday glaucoma clinics, experts' speculations suggest that this might not be ideal. Potential issues might depend on the structural and functional indices

considered as well as the approach employed to establish topographic agreement (Hood and De Moraes, 2018a). There is the risk that clinicians only perform a visual comparison between the two results, largely relying on self-developed explicative criteria. Such an approach has been reported to be rudimentary, highly subjective and potentially fragmented (Tatham et al., 2014). As shown in different areas of glaucoma diagnosis, the coupling between variability in results and application of subjective criteria of graders may yield suboptimal concordance and reproducibility among clinicians. Similar patterns might be expected in the evaluation of structural and functional data. Only a few groups have explored the reliability of glaucoma clinicians to relate structure and function, suggesting that the within- and between-observer agreement might be moderate to poor (van der Schoot et al., 2013, Moreno-Montanes et al., 2017). Additionally, results suggest that consideration of the structure-function relationship as currently determined may not necessarily lead to a greater accuracy in diagnosis of glaucoma, compared to the use of information from one domain in isolation (van der Schoot et al., 2013). On the whole, although more detailed investigation on the attitude of clinicians to combine these domains is warranted, concerns about the consistency and repeatability in establishing structure-function relationships in clinics appear to be grounded.

1.5b Topographic structure-function relationship

When structural parameters, such as OCT-measured retinal thickness around the ONH, need to be related with the corresponding functional status, a major issue arises from the indirect topographical correspondence between the retinal lesion and the served functional area. Structural measurements in glaucoma have typically been performed by assessing the neuro-retinal rim, and subsequently by measuring the cpRNFL thickness. There has, therefore, been considerable interest to investigate and develop maps enabling comparison between ONH regions and the corresponding VF locations on the retina (Yohannan and Boland, 2017, Denniss et al., 2018).

A milestone is represented by the work of Garway-Heath et al, presenting a clinically-friendly structure-function map to be implemented in both clinics and research (Garway-Heath et al., 2000b). Garway-Heath et al. superimposed the

SAP 24-2 grid on fundus photographs of 69 normal tension glaucoma eyes and subsequently adopted hand-tracing techniques of visible RNFBs to map each 24-2 location to the ONH. This resulted in clusters of VF locations mapped to six wedge-shaped sectors of the ONH. One of the shortcomings of the derived map is that it is limited to the 24-2 grid points, therefore not usable with different VF patterns (Denniss et al., 2018). In a partially similar approach, hand-tracing of visible RNFBs in fundus photographs was later used to derive a mathematical model able to map any single retinal location to the optic disc, (Jansonius et al., 2009, Jansonius et al., 2012). Schematic representations of both the Garway-Heath and the Jansonius maps are given in Figure 1.14.
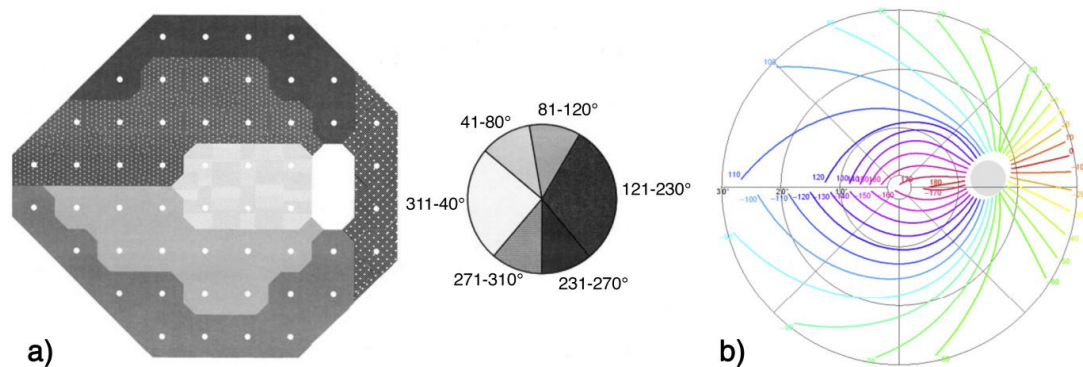
Figure 1.14 Structure-function mapping models proposed by Garway-Heath et al. (a) and Jansonius et al. (b), to relate VF locations to specific sectors or regions of the ONH. In the Garway-Heath map, 24-2 locations are grouped into 6 clusters and related to ONH sectors. According to the Jansonius model, any retinal location can be related to an ONH location, with 1° steps. In the scheme the trajectories corresponding to 10° steps on the ONH are shown. Reproduced from (Garway-Heath et al., 2000b, Jansonius et al., 2012).

Although representing a crucial first step in this area, both the Garway-Heath and Jansonius maps rely to varying extent upon hand-tracing techniques of RNFL bundles from fundus images, which may be a suboptimal method to map VF locations to the ONH. Firstly, variability in subjective hand-tracing of visible RNFBs by clinicians should be considered. Whereas one study found a small contribution of this component the overall variability of bundles trajectory (Jansonius et al., 2012), later work including a larger number of clinicians showed considerable between- and within-operator variability of the mapped ONH insertions, with the uncertainty increasing with distance from the ONH (Denniss et al., 2014c). Also, using fundus images acquired with scanning laser ophthalmoscopy (SLO) may limit the assessment to the superficial bundles of the RNFL, whose trajectories might differ to a varying

extent from ones found at greater depths (Ogden, 1983). One additional limitation of some of the models is their assumption of average anatomical parameters for all the assessed eyes. Hence, these maps are likely to provide accurate predictions in eyes with population-average morphology (Malik et al., 2012, Denniss et al., 2018), and thus defined as "population-maps" (Denniss et al., 2018). Alternative population-maps have been proposed in the literature, derived with several methods, such as the highest level of correlation between ONH and VF locations or axonal growth models. Overall, several structure-function maps showed a broad agreement with the one proposed by Garway-Heath, which has been often used as comparison (Denniss et al., 2018, Strouthidis et al., 2006). However, it should be noted that using such reference standard cannot be considered unbiased, since the reference itself lacks adequate validation (Turpin and McKendrick, 2021).

A critical shortcoming of a population map is the effect of anatomical variability on axonal pathways linking RGCs and the related VF locations to the ONH (Lamparter et al., 2013, Denniss et al., 2014a, Denniss et al., 2012, Jansonius et al., 2012). Indeed, it has been estimated from fundus photographs that the mapped insertion to the ONH of a VF location might vary on average between 20º and 30º, compared to population maps (Malik et al., 2012). Among ocular parameters, axial length, ONH position as well as the fovea-raphe angle seem to affect the mapping to the ONH (Denniss et al., 2014a, Denniss et al., 2012). To account for the effect of anatomic variability on bundle trajectories, individual-specific maps can be developed, aiming to more precisely establish correspondence between structure and function. These customised maps, which control for the individual anatomy, might be able to reduce the noise in the structure-function relationship arising from comparison of disparate VF and ONH locations (Malik et al., 2012).

Denniss et al. published details of a computational model able to combine individuals' anatomy with an axonal growth model to map any desired location of the central VF to the optic disc (Denniss et al., 2012). Briefly, axons were directed to an ONH location, based on the shortest path towards sectors with still space available to house the bundles. Additionally, a hierarchical order of filling the ONH sectors was included, where allocation of foveal and macular

axons was prioritised compared to other retinal areas. Besides being a structure-function map tailored on the individual anatomy, the model is independent from hand-tracing of RNFBs. When considering average ocular parameters, similar results to widely used population-maps were achieved (Denniss et al., 2012). Additional work performed by the same study group on the applicability of their model, assessed the effects of measurement errors on the accuracy of ONH mapped locations. As a result, the use of 30°-wide sectors centred on the ONH-fovea angle was suggested, representing a suitable compromise between precision in repeated measures and resolution (Denniss et al., 2014a). Indeed, whereas narrower sectors might encroach variability boundaries, the adoption of wider ones would require an averaging process where localised defects of the cpRNFL are likely to be hidden by neighbouring healthy neural tissue. A representation of the Denniss et al. model for varying anatomic parameters is given in Figure 1.15.
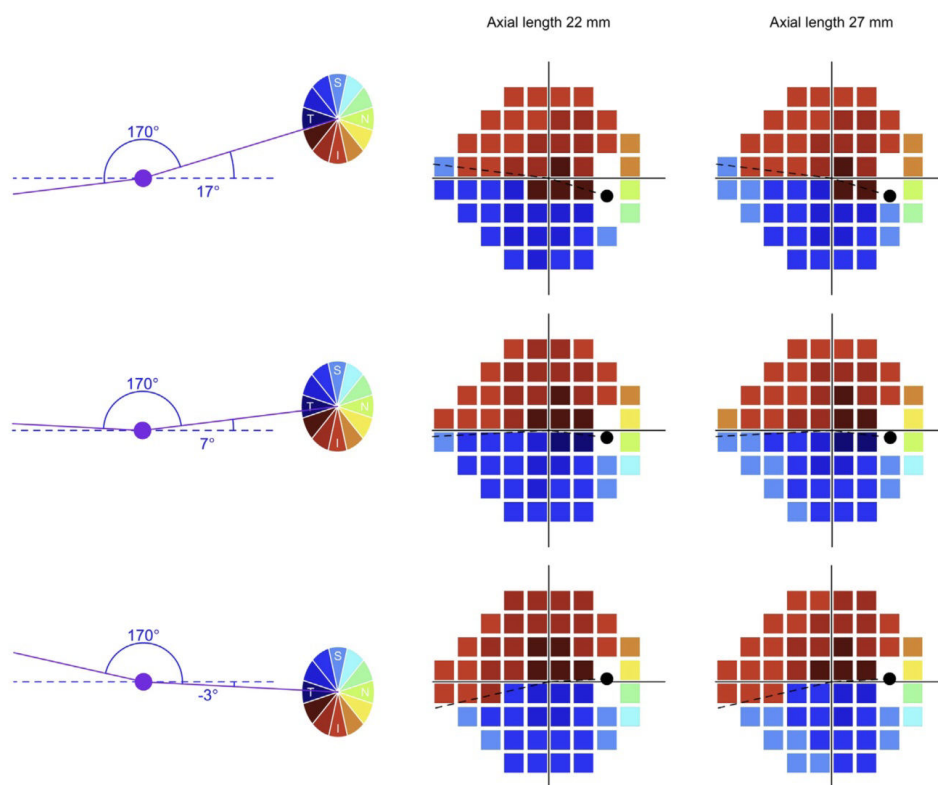
Figure 1.15 Diagrammatic representation of the 24-2 VF grid mapped to the ONH according to the Denniss et al. customised map (Denniss et al., 2012). The figure enables one to observe the variability of the mapped locations at the ONH, as a function of ONH position relative to the fovea (left column, varying fovea-ONH angle) and axial length (middle and right-most columns). Figure reproduced from (McKendrick et al., 2017).

As a result of the increased focus on the macula, coupled with advancements of segmentation algorithms in OCT devices, there has been increased interest in the macular structure-function relationship (Yohannan and Boland, 2017). Peculiar properties of topographical correspondence between the two domains at the macula require, however, consideration. Macular RGCs and photoreceptors are displaced to address physiologic requirements of detailed vision and maps controlling for this incongruity have been proposed in the literature (Denniss et al., 2018). Displacement magnitude is defined by the length of Henle fibres, which has been measured by histologic studies, enabling development of models of RGC displacements (Drasdo et al., 2007, Sjöstrand et al., 1999). In turn, these models have been adopted to relate structure and function in the macula (Raza et al., 2011, Hood and Raza, 2011). Generally, the displacement is maximal at 1º-3º eccentricity with values between 1º-2º (Figure 1.16), whereas displacement approaches negligible magnitudes in the periphery of the macula (Turpin et al., 2015).
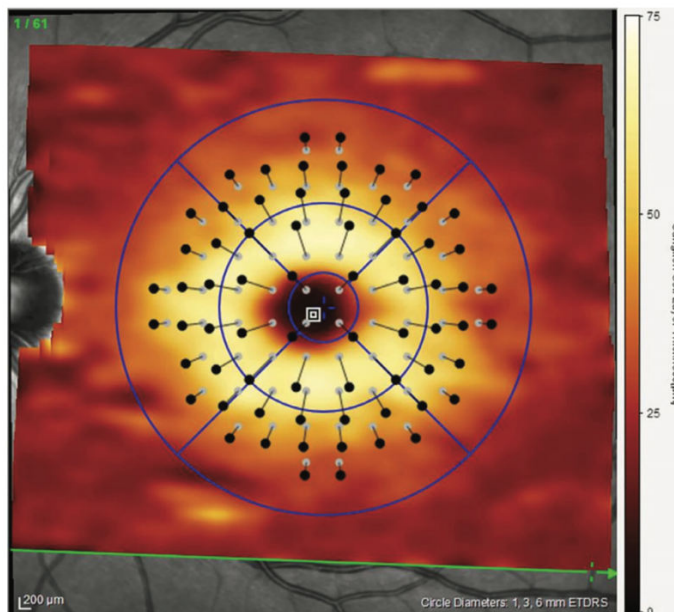
Figure 1.16 Model to relate VF locations (from a 10-2 grid, grey dots) to corresponding retinal location, accounting for RGC displacement (black dots) as described in (Drasdo et al., 2007). The model is superimposed to a macular OCT heat-map showing the Ganglion Cell Layer thickness. The circles and grids correspond to the macular areas classification as proposed in the ETDRS. Figure reproduced from (Denniss et al., 2018).

1.5c Recent research in structure-function

There has been sustained interest in structure-function over the years, resulting in significant production of scientific literature (Malik et al., 2012, Hood and Kardon, 2007, Sung et al., 2019). An overall assessment of recent studies is, however, complicated by the abundance of approaches used to explore this relationship. Heterogeneity across studies might arise from the

structural and functional parameters considered, the characteristics of samples studied, as well as the statistical approaches selected to explore the relationship. All these elements have been inconsistent across studies and have the potential to affect the strength of the relationship observed (Hood and De Moraes, 2018a). An additional methodological element to consider in structure-function studies, is whether the proposed design combines data from eyes with and without glaucoma. In fact, structure-function correlation in healthy subjects seems to be absent or poor (Araie et al., 2014, Chu et al., 2018, Denniss et al., 2014b), suggesting that inclusion of healthy and glaucoma eyes together could also alter the observed relationship. It should also be noted that the concoction of data from normal regions and ones with glaucomatous defects could be induced by the specific analysis, in studies only assessing eyes with glaucoma. For instance, a *generalised approach* to structure-function analysis, achieved by relating average cpRNFL thickness and SAP MD, as often proposed (see below), is similarly likely to assess the relationship between largely spared sectors of the ONH and intact VF areas. This could happen even if examining a group of glaucoma eyes only, especially if presenting early-moderate defects.

Structure-function at the ONH

Many studies have explored structure-function correlation implementing primarily structural measures of the ONH and surround. The large majority of reports sought relationship between the average or sectorial cpRNFL thickness and various SAP indices. Functional measurements frequently computed were the mean sensitivity of either the whole VF or the superior/inferior hemispheres. In other studies, MD, PSD or visual field index were alternatively adopted to this end. This approach of relating structure and function can be considered as *generalised*. In fact, it does not fully account of the individual variability, both in terms of ocular physiology and specific configuration of glaucomatous damage. Similarly, the typical glaucoma feature of developing localised lesions rather than widespread loss is not entirely considered. The strongest correlation indices achieved with such designs range between Spearman's ρ of 0.45-0.75 (Horn et al., 2014, Pinto et al., 2014). Available results are also consistent in showing stronger relationship

between inferior-temporal ONH sectors and superior-temporal corresponding VF regions, when ONH sectors were considered. Yet, the ranges of the numerous correlation indices found by these studies showed inferior boundaries as low as 0.1-0.3 (Pinto et al., 2014, Hollo, 2017, Chu et al., 2018).

A different approach is the one proposed in a study assessing the effect of increased measurement accuracy on structure-function correlation (Ballae Ganeshrao et al., 2015). The authors took several steps to reduce the variability of cross sectional measurements of each domain. Individual variability was also considered. Firstly, the most prominent defect in the cpRNFL was identified, and 30º wide ONH sectors were customised to be centred on the most damaged area. Additionally, ONH sectors were mapped to VF locations via an individual-specific structure-function map (Denniss et al., 2012). Gains in strength of correlation brought by each step of enhancement were assessed individually and compared with a "baseline" condition embodied by the relationship between single scan-measured cpRNFL thickness of Garway-Heath map sectors and the average sensitivities of the corresponding VF clusters. The highest correlation with baseline settings was 0.52 (95%CI: 0.13-0.78) at the superior-temporal ONH sector. Remarkably, no significant increase in correlation was observed with various enhancements except when the customisation of sectors was added. In this latter case, highest correlation was 0.77 (95% CI: 0.47-0.92), resulting from all enhanced measures combined with the custom structure-function map and ONH sectors centred on the most abnormal cpRNFL location (Ballae Ganeshrao et al., 2015). When adopting the same measurements, but considering standardised Garway-Heath map's sectors, the correlation was 0.62 (95% CI: 0.20-0.85). As also underlined by the authors, an averaging process is expected when a focal RNFL defect falls within a large and fixed sector, as the 40º to 90º wide sectors with fixed boundaries of the Garway-Heath map (Figure 1.14a). Also importantly, narrow sectors with fixed boundaries might split defects when they are localised at their borders, resulting in both instances in inadequate representation of the disease (Ballae Ganeshrao et al., 2015).

Although the proposed strategy is open to further refinement, the novel approach to minimise mixture of spared regions with locations actually affected by glaucoma is considerable. Indeed, by considering the most abnormal cpRNFL sector, it could be expected that focus is directed on a region of the ONH highly likely to be damaged by the disease. The benefits carried by such an approach could be significant, not only in terms of strengthening the structure-function relationship, but even in an innovative approach to the assessment of cpRNFL.

Structure-function at the macula

Among studies assessing macular structure-function relationships, early work employed measures of the whole retinal thickness. A characteristic approach involved averaging thickness measures over the entire macula, or in hemispheres and/or sectors, and then relating it to central visual function as measured with SAP 10-2. As an alternative, the central 12 or 16 locations of the 24-2 grid have been considered, representing 3 or 4 test points per VF quadrant, respectively. Overall, these studies showed a stronger correlation between the inferior-temporal macula sector on OCT with corresponding function at the superior-nasal VF (Boling et al., 2012, Mota et al., 2016, Liu et al., 2017). Such relationship was however imperfect, with best coefficients of 0.49-0.57 (Liu et al., 2017, Rolle et al., 2016). In fact, weaknesses such as the inclusion of retinal layers which are not primarily involved in glaucoma in the analysis have to be considered. Strategies to increase the strength of macular structure-function relationships include the focus on retinal layers more affected by glaucoma, such as the RNFL, the GCL and the contiguous IPL (Lee et al., 2017, Hirooka et al., 2016, Sato et al., 2013). Studies including such implementation overall achieved stronger correlation coefficients, with values around 0.7 (Lee et al., 2017, Hirooka et al., 2016, Sato et al., 2013).

As described above, RGC displacement in the macula must be accounted for when relating structure-function in this area. This may be especially true when the analysis includes or is limited to the RGC-related layers, i.e. the GCL, the IPL or their combinations (Ohkubo et al., 2014). Accordingly, several studies have included RGC displacements in structure-function analysis (Lee et al.,

2017, Rao et al., 2015), with some findings suggesting a noticeable impact on the strength of correlation for the very central locations of the macula, between 1º and 3º eccentricity (Ohkubo et al., 2014).

Limitations of structure-function in glaucoma

Available evidence leads to the current belief that, according to the same causative factor, structural and functional changes in glaucoma develop following a common timescale. It is likely that the inconsistencies are explicable by technological paradigms used to test and relate structure and function. These include the testing strategies, metrics and scales used to perform the measurements and their associated errors, as well as baseline individual variability (Denniss et al., 2018, Malik et al., 2012, Hood and De Moraes, 2018a). As a result, the relationship between VF sensitivity and corresponding structural damage remains imperfect. A number of factors are described as contributing to the scattered relationship. Although these have been partly introduced earlier, a brief summary follows (Denniss et al., 2018, Malik et al., 2012, Lucy and Wollstein, 2016).

- Measurements of structure and function have different dynamic range, with VF sensitivity theoretically able to reach 0dB and beyond, whereas a considerable floor effect affects structural measures (Mwanza et al., 2015, Bowd et al., 2017). At advanced stages of damage, although a substantial thickness is still measurable, this is deemed to represent non-neural components of the examined tissue while the eye might have lost all corresponding function.
- Test-retest variability can be substantial in perimetry, especially for damaged areas where a large part of the dynamic range might be consumed by measurement error. Despite structural measurements being generally less variable than SAP, the effect of poor image quality, media opacities, and segmentation errors on OCT indices should also be considered.
- The 24-2 testing grid might limit structure-function relationships due to its inadequate sampling in retinal regions with greater RGC density.

- Although associated to remarkable resolution, current structural measurements represent a surrogate of the primary object of glaucoma damage, which is ultimately the RGC. Indeed, some inaccuracy can be found in the correspondence between these two variables.

- Structural measurement scales, such as μm, are linear, whereas light sensitivity is typically measured in dB, a logarithmic scale. A stronger and more linear relationship has been suggested between structure-function when both are related through linearised scales.

- Apparent topographic disagreement between structural and functional lesions might be the consequence of the current mapping models adopted.

- Cross-sectional designs might be a suboptimal approach to assess the structure-function relationship, due to considerable between-subject variability. For instance, longitudinal data might offer the opportunity to normalise for baselines values and measure defects according to differences from baseline (Goren et al., 2013, Mohammadzadeh et al., 2020).

- Lastly, it should be considered that visual function assessed in VF is determined by the whole visual pathway, and its relationship with structural changes measured only in the retina could be imperfect even with flawless measurements in both domains. For instance, analyses limited to the RNFL may fail to capture structural consequences of glaucoma in their entirety (Frezzotti et al., 2014). Compensatory neural mechanisms could also be established, to reduce the impact of RGC damage on visual function (Bham et al., 2020).

1.5d Future directions

As discussed above, the lack of perfect agreement between structure and function can, at least in part, be explained by current methods of examination and their interpretation. Suggestions by Hood and De Moraes to obtain an improved structure-function relationship include: i) a stronger consideration of macular structural lesions; ii) use of denser VF grids, at least in the macular region, to depict early functional loss with greater sensitivity; and iii) a more substantial use of a topographical approach, including normalisation for the

baseline anatomy of the specific patient (Hood and De Moraes, 2018a). Overall, the proposed guidelines seem to point toward a greater emphasis on glaucomatous defects, achieved through a focus on damaged locations in both domains, while excluding from analysis spared regions, which are affected by considerable individual variability. Building on available models (Ballae Ganeshrao et al., 2015), standardised methods to prioritise such approaches should be further developed and tested in order to explore their benefits.

One way to exploit the structure-function relationship in glaucoma might be the development of newer VF paradigms able to combine information from the two domains, aiming either to reduce measurement error or to be more informative of the disease status (Denniss et al., 2018). Among a variety of applications of custom perimetry, one approach aims to identify patient-specific regions of interest, deemed to deserve a more accurate exploration of functional status. For instance, this might be the case of a localised and obvious cpRNFL defect, whose corresponding function could be assessed by a denser test pattern to increase confidence in glaucoma diagnosis (Denniss et al., 2018). On this line of research, Ganeshrao et al. sought to increase the number of VF locations corresponding to the most abnormal ONH section as observed on cpRNFL scans (Ballae Ganeshrao et al., 2018). In their study, the ability to spatially detect VF abnormalities for different grids was compared. The analysis included a modified version of the 24-2 grid, where half of the locations were selected from a pool of 2° dense VF points in relation to the most prominent cpRNFL lesion in each individual. A dense supra-threshold strategy served as a reference standard in every participant, then the number of defects identified by the grids (all fixed at 52 locations) was considered as an outcome measure. Significant differences in performance between different test patterns were found in 14 participants out of 23, among which the approach tailored to the patient's defect found more abnormal locations than alternatives in 9 cases. Two considerable advantages were introduced in this work, including an automated approach to select alternative locations to test and the number of tested locations of the customised grid kept within current SAP limits. Nevertheless, the selection of retinal areas to be tested was not completely structurally driven, with half of the original 24-2 locations preserved in the

strategy, addressing the variability of structural measurements (Ballae Ganeshrao et al., 2018). Although this might have been able to depict some unpredicted damaged areas, an abnormal ONH sector might have been not fully assessed functionally.

In the aforementioned report, regions more likely to present glaucomatous defect were identified in circumpapillary OCT scans (Ballae Ganeshrao et al., 2018), which might have underestimated defects observed the macula. Moreover, structure-function correspondence was established by means of structure-function mapping models whose full validation is still unavailable because of the lack of reference standards. The use of wide-field en face OCT to assess reflectance abnormalities of the RNFL (Ashimatey et al., 2018a) was discussed previously (section 1.3d). This approach minimises the role of structure-function maps, and may represent a considerable resource in custom perimetry. Alluwimi and colleagues presented works showing significant spatial agreement between structural lesions, as determined by en face images, and custom perimetry at corresponding locations (Alluwimi et al., 2018a, Alluwimi et al., 2018b). Although both studies were limited by a subjective – and therefore potentially inconsistent – definition of structural lesions and a perimetric strategy not validated elsewhere, their findings are promising, making further examination of more repeatable strategies highly valuable.

Before implementation of custom perimetry can take place in glaucoma clinics, several possible constraints must be considered and addressed. For instance, exploring the functional status with denser grids requires the use of additional presentations. This is likely to result in longer test strategies, known to increase observer fatigue, leading to greater response variability. Additionally, the allocation of supplemental VF locations has often been inconsistent; clinicians and researchers might have added extra VF points according to self-developed and potentially poorly repeatable criteria. To tackle such limitations, strictly defined criteria to identify locations to be tested – ideally by using automated algorithms – are desirable. Similarly, additional strategies to interpret results at population level must be available when patients will be tested according to different test patterns and this might also result in auxiliary

burden for clinicians (Ballae Ganeshrao et al., 2018). Further considerations might involve how the test pattern will be updated for its use to collect longitudinal data. In fact, having a fixed region of interest throughout the whole course of the disease would be unlikely. Instead, it could be expected for any region of interest to be itinerant as a function of glaucoma progression. Lastly, as underlined above, the merit of spending extra presentations at locations that are likely to be damaged and therefore associated to more variable results is questionable. A more detailed testing of these areas might be highly valuable for the diagnosis of the disease, where topographic agreement between structure and function is likely to lead to increased confidence of diagnosis. Similarly, high density grids of the borders of a defect could prove beneficial to assess progression in early glaucoma. However, the same approach is likely to be suboptimal when focussing on the detection of glaucoma progression at moderate to advanced stages of disease, when a substantial portion of the VF may be damaged. In these cases, predicted spared regions might be more informative for a prompt and accurate detection of the enlargement or the development of new defects.

1.6 Overall aim and research questions

As discussed in details in previous sections glaucoma demands an early and accurate diagnosis. OCT is increasingly considered an essential tool in the assessment of glaucoma (Tatham et al., 2015, World Glaucoma Association, 2016), and although it has brought significant support to clinicians in the diagnosis of glaucoma (Jindal et al., 2019), detection of early cases through single cross-sectional exams remain challenging (Founti et al., 2018, Hohn et al., 2018, Karvonen et al., 2018). As such, there remains a strong rationale to develop methods for more confident glaucoma diagnosis, especially with high specificity to minimise false positives.

En face OCT imaging holds promise in this regard. The technique now enables evaluation of RNFB reflectance in greater detail compared to previous imaging modalities, resulting in particular interest in glaucoma. In fact, en face OCT provides a method to relate structural and functional data independently from retina to optic disc maps, and this is meaningful in the exploitation of structure-function relationship via approaches such as custom perimetry. Furthermore,

evidence from on animal models indicates that loss of RNFL reflectance may precede changes of RNFL thickness. Accordingly, evaluation of RNFBs reflectance in en face imaging might provide clinicians with a new tool for prompter and more accurate detection of glaucoma.

Analysis of RNFB reflectance in en face OCT images is, however, a relatively new area in glaucoma, with little research available. Pending gaps include an incomplete understanding of RNFB arrangement in healthy and glaucoma eyes, as well as the relationship between reflectance modifications and RNFL thickness in clinical settings. There is also no consensus on the ideal approach to extract reflectance defects in en face images, with criteria driving the choice of slab characteristics that are poorly defined. For instance, current studies do not allow complete understanding of how parameters of slab construction could affect the ability to detect glaucoma defects, and how these should be optimally set. As such, this project aimed to better understand RNFB characteristics in healthy and glaucoma eyes and to identify optimal strategies for automated and objective extraction of reflectance defects from en face OCT images. This is essential for wider and more consistent adoption of en face analyses in glaucoma, with less dependence on clinicians' own judgement. To this end, the following research questions were addressed:

1. How does configuration of RNFBs change across the retina and in different eyes, and what is the relationship with conventional measures of RNFL thickness?

2. Can measures of RNFB reflectance be used to detect glaucoma defects with greater accuracy than conventional RNFL thickness?

3. What is the effect of slab construction on the objective and automated extraction of reflectance defects in en face OCT images?

4. What is the relationship between automatically extracted RNFB en face defects and conventional measures of glaucoma damage such as thinning of the cpRNFL and visual loss in perimetry?

2. Depth-resolved variations in visibility of retinal nerve fibre bundles across the retina in en face OCT images of healthy eyes

2.1 **Abstract**

Background: Recent SD-OCT developments enabled direct en face visualisation of RNFB loss in glaucoma, however, the optimum depth at which to visualise RNFBs across the retina is unknown. We aimed to evaluate the range of depths and optimum depth at which RNFBs can be visualised across the retina in different eyes.

Methods: High-resolution wide-field scans of the central retina in 10 healthy adults (median age 68.5 years, range 57 to 75 years) were acquired using SD-OCT. To minimise artefacts, slab images of the maximum axial resolution (3.87µm) containing depth-resolved attenuation coefficients were extracted from 0 to 193.5µm below the ILM. Bundle visibility within 10 sectors of a superimposed grid was assessed subjectively by trained optometrists (n=8). Anterior and posterior limits of RNFB visibility and depth of best visibility were identified for each grid sector. Effects of retinal location and individual eye on RNFB visibility were explored using linear mixed models. Intraclass correlation coefficient (ICC) was used to measure overall agreement and repeatability of grading. Spearman's correlation was used to measure correlation between depth range of visible RNFBs and RNFL thickness.

Results: Retinal location and individual eye affected anterior limit of visibility ($\chi^2_{(9)}$ = 58.6 & 60.5, both p<0.0001). Yet, none of the differences exceeded instrument resolution making anterior limit consistent across the retina and different eyes. Larger differences were observed in the posterior limit of visibility across retinal areas ($\chi^2_{(9)}$ = 1671.1, p<0.0001) and different eyes ($\chi^2_{(9)}$ = 88.7, p<0.0001). Depth of best RNFB visibility was around 20µm below the ILM in most regions. It varied slightly with retinal location ($\chi^2_{(8)}$ = 58.8, p<0.0001), but was not affected by individual eye ($\chi^2_{(9)}$ = 10.7, p=0.29). RNFB visibility showed good agreement between graders (ICC 0.89, 95%CI 0.87-0.91), and excellent repeatability (ICC 0.96 to 0.99). Depth range of visible

RNFBs was highly correlated with RNFL thickness (ρ = 0.9, 95%CI: 0.86-0.95).

Conclusions: The range of depths below the ILM with visible RNFBs varies markedly across the healthy retina, and changes are consistent with RNFL thickness. Optimal depth for RNFB visualisation varies across the retina, being at around 20µm below the ILM in most locations. To extract all RNFB information consistently across the retina, slab properties should account for differences across retinal locations and between individual eyes.

2.2 Introduction

Even though the use of en face OCT images is a promising perspective in glaucoma care, current understanding of the arrangement of RNFBs in healthy and glaucoma eyes is limited. A thorough understanding of the normal appearance of RNFBs in en face OCT images is important for future development of objective methods that aim to detect glaucomatous defects in these images. For instance, it remains unclear whether visibility of RNFBs in en face images varies across the healthy retina and between individual eyes. It is, therefore, also unclear what depths should be considered to capture a consistent portion of the RNFL through the retina.

As described earlier (see section 1.3d), en face analysis is often accomplished by collapsing the reflectivity information from volumetric OCT scans over a certain range of depths into a 2D image. The resulting transverse slab is frequently derived from a fixed retinal thickness (often 50µm) beginning anteriorly from the ILM (Hood et al., 2015, Sakamoto et al., 2019). The use of a fixed thickness slab to examine retinal areas with varying morphology may be expected to lead to an uneven composition of en face images. Indeed, some areas may contain RNFL only, whereas some also include deeper retinal layers with different intrinsic reflectance. It could be expected that combining retinal layers in this way would result in different appearance and detectability of glaucoma defects across the retina (Ashimatey et al., 2018a). Additionally, there may be situations where it is desirable to use a single slab thickness across the whole retina examined. In these cases, it could be expected that the slab's axial positioning within the thicker regions of RNFL may affect the

ability to detect changes due to glaucoma. As such, it may be convenient to know whether there exists an optimum depth for visualisation of RNFBs and how this varies across the retina. Examining glaucoma defects at this depth may provide higher specificity than other depths if it results in a lower chance of misidentifying normal variation as a defect.

This study aimed to develop preliminary understanding of the configuration of RNFBs in healthy eyes as observed subjectively by clinicians in en face OCT images. Specifically, the study sought to identify the range of depths and the optimal depth for subjective visualisation of RNFBs across different retinal regions in eyes of healthy adults. Further, agreement between- and repeatability within-clinicians in determining RNFB visibility was assessed. Since RNFL thickness varies across the healthy retina (Varma et al., 1996), RNFBs could be expected to be visible over a varying range of depths at different retinal locations and in different eyes. The results contribute to identifying the range of depths that future automated image analysis methods should consider to consistently assess the RNFL.

2.3 Methods

Settings and participants

Healthy volunteers aged over 50 years were recruited from the University Eye clinic (School of Optometry and Vision Science, University of Bradford) and through adverts in newspapers, high-street optometric practices and hospital eye departments. Older adults were targeted to include an age range relevant to primary open angle glaucoma (Rudnicka et al., 2006). One eye per participant was included, and if both eyes were eligible, the included eye was selected at random. The data collection process was part of a larger project, where part of the testing procedures was joint with a second experiment. This, as well as additional time needs, might have resulted in two different visits being required to collect all experimental data for each participant. In cases where two appointments were run, these took place within no longer than 4 weeks. Written informed consent was given by each participant and the study achieved ethical approval from the National Health Service's Research Ethics Service.

To establish eligibility, a comprehensive examination was performed, including refraction, GAT and slit lamp examination assessing media transparency and ocular health. Additionally, perimetry was performed on all participants (24-2 SITA-Standard, Humphrey Field Analyser, Carl Zeiss Meditec, Dublin, CA) and the ONH and macula were further assessed with OCT (Spectralis, Heidelberg Engineering GmbH, Germany). Willing volunteers were included if presenting refractive error within ±6.00DS and ±3.00DC, with best corrected visual acuity equal to or better than 6/9.5 (0.20 LogMar) and with uncomplicated cataract surgery or clear lens as defined by a Lens Opacity Classification System (LOCS III) score equal to or below NO3, NC3, C2 and P2 (Chylack et al., 1993). Additionally, all participants had normal VF, defined by GHT within normal limits combined with a normal MD (p>0.05) and the absence of a VF defect (defined as three contiguous non edge points <5% on the pattern deviation plot). Participants with positive history of glaucoma or presenting other conditions that could affect vision were excluded. Similarly, eyes with tilted disc or any other anomalies affecting the optic nerve, or high IOP in either eye (>21mmHg) or significant difference between the two eyes (>4mmHg) were also excluded.

OCT imaging and processing

Multiple OCT high-density high-speed scans were acquired, encompassing overall the central ±25º volume of the retina. By adopting the OCT's built-in fixation dislocation tool, seven volume scans were performed in different positions of gaze, all of which consisted of 30µm separated B-scans. Such resolution resulted in an average of 9.65 B-scans per degree, whereas every B-scan had a density of ~25 A-scans/degree. To reduce measurement noise, all performed acquisitions were supported by the use of the Spectralis eye tracker (Automated Retinal Tracker) which was set to 16 B-scans per location, meaning that for each retinal location 16 B-scans were averaged. As shown in Figure 2.1, scan width and the orientation of B-scans was adjusted to yield full retina coverage and high image resolution in a shorter time. It should also be noted that peripheral scans were externally shifted in relationship to the acquisition window (Figure 2.1a), aiming to maximise retinal coverage while minimising overlaps between neighbouring scans. The protocol resulted in

considerable overlap between superior scans and inferior scans as well as between superior/inferior and central scans. The overlap between central scans was instead marginal. Dilation was performed only if inadequate scan quality was achieved in physiological conditions (i.e. below 25db as suggested by the manufacturer and as proposed elsewhere (Ctori and Huntjens, 2015)). All images were acquired with signal to noise ratio above 20dB as per the manufacturer's instructions.

Conventional OCT imaging of the ONH was also performed via a circle B-scan (3.5mm diameter) around the ONH which provided the mean cpRNFL thickness as automatically computed by the Spectralis built-in software.
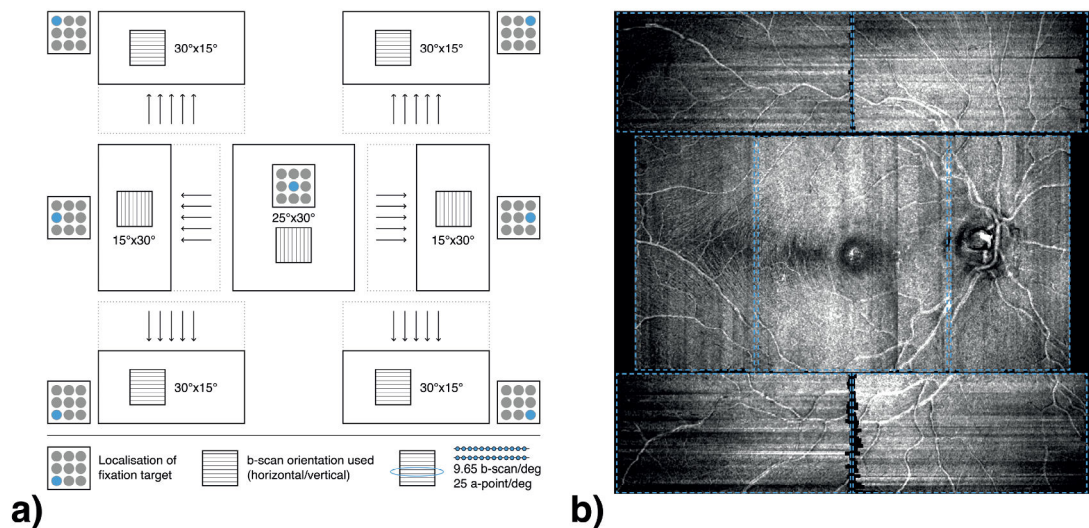
Figure 2.1 In (a), the wide field OCT acquisition protocol comprising of 7 volume scans (black boxes) reproduced within the acquisition window (dashed boxes) at each scan location. Scan dimensions and orientation are shown for every location. The arrows indicate the displacement of scan positions from the manufacturer's default positions. (b) shows an example of the resulting attenuation coefficient pseudo-montage for one eye 20µm below the ILM. Individual volume scan positions are shown by the blue boxes.

Code developed in *R,* version 3.5.1 (R Core Team, 2020) was used to extract single-pixel thick slabs containing depth-resolved attenuation coefficients (Vermeer et al., 2013) from volumetric data. Slab images encompassed a volume from 0 to 193.5µm below the ILM, as automatically segmented from the OCT's built in software. Manual correction of segmentation was performed in case of inaccuracy and errors. Each image resulted in a 3.87µm thick slab, corresponding to the maximum digital axial resolution of the instrument (Figure 2.2). The greatest depth explored with this method (193.5µm) was selected in order to approximately include the maximum expected cpRNFL thickness for

a 50 year old healthy eye. This was defined as the superior 95th percentile of the instrument reference database for the thickest region of the cpRNFL (i.e. superior temporal and inferior temporal sectors). As proposed elsewhere (Ashimatey et al., 2018a), attenuation coefficients, rather than raw intensity of each pixel, were extracted by using equations 17 and 18 in (Vermeer et al., 2013). Attenuation coefficients are thought to describe how quickly incident light is attenuated when passing through the retinal portion of interest, independently from the amount of light received (Chang and Bowden, 2019). These coefficients were originally developed to minimise shadowing effects of blood vessels on underlying retinal tissue, however, attenuation coefficients have also been used to diminish artefacts of reflectivity due to media opacity and poor quality B-scans (Ashimatey et al., 2018a). Once raw images were converted, data containing attenuation coefficients for each retinal area at different depths were imported into MATLAB for image processing (Version 9.6.0, The MathWorks Inc., Natick, Massachusetts).
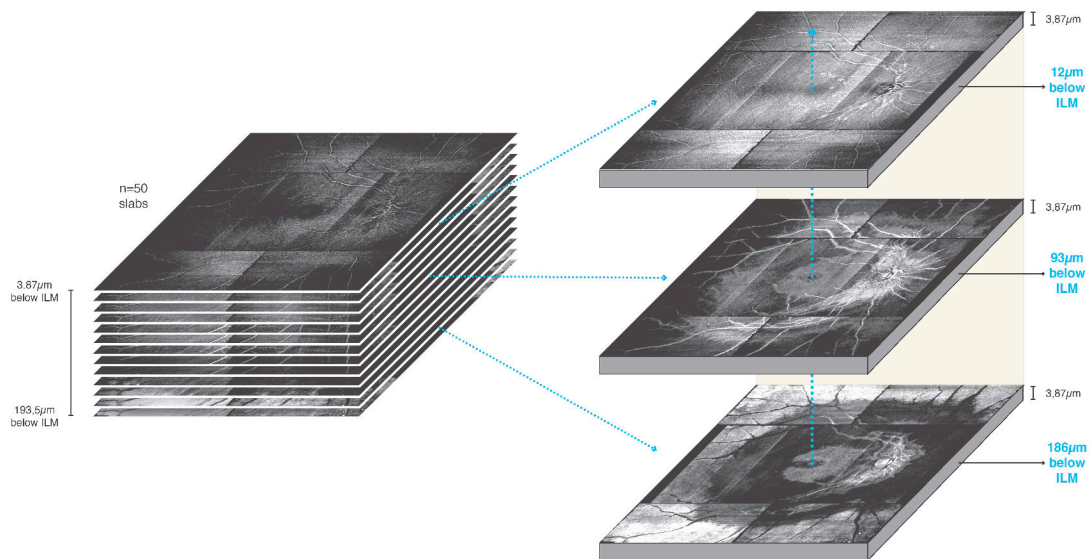
Figure 2.2 Schematic representation of the OCT volume acquired and analysed.

Imported single images were rotated and resized as needed, in order to reproduce the original retinal proportion. Bi-cubic spatial interpolation was used to account for the asymmetric resolution of vertical and horizontal A-scans. Subsequently, at each depth and for each individual eye the 7 single images were collated together into a single pseudo-montage. As a consequence of some overlap between individual scans, the adopted method

resulted in partial duplication of information from certain retinal areas that was accounted for at later stages (see below). Minimal image processing was then applied to optimise image visualisation without altering the relative intensities of different slabs within the same participant's eye, according to the following steps (examples of histograms corresponding to each transformation are showed in Figure 2.3):

- *Noise reduction:* A smoothing filter (3x3 Wiener filter) was applied to all images to reduce noise. The selected filter adopts a deconvolution approach and was applied at earlier stages of image processing transformation to minimise the effect of noise on subsequent transformations.

- *Partial threshold:* Considering only the slab image 35μm below the ILM, a ~100x50 pixel rectangle in the raphe area with minimal to no blood vessels was extracted from the composite image of every participant. This area can be deemed to be lacking any bundles, and was used to determine a background reference value used to reset the inferior limit of the image's range (Ashimatey et al., 2018a). A value 30% below the median pixel value within this rectangle was computed as 'background'. This additional reduction was adopted to identify an intensity measure sufficiently below those of the hypo-reflective RNFBs. A partial threshold transformation was then applied. Firstly, this background value was subtracted from all pixels of all slabs, then all values falling below 0 were clipped to 0. This resulted in the background value being set as the new zero of the image.

- *Normalisation:* Lastly, for each participant, the median of the 99th percentiles of all 50 slabs was computed. The identified value, corresponding to an average measure of the most intense pixels among extracted images, was used to normalise each participant's slabs by division, with values above 1 being clipped to 1. This process left all pixel values within a 0 to 1 range, before 8 bit images were extracted for display in the grading task. Using the same value across all the slabs aimed to preserve a meaningful comparison between different depths.
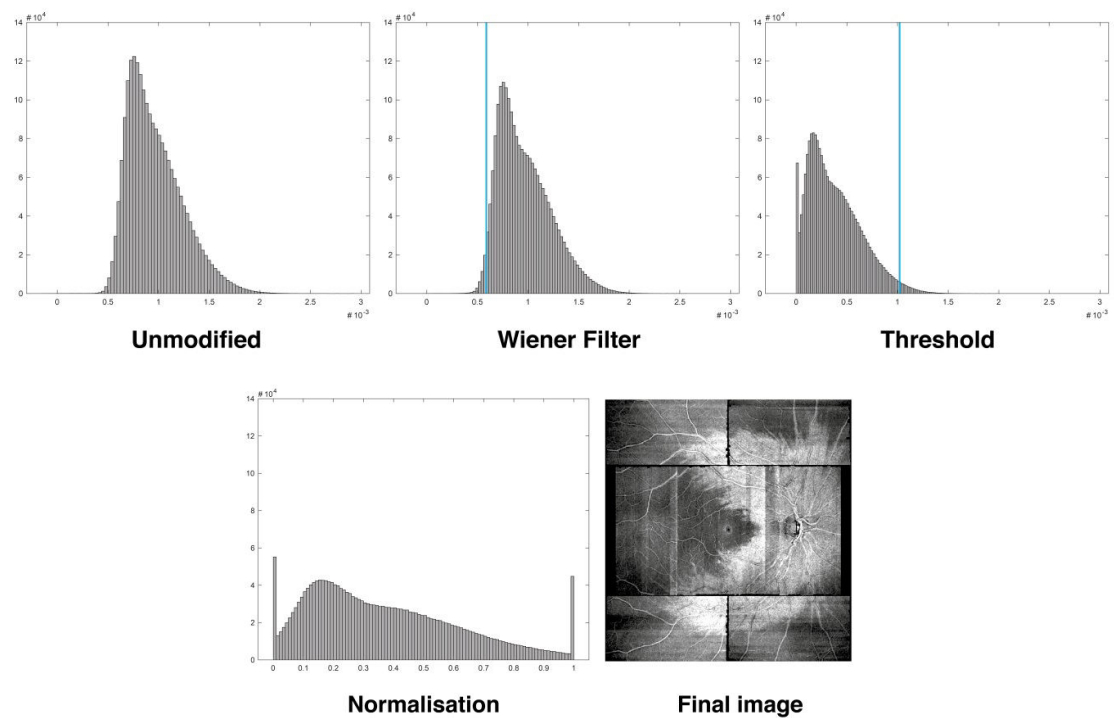
Figure 2.3 Example of the image processing for a healthy eye and its slab image (39μm below the ILM). Corresponding transformations are shown below the histograms as well as the values of background (blue line in the Wiener filter histogram) and superior limit (blue line in threshold histogram).

Grading task

Images from the same participant were arranged in a single presentation file to facilitate a quick examination of different depths while preserving consistency between image positioning. Subjective visibility of RNFBs was rated over 10 retinal regions of interest, each approximately consisting of a 15x15° square apart from the macular area which was further split into temporal and nasal regions (Figure 2.4). To account for anatomical variability between different eyes (Bedggood et al., 2017, Denniss et al., 2012), the central row of the grid was tilted as needed to follow the ONH-fovea angle as well as the fovea-raphe angle in each individual eye. The number of retinal regions identified represented a trade-off between higher resolution and time required to complete the task within manageable limits. To reduce the risk of bias from duplicated information, overlapping areas from different scans were obscured and graders were instructed to ignore those areas during the task (Figure 2.4). An example of the presentation file showed to clinicians is reported as supplementary material (Appendix A, Supplementary File 2.1).
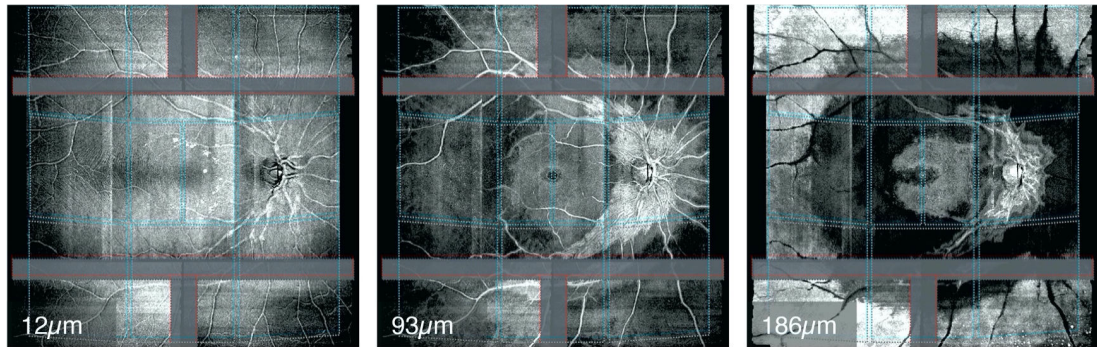
Figure 2.4 Example of the grid (blue lines) superimposed onto 3 of the 50 pseudo-montaged slab images for one eye, delimiting the regions of interest in the retina. Grey patches within red dashed bounding lines indicate the overlapping areas that graders were instructed to ignore. From left to right, the images correspond to increasing depth below the ILM (reported in μm), showing how the visible presence of RNFBs changes moving further down from the ILM. At 12μm below the ILM, RNFBs are visible almost throughout the whole retina, with the exception of the raphe, the fovea and peripheral portions of the inferior temporal, nasal and superior temporal regions. The reduction of reflectivity in these areas can be interpreted as reduced density or absence of RNFBs. At 93μm below the ILM, RNFBs are present only in the ONH region, with arcuate wedges of RNFBs in the inferior and superior central retina following the vascular arcades. Lastly, at 186μm below the ILM, it is possible to appreciate hyper-reflectivity of the retinal pigment epithelium in almost all peripheral regions and the macula, whereas only a few visible bundles are present around the ONH.

Images were displayed on a 13" laptop computer (MacBook Pro 2017 version, Apple Inc, USA) under standardised lighting conditions. To minimise the impact of any learning effect, data from different healthy participants were shown in random order. Before performing the task, all graders were provided with the same written instructions (Appendix A, Supplementary File 2.2), which included information on characteristic appearance of RNFBs, visible artefacts and potential confounders such as glial alterations (Ashimatey et al., 2018b). Instructions were repeated during the task in response to grader's queries. Subsequently, graders were allowed to scroll within participant's images unlimited times, without any time limit. The following subjective judgements were collected for each region of each imaged eye:

i) Boundaries of RNFB visibility: The first depth below the ILM at which RNFBs become visible in that specific region (anterior limit, μm) and the depth below the ILM at which RNFBs are last visible in that specific region (posterior limit, μm). Graders were instructed to consider RNFBs to be present when 25% or more of the region was occupied.

ii) Best visibility of RNFBs: The depth of greatest RNFB visibility (μm) according to features such as intensity and sharpness while ignoring the proportion of the region occupied from RNFBs. Since the best

visibility value aims to identify a single pixel depth for optimum observation of bundles, in the case of equal visibility of RNFBs across multiple depths graders were invited to report '*none*'.

Eight graders, including the author and the supervisor (JD) (median age 29.5, range 26 to 41; median years since qualification 7, range 2 to 16) completed the grading task. They all were optometrists, from the staff and post-graduate members of the School of Optometry and Vision Science at the University of Bradford. The optometrists had varying level of post-qualification training (4 out of 8 had undertaken specific training in glaucoma), but apart from the two member of the research team (RC, JD) they were all naïve to en face OCT images of RNFB reflectivity. All were pre-presbyopic, with self-reported normal vision, and were invited to perform the task with their habitual refractive correction. On a subgroup of graders (n=2), the task was repeated 3 times overall on 3 randomly selected participants and with no information regarding previous grades available to assess repeatability. Repeated measures were taken 1 week apart and graders were monitored throughout the execution of the task.

Data analysis

Data on RNFB visibilities (anterior limit, posterior limit and best visibility) were characterised by multiple dependencies, since visibility was assessed at multiple locations for each individual eye. Also, since measurements in different individuals were expressed by the same graders, these were not fully independent. To account for the inter-related nature of the data, we used linear mixed models to explore the effect of retinal location and individual eye on anterior limit, posterior limit and best RNFB visibility. Linear mixed models largely rely on principles of linear models, where the effect of an explanatory variable ('*fixed effect*') on the outcome of interest is estimated. Typically, explanatory variables selected here have a systematic and predictable effect on the outcome. However, within the mixed model framework it is possible to account for variables that are likely to affect the outcome, yet in a less predictable and random fashion ('*random effect*'). Although being driven by the same ground principle of linear regression, such an approach allows for

greater flexibility while taking into account all the data available. The lme4 *R* package was used (Bates et al., 2014). Means were computed to summarise ratings from all graders at each region of each eye. For greater robustness of results, best visibility analysis was limited to cases where the proportion of 'none' ratings was below 30%. The first model tested whether retinal region affected boundaries of visible bundles and best visibility, accounting for random effects of eye and grader. This model had the form:

$$y \sim 1 + region + (1|eye) + (1|grader) + \varepsilon \qquad (1)$$

where y signifies the measure of interest (e.g. best visibility depth), 1 signifies the intercept and ε signifies random error. Second, effects of individual eye on RNFB visibility were tested accounting for different retinal regions and graders using a model of the form:

$$y \sim 1 + eye + (1|region) + (1|grader) + \varepsilon \qquad (2)$$

whose symbols are as defined for equation 1. The significance of an effect within linear mixed models is usually tested comparing the improvement in the fitting of available data yielded by the inclusion of the variable of interest in the model. A null model was first created with only a fixed intercept and the same random effects included in the final one. Subsequently, the full model is compared with the null model using a Chi-squared likelihood ratio test and p=0.05 considered as the significance level. In case of a significant effect of the variable tested, the reduced and complete model should differ significantly. In such instance, post-hoc analysis was performed to test pairwise differences, adjusting for multiple comparisons with the Tukey method.

Intraclass correlation coefficient (ICC) according to (Shrout and Fleiss, 1979) was used to explore overall agreement between graders and repeatability within graders, using the *R* package *psyc* (Revelle, 2018). We used ICC classes ICC(2,1) and ICC(3,1) for a single rating to estimate reliability and repeatability, respectively. ICC is obtained combining information from both bias and association between ratings. Thus, a low ICC might originate from either a systematic difference between graders or a lack of correlation. In case of suboptimal ICC (i.e. <0.9), the source of scarce agreement was further

explored with Pearson's correlation. Low ICC and poor correlation would suggest that clinicians did not interpret the trait consistently (i.e. trait not discernible). Conversely, low ICC and high correlation would determine a different but consistent mean of ratings. ICC is strongly determined by the variance of data in the sample studied. As a result, highly consistent ratings on a measure with low variability would lead to a low ICC (Finn, 1970). This shortcoming of ICC would prevent exploration of differences among subgroups of data with inconsistent variances. In fact, clustering data for different tasks as well as for varying retinal regions, might result in subsamples having different variances. Therefore, to further evaluate variability between graders for different tasks and retinal regions we also computed the central 90% range width (difference between 5th and 95th percentiles) of ratings at each region of each image. This was done for ratings of anterior limit, posterior limit and best visibility. This measure is referred to as between-grader variability henceforth. As above, linear mixed models were used to explore the effects of retinal region and individual task on agreement. The model testing the effect of retinal region on between-grader variability included task and individual eye as random effects and had the form:

$$variability \sim 1 + region + (1|task) + (1|eye) + \varepsilon \qquad (3)$$

where 1 signifies the intercept and ε signifies random error. Effects of task were assessed accounting for retinal region and individual eye:

$$variability \sim 1 + task + (1|region) + (1|eye) + \varepsilon \qquad (4)$$

where terms are as per equation 3.

Lastly, as further validation that the object of the subjective grading task was actually the RNFBs visibility, and not a different retinal image feature, the correlation (Spearman's ρ) between thickness of visible RNFBs and mean RNFL thickness in the corresponding region was also explored. The thickness of visible RNFBs was computed as the axial distance (µm) between the subjectively-defined anterior and posterior limits of visible RNFBs, in each region and in each eye. RNFL thickness data as automatically segmented by Spectralis OCT were extracted from the wide-field scans and the mean

thickness was computed along the same grid adopted for the grading task. Lastly, correlation between thickness of visible bundles and between-grader variability was tested.

All data were anonymised before analysis, performed with the statistical software *R,* version 3.5.1 (R Core Team, 2020). As proposed before (Denniss et al., 2014c), the selected combination of number of graders, images and repetition should produce 95% CI within 0.1 for agreement ICC for any value of ICC and for ICC >0.79 for the repeatability analysis.

2.4 Results

Images of 10 eyes from 10 healthy participants were included in the study (5 females, median age 68.5 years, range 57 to 75), whose demographics and clinical details are reported in Table 2.1. Overall, the grading task required a median total of 73 minutes (range 51 to 144).

Table 2.1
Demographics and RNFL parameters for the 10 imaged eyes.

Participant	Eye	Age (y)	Mean cpRNFL thickness (µm)	Mean thickness of visible RNFBs (µm)
1	R	57	92	38
2	R	63	108	48
3	R	66	90	43
4	L	72	97	42
5	R	69	111	50
6	R	73	98	43
7	L	67	99	48
8	R	68	96	42
9	L	75	78	38
10	R	73	103	46

The grand mean and (±) standard deviation (SD) for anterior limit of visibility was 9.9±0.8µm (range 8.8±1.5µm for superior nasal region to 11.3±1.5µm for ONH region) and the grand mean for posterior limit of visibility was 53.7±30.6µm (range 28.1±1.2µm for raphe region to 131.6±13.7µm for ONH region).

As shown in Figure 2.5, the anterior limit of RNFB visibility was affected significantly by retinal region ($\chi^2_{(9)}$ = 58.6, p<0.0001), and the greatest difference was found between the optic disc region and the superior nasal region (2.5µm, p<0.0001). Nonetheless, this difference was below the instrument's digital axial resolution, therefore not clinically significant. Similarly, there were significant differences between individual eyes in anterior limit of RNFB visibility ($\chi^2_{(9)}$ = 60.5, p<0.0001). However, these differences also did not exceed digital axial resolution of the instrument so are not clinically significant (greatest pairwise difference: 2.5µm, p<0.0001). On the contrary, differences in posterior limit of RNFB visibility across both retinal regions ($\chi^2_{(9)}$ = 1671.1, p<0.0001) and individual eyes ($\chi^2_{(9)}$ = 88.7, p<0.0001) were both statistically and clinically significant. The mean difference between retinal regions was 31.1µm (range 0.3 to 103.5 µm), whereas the mean difference between individual eyes was 4.6µm (range 0.2 to 12µm). Mean limits of RNFB visibility for individual regions are shown in Figure 2.5.
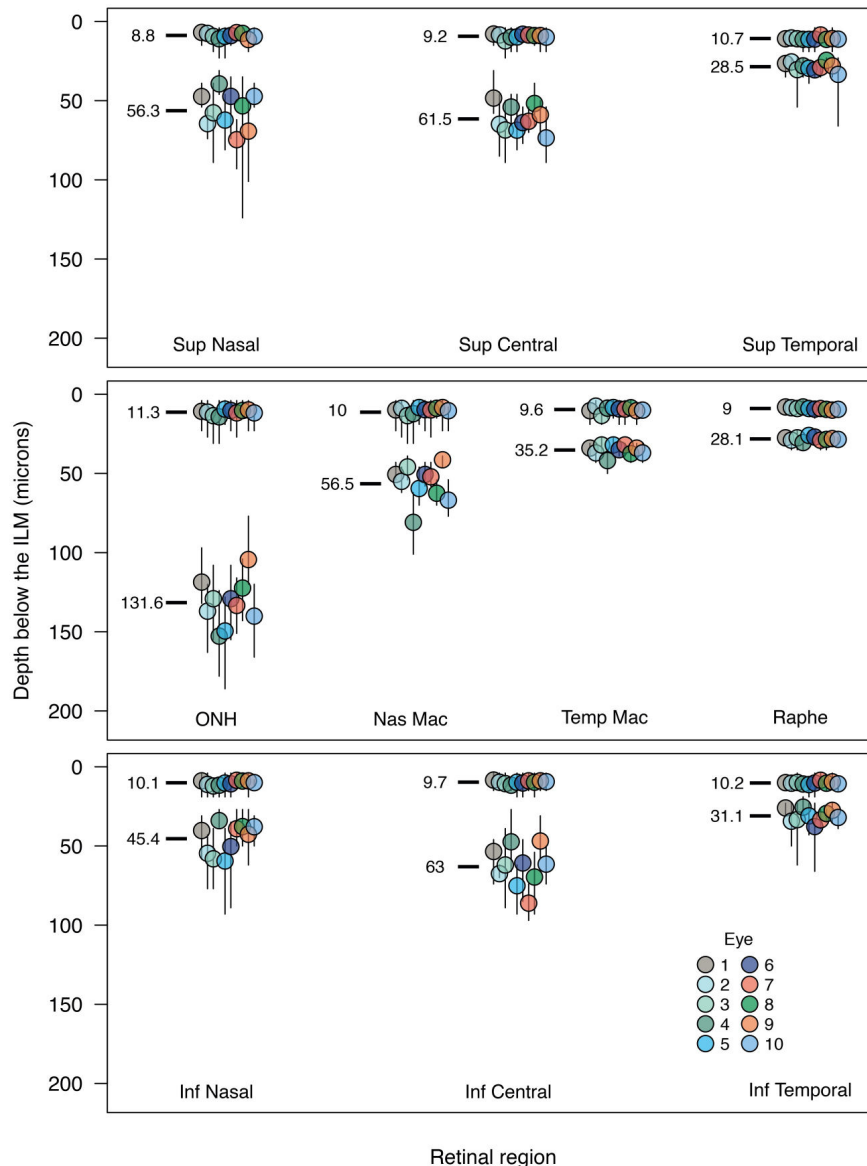
Figure 2.5 Anterior and posterior limits of RNFB visibility (µm) at different retinal regions. The top, middle and lower panels correspond to the superior, central and inferior retina respectively. Points and error bars show respectively the mean and range of all ratings for each of the 10 individual eyes. Black lines and numbers on the left of each cluster of points show the group mean limit of visibility across all eyes. Points are colour-coded according to individual eyes numbered as in Table 2.1. Sup = superior, Inf = inferior, Nas = nasal, Temp = temporal, Mac = macula.

As shown in Figure 2.6, optimum depth for best RNFB visibility was affected by retinal region ($\chi^2_{(8)}$ = 58.8, p<0.0001). The rate of ungradable judgements for each retinal region is also shown in Figure 2.6 and was within 30% for most retinal regions. In the ONH region, 40% of the time graders could not identify a single depth with best visibility of bundles so for greater robustness, this region was excluded from analysis. The grand mean for best RNFB visibility was 20.3±1.9µm (range 17.4±1.3µm for the raphe region to 22.8±2.0µm for the nasal macula), whereas mean limits for individual regions are shown in

Figure 2.6. Pairwise analysis showed most of the significant differences to be just above the instrument's digital axial resolution (greatest difference between nasal macula and raphe regions, 5.4µm, p<0.0001). In contrast, differences in optimum depth for best RNFB visibility between eyes were not significant ($\chi^2_{(9)}$ = 10.7, p=0.29).
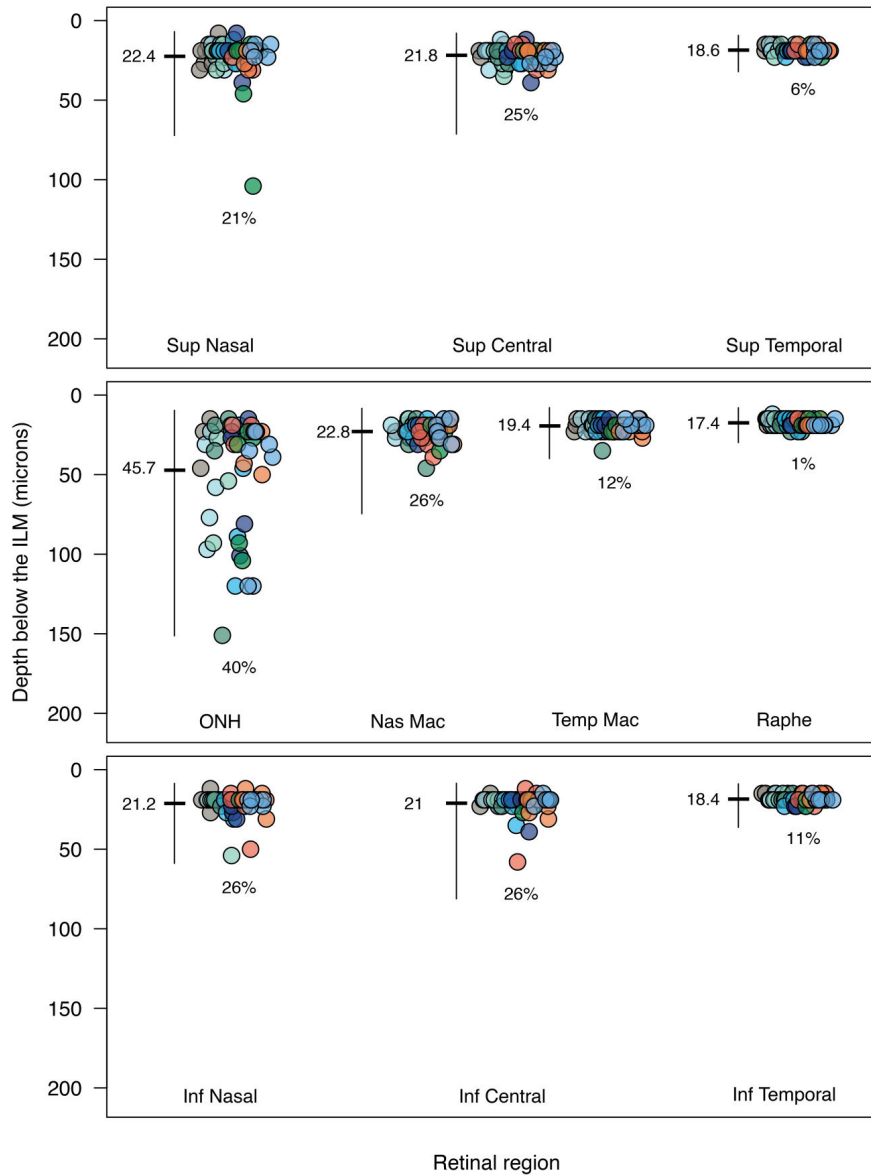
Figure 2.6 Optimum depth for best RNFB visibility in different retinal regions. Individual points show individual graders' ratings of individual eyes with eyes colour-coded as in Figure 2.5. Horizontal black bars and numbers show mean ratings across all graders and eyes. Vertical bars indicate depths where RNFBs were visible (5th and 95th percentiles of overall anterior and posterior limit of visibility respectively). Percentage of ungradable ratings (no single best depth identified) is shown for each location. Panel arrangement and abbreviations are as in Figure 2.5.

Overall, we observed good agreement between graders' ratings of RNFB visibility (ICC(2,1) = 0.89, 95%CI: 0.87 to 0.91). Similarly, graders' estimates

were highly repeatable (grader 1, ICC(3,1) = 0.96, 95%CI: 0.95 to 0.97; grader 2, ICC(3,1) = 0.99, 95%CI: 0.98 to 0.99). Analysis of between-grader variability, broken down by retinal region and specific grading task is shown in Figure 2.7. Retinal region had a significant effect on between-grader variability ($\chi^2_{(9)}$ = 158.3, p<0.0001), with the greatest differences found between the ONH region and other retinal regions (max difference: ONH/raphe = 35.8µm, p<0.0001). Moving towards the nasal regions of the retina, gradings appeared progressively more variable in comparison to temporal regions (superior nasal-superior temporal: 11.74µm, p=0.003; ONH-raphe: 35.76µm, p<0.0001; and inferior nasal-inferior temporal: 7.04µm, p=0.33). Similarly, the specific grading task significantly affected between-grader variability ($\chi^2_{(2)}$ = 38.5, p<0.0001), with the identification of the posterior limit of RNFB visibility being the most variable (mean between-grader variability 20.7µm) compared to anterior limit (10.3µm) and best visibility (14.9µm). Differences in between-grader variability among tasks were all statistically significant once adjusted for multiple comparisons (anterior and best: -4.63µm, p=0.012; anterior and posterior: -10.44µm, p<0.0001; best and posterior: -5.81µm, p=0.001).
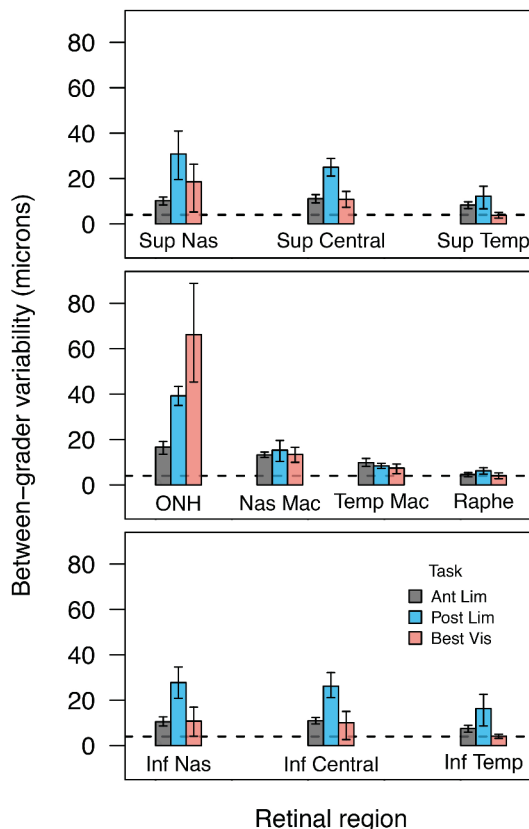
Figure 2.7 Between-grader variability in ratings for each task and retinal region. Between-grader variability was computed as the central 90% range width (difference between 5th and 95th percentiles) of ratings at each region of each image. Panel arrangement and abbreviations as in Figure 2.5 and 2.6. Error bars represent bootstrap 95% CI. Horizontal dashed lines indicate the instrument's digital axial resolution.

As shown in Figure 2.8, thickness of visible RNFBs (axial distance between anterior and posterior limits of visibility) and RNFL thickness as measured by the Spectralis OCT were strongly correlated (Spearman's ρ: 0.9, 95%CI: 0.85 to 0.93, p<0.0001). For all tasks, between-grader variability showed strong correlation with thickness of visible RNFBs (Spearman's ρ: 0.66, 95%CI: 0.58 to 0.72, p<0.0001). In other words, greater consistency (smaller variability) among clinicians was observed in regions with RNFBs visually present at a narrower range of depths.
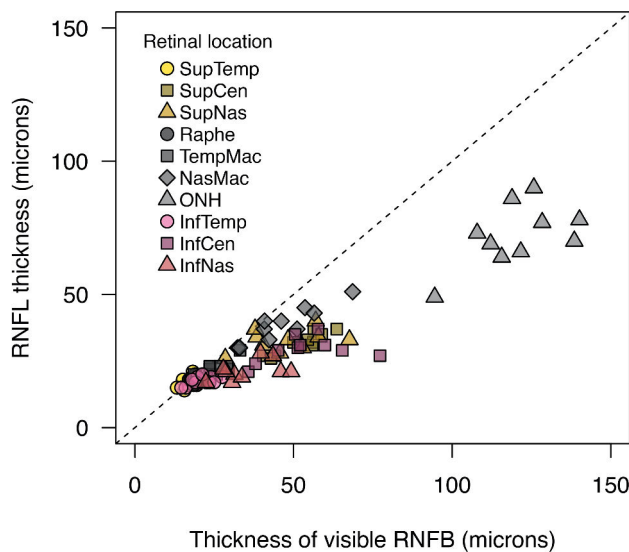
Figure 2.8 Thickness of visible retinal nerve fibre bundles (RNFB) vs mean RNFL thickness in the corresponding retinal sector. Points are coded according to the retinal location as shown in the key. The dashed line represents a 1:1 relationship.

2.5 Discussion

Detecting abnormalities of RNFB reflectivity in OCT en face images is a promising approach to diagnose and monitor glaucoma, complementary to conventional morphological examination (Hood et al., 2015, Gardiner et al., 2016, Ashimatey et al., 2018a, Sakamoto et al., 2019, Vermeer et al., 2012). Yet, more robust employment of this technique in clinics is limited by a coarse understanding of the normal appearance of RNFBs at different depths. Additionally, methods to objectively define defects within this domain are also lacking. To accomplish both reliable imaging and defect definition, ideally achieved by objective and quantitative methods, adequate understanding of RNFB configuration in healthy eyes is required. These include the depths at which RNFBs are visible across the retina and the variability between different eyes. Such information was not previously available and therefore was the objective of this investigation.

Findings of this analysis suggest that the first depth below the ILM at which bundles become visible is consistent throughout the retina and in different eyes. On the contrary, the posterior limit of visibility changed significantly across retinal regions and in different eyes. The effects were statistically and clinically significant, requiring consideration when extracting en face slab images. It should be noted that pairwise analysis suggested that not all retinal regions examined here had a different posterior limit of visibility, identifying some areas with a similar range of depths of visible bundles (Figure 2.9). The depth of best visibility of RNFBs varied with retinal region but was similar across different eyes. Proportion of ungradable ratings changed across the retina, exceeding the 30% limit in the ONH region only, where optimal visibility was also found at a wider range of depths. As reported in Figure 2.9b, best depth in the remaining retinal regions was often found at around 20μm below the ILM (mean 20.3μm, range 17.4μm to 22.8μm). Pairwise differences were statistically significant for a few regions (Figure 2.9b), however, these were close to the OCT instrument's digital axial resolution (greatest difference: 5.4μm, $p < 0.0001$), indicating that depth differences were 1-2 pixels at most. Estimate of optimum depth for best bundle visibility provides a single depth at which visualisation of RNFBs appeared optimal across the retinae of studied eyes. The associated enhanced value in detecting defects of this specific depth below the ILM could be further explored in future studies of glaucoma.
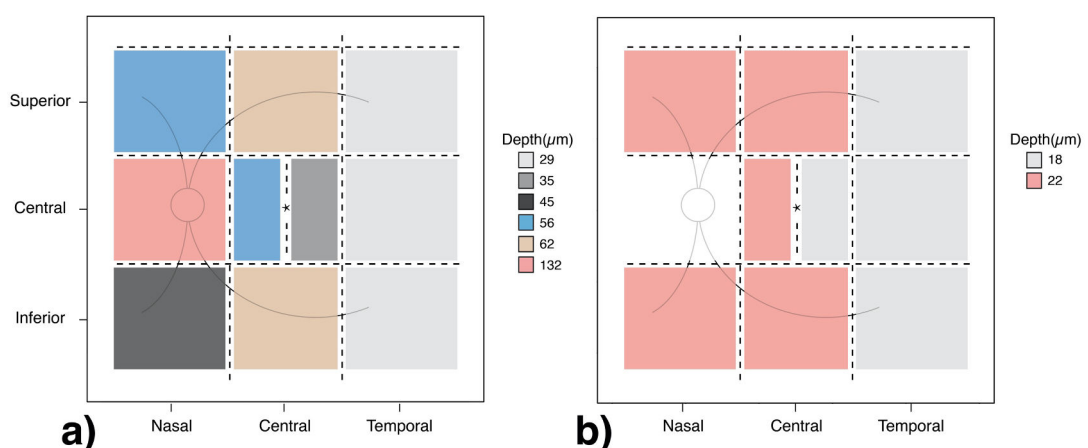
Figure 2.9 Schematic representations of pairwise differences between retinal regions for: (a) posterior limit of visibility; and (b) the optimum depth of visibility. Retinal regions are coloured differently in cases of a statistically significant difference exceeding the instrument digital axial resolution (4μm). This analysis identified 6 and 2 macro-regions for the posterior limit and best visibility respectively. The anterior limit of visibility was omitted since differences between retinal regions were all below 4μm.

Overall, agreement analysis showed good repeatability and conformity between graders for identification of RNFB visibility. Results were similar to findings from a recent study where graders consistently reported the presence or absence of bundles in en face images of the macula (Iikawa et al., 2020). A more detailed analysis of agreement with ICC was prevented by its susceptibility to the variance of the sample tested (Finn, 1970), which changed significantly across tasks and retinal regions (Figures 2.5 and 2.6). As such, between-grader variability was further evaluated by computing the 90% range width of gradings, as a surrogate measures of concordance. Greater variability was observed for best visibility and posterior limit of visibility compared to anterior limit of visibility. Further, between-grader variability changed with retinal region, becoming progressively greater in nasal regions, where the RNFL thickness is known to increase (Hondur et al., 2018, Varma et al., 1996). It is possible that the presence of visible bundles at an increased range of depths contributed to the increase in variability between-graders. Indeed, a wider range of depths with present RNFBs would result in a greater number of possible answers, contributing to greater variability between graders. This hypothesis was also supported by the strong correlation (ρ=0.66) between between-grader variability and the thickness of visible bundles in the corresponding region. Nonetheless, even though greater variability between clinicians was observed in certain retinal regions, the overall high to excellent ICCs support the validity of the visible ranges identified.

No previous studies have characterised RNFB visibility of the central ±25° of the retinae of healthy adults. Additionally, this study was the first to consider single slabs of the maximum axial resolution achievable with the instrument as the unit of assessment of bundle reflectivity. Most of the work available has usually collapsed a retinal volume of a fixed thickness into a single 2D slab image to examine reflectivity loss in eyes with glaucoma (Sakamoto et al., 2019, Ashimatey et al., 2018a, Hood et al., 2015). Among the earliest work in en face OCT imaging, Hood et al. assessed reflectance loss in a small group of glaucoma eyes by extracting fixed thickness slabs from a 9x12mm scan centred on the ONH (Hood et al., 2015). En face images were obtained by averaging intensity values from the ILM up to 52µm below it (roughly

corresponding to the average of 14 slabs in our settings). As discussed by the authors, slab axial depth represented a compromise between a narrow thickness, able to detect local changes of reflectivity, and a large enough number of pixels to enhance signal to noise ratio. The authors also acknowledged that in certain regions with glaucomatous damage, the RNFL might have been as thin as 20-30µm, resulting in the inclusion of hypo-reflective pixels located in the GCL and/or the IPL. One additional limitation that could have affected this approach is an incomplete capture of the full depth of RNFBs in some regions. In a larger case-control study (Ashimatey et al., 2018a), central retinae of 31 healthy participants and 33 eyes with glaucoma were imaged to evaluate reflectance defects and explore the relationship with cpRNFL thickness loss. The authors considered several methods of slab extraction, including 52µm fixed thickness approach as proposed previously (Hood et al., 2015) and a slab of varying thickness aimed to minimise artefacts from hyper-reflective glial alterations found immediately below the ILM. The latter slab extraction method averaged pixels from 24 to 52µm below the ILM in the ONH area, from 24 to 36µm in the region between ONH and fovea and reduced the sampled depths to 16-24µm in the temporal macula-raphe area. Defects were then defined as retinal areas where reflectivity was 2.34 SD smaller than data from healthy controls. Although good agreement between en face and cpRNFL thickness changes was found in a large proportion of glaucoma eyes, all RNFB lesions were only identified once inspection was extended beyond 52µm below the ILM and generalised loss was considered. As also reported by the authors, limiting the extraction to 52µm below the ILM and not accounting for either retinal location nor individual changes should be considered as potential limitations of the strategy employed to identify glaucomatous defects (Ashimatey et al., 2018a). Another study explored the structure-function relationship in the macula of eyes with advanced glaucoma, using en face visualisation of preserved RNFBs (Sakamoto et al., 2019). In this work, 6x6mm cube scans centred on the macula were collected and 50µm fixed-thickness slabs from the ILM were extracted averaging the A-scan intensity at each location. According to the authors, the 50µm axial depth of slab images was selected upon evaluation of a variety of fixed thicknesses in both healthy and glaucoma participants (10 to

250µm; see Figure 2.10, reproduced from Sakamoto et al.). In healthy maculae, optical intensity of en face slabs steadily increased with thickness up to 50µm where a plateau was reached until 100µm below the ILM. Subsequently, intensity decreased with greater thicknesses up until about 250µm. This depth approached the retinal pigment epithelium, which is reported as the most hyper-reflective retinal layer (250µm panels in Figure 2.10). Sakamoto et al. argued that the adopted axial depth would include approximately 20µm from the GCL beneath a thinned RNFL but that this would be unlikely to alter reflectance results due to the hypo-reflectance of the layer.
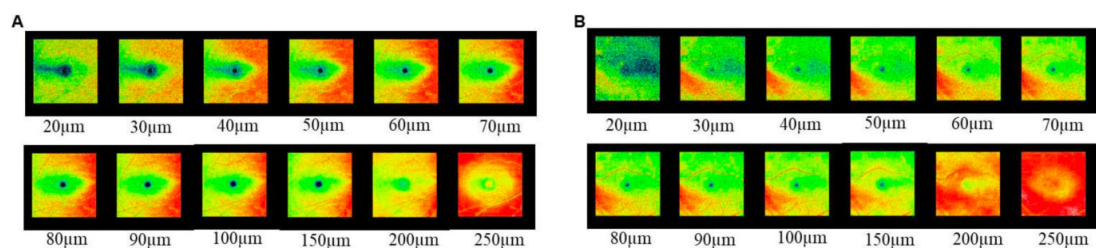
Figure 2.10 Figure reproduced from Sakamoto et al., showing reflectivity of macular slabs in a variety of slab thicknesses in a healthy macula (a) and the macula of an eye with advanced glaucoma (b). After slab extraction, reflectance was colour-coded using warm and cold colours for high and low reflectivity values, respectively. Axial depth from the ILM of each slab is reported in µm (Sakamoto et al., 2019).

Although limited to the macula, data from the work of Sakamoto et al. (2019) appears in agreement with results of our study, where bundles in the nasal and temporal macula were visible up to 56.5µm and 35.2µm, respectively. As shown in Figure 2.10a, when those depths are surpassed, reflectivity starts to decrease progressively as represented by cooler colours. This could be expected to follow the averaging process with hypo-reflective pixels found posteriorly to the RNFL. Notably, the example reported by Sakamoto et al. is also consistent with the nasal/temporal asymmetry of depth of visually present RNFBs that we found at the macula (see Figures 2.5 and 2.9). In addition, the extremely nasal portion of the slabs in Figure 2.10a (right hand-side, approaching the ONH) preserves highly warm tonalities at greater depths as suggested by visible RNFBs in the ONH region until >130µm below the ILM in our sample.

The use of a fixed depth slab to examine regions of the retina presenting different RNFL thickness may be unable to capture all glaucomatous defects. Even though most studies using en face OCT imaging in glaucoma

acknowledged this as a limitation, neither differences between retinal regions nor differences between different eyes were fully accounted for in these analyses (Ashimatey et al., 2018a, Hood et al., 2015, Sakamoto et al., 2019, Iikawa et al., 2020). A conventional choice in the available studies has been the use of a fixed 50µm thick slab, however, according to our results, a similar approach may not be inclusive of all RNFBs in all retinal regions. In fact, in the nasal macula, the superior/inferior central retina and the ONH region, information regarding the status of bundles might have been overlooked by those approaches. With this in mind, it is unsurprising that the work from Ashimatey et al. was only able to fully identify en face defects when extending the analysis to greater depths, below the original limit of 52µm (Ashimatey et al., 2018a). It is also worth considering that such methods would likely exceed the RNFL in its thinner regions, resulting in the inclusion of deeper, hypo-reflective retinal layers. Therefore, lower intensities of RNFBs could be observed not only in case of a primary loss of reflectivity but also from the inclusion of deeper layers in the slab. Overall, the ability to identify en face defects may vary according to slab thickness and composition. Since RNFBs are visible at different depths in different areas of the retina, slab thickness should also vary across the retina to include a consistent proportion of the RNFL.

It has been suggested that glaucomatous changes of reflectivity might not necessarily be met by a reduction in thickness (Hood et al., 2015, Gardiner et al., 2016). As a consequence, reflectivity could provide an additional body of information on RNFL status in glaucoma (Vermeer et al., 2012, Gardiner et al., 2016). Several sources of disagreement between these two approaches should be considered, including the method used to generate the data. Thickness analysis relies fully on segmentation which may sometimes be inaccurate, especially for the proximal RNFL boundary (Hood et al., 2015). Conversely, en face imaging depends only on the vitreous-ILM segmentation, described as one of the easiest surfaces to be automatically detected (Yang et al., 2010). Further, blood vessels, while being a significant source of artefacts in thickness analysis, might be easily differentiated from preserved bundles in glaucoma (Hood et al., 2015). Lastly, and perhaps more

importantly, current en face images are determined by both thickness and intensity properties of the RNFL, which might show evidence of glaucoma lesions on a different timescale (Hood et al., 2015, Liu et al., 2014, Huang et al., 2011). Despite possible disagreement in glaucoma, RNFL thickness and the range of depths across which bundles are visible were highly correlated in the healthy eyes assessed in our study (Figure 2.8). Although highly correlated, the slope of the points in the scatterplot is below the 1:1 line, suggesting that the range of depths across which bundles are visible overestimates RNFL thickness. It is likely for this overestimate to be an artefact of the instructions provided to graders. Bundles were considered to be visible when occupying more than 25% of the region and this could have resulted in discrepancies with RNFL thickness, calculated as the mean across the region – more likely to represent the depth with visible bundles in 50% of the region.

Considering the strong correlation between the range of depths with visible bundles and RNFL thickness, a further term of comparison for our results could be gained from studies assessing RNFL thickness in wide-field retina scans. Although technological advancements have enabled such evaluations, there is a paucity of reports available which have extended RNFL thickness analysis over the macula and the ONH (Hondur et al., 2018, Kim et al., 2021, Bogunovic et al., 2014). In the commercially available '*Hood Report*' from the 3D-OCT (Topcon Corp, Tokyo, Japan) RNFL thickness over a 12x9mm volume scan centred on the ONH is compared to normative data from 57 healthy eyes (Hood et al., 2016). Yet, the database was not made publically available, and the area analysed does not fully cover the area tested in our study. A more recent work from Kim and colleagues adopted the same OCT device and scan settings to establish normative RNFL thickness data from 220 healthy eyes (Kim et al., 2021). The sample studied had a slightly younger age compared to the healthy eyes included in our study (mean age 56.2 years, SD 16.6 years), however, qualitative observation of their data appear in good concordance with our findings. The authors reported an average thickness of 25-40µm in the temporal macula which increased moving toward the ONH with values of 50-60µm in the arcuate regions and ~100µm in the temporal superior and temporal inferior surroundings of the ONH. In close proximity of the ONH,

peaks up to 150μm were observed (Figure 1 from Kim et al., 2021). In other studies including wide-field OCT imaging, specific questions such as RNFL thickness modifications over time and structure-function relationship limited study design and presentation with no normative data reported (Hondur et al., 2018, Bogunovic et al., 2014, Guo et al., 2017). Although findings from these works cannot be directly compared with our study, there seems to be an overall concordance between the RNFL thickness pattern presented and the changes of RNFB visibility across the retina observed in our sample.

Several limitations of this study should be considered. Firstly, only a small number of healthy eyes, of a single ethnic group (9/10 participants were Caucasian), were evaluated. Although this provided adequate power to assess agreement among graders, it is possible that some more unusual RNFL configurations or thicknesses may not be captured in our sample. Similarly, the opportunistic selection of a small number of eyes may limit generalisability to different populations. Nonetheless, the included age group was selected to be similar to that of open angle glaucoma (Rudnicka et al., 2006), limiting the impact of healthy ageing on the applicability of our findings.

In this study we computed depth-resolved attenuation coefficients (Vermeer et al., 2013) which, unlike raw reflectivity values, are supposed to represent an intrinsic optical property of the tissue of interest (Chang and Bowden, 2019, Vermeer et al., 2013, Vermeer et al., 2012), and reduce the impact of artefacts on en face images (Ashimatey et al., 2018a). Earlier work considered raw OCT intensities to explore reflectivity defects of the RNFL (Vermeer et al., 2011), yet, factors such as incident light strength, media quality and age were reported to affect measurements (Vermeer et al., 2012, Chen et al., 2013, Tappeiner et al., 2008, Chen et al., 2016). To account for the variable amount of incident light, normalisation with a reference layer had been used (Vermeer et al., 2012, Mehta et al., 2018, Gardiner et al., 2016). Alternatively, attenuation coefficients can be computed to determine optical properties independent from the amount of light received, and from additional segmentation requirements (Vermeer et al., 2013, Thepass et al., 2017, Chang and Bowden, 2019). Whilst this method is expected to reduce artefacts, one study suggested that age and scan quality may still have a significant

effect on attenuation coefficients (Thepass et al., 2017). Several reasons could justify a relationship with age, including an overall poorer optical quality as well as a physiological reduction in bundle density. Yet, relation with scan quality was not expected (Thepass et al., 2017), and the preserved dependency needs to be corroborated in further studies. Notwithstanding, usage of attenuation coefficients, instead of raw intensities, is likely to represent the best available method to mitigate the effects of confounders of this technique. It should also be considered that these effects were likely to be minimised in this study by the qualitative nature of the grading task, the inclusion of participants only within the age range of interest and the inclusion only of scans well above the manufacturer's recommended minimum quality.

Further, the image processing methods applied in our work had some drawbacks. For instance, as proposed elsewhere (Ashimatey et al., 2018a) we accounted for the background intensity of individual images by extracting the value from the raphe region. However, this relies on the assumption of constant background throughout the whole retina which may not necessarily hold for all participants. In addition, software to perform image registration and montaging of the scans was not available at the time of analysis. Consequently, overlapping areas of adjacent images were concealed with grey patches for display purposes (Figure 2.4) and clinicians were instructed to ignore those regions while assessing RNFB visibility. Although minimal duplication of information might have occurred in certain retinal locations, the highly consistent and repeatable data suggest that the graders were able to overcome the drawbacks to produce reliable measures. As such we do not believe the results of this study would be significantly altered by improved montaging of the individual images.

Lastly, the typical hyper-reflective appearance of RNFBs captured in en face imaging is determined by the cylindrical conformation of bundles (Huang et al., 2006). The effect of different directions of incident light beam might be substantial on the measurements and should be considered in this analysis (Huang et al., 2016). Although directional reflectivity might be of particular importance, especially in context where wide-field imaging is used, no means of compensation were available.

2.6 Conclusions

The range of depths with visible bundles varied markedly throughout the retina of the healthy participants in our study, consistent with the RNFL thickness. The optimal depth for visualisation of RNFBs was found at around 20µm below the ILM in most retinal locations with the exception of the ONH region, where visibility was close to optimal at a wider range of depths. As the range of visible RNFBs changes significantly, to extract all RNFB information consistently across the retina, slab properties should account for changes in retinal location and differences between individual eyes.

3. Simple subjective evaluation of en face OCT reflectance images can be used to reliably identify RNFL loss in glaucoma

3.1 Abstract

Background: This study aimed to evaluate a simple approach for subjective identification of glaucomatous RNFB loss in en face images, and to assess its diagnostic performance. We also tested the hypothesis that if reflectivity changes precede thickness changes in glaucoma there should be reduced correlation between the modalities in glaucoma compared to controls.

Methods: 20 glaucoma participants and 20 age matched controls were imaged with high-resolution OCT and conventional optic disc imaging. 3.87µm thick en face slabs containing attenuation coefficients were extracted from the ILM to 194µm below it. Depths and corresponding angles of *first gap* in visible RNFBs and *last visible bundle* were subjectively evaluated in 6 sectors of a 3.5mm circle around the optic disc by one clinician. The RNFL thickness along the same circle was extracted at the corresponding angles of en face indices. Effect of glaucoma on visible RNFBs was explored with linear mixed models. Diagnostic capability was evaluated with partial receiver operating characteristic area (pAUC). Repeated measures correlation was evaluated between en face parameters and corresponding RNFL thickness.

Results: Glaucoma affected depth of visible presence of RNFBs, with *first gap* and *last visible bundle* being closer to the ILM in glaucoma eyes (mean difference: 39.1µm 95%CI: 33.0 to 45.3 and 48.1µm 95%CI: 38.8 to 57.4 respectively, both p<0.0001). Many en face indices showed excellent diagnostic performance (pAUCs 0.63 to 1.00), yet were similar to conventional RNFL thickness parameters (pAUCs 0.63 to 0.95). Depths of visible RNFBs were strongly correlated with RNFL thickness counterparts, with the strength of correlation being similar or only slightly poorer in glaucoma eyes (r = 0.73 to 0.80) compared to control ones (r = 0.81 to 0.92).

Conclusions: This simple method reliably identified glaucomatous defects in en face images with diagnostic performance equivalent to existing thickness

indices. In this sample, strong evidence of reflectivity loss without corresponding thickness loss was not found. Development of more sensitive automated analyses and integration with perimetry may realise further potential of en face OCT images in glaucoma.

3.2 Introduction

As detailed in previous chapters, the early diagnosis of glaucoma is critical to minimise visual impairment (Prum et al., 2016, National Institute of Health and Care Excellence, 2017). Concurrently, the burden linked to lifelong treatment initiated in glaucoma patients demands accurate diagnosis with minimisation of false positives (Mwanza et al., 2018a, Fallon et al., 2017). Although OCT has brought significant support in the assessment of glaucoma (Tatham et al., 2015, World Glaucoma Association, 2016, Jindal et al., 2019), the early detection through single cross-sectional exams is still imperfect (Founti et al., 2018, Hohn et al., 2018, Karvonen et al., 2018). Analysis of reflectivity might provide an additional source of information on ganglion cells' axonal status in glaucoma beside conventional RNFL thickness measures (Fortune, 2015, Vermeer et al., 2012). To this end, en face images allow direct observation of RNFBs, which according to animal models mentioned in previous sections (Huang et al., 2011, Fortune et al., 2013), might be lost early in glaucoma. Observation of RNFB reflectance properties might be exploited subjectively by clinicians (Cheloni and Denniss, 2021, Iikawa et al., 2020) and could provide additional, potentially earlier, markers of glaucomatous changes (Hood et al., 2015).

Limitations of early glaucoma detection by OCT may in part arise from OCT data analysis. It has been suggested that the current main focus on summary indices and red/green classification might restrict accurate diagnosis by failing to make full use of information available (Sayed et al., 2017, Chong and Lee, 2012, Hood and De Moraes, 2018a). Accordingly, clinicians should also look in greater detail at B-scans for evidence of glaucoma damage missed by OCT thickness analysis (Hood and De Moraes, 2018b, Hood and De Moraes, 2018a). En face images might not only provide a method to detect earlier signs of the disease, but even facilitate the observation of glaucomatous lesions missed by the conventional thickness approach. Hood et al. suggested that

defects observable in en face images might be present even in ordinary B-scans, but would most likely be neglected by thickness analysis because of limitations of segmentation algorithms and/or artefacts (Hood et al., 2015).

Though analysis of en face images is promising, its usability in clinics remains poor, because of the lack of established methods available to clinicians to use it. No accepted objective criteria to define defects in this domain are available, and subjective analysis, though proposed at times (Hood et al., 2015, Sakamoto et al., 2019), has not been validated and nor has its actual diagnostic value been explored. The purpose of this study was, therefore, to evaluate a simple subjective approach for subjective identification of RNFB reflectance loss in glaucoma (Figure 3.1), and to quantify the associated diagnostic capability. Further, since the hypothesis that reflectance loss occurs before thickness changes has been minimally investigated in humans, we aimed to test for discordance between RNFB reflectivity and RNFL thickness changes that may indicate a temporal decoupling between these parameters that could be exploited for glaucoma diagnosis.
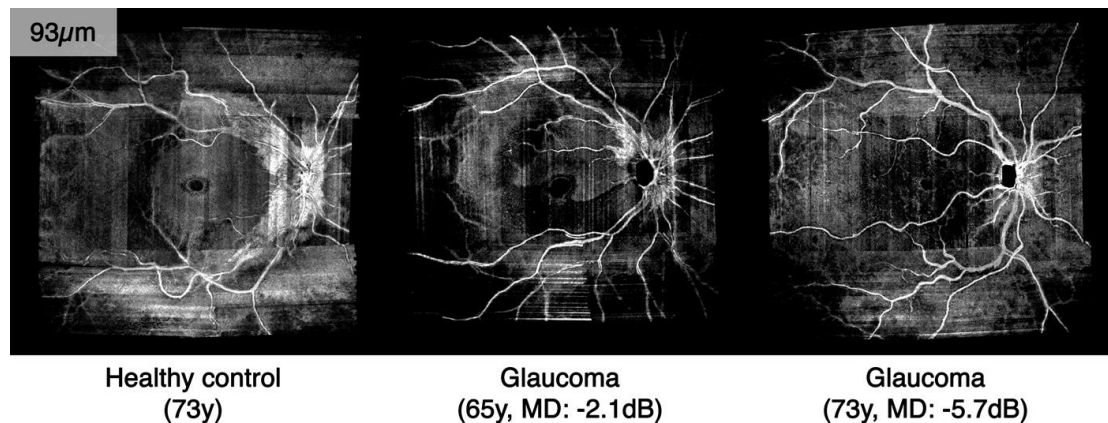
Figure 3.1 Example of how visible presence of RNFBs at 93μm below the ILM differs between a healthy eye different stages of glaucoma. The images are single pixel deep en face OCT images without depth-averaging. At this depth, RNFBs are still visible all around the optic disc in the healthy eye, whereas RNFBs have already disappeared in the rest of the retina where the slab encompasses deeper, hypo-reflective retinal layers. In the early glaucoma eye (central panel) a substantial loss of RNFBs can be seen in the temporal and temporal inferior sectors, with no visible presence of RNFBs. In the more advanced glaucoma eye (right panel) no bundles are visible around the optic disc or elsewhere, with the only hyper-reflective elements provided by blood vessels. An animated version of this figure, showing a range of depths below the ILM, is provided in Appendix A (Supplementary File 3.1). MD = mean deviation.

3.3 Methods

Settings and participants

Participants were recruited as part of an ongoing case-control study exploring structure-function relationships in glaucoma at the School of Optometry and Vision Science, University of Bradford. Informed consent was collected from every participant and the study achieved ethical approval from the National Health Service's Research Ethics Service.

Details of inclusion and exclusion criteria have been presented in detail earlier (section 2.3). Briefly, we included glaucoma participants if 40 years and older and with a confirmed medical diagnosis of open angle glaucoma. To be eligible for this specific analysis, participants required evidence of both a structural and a functional defect. The former was defined as at least one abnormal ONH sector ($p<1\%$) from the 3.5mm diameter cpRNFL thickness Spectralis classification. Participants were tested with Humphrey VF (SITA-Standard 24-2), and a VF defect was defined by the presence of three contiguous non-edge points with $p<5\%$ on the pattern deviation plot. Additional inclusion criteria were a refractive error within ±6.00DS and ±3.00DC, clear optical media or uncomplicated cataract surgery and not presenting other disease except glaucoma that could affect vision. Age-similar healthy controls were included if presenting with no eye conditions, including ocular hypertension or different IOP between eyes (>4mmHg). Healthy participants were eligible if presenting a normal VF (normal MD, with $p>5\%$, GHT within normal limits and no VF defect) and if they had visual acuity ≤0.20 logMAR (6/9.5 Snellen) measured with appropriate refractive correction. One eye per participant was included. If both eyes were eligible, the included eye was selected at random in controls, whereas the eye with milder defect (as identified by a less negative MD) was included in glaucoma participants.

OCT imaging

The OCT imaging procedure was described in detail previously (section 2.3). This consisted of multiple high-density, high-speed OCT scans (with 30µm separation between B-scans), collected in every participant over the central ±25° of the retina. All images were acquired with signal to noise ratio above

20dB as per the manufacturer's instructions. Single slab images of the maximum digital axial resolution (3.87µm), containing depth-resolved attenuation coefficients, according to Vermeer et al. (2013), were extracted from 0 to 193.5µm below the ILM. Besides OCT en face imaging, conventional ONH analysis was also performed, by assessing the cpRNFL thickness at the 3.5mm diameter circle around the ONH as segmented by the device's inbuilt software.

Attenuation coefficients were imported into MATLAB (Version 9.6.0, The MathWorks Inc., Natick, Massachusetts) for montaging and image-processing. Different to the previous chapter (section 2.3), where single images were collated together in a pseudo-montage, custom software was used to compute a semi-automated montage of each single image. The code adopted MATLAB's built-in control points selection tool (*cpSelect* function), which employs a combination of translation, rotation and scaling to generate a geometric transformation to register individual images to a reference image according to a set of selected corresponding points. The macular scan was selected as the reference image. Coordinates of corresponding points of each individual image with the macular image were identified from Spectralis reference SLO images. In turn, a minimum of 9 points representing the same retinal detail (e.g. a blood vessel intersection) were manually selected from the overlapping region between the macula and each other image. The identified points fed the control point selection tool resulting in geometric projective transformations that were subsequently applied to align and montage OCT data at each single depth. The highest pixel value was selected from overlapping regions. The task was repeated for each different retinal scan, and served to extract a specific transformation to align and montage OCT data at every single depth.

For each participant, obtained montages underwent image-processing to optimise visualisation, the details of which have been presented earlier (section 2.3). Briefly, the intensity of an area within the raphe region, 35µm below the ILM, with no RNFBs was set as background by subtracting this lower limit from all pixels and clipping negative values to zero. Then, the average 99th percentile from all depths was used to normalise the attenuation

coefficients arrays in 0-1 range images. Figure 3.1 above shows examples of the final montage.

Data extraction

Consecutive images pertaining to the same eye were arranged in a presentation file allowing obseration of consecutive slabs with perfect spatial alignment. Visible presence of RNFBs was evaluated around the ONH in 6 wedge-shaped sectors corresponding to those of the Spectralis 3.5mm diameter circle scan. The sectors adopted (Garway-Heath et al., 2000b), were the temporal (90º), nasal (110º), two superior and two inferior sectors, split into temporal and nasal (40º per each of 4 sectors). A sector grid (Figure 3.2) with fixed and standardised dimension was overlaid onto en face images of all participants, at each depth. Aiming to adopt the same circle size of cpRNFL thickness analysis, the corresponding SLO image of Spectralis analysis from one participant was used to drive the construction of the grid, by overlapping correspondent retinal structures. The resulting grid was subsequently verified on a second participant. Eventually, grid dimensions were preserved unchanged and applied to all assessed eyes. Conversely, the sector grid was tilted in order to follow the individual fovea-disc angle as subjectively identified in the en face image. Accordingly, the temporal sector of each individual eye was centred on the fovea-disc axis, mimicking the arrangement proposed in the Spectralis cpRNFL analysis and, therefore, facilitating comparison.

Visible presence of RNFBs was recorded subjectively by the author in standardised lighting conditions, observing images on a laptop computer (MacBook Pro 13" 2017 version, Apple Inc., Cupertino, California). Both the depth and corresponding angle (with 0° at the fovea-disc axis and angles increasing clockwise for right eyes and anti-clockwise for left eyes) of two en face indices were extracted at each ONH sector (Figure 3.2):

> i) *First gap in visible bundle* (Figure 3.2 a, c): For this parameter, subsequently referred to as *first gap*, the grader recorded the first (most anterior) depth at which a gap between RNFBs can be seen crossing the 3.5mm circle in the sector of interest (i.e. the first depth of a single perceivable missing bundle in the sector); and

ii) *Last visible bundle* (Figure 3.2 b, d): representing the most posterior depth at which one or more visible bundles crosses the 3.5mm circle in the sector of interest (i.e. the last depth containing at least one single bundle).
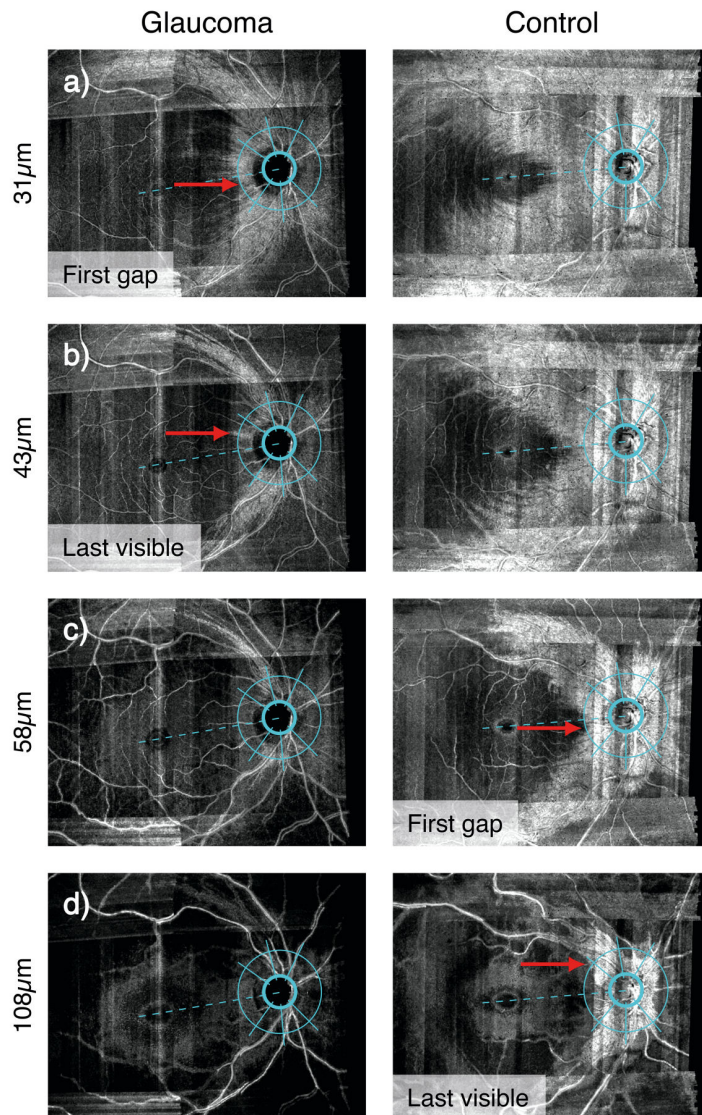
Figure 3.2 Example of the task in the temporal sector for a glaucoma eye (left panels) and an age-similar healthy participant (right panels). In (a) the red arrow shows the first gap for the glaucoma eye at 31μm below the ILM, whereas the corresponding depth for the healthy eye is reached at 58μm below the ILM (red arrow in c). The depth of last visible bundle (Last visible) was 43μm (red arrow in b) and 108μm (red arrow in d) below the ILM for the glaucoma and healthy eye respectively.

To reduce measurement bias, the grader was blind to the depth value until a final decision of the relevant image was made for the specific en face index. Nonetheless, it was not possible to mask the grader to the disease status, since the typical glaucomatous arcuate defects originating from the ONH were easily observable while viewing en face images. However, to minimise potential effects of preconception, the grading task was performed first in eyes

with glaucoma, and only subsequently in healthy controls. Hyper-reflectivity from blood vessels was ignored in performing the judgement.

In addition to the en face indices, RNFL thickness along the same 3.5mm circle was extracted at the angles of the en face parameters in each ONH sector (Figure 3.3). Conventional cpRNFL parameters including mean sector thickness and global thickness were also extracted. As per the en face measurements, thickness data were adjusted for the fovea-disc angle, as automatically measured by the Spectralis. These data served to establish a comparison between the en face parameters (first gap and last visible bundle) and the conventional thickness measures in our sample.

Lastly, the mean first gap and last visible bundle were computed for each eye, weighting for the width of each ONH sector by multiplying the indices by the width of the corresponding sector, summing and then dividing by 360°. Weighted means of the RNFL thickness at first gap and last visible bundle angles were computed the same way.
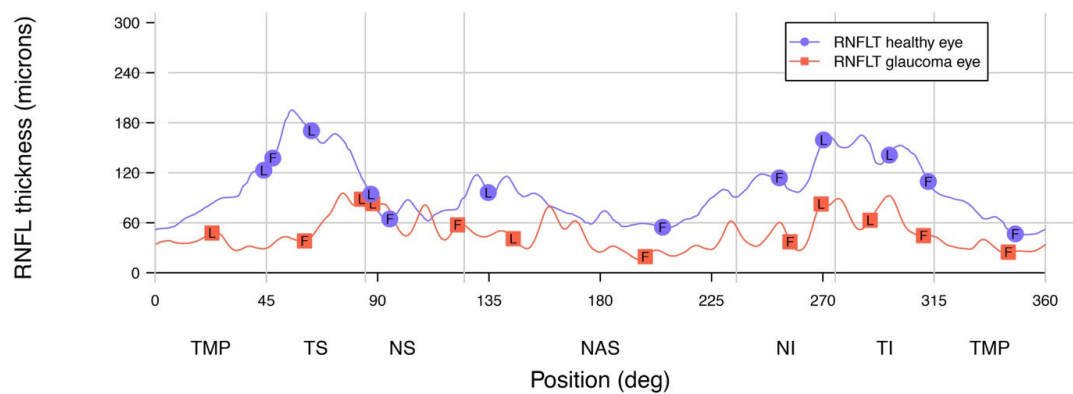
Figure 3.3 RNFL thickness (RNFLT) profiles for the same control and glaucoma eye as in Figure 3.2. Points marked by F and L represent the RNFL thickness at the angles corresponding to first gap and last visible bundle, respectively. TMP = temporal, TS = Temporal superior, NS = Nasal superior, NAS = nasal, NI = Nasal inferior, TI = Temporal inferior.

Data analysis

All analyses were conducted in the open-source environment *R,* version 3.5.1 (R Core Team, 2020). After grouping data according to disease status, distributions were explored visually and normality was tested with the Shapiro-Wilk tests. Most of the en face and RNFL thickness data were normally

distributed (p<0.05 in 4 and 3 out of 28 sets of data, respectively). As such, parametric statistics were used in subsequent analysis.

Linear mixed models (Bates et al., 2014) and likelihood ratio tests were used to evaluate the overall effect of glaucoma on en face first gap and last visible bundle, while accounting for repeated-measures from 6 ONH sectors. We tested whether the depth of en face indices was affected by glaucoma (fixed effect), while accounting for the individual eye and the ONH sector (random effects). Analysis was limited to sector data (i.e. sectors-average first gap and last visible bundle were censored). The model took the form:

$$y \sim 1 + Disease\ Status + (1|eye) + (1|ONH\ sector) + \varepsilon \qquad (1)$$

where y signifies the measure of interest (e.g. first gap), 1 signifies the intercept and ε signifies random error. A model of the same form was applied to RNFL thickness counterpart data. Post-hoc analysis was performed with independent t-tests to evaluate pairwise differences, adjusting for multiple comparisons with Bonferroni correction.

Diagnostic capability was quantified with standardised partial receiver operating characteristic area (pAUC), as computed with the *pROC R* package (Robin et al., 2011). To focus on the highest levels of specificity (Fallon et al., 2017, Mwanza et al., 2018a), pAUCs were calculated at specificity between 90% and 100%, with the trapezoid method. Values were standardised to range on a 50-100% scale and bootstrapped 95% CI were also computed. Where relevant, pAUCs of en face indices and the corresponding RNFL thickness parameters were compared with the DeLong method (DeLong et al., 1988).

An overall measure of the strength of correlation between depth of visible presence of RNFBs and RNFL thickness was estimated with repeated measures correlation, using the *rmcorr R* package (Bakdash and Marusich, 2017), according to Bland & Altman (Bland and Altman, 1995a, Bland and Altman, 1995b). The method allowed us to account for non-independency of our data when pooling data from the 6 ONH sectors of the same eyes, and provides a measure of strength of inter-individual association (*r*), interpretable as a Pearson's coefficient. For consistency, Pearson's correlation was used to

assess the strength of en face-thickness relation at each ONH sector. Correlation analyses were further explored according to disease status, with the hypothesis that if reflectivity loss precedes thinning, correlation could be expected to be poorer in glaucoma as compared to control eyes.

According to recent data on global cpRNFL thickness in healthy and glaucoma eyes (Verticchio Vercellin et al., 2018), two groups as small as n=5 would suffice to discover differences of similar effect size at 90% power and alpha level of 0.05. Data from the overall group (n=40) would allow to identify correlation (*r*) equal or greater than 0.48, at 90% power and α=0.05 (Champely, 2020). On the other hand, when grouping data for disease status (n=20), a correlation (*r*) equal or greater than 0.65 could be identified, at similar power and alpha level (Champely, 2020).

3.4 Results

20 glaucoma participants and 20 age-similar healthy controls were included in this study, whose demographics are reported in Table 3.1. On average, glaucoma participants presented an early-moderate defect, with the majority (17/20) showing an MD equal or better than -6dB. The remaining 3 participants presented MD of -6.1dB, -8.6dB and -14.9dB respectively.

Table 3.1
Demographics of included participants. Continuous data are summarised as mean and (SD).

	Control	*Glaucoma*
N	20	20
Male/Female	8/12	9/11
Mean Spherical equivalent (SD)	+0.7(2.0)	+0.0(1.4)
SAP Mean Deviation (dB)	0.6 (1.1)	-4.5 (3.1)
Age (years)	68.6 (5.0)	69.3 (5.1)
Average cpRNFL thickness (µm)	95.1 (9.3)	66.3 (9.4)

Overall, disease status had a significant effect on visible presence of RNFBs ($\chi^2_{(1)}$ = 63.3 & 51.6, both p<0.0001) with both first gap and last visible bundle that were closer to the ILM in glaucomatous eyes (mean difference: 39.1µm, 95%CI: 33.0 to 45.3 and 48.1µm, 95%CI: 38.8 to 57.4, respectively, both p<0.0001). Pairwise differences (Figure 3.4) showed both en face indices to

be significantly smaller in eyes with glaucoma compared to healthy controls across all ONH sectors (all p<0.0036). The greatest separation was found in the temporal inferior sector for both first gap parameters (difference: 80.8µm, 95%CI: 62.3 to 98.7; $t_{35.2}$=9.1, p<0.0001) and last visible bundle parameters (82.9µm, 95%CI: 66.0 to 99.9; $t_{34.2}$=9.9, p<0.0001).
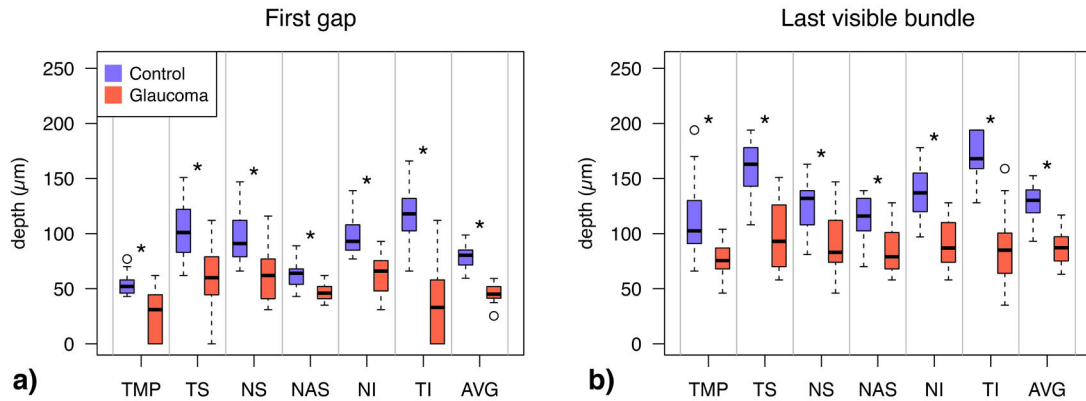
Figure 3.4 Boxplots showing differences between glaucoma and control eyes for the first gap (a) and last visible bundle (b) for every ONH sector and the sectors-average. At each ONH sector, control and glaucoma data are reported by the left-most and right-most box respectively, and colour-coded accordingly. After Bonferroni correction (14 comparisons), pairwise differences were considered significant when p<0.0036, and flagged with (*). Boxes report medians and 25th to 75th percentiles. Whiskers represent maximum and minimum values of data within 1.5x interquartile range above or below the limits of the box. Unfilled symbols represent outliers. ONH sectors acronyms as per Figure 3.3; AVG = average.

Linear mixed model analysis showed that RNFL thickness at angles corresponding to en face first gap and last visible bundle were also significantly smaller in glaucoma ($\chi^2_{(1)}$ = 59.6 & 37.4, both p<0.0001). Pairwise differences among RNFL thickness parameters are reported in Figure 3.5.
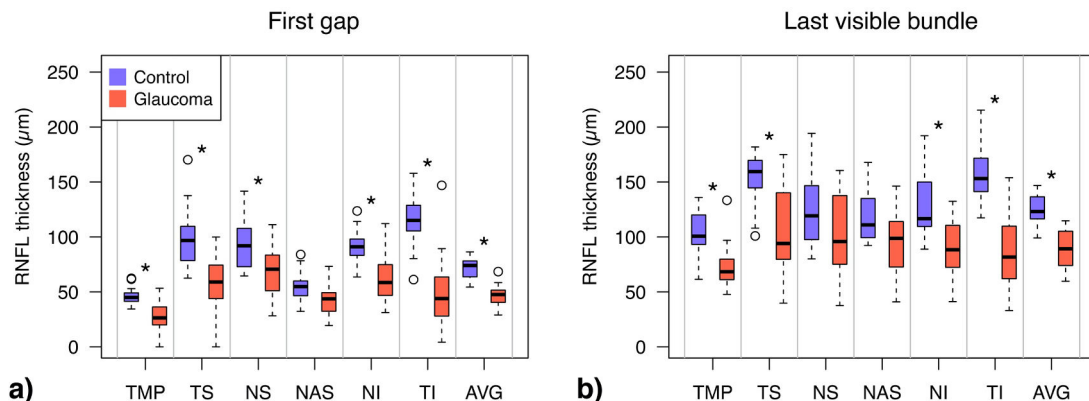
Figure 3.5 Boxplots showing pairwise differences between glaucoma and control eyes for RNFL thickness at corresponding angles of en face first gap (a) and last visible bundle (b). As per Figure 3.4, differences were considered significant when p<0.0036 (after Bonferroni correction, 14 comparisons), and flagged with (*). Data are reported for every ONH sector and the sectors-average. ONH sector acronyms as per Figure 3.4.

Diagnostic performance of visible presence of RNFB indices and RNFL thickness parameters at corresponding angles is shown in Table 3.2. The performance of conventional cpRNFL thickness analysis is also reported to establish a comparison and to assess the level of generalisability of this sample to clinical settings. Several en face indices showed excellent diagnostic capability (pAUCs > 0.9), yet similar to RNFL thickness counterparts. In fact, although en face first gap indices with best diagnostic performance (inferior temporal and sectors-average) were slightly higher than RNFL thickness counterparts, they were statistically similar (p=0.18 and p=0.16). Similarly, best en face indices for last visible bundles (inferior temporal and sectors-average) were marginally higher than corresponding RNFL thickness parameters, but differences were not statistically significant (p=0.33 and p=0.30).

Table 3.2
Diagnostic performance (standardised pAUC at specificity 90-100% with 95% CIs) of en face indices, RNFL thickness (RNFLT) at corresponding angles and conventional cpRNFL thickness measurements. ONH sector labels as per Figure 3.3 (TMP = temporal, TS = Temporal superior, NS = Nasal superior, NAS = nasal, NI = Nasal inferior, TI = Temporal inferior, AVG = average.).

ONH sector	En face First gap	RNFLT at First Gap Angle	En face Last Visible bundle	RNFLT at Last Visible Angle	cpRNFL thickness
TMP	0.86 (0.75, 0.95)	0.82 (0.71, 0.95)	0.63 (0.53, 0.87)	0.67 (0.53, 0.92)	0.70 (0.59, 0.95)
TS	0.79 (0.67, 0.92)	0.79 (0.68, 0.92)	0.80 (0.68, 0.95)	0.76 (0.66, 0.90)	0.74 (0.63, 0.92)
NS	0.82 (0.71, 0.92)	0.67 (0.58, 0.82)	0.76 (0.65, 0.90)	0.65 (0.53, 0.79)	0.72 (0.61, 0.84)
NAS	0.67 (0.55, 0.91)	0.63 (0.53, 0.79)	0.71 (0.57, 0.90)	0.70 (0.59, 0.83)	0.70 (0.55, 0.87)
NI	0.91 (0.83, 0.99)	0.84 (0.71, 0.97)	0.83 (0.71, 0.95)	0.76 (0.66, 0.92)	0.92 (0.82, 1)
TI	0.94 (0.84, 1)	0.88 (0.76, 1)	0.95 (0.87, 1)	0.92 (0.84, 1)	0.95 (0.87, 1)
AVG	1 (1, 1)	0.95 (0.87, 1)	0.90 (0.74, 1)	0.83 (0.72, 0.97)	0.95 (0.87, 1)

To explore diagnostic capability in settings potentially more relevant to clinical practice, diagnostic accuracy analysis was repeated in a subgroup of glaucoma participants with MD equal or better than -4.0 dB (n=11, Table 3.3).

Restricting the analysis to participants with glaucoma at earlier stages aimed to ignore more advanced cases which are deemed 'easier' to diagnose in practice (Chen et al., 2018). Nonetheless, pAUCs in the earlier glaucoma group were similar to the ones identified in the overall sample, suggesting no loss of diagnostic capability in our settings. Indeed, the best en face and RNFL thickness parameters (temporal inferior and sectors-average) were similar between overall group and the subgroup of glaucoma eyes with MD better than -4dB (all p>0.05).

Table 3.3
Diagnostic accuracy analysis (standardised pAUC at 90-100% specificity with 95% CIs) repeated in a subgroup of early glaucoma participants with MD better than -4.0dB (n=11). ONH sector labels as per Table 3.2.

ONH sector	En face First gap	RNFLT at First Gap Angle	En face Last Visible	RNFLT at Last Visible Angle	cpRNFL thickness
TMP	0.83 (0.68, 0.95)	0.81 (0.67, 0.98)	0.55 (0.47, 0.88)	0.67 (0.52, 0.93)	0.67 (0.53, 0.96)
TS	0.70 (0.55, 0.89)	0.71 (0.57, 0.90)	0.73 (0.57, 0.92)	0.67 (0.52, 0.86)	0.62 (0.52, 0.86)
NS	0.76 (0.62, 0.90)	0.62 (0.47, 0.79)	0.69 (0.55, 0.86)	0.61 (0.52, 0.79)	0.65 (0.52, 0.81)
NAS	0.65 (0.52, 0.94)	0.64 (0.52, 0.83)	0.65 (0.51, 0.86)	0.64 (0.52, 0.83)	0.65 (0.52, 0.86)
NI	0.84 (0.69, 0.98)	0.76 (0.57, 1)	0.71 (0.57, 0.90)	0.69 (0.57, 0.90)	0.90 (0.76, 1)
TI	0.98 (0.90, 1)	0.90 (0.76, 1)	1.00 (1, 1)	0.95 (0.86, 1)	0.98 (0.90, 1)
AVG	1.00 (1, 1)	0.98 (0.9, 1)	0.86 (0.66, 1)	0.76 (0.62, 0.95)	0.95 (0.86, 1)

Overall, depth of visible presence of RNFBs was strongly related to the corresponding RNFL thickness. Repeated measures correlation analysis showed a strong relationship between depth of first gap and the RNFL thickness at the same angle ($r_{df=199}$=0.87, 95%CI: 0.83 to 0.90, p<0.0001). Depth of last visible bundle and corresponding RNFL thickness were also strongly related ($r_{df=199}$=0.78, 95%CI: 0.72 to 0.83, p<0.0001). The relationship between en face indices and corresponding RNFL thickness in each ONH sector is shown in Figure 3.6. For first gap, the strongest correlation was found in the temporal sector (r = 0.92, 95%CI: 0.86 to 0.96, p<0.0001), whereas the

temporal inferior sector showed the strongest correlation for last visible bundle (r = 0.92, 95%CI: 0.85 to 0.96, p<0.0001).
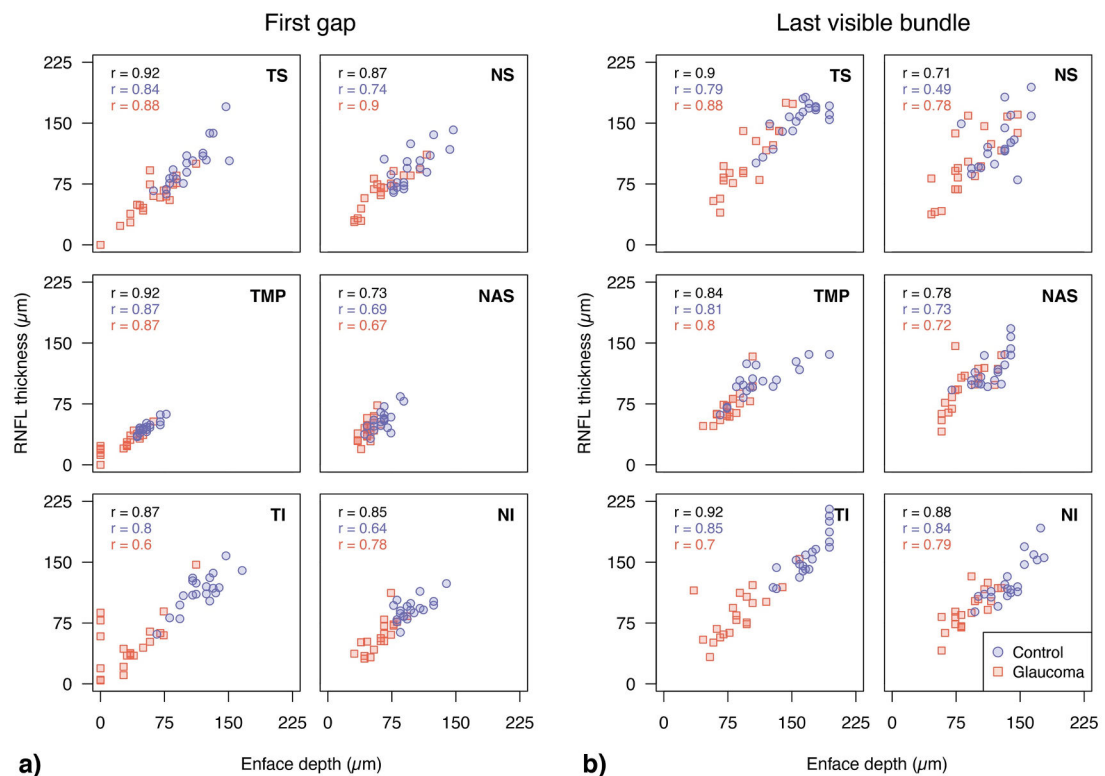
Figure 3.6 Relationships between (a) first gap & (b) last visible bundle with RNFL thickness at the corresponding angle in each sector of the ONH. Points are colour coded and shaped according to disease status. The Pearson correlation coefficients are computed with data from both glaucoma and healthy eyes combined (black), as well as grouping data according to disease status (colour coded accordingly). In the combined group, all correlation coefficients were p<0.0001. All correlation coefficients from data grouped according to disease status were p<0.001, with the exception of last visible bundle at NS and NAS in controls (p=0.02 and 0.002, respectively) and NAS and NI first gap in glaucoma eyes (p= 0.012 and 0.005 respectively). ONH sectors are labelled as in Figure 3.3.

To explore the hypothesis that loss of reflectivity might precede thinning of the RNFL, the strength of correlation was analysed by grouping data according to disease status. Across all sectors, repeated measures correlation for first gap was slightly stronger in healthy eyes compared to glaucoma eyes ($r_{df=99}$=0.92, 95%CI: 0.89 to 0.95, and $r_{df=99}$=0.80, 95%CI: 0.73 to 0.87, p<0.0001, respectively). On the other hand, although overall correlation between last visible bundle and corresponding RNFL thickness was marginally higher in healthy eyes ($r_{df=99}$=0.81, 95%CI: 0.74 to 0.88, p<0.0001) than glaucoma eyes ($r_{df=99}$=0.73, 95%CI: 0.62 to 0.81, p<0.0001), 95% CI limits overlapped. A more detailed analysis of sector-wise difference in strength of correlation between glaucoma and healthy eyes is reported in Figure 3.7. Coefficients were similar in many ONH sectors for both en face parameters, being sometimes greater

in healthy eyes (e.g. TI, first gap) and vice-versa in other ONH sectors (e.g. NS & NI first gap).
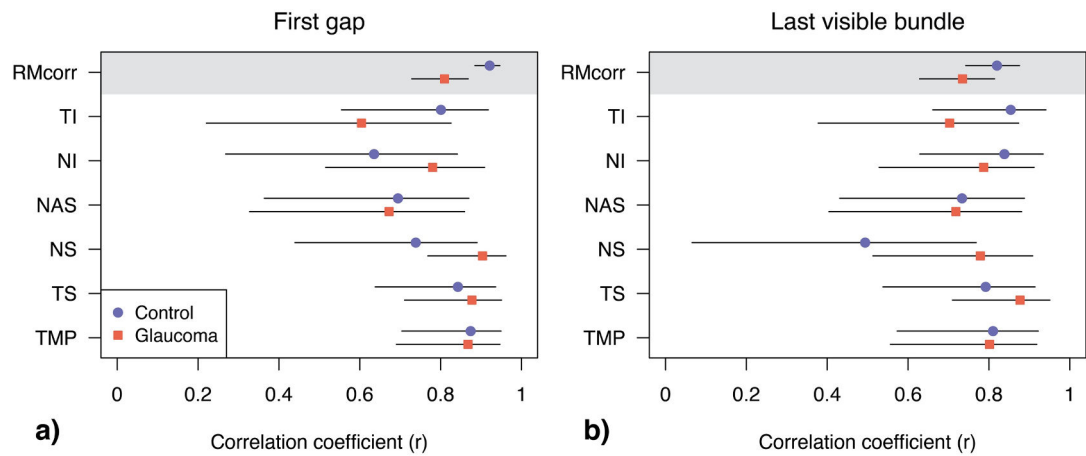
Figure 3.7 Pearson's correlation coefficients and their 95% CI for first gap (a) & last visible bundle (b) and corresponding RNFL thickness at each ONH sector, computed by grouping data according to disease status. Top panel in each plot reports the overall correlation and its 95% CI limits, computed with repeated measure correlation (RMCorr). ONH sectors are labelled as in Figure 3.3.

3.5 Discussion

Exploiting RNFL reflectance for prompter glaucoma diagnosis has attracted significant interest in glaucoma research (Pons et al., 2000, Liu et al., 2014, Vermeer et al., 2012, Gardiner et al., 2016, Tan et al., 2020, Thepass et al., 2017), and en face imaging now provides clinicians with a potentially powerful tool to explore it (Hood et al., 2016, Ashimatey et al., 2018a, Sakamoto et al., 2019, Iikawa et al., 2020, Leitgeb, 2019). Indeed, the technique might show RNFL changes before measurable thinning (Huang et al., 2011, Fortune et al., 2013). Additionally, en face images could also facilitate a more detailed clinical approach to OCT in glaucoma as opposed to the sole consideration of thickness summary measures (Sayed et al., 2017, Chong and Lee, 2012, Hood and De Moraes, 2018a, Hood et al., 2015). Yet, the clinical usability remains poor, with most objective and subjective methods for the assessment of reflectivity currently confined to research settings (Pons et al., 2000, van der Schoot et al., 2012, Vermeer et al., 2012, Ashimatey et al., 2018a, Tan et al., 2020, Thepass et al., 2017, Sakamoto et al., 2019, Miura et al., 2017). As such, we sought to develop a simple and clinically usable method for the evaluation of glaucomatous changes in en face images, which focussed on the subjective assessment of visible presence of RNFBs around the ONH.

En face parameters were able to identify glaucomatous changes. Indeed, both first gap and last visible bundle were significantly closer to the ILM in glaucoma, with temporal inferior sector and sectors-average measures showing greatest differences (Figure 3.4). This was not surprising since RNFL thinning is a hallmark of glaucoma, and since such changes might be predominantly noticeable at some ONH sectors, including the temporal inferior (Hood, 2017, Kansal et al., 2018, Chen et al., 2018, Weinreb et al., 2016, Jonas et al., 2017). In cases of thinner RNFL, deeper and hypo-reflective retinal layers such as the GCL and the IPL would be encountered at depths closer to the ILM. Further, some preserved RNFBs in glaucoma eyes might show reduced reflectivity, hence mimicking lack of bundles and contributing to smaller en face depths in our study (Hood et al., 2015, Cheloni and Denniss, 2021).

Previous studies in this area focussed on quantitative assessment of RNFL reflectance and also found a significant effect of glaucoma, suggesting agreement with our findings. Indeed, lower reflectivity of the RNFL was shown with time-domain OCT (Pons et al., 2000), and results replicated later with more advanced instruments (Vermeer et al., 2012, van der Schoot et al., 2012, Thepass et al., 2017). In these studies, the raw and normalised reflectivity as well as the attenuation coefficients were used to quantify reflectivity. Yet, irrespectively of the analysis performed, reflectivity of the RNFL was reduced in eyes with glaucoma compared to healthy eyes, and increasingly so with more severe disease (Vermeer et al., 2012, van der Schoot et al., 2012, Pons et al., 2000, Thepass et al., 2017).

A second purpose of the study was to assess the diagnostic capability of en face measures. Discrimination performance was excellent in many ONH sectors (pAUCs ≥ 0.9, Table 3.2), yet statistically similar to corresponding RNFL thickness parameters. The high accuracy of conventional cpRNFL thickness (e.g. global cpRNFL pAUC: 0.95) suggests that glaucomatous defects in this sample were already well captured by conventional morphological OCT analysis. This is unsurprising given that our inclusion criteria required a structural defect, nonetheless it is notable that en face indices performed similarly to or better than conventional thickness

measurements (pAUC higher in 12 of 14 comparisons, though all differences p>0.05). Estimated diagnostic performances, therefore, might not necessarily be representative of clinical settings, where clinicians aim to diagnose the earliest glaucoma cases, and conventional OCT analysis is imperfect (Stagg and Medeiros, 2020, Virgili et al., 2018, Karvonen et al., 2020, Michelessi et al., 2020). A few recent studies reported OCT diagnostic performance in clinically relevant populations, also avoiding some of the recurrent limitations of diagnostic accuracy studies in this area, such as case-control design and weak reference standard (Stagg and Medeiros, 2020, Virgili et al., 2018, Karvonen et al., 2020, Michelessi et al., 2020). Works conducted in the UK and Finland explored OCT diagnostic accuracy in large samples, representative of glaucoma patients potentially seen in practice (Virgili et al., 2018, Karvonen et al., 2020, Michelessi et al., 2020). Consistently with our findings, best diagnostic performance was achieved by the global and temporal inferior cpRNFL thickness, yet, with much lower sensitivity (range of highest sensitivity at 95% specificity: 0.53 to 0.55) and AUCs (range of highest AUCs: 0.76 to 0.84). Additional analysis from Stagg and Medeiros adopted a longitudinal design and defined glaucoma as a progressive ONH and/or RNFL defect in fundus stereo-photos in a group of glaucoma suspects with suspicious disc but normal VF at baseline (Stagg and Medeiros, 2020). Global cpRNFL thickness performed better than individual-sector indices and showed sensitivity of 0.60 at 95% specificity and an AUC of 0.89 (Stagg and Medeiros, 2020). The studies discussed above used overall AUCs and sensitivity at high specificity to report diagnostic accuracy; both parameters not included in our original analysis. To facilitate comparison, sensitivity at 95% specificity of global cpRNFL thickness from our data was 0.95 (95%CI: 0.75 to 1) in the whole sample, remaining as high as 0.91 (95%CI: 0.73 to 1) in the early glaucoma subgroup. Further, AUCs of global cpRNFL were close to perfect, irrespective of subgroup analysis (0.99, 95%CI: 0.96 to 1).

These findings confirm that research on improvements of OCT diagnostic accuracy in glaucoma retains a strong rationale. They also suggest that to thoroughly understand the actual value of new clinical procedures, after a first validation such procedures also need evaluation in settings and populations

highly representative of clinical practice (Stagg and Medeiros, 2020, Michelessi et al., 2020). While considering this limitation, some of the en face parameters presented hold promise for early glaucoma detection. For instance, sectors-average first gap discriminated glaucoma perfectly in this sample, making it pertinent to further explore the related diagnostic capability in settings more closely replicating clinical practice. To this end, we also explored diagnostic accuracy in a *'more difficult'* subgroup of glaucoma participants (MD better than -4dB). Yet, diagnostic performance was similar to the one in the overall sample for all indices (Table 3.3). Reduction of diagnostic accuracy with progressively earlier stages of disease is well known in the glaucoma literature (Kansal et al., 2018, Chen et al., 2018, Rao et al., 2011, Leite et al., 2010), and the lack of a similar pattern herein might be artefactual of our experimental settings. Indeed, our inclusion criteria required at least one abnormal sector from cpRNFL analysis. This could explain why diagnostic accuracy of OCT measures remained strong, irrespectively of subgroup analysis.

Among studies employing RNFL reflectivity for glaucoma diagnosis, none has adopted similar approach to the one proposed here. A few reports conducted quantitative analysis of RNFL reflectance for the discrimination of glaucoma and showed mixed results (Liu et al., 2014, Thepass et al., 2017, Tan et al., 2021). Liu et al. studied 34 healthy eyes, 18 suspects (either pre-perimetric glaucoma or ocular hypertension) and 24 definite glaucoma participants, and compared cpRNFL thickness and a pigment epithelium normalised reflectance index of the cpRNFL for glaucoma diagnosis (Liu et al., 2014). Conventional thickness showed similar accuracy to reflectance indices to detect definite glaucoma, whereas some superiority of reflectivity analysis (0.05 differences in AUCs) was found in the discrimination of glaucoma suspects (Liu et al., 2014). Aside from being a case-control analysis, the study was limited by a fragmented design which forced the use of two different OCT devices and included controls significantly younger than both groups of glaucoma participants. A similar normalised reflectance index was employed later by a study aiming to detect glaucoma progression, rather than diagnosis (Gardiner et al., 2016). Accordingly, reflectance analysis did not outperform cpRNFL

thickness to predict functional progression but, for a fixed amount of thinning, loss of reflectivity related to more rapid VF degradation. A more recent work further refined normalised reflectance indices and tested the related diagnostic capability as compared to cpRNFL thickness, in both pre-perimetric and perimetric glaucoma (Tan et al., 2021). Though AUCs suggested similar discrimination capability, reflectance analysis outperformed thickness in terms of sensitivity at 99% specificity in both glaucoma groups (best reflectance & best RNFL thickness: 0.53 vs 0.23, p=0.027 and 1.0 vs 0.80, p<0.001 in pre-perimetric group and perimetric group, respectively). Some limitations demand consideration, however. The included pre-perimetric glaucoma group had surprisingly low MDs (range: -7.3dB to 2dB), which might not be representative of a target population with undetected functional deterioration. Further, settings were likely to favour performance of reflectance over thickness, with the risk of data overfitting. While cpRNFL thickness was used as a continuous variable, reflectance analysis focussed on super-pixels with 'low-reflectance', considered such if the intensity was below the 5th percentile of control data. Accordingly, reflectance parameters were derived to work ideally in this specific setting, whereas the same was not true for thickness analysis. Lastly, in a case control study by Thepass and colleagues (Thepass et al., 2017), conventional thickness analysis (global cpRNFL) performed significantly better than refined measures of RNFL reflectance over the same OCT scan, both in terms of AUC (0.97 vs 0.83) and sensitivity at 90% specificity (97% vs 60%). Our results seem to align with the current evidence, which overall suggests that reflectance analysis performs well for glaucoma detection, though is not likely to be substantially superior to thickness analysis. Further studies are needed to test whether reflectance information can be combined with thickness measurements to further improve OCT diagnostic accuracy. It is worth noting, though, that there remains scope for improvement in the observation and quantification of defects in en face images, and, given the strong performance of simple approaches such as the one reported in this study, such improvements may yield greater diagnostic performance.

Our final analysis evaluated the relationship between en face parameters and corresponding RNFL thickness (Figure 3.6), revealing a strong correlation.

These findings are in concordance with those in the literature, though published data were generated with dissimilar approaches to the one used here. Pons et al. were the first to show a strong relationship (R^2: 0.61) between RNFL raw intensity and thickness in manifest glaucoma (Pons et al., 2000). More recently, Thepass et al. measured RNFL reflectance with attenuation coefficients and confirmed a significant relationship with RNFL thickness (Thepass et al., 2017). Similarly, Ashimatey and colleagues found a strong correlation (Spearman's ρ 0.74 to 0.82) between the fraction of abnormal en face reflectivity in the central retina and cpRNFL thickness of corresponding ONH sector (Ashimatey et al., 2018a). A strong relationship (Pearson's r 0.79 to 0.85) was also found in a study relating abnormalities of normalised RNFL reflectance and average thickness of the cpRNFL (Tan et al., 2021). The en face-thickness relationship we found was strong but imperfect, and several reasons for incongruence should be considered. For instance, blood vessels could be expected to have a larger impact on thickness measurements, since these could be distinguished subjectively from RNFBs in en face analysis. An estimate of such effect could be inferred from the slightly poorer correlation in last visible bundle compared to first gap parameters (~0.1), arguably imputable to the presence of blood vessels. In fact, major blood vessels are usually located in regions with thicker RNFL (Patel et al., 2011, Hood et al., 2008), where also RNFBs are expected to be visible at greater depths. In contrast, first gap more often coincided with regions of thinner RNFL which are also more likely to be areas free from major blood vessels. Although methods for removal of blood vessels from OCT scans exist (Patel et al., 2011, Strouthidis et al., 2009, Ballae Ganeshrao et al., 2015), these are not routinely adopted in clinics, and their usage here would likely result in further improvement of an already strong correlation. Segmentation inaccuracies of the proximal RNFL boundary could also play a role. This has been shown to be a difficult surface to segment (Yang et al., 2010), especially in areas with established damage (Vermeer et al., 2012). Since en face images as used here only depended on the vitreous-ILM surface segmentation, these inaccuracies only affected thickness measures. Lastly, since images feeding en face and thickness analyses were obtained with different scans, some of the differences might be

explained by small incongruences of corresponding angles and sector boundaries among the two domains. Besides these factors, some genuine differences between reflectivity and thickness properties should also be considered, but the overall strong correlation suggests they might have been small in our study.

To explore the hypothesis of reflectivity loss preceding thinning of the RNFL, the strength of correlation was evaluated in control and glaucoma eyes separately. A slightly poorer overall correlation in glaucoma was found only between first gap and corresponding RNFL thickness (Figure 3.7a). A more detailed look at sector data suggests that correlation was similar or even marginally stronger in glaucoma eyes in every sector but the temporal inferior. This was confirmed by similar repeated-measure correlation between the two groups found once censoring the temporal inferior sector from first gap data (glaucoma: $r_{df=79}$=0.89, 95%CI: 0.84 to 0.93, p<0.0001; and controls: $r_{df=79}$=0.91, 95%CI: 0.87 to 0.94, p<0.0001). The scatterplot corresponding to this relationship (Figure 3.6a, bottom left subplot) shows the presence of three outliers in an otherwise almost perfect relationship. En face images of these participants showed first gap in the temporal inferior sector at 0µm below the ILM, meaning that in part of this sector there were no visible bundles albeit still with measurable RNFL thickness at corresponding angles. Yet, two of the three eyes (Appendix A, Supplementary Figures 3.2, 3.3) presented prominent thinning of the temporal inferior RNFL and angular incongruences from experimental settings and/or effect of blood vessels on segmentation might have caused the observed differences. These areas with thinned RNFL (up to 20µm) would be more likely to correspond to en face first gap measures, hence leading to greater en face-thickness concordance. Conversely, in one case (Appendix A, Supplementary Figure 3.4) we observed some genuine differences between the en face first gap and inferior temporal RNFL thickness, which remained around 50µm. Nonetheless, it must be noted that the RNFL thickness was markedly outside normal limits, allowing the device's classification system to flag this area as a defect. On the whole, data from this sample did not seem to provide compelling evidence supporting loss of reflectivity without loss of thickness of the RNFL.

Despite considerable research interest, the temporal relationship between changes of reflectivity and thickness of the RNFL is not fully understood. Findings from models of experimental glaucoma (Huang et al., 2011, Fortune et al., 2013, Dwelle et al., 2012, Fortune et al., 2012, Fortune et al., 2015) suggest that reflectivity deteriorates earlier than a measurable thinning of the RNFL. Nonetheless, only part of these studies were conducted on non-human primates and reports are further limited by small sample sizes and incomplete or fragmented follow-ups imposed by experimental settings (Huang et al., 2011, Fortune et al., 2013, Dwelle et al., 2012). As such, evidence supporting nonparallel changes of reflectance and thickness is not yet fully compelling and the time delay between occurrence of changes in the two domains remains undetermined in humans. Indeed, the reflectivity-thickness relationship is still minimally investigated in clinical settings. No studies monitoring reflectance and thickness before and after conversion to glaucoma are available with only one published report that adopted reflectivity to detect progression of established glaucoma (Xu et al., 2013). In that study, Xu et al. measured retardance (scanning laser polarimetry) and cpRNFL thickness with OCT longitudinally. Accordingly, the majority of cases showed progression by RNFL thickness and in all cases thickness progression preceded retardance changes, by an average lag time of 13.4 months (Xu et al., 2013). These findings appear in partial disagreement with previous animal models using similar instrumentation, but, some caveats must first be considered. Measurements of retardance were more variable than RNFL thickness, hence required bigger changes to be marked as progression. Further, only a minority of participants actually progressed, and no gold standard of progression was considered by the study design. As such, evidence of earlier loss of reflectivity than thickness has still to be verified in clinical settings, as has the extent and significance of the timeframe interposed between changes in the two domains.

Overall these results fit well with findings from diagnostic accuracy studies, suggesting that analysis of RNFL reflectivity might be meaningful especially at the earliest stages of glaucomatous changes, becoming progressively less valuable when the disease is established and/or progressing (Liu et al., 2014, Xu et al., 2013, Fortune et al., 2015, Gardiner et al., 2016). According to our

results and to the available literature, once the disease is well established reflectivity information might strongly relate to RNFL thickness (Pons et al., 2000). This, however, does not preclude other uses of en face imaging providing additional value. For example, en face imaging may be useful in combination with other test modalities such as VF, enabling direct exploitation of the structure-function relationship without the need for spatial structure-function mapping, and in facilitating custom-perimetry based on structural data (Denniss et al., 2013, Montesano et al., 2018, Alluwimi et al., 2018b, Turpin et al., 2018, Ganeshrao et al., 2015).

This study has some limitations. Though we aimed to introduce a method for en face analysis immediately usable in clinics, subtle changes of reflectivity might be overlooked by our subjective approach. More sensitive and quantitative methods might be required to fully exploit the value of en face images (Ashimatey et al., 2018a, Vermeer et al., 2012). Second, we studied established glaucoma eyes and the hypothesis of existent discrepancies between RNFL reflectance and thickness at earlier stages of glaucoma has to be considered. A fairer analysis on the reflectivity-thickness relationship should, perhaps, be longitudinal and focus on conversion to glaucoma of healthy or glaucoma suspect eyes. Cross-sectional approaches as employed here, and commonly in the literature (Michelessi et al., 2020), will generally adopt a reference standard for glaucoma diagnosis requiring obvious signs (e.g. defined RNFL thinning) and, therefore, be biased toward tests from the same analysis. Accurate glaucoma detection appeared easy in this cohort and this was likely not representative of the actual clinical scenario of glaucoma diagnosis (Stagg and Medeiros, 2020, Virgili et al., 2018).

Lastly, the image processing and the grading task performed might present some constraints. We did not include correction for light beam incident angle, which is among the determinants of RNFL reflectance (Huang et al., 2016). The incidence angle is known to change in circumpapillary scans, and could be expected to play an even bigger role in wide-field imaging seeded by multiple OCT scans (Tan et al., 2021, Knighton and Huang, 1999). Nonetheless, these caveats may have greater impact on a purely quantitative approach, but may have less effect on subjective evaluation of RNFB visible

presence. It should also be noted that extraction of en face parameters considered here could be accomplished with a single cube scan centred on the ONH, further confining such limitations. Additionally, images feeding the en face and thickness domains were obtained from different scans, which were not mutually registered. To minimise disagreement, data in each domain were adjusted for the individual fovea-disc angle and the strong correlation found might suggest that any angular incongruences, if preserved, were small. Concerning the grading task, the order of en face images was not randomised between glaucoma and controls, and the presence of visible glaucomatous changes in many en face images precluded masking of the grader to disease status. Further, the task was performed by a single observer and not repeated over time. However, previous work showed both consistency between observers and repeatability within observers to be excellent in the assessment of visible presence of RNFBs (Cheloni and Denniss, 2021, Iikawa et al., 2020). Compared to the grading task mentioned above (see also chapter 2), it could be speculated that the task performed here left less room for individual interpretation, resulting in similar or better agreement and repeatability.

3.6 Conclusions

This simple method to observe visible presence of RNFBs reliably identified glaucomatous defects in en face OCT images, with diagnostic performance at least equivalent to existing thickness parameters. The abnormalities found were already well captured by conventional thickness measurements, and no strong evidence of reflectivity loss without corresponding thickness loss was found. However, alternative ways to detect abnormalities in en face images (e.g. automated methods, artificial intelligence) and different glaucoma populations may produce different results. Development of more sensitive automated analyses and integration with perimetry may realise further potential of en face OCT images in glaucoma.

4. Enhanced objective detection of retinal nerve fibre bundle defects in glaucoma with a novel method for en face OCT slab image construction and analysis

4.1 Abstract

Background: Optimal methods to extract en face OCT slab images to detect RNFB glaucoma defects remain undetermined. We aimed to introduce a novel method for en face slab image construction (SMAS) which considers varying individual anatomy and configuration of RNFBs through the retina. We also evaluated the ability of SMAS and several other methods for slab construction to objectively identify RNFB reflectance defects in en face images.

Methods: Dense OCT scans of the central retina were acquired in 16 participants with glaucoma (median age: 70, range 61 to 77) and 19 age-similar controls. En face slab images were generated according to several methods, averaging reflectivity over different depths below the ILM. SMAS considered multiple 16μm thick slabs from 8 to 116μm below the ILM, whereas 5 alternative methods considered single summary slabs of various thickness. Superpixels in glaucoma eyes were considered abnormal if below the 1st percentile of fitted distributions to controls data. In SMAS, defects were defined when occurring in any slab. All methods adjusted for the individual position of the raphe, fovea and optic disc. Ability to detect glaucoma defects was measured by the proportion of abnormal superpixels and locations tested. Proportion of superpixels below the fitted 1st percentile in controls was used as surrogate false positive rate. The effect of slab methods on performance measures was evaluated with linear mixed models.

Results: The ability to detect glaucoma defects changed significantly between different slab methods ($\chi^2_{(5)}$ = 120.9, p<0.0001), with SMAS showing proportion of abnormal superpixels 0.05 (5%) to 0.09 (9%) larger than alternatives (all p<0.0001). No slab method found superpixels in controls below the 1st percentile suggesting similar specificity.

Conclusions: Ability to detect glaucoma abnormalities in en face images changed between slab methods. SMAS evaluates all depths with potential

RNFB presence by combining several thin slabs at each location and resulted in greater detection of glaucomatous reflectance abnormalities.

4.2 Introduction

Though direct observation of RNFBs with en face OCT imaging may be linked to multi-fold benefits in the assessment of glaucoma, the ideal method to construct and analyse slab images remains undetermined. As a consequence, objective and automated criteria of defining defects in this domain are also lacking (Cheloni and Denniss, 2021).

As detailed above, en face images are usually generated from dense volumetric scans of the area of interest followed by projection of pixel intensities from a certain range of depths within each A-scan into a transverse slab image (Leitgeb, 2019). Healthy RNFBs appear hyper-reflective because of the ordered structure of their axonal cytoskeleton (Huang et al., 2006), whereas in damaged bundles hyper-reflectivity is lost according to typical glaucomatous patterns such as arcuate and wedge-shaped defects (Ashimatey et al., 2018a). Besides a primary loss of reflectivity due to axonal disruption and reduced bundle density, thinning of the RNFL leads to inclusion of deeper, hypo-reflective retinal layers in the slab image, also reducing reflectivity (Sakamoto et al., 2019, Hood et al., 2015).

The construction and analysis of en face images can be accomplished through several approaches, posing the selection of multiple parameters. These range from the retinal surface imaged to the actual slab composition, as defined by the depths below the ILM that are combined. Furthermore, several arithmetic methods could be considered to project 3-dimensional data into transverse images, and different approaches to account for individual anatomy could also be applied. We showed previously (see chapter 2) that the configuration of RNFBs varies with retinal location and individual anatomy (Cheloni and Denniss, 2021). Accordingly, the final appearance and diagnostic utility of en face slab images is likely affected by details of slab construction. The magnitudes of these effects have been minimally investigated and methodological choices in the slab extraction process are currently poorly informed. Previous works in this area have often averaged the first 50µm

below the ILM into a single slab image (Hood et al., 2015, Sakamoto et al., 2019, Iikawa et al., 2020, Miura et al., 2017). Alternatively, Ashimatey et al. proposed to adjust slab thickness at different retinal locations, but still limited their analysis to a 50μm depth (Ashimatey et al., 2018a). Consistently with RNFL thickness (Varma et al., 1996, Mauschitz et al., 2018, Kim et al., 2021), RNFBs are present at depths larger than 50μm proximal to the ONH (Cheloni and Denniss, 2021), and limiting en face analysis to this depth might prevent capture of glaucoma defects in their entirety (Ashimatey et al., 2018a). Further, although many authors have recognised the requirement to adjust slab characteristics to individual anatomy and the varying morphology across the retina (Hood et al., 2015, Sakamoto et al., 2019, Iikawa et al., 2020, Ashimatey et al., 2018a), these features have not been fully considered.

In this study we introduce 'Summary of Multiple Anatomically-adjusted Slabs' (SMAS), which is a novel method for the construction and analysis of slab images. SMAS aimed to address some of the current limitations of en face imaging, including: i) adapting to individual anatomy; ii) considering all depths with present RNFBs, and iii) adjusting for different layer morphology across the retina. We also aimed to evaluate the ability of several methods of slab construction to detect en face glaucoma defects compared to SMAS. Collapsing en face information into a single slab has the potential to facilitate an immediate observation of glaucoma status, however, we expect this benefit to be counteracted by a potential underrepresentation of defects. The results will provide insights on how slab parameters affect the ability to retrieve reflectance anomalies, resulting in preliminary evidence to inform the choice of the ideal strategy to automatically extract defects in this domain.

4.3 Methods

Settings and participants

22 participants with open angle glaucoma and 19 age-similar healthy controls had usable en face montages and were included. The study achieved ethical approval from the National Health Service's Research Ethics Service and all participants provided written inform consents in adherence to the Declaration of Helsinki.

Details of inclusion and exclusion criteria have been presented earlier (section 2.3). Briefly, participants with glaucoma were only included if older than 40 years and presenting a clinical diagnosis of open angle glaucoma. In addition, inclusion required evidence of structural damage defined as at least one ONH sector with $p<1\%$ from the Spectralis cpRNFL thickness analysis. No VF inclusion criteria were applied to the glaucoma group in order to include the earliest cases. Participants with refractive error greater than ±6.00DS and ±3.00DC, evidence of lens opacification (Chylack et al., 1993) or other eye conditions except glaucoma were excluded. Healthy participants were included if they had no eye diseases and normal VF as defined before (see section 2.3). One eye per participant was included. If both eyes were eligible, the tested eye was selected at random in healthy controls, whereas the one with milder defect (as identified by a less negative MD) was included among glaucoma participants.

OCT imaging and image processing

All participants were imaged with multiple high-speed dense OCT scans (30μm separation between B-scans) of the central ±25° of the retina. The OCT imaging procedure has been described in detail earlier (section 2.3). Besides en face imaging, to evaluate RNFL thickness conventional circumpapillary analysis was also conducted by means of the 3.5mm diameter circle B-scan as automatically segmented by the OCT built-in software.

Single-pixel deep slabs (n=50) of the instrument's maximum digital axial resolution (3.87μm), ranging from the ILM to 193.5μm below it, were extracted from individual B-scans. Slab images were converted to depth-resolved attenuation coefficients (Vermeer et al., 2013) and data were imported into MATLAB (Version 9.6.0, The MathWorks Inc., Natick, Massachusetts) for montaging and image-processing.

Different to the previous chapters (section 2.3 and 3.3), gamma-correction was used to smooth intensity differences between OCT images from different retinal locations before montaging (Xiong and Pulli, 2010). Several image processing strategies were explored, including linear scaling and histogram equalisation (Gonzalez et al., 2004), however, gamma-correction provided the

best equalisation while preserving the initial contrast and appearance of original images. The transformation aimed to match the luminance of overlapping regions of neighbouring images at each depth using the central macular image as the reference. The ratio between median intensities of the individual slab and the macular image in corresponding overlapping regions was used as the gamma coefficient. Accordingly, single images at each of the 50 depths were normalised with the median of the 99th percentile of all attenuation coefficients at each depth. This was computed by dividing the intensities of each pixel by the median of the 99th percentiles. Subsequently, regions of overlap between the macular scan and other scans were identified for all images (i.e. Inferior Nasal, Inferior Temporal; ONH, Raphe; Superior Nasal and Superior temporal, see Figure 4.1). Median intensities of the regions of overlap were extracted for each individual image and the macular image at each depth. The ratio between the average intensity of the individual slab and the corresponding macula region was then computed and used as the gamma coefficient in the transformation.
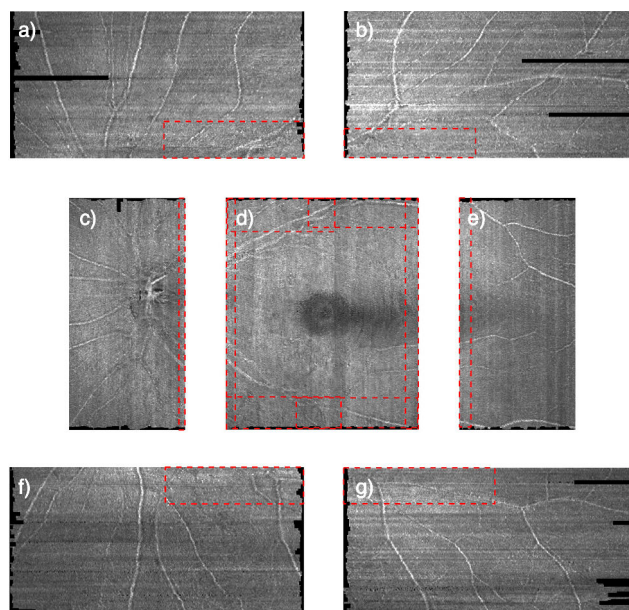
Figure 4.1 Slab images of a healthy eye (4µm below the ILM) containing attenuation coefficients normalised by the median of 99th percentiles. Red dashed lines indicate overlapping regions between the macula and other slabs.

For example, to identify the transformation coefficient for the raphe image at a specific depth, the median intensity of the area of overlap in the raphe scan (Figure 4.1e) was divided by the median intensity of the corresponding area of overlap in the macula scan (Figure 4.1d – right most rectangle). The gamma correction used ratio coefficients computed specifically for each depth and

each individual eye. Images with overlapping areas brighter than the macula led to coefficients greater than 1, which would result in gamma corrections reducing the luminance of the image. Gamma corrections with ratios smaller than 1 increased the luminance of the image. Following gamma correction, slab images at each depth were converted back to corrected attenuation coefficients by multiplication by the same parameter used earlier for normalisation. We then montaged the images using custom software as per section 3.3, using the macular scan as a reference image. Lastly, image processing described in detail earlier was applied to the montaged en face images (section 2.3).

Adjusting for individual anatomy

We aimed to minimise the impact of individual anatomy by adjusting en face images to the fovea-disc and fovea-raphe angles. This was achieved by using geometric image transformations to align the ONH, fovea and raphe along a common horizontal axis (Figure 4.2). Custom MATLAB software was used to perform all transformations. Left eyes were flipped to right eye format and the slab image offering best visibility of the foveal pit was used to manually extract the coordinates of the fovea and the centre of the ONH. The fovea-disc angle was defined as a straight line between these two points. The orientation of the raphe is known to change with individual anatomy and can be measured in both healthy and glaucoma eyes (Bedggood et al., 2017). The fovea-raphe angle was extracted following an existing method (Bedggood et al., 2016). By using a single slab image with optimal visibility of RNFBs of the raphe (median 15.5μm below ILM, range 11.6 to 27.1μm), the fovea-raphe angle was measured by tracing lines connecting the fovea to 5 manually-selected points in the raphe gap region. The fovea-raphe angle was ultimately considered as the average of the 5 angles identified (Bedggood et al., 2016). Shear transformations of the vertical image dimension were then applied separately to the temporal and nasal retina, leaving the horizontal dimension unmodified. This approach considered the fovea and the vertical axis running through it as the centre of the transformation, leaving this region unmodified whilst aligning the raphe, fovea and ONH along a horizontal line (see Figure 4.2 for examples). Shear transformations enable the shift of a single dimension of the

image (vertical in this case) by a given angular value, leaving the other (horizontal) dimension unmodified. This approach allowed us to evaluate reflectivity over a square array of superpixels (see below), with landmark retinal locations laying on a common horizontal axis (Figure 4.2).
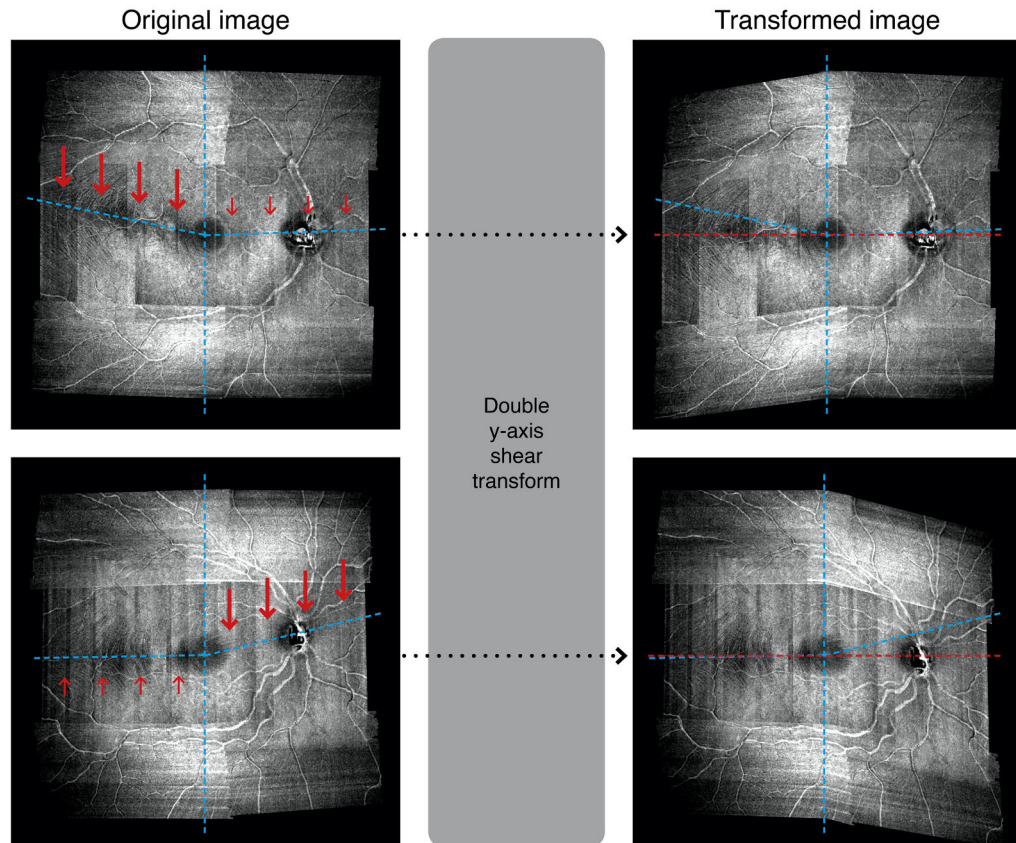
Figure 4.2 Example of the double vertical shear transformation applied to the en face images of two healthy controls. For the image shown in the top panels, the major transformation (thick red arrows) was applied to the temporal retina whereas a smaller transformation (small red arrows) was applied to the nasal retina. The opposite applies for the image shown in the bottom panels. Irrespectively of the original anatomy (blue dashed lines), transformed images (right column) align the raphe, fovea and ONH along a horizontal line (red dashed lines).

Extraction of reflectance abnormalities with SMAS

With the objective of considering all depths with present RNFBs and the different layer morphology across the retina, 4 single-pixel slabs were averaged together starting from 7.8µm up to 193.5µm below the ILM. The first two depths (i.e. up to 7.7µm below the ILM) were excluded from slab construction since, as observed previously (chapter 2), they do not present visible RNFBs (Cheloni and Denniss, 2021). This censoring should also reduce the impact of vitreous interface and glial artefacts, which predominantly affect depths immediately below the ILM (Ashimatey et al., 2018b, Ashimatey

et al., 2018a). The averaging of 4 slabs together aimed to combine the greatest number of single-pixel slabs while minimising the mixture of retinal layers, i.e. RNFL with deeper layers such as GCL and IPL (Cheloni and Denniss, 2021). In fact, RNFBs are visually present at the narrowest range of depths in the temporal retina, and the averaging used by SMAS should allow inclusion of all bundles from these regions in the first slab (7.8μm to 23.2μm below the ILM). This should also be achieved in eyes with particularly thin RNFL (Cheloni and Denniss, 2021). Overall, SMAS yielded 12 slabs (≈15.5μm thick) from 7.8μm to 193.5μm below the ILM (Table 4.1).

Table 4.1
Detailed parameters of the 12 compounded slabs generated with SMAS

Slab n°	From (μm below ILM)	To (μm below ILM)	Included in final analysis	Slab n°	From (μm below ILM)	To (μm below ILM)	Included in final analysis
1	7.8	23.2	Yes	7	100.7	116.1	Yes
2	23.3	38.7	Yes	8	116.2	131.6	No
3	38.8	54.2	Yes	9	131.7	147.1	No
4	54.3	69.7	Yes	10	147.2	162.5	No
5	69.8	85.1	Yes	11	162.6	178	No
6	85.2	100.6	Yes	12	178.1	193.5	No

Analysis of reflectivity was performed on a superpixel grid centred on the fovea, with each superpixel comprised of a number of individual pixels in a n by n pixel configuration. Intensity of each superpixel was the mean of its constituent pixels. Additional mitigation of anatomical variability was achieved by controlling for the varying distance between the fovea and ONH by adjusting the size of superpixels in individual images such that a fixed 20 superpixels separated the fovea and ONH. This number was chosen to target a superpixel dimension of 20x20 pixels, previously suggested to represent a suitable compromise between between-subject variability and sufficient resolution to detect wedge shaped defects (Ashimatey et al., 2018a). Henceforth, the dimensions of a superpixel in an individual participant were determined by dividing the fovea-ONH distance (in pixels) by 20. Superpixels in individual images averaged median 20x20 pixels (range 17x17 to 23x23) and this system of coordinates maximised consistency between retinal

locations among different eyes. For instance, [x=20, y=0] and [x=0, y=0] would always represent the centre of the ONH and the fovea respectively, for all participants. Figures 4.3 and 4.4 show examples of the superpixel grid.

Data from all controls were extracted at all depths on the individualised superpixel grid. Distributions of superpixel intensities were visually explored at different retinal locations (Appendix A, Supplementary Figure 4.1) and Shapiro-Wilk tests for normality were performed. The majority of distributions were either multimodal or significantly skewed. To account for observed distributions and the modest sample size, summary statistics and limits of normality were derived from kernel density-estimated frequency distributions rather than sample data (Scott, 1992). Alongside the median from superpixels of all control eyes (Figure 4.3), we extracted the estimated 10th, 5th and 1st percentiles at all depths of all retinal locations.
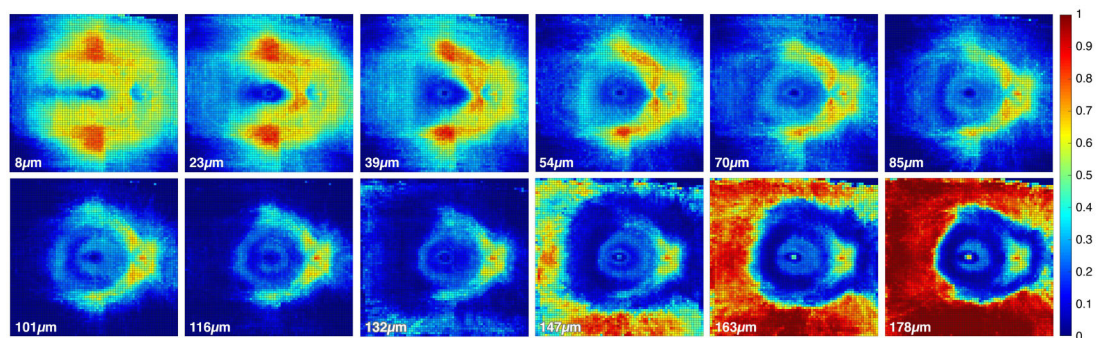
Figure 4.3 Heatmaps of median normative data for the 12 slabs generated in the SMAS method. It is possible to identify the main retinal structures such as the temporal raphe, fovea and ONH. At greater depths the hyper-reflectivity of the retinal pigment epithelium becomes visible. Depths (μm) shown correspond to the anterior depth at which each of the 15.5μm thick slabs commenced. The heatmaps also show hyper-reflective artefacts in the superior- and inferior regions (first 3-4 slabs), likely due to the varying beam-light incident angle from wide-field OCT imaging, see Discussion (Huang et al., 2016).

As reported previously (chapter 2), and as shown in Figure 4.3, the presence of RNFBs changes with retinal location and with the depth below the ILM. Evaluation of reflectance in locations where RNFBs are not expected to be present even in healthy retinae is unlikely to have any additional diagnostic value, therefore these areas were censored from analysis. Accordingly, quantification of reflectance loss was restricted to the first 7 slabs (up to 116μm below the ILM). In fact, at greater depths (8th to 12th slabs, Figure 4.3) only a minority of the surfaces examined presented RNFBs and even further below the ILM the retinal pigment epithelium would likely confound the investigation.

To further confine analysis to retinal regions where RNFBs may be present, we manually identified regions of interest in each of the 7 normative slabs in which RNFBs were visible in healthy eyes. Within these regions, mean and SD of visible RNFB intensities were computed. The grand mean of average bundle intensities at each depth was computed to determine a suitable cut-off for a threshold transformation to exclude regions with no RNFBs in healthy eyes. As such, different levels of intensity below the grand mean (1.5, 2, 2.5, and 3 SD) were visually explored to identify the best compromise between reduction of the area evaluated and the adequate exclusion of regions with no visually present RNFBs (Figure 4.4). 2.5 SD below the grand mean of RNFB intensity was selected. Regions with lower mean intensity in control eyes were excluded from analysis in all images. This censoring was applied to each depth individually.
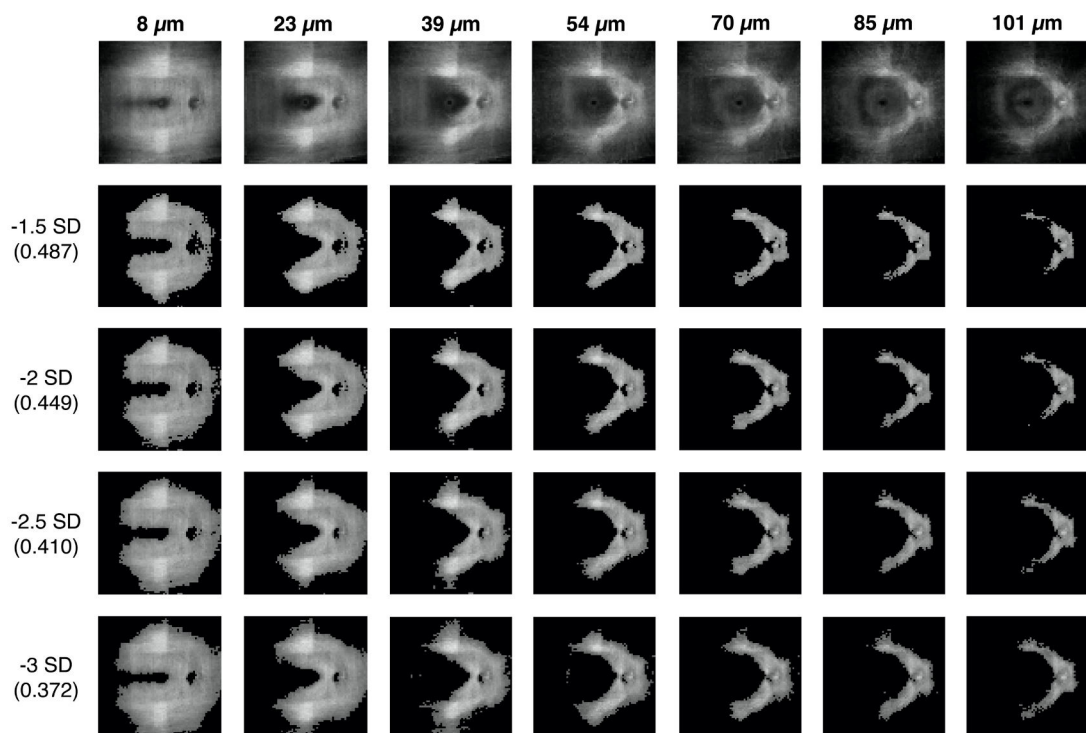
Figure 4.4 Different threshold levels (SDs below the grand mean intensity of visually present RNFBs) evaluated to censor retinal regions with no visible RNFBs. The top panel shows normative median intensity from all controls. 2.5 SD below the grand mean intensity was selected as the final level. These regions can be seen in the corresponding panel as shades of grey, whereas regions in full black were ignored. SD = Standard Deviation.

Finally, slab images according to the SMAS approach were extracted in all glaucoma participants. Superpixel values at all locations and depths were compared with corresponding estimated normal limits from controls to

establish reflectance abnormalities. Data on abnormal reflectivity for each glaucoma participant were reported in single depth deviation maps, at the 10%, 5% and 1% level of significance (Figure 4.5).

Additionally, the 7 deviation maps from each single slab were combined into a summary deviation map reporting abnormal superpixels (below 1%) identified in any slab (Figure 4.5, bottom right).
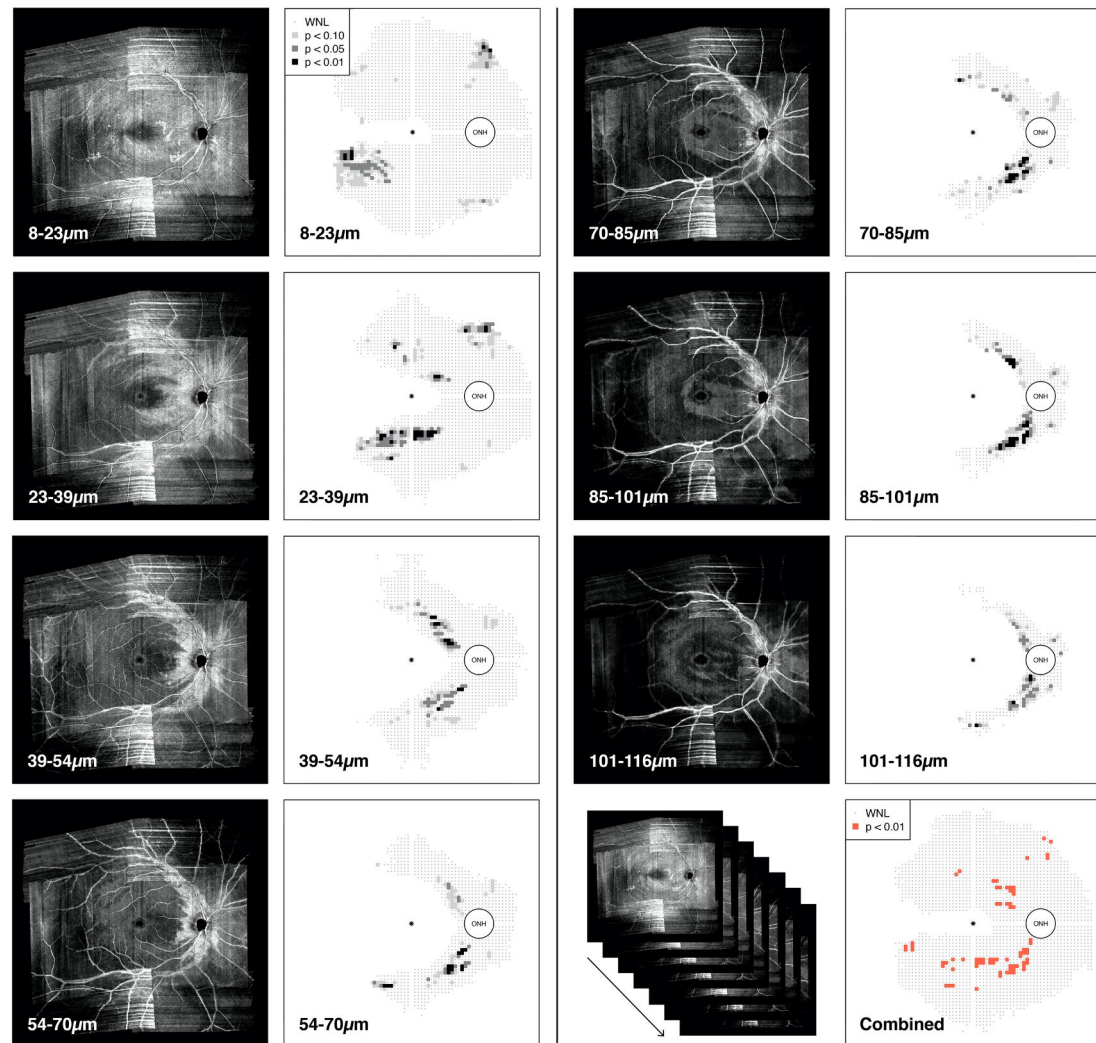
Figure 4.5 Example of the multiple slab images generated with the SMAS method for one glaucoma participant, as well as the combined deviation map reporting all reflectance abnormalities detected in any slab. At each specific slab depth, whose starting and ending depths are reported in µm, the left-most image shows the actual slab image and the right-most image shows the corresponding deviation map. In deviation maps, grey points indicate superpixels found within normal limits (WNL), and superpixels below the 10th, 5th and 1st percentiles are reported as squares colour-coded according to the level of significance. The bottom-right panel shows the combined deviation map with abnormal superpixels (below 1%) identified at any depth. In deviation maps each data point corresponds to 1 superpixel.

Extraction of reflectance abnormalities with alternative en face slab methods

En face images for all participants were also generated using several other slab methods based on our previous experiment (chapter 2) and published studies (Hood et al., 2015, Ashimatey et al., 2018a, Cheloni and Denniss, 2021). For each method explored, transformed single-pixel slabs of individual eyes were averaged together over a specific range of depths. Normative data from controls were then extracted as described earlier and reflectance abnormalities were evaluated in glaucoma eyes. Adjustments for individual anatomy made for SMAS were applied identically for other slab methods, as was the positioning and spacing of the superpixel grid. Therefore, different slab methods differed only for the retinal depths averaged (Figure 4.6). All slab methods were evaluated over the same region of the retina tested by SMAS.

Hood slab:

This method was similar to that of Hood and colleagues in which pixel intensity was averaged over a 52µm deep slab starting from the ILM (Hood et al., 2015). We averaged the first 13 single-pixel slabs together, encompassing depths from the ILM to 50.3µm below it as the closest possible match to the method of Hood et al.

Ashimatey slab:

Ashimatey et al. used a slab with decreasing thickness from the ONH to the temporal retina (Ashimatey et al., 2018a). They averaged pixel intensity from 24 to 52µm below the ILM in the optic disc region, from 24 to 36µm in the central retina and from 16 to 24µm in the temporal macula and raphe region. To reproduce a similar slab configuration, we considered three vertically-separated regions with different thickness (Figure 4.6): 7-13th pixels in the ONH region (≈ 23.3 to 50.3µm), 7-9th pixels in the macular area (≈ 23.3 to 34.8µm), and 5-6th pixels in the raphe area (≈ 15.6 to 23.2µm).

Best Visibility slab (BestVis):

This slab was centred on the depth of best RNFB visibility across the retina in healthy eyes, which, as reported in chapter 2, we found at an average of 20.3µm (SD: 1.9µm) below the ILM, with slight differences between temporal and nasal retina (Cheloni and Denniss, 2021). Accordingly, the single-pixel slab at the depth of best RNFB visibility alongside the one above and the one below were averaged together in the *BestVis* slab (Figure 4.6). The 4-6th pixels were included in the raphe and temporal macula (≈ 11.7 to 23.2µm, best = 18µm), whereas the 5-7th pixels were averaged in the rest of the retina (15.6 to 27.1µm, best = 22µm).

All visible RNFBs slab (AllVis):

As per the *BestVis* slab, this approach considered our previous work and averaged all depths expected to contain visible RNFBs in healthy eyes (see chapter 2). Differences in RNFB visibility across the healthy retina were accounted for by averaging varying depths in different regions of the retina (Figure 4.6). Hence, the pixel depths included were: 3-8th in the raphe (≈ 7.8 to 31µm); 3-10th in the temporal macula (≈ 7.8 to 38.7µm); 3-12th in the inferior nasal quadrant (≈ 7.8 to 46.4µm); 3-15th in the central and superior-nasal retina (≈ 7.8 to 58.1µm); and 3-21st in the ONH region (≈ 7.8 to 81.3µm).

Deep slab:

The *deep slab* included greater depths below the ILM than considered by most of the methods above and was included for completeness, rather than expected greater detection capability. The *Deep slab* averaged intensity starting from depths close to the posterior limit of the *Hood slab* and the *Ashimatey slab* up until the greatest depths at which arcuate regions and the nasal retina around the ONH still present RNFBs in controls, as per Figure 4.3 and findings from chapter 2. The slab averaged the 10-20th pixels through the whole retina, corresponding to 34.9 to 77.4µm below the ILM.
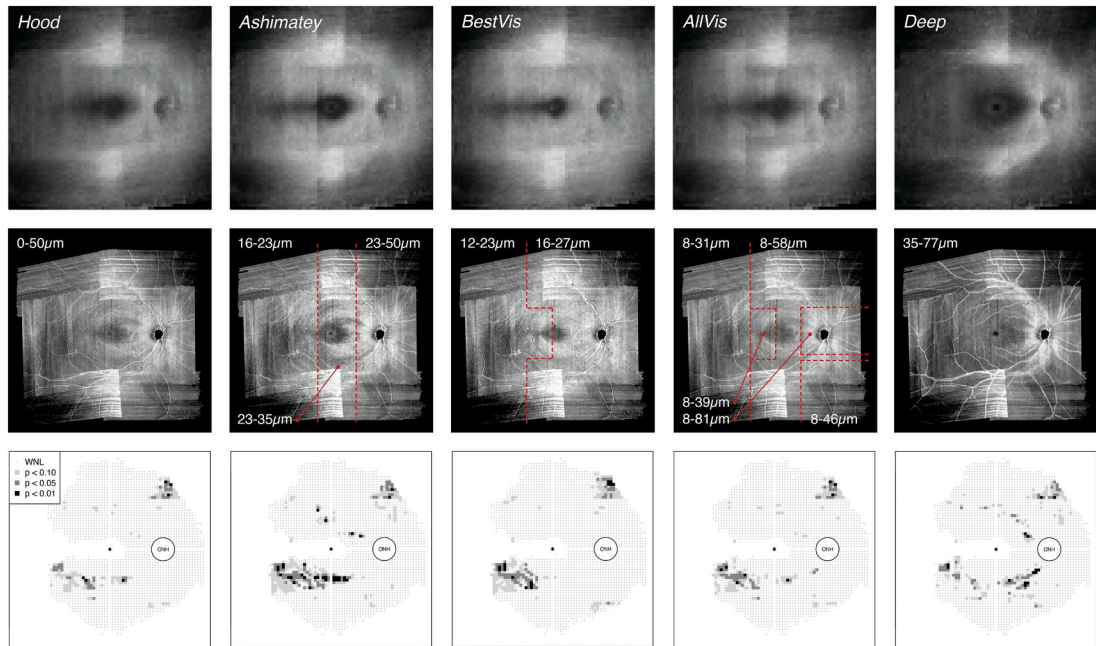
Figure 4.6 Alternative methods of slab construction explored in this study in addition to SMAS. Top row: Method-specific normative en face slabs derived from control eyes. Middle row: Individual slab images for the same glaucoma participant as shown in Figure 4.5. Red dashed lines demarcate different regions of the slab characterised by different depths considered (see text), as labelled in µm. Bottom row: Corresponding deviation maps for different slab methods. Format of deviation maps as per Figure 4.5.

Data analysis

All extracted slabs and corresponding deviation maps were examined by two investigators (RC, JD) for the impact of artefacts. Either whole images or specific regions from participants with substantial effects of artefacts were excluded from further analysis. Artefacts of en face images could arise from low-quality B-scans, floaters and glial cell alterations (Ashimatey et al., 2018b). Joint discussion of single cases was performed until consensus on data exclusion was reached.

Performance of different slab methods was compared by the proportion of abnormal superpixels identified in each glaucoma participant. This was computed as the number of superpixels below the 1st percentile of the corresponding normative data divided by the number of tested superpixels. For SMAS, the combined deviation map was considered. Differences in proportion of abnormal superpixels between slab methods were explored with linear mixed models and chi-squared likelihood ratio tests, using the *lme4* package in R (Bates et al., 2014). Slab method was considered as a fixed effect, whereas individual participants were modelled as a random effect to

124

account for repeated measures. Statistical significance was considered at p<0.05 and the model had the following form:

$$y \sim 1 + Slab\ Method + (1|Eye) + \varepsilon \qquad (1)$$

where y signifies the outcome of interest (e.g. proportion of abnormal superpixels), and 1 and ε signify intercept and random error respectively. Pairwise differences were tested post-hoc with t-tests, adjusting for multiple comparisons with the Tukey method.

To further characterise individual slab methods, median distance of abnormal superpixels from the ONH was extracted from all glaucoma participants for all different slab construction methods. As specific methods might include different retinal sections across the area examined, the resulting slab composition and therefore the ability of detecting abnormalities could also change with retinal location. Further, false positive rates of different slab methods could not be directly evaluated due to the lack of an independent reference standard. As a surrogate measure, we explored the rate of abnormal superpixels in control eyes at the 1% level of significance (derived from kernel density estimation as described earlier). Linear mixed models of the form above were used to evaluate differences of distance from ONH and surrogate false positive rate among different slab methods. Lastly, we tested whether differences in performance between SMAS and alternatives were related to the severity of reflectance defects. As such, we computed coefficients of determination (R^2) between the mean and standardised difference in proportions of abnormal superpixels between each slab method and SMAS. The latter was computed as:

$$(SMAS - Alternative\ slab\ method) / SMAS \qquad (2)$$

The use of standardised differences aimed to control for varying levels of reflectance loss, considering relative rather than absolute difference in proportion of abnormal superpixels.

As estimated with the *simr R* package (Green and MacLeod, 2016), this study had 91% power (95%CI: 89.2 to 92.8) to detect a 0.02 (2%) difference of the proportion of abnormal superpixels at the 0.05 alpha level.

4.4 Results

Images from 6 glaucoma eyes (median age 69, range 67 to 78; median MD -6.0dB, range -1.6 to -12.8) contained significant artefacts and were excluded from the main analysis. Further, parts of the en face images of 3 glaucoma participants were censored for similar reasons. For one participant the whole upper hemi-field was excluded, whereas a horizontal band in the upper retina and the inferior temporal retina were censored in the remaining 2 cases.

Overall, 19 controls (median age: 68, range 56 to 75) and 16 glaucoma participants (median age: 70, range 61 to 77) were included. All participants in the glaucoma group but one had a VF defect according to the criteria presented earlier (see section 2.3). The remaining glaucoma participant had three contiguous defective points but one was an edge location. Table 4.2 reports detailed participant demographics.

Table 4.2
Demographics of included participants. Continuous data are reported as median and (interquartile range).

	Control group	Glaucoma group
N	19	16
Age (years)	68 (6)	70 (8.25)
Caucasian/Other ethnicity	18/1	16/0
Male/Female	8/11	8/8
SAP Mean Deviation (dB)	0.8 (1.4)	-3.3 (2.2)
Average cpRNFL thickness (µm)	98 (11.5)	68 (14)
Axial length (mm)	23.26 (0.50)	24.12 (0.96)

The proportion of abnormal superpixels identified by different slab methods for all glaucoma participants is reported in Figure 4.7. SMAS found a greater proportion of abnormal superpixels than all alternative slab methods in all participants with glaucoma.
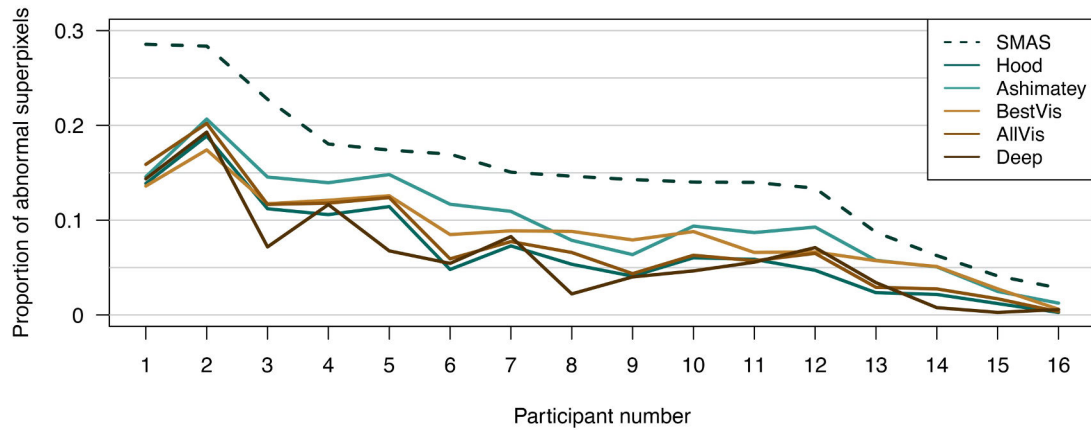
Figure 4.7 Proportion of abnormal superpixels identified by different slab methods in all glaucoma participants. This was computed as the ratio between the number of superpixels with p<1% and the total number of superpixels. Participants 1, 6 and 8 had images partially censored due to artefacts.

There were significant differences between slab methods in the detection of glaucoma defects ($\chi^2_{(5)}$ = 120.9, p<0.0001). Pairwise differences between SMAS and each other slab method are shown in Figure 4.8. All other methods retrieved smaller proportions of abnormal superpixels compared to SMAS (all p<0.0001). The smallest difference in proportion of abnormal superpixels, i.e. the slab with most similar performance to SMAS, was found for the *Ashimatey slab* (-0.051, 95%CI: -0.063 to -0.039, p<0,0001), whereas the *Deep slab* showed the largest difference (-0.086, 95%CI: -0.098 to -0.074, p<0,0001). Among alternative methods, the *Ashimatey slab* performed significantly better than the *Hood*, *AllVis* and *Deep slabs* by 0.03 (95%CI: 0.018 to 0.042), 0.022 (95%CI: 0.01 to 0.034), and 0.035 (95%CI: 0.023 to 0.047) respectively (all p<0.05), whereas the *BestVis slab* outperformed the *Deep slab* by 0.023 (95%CI: 0.011 to 0.035, p=0.005).
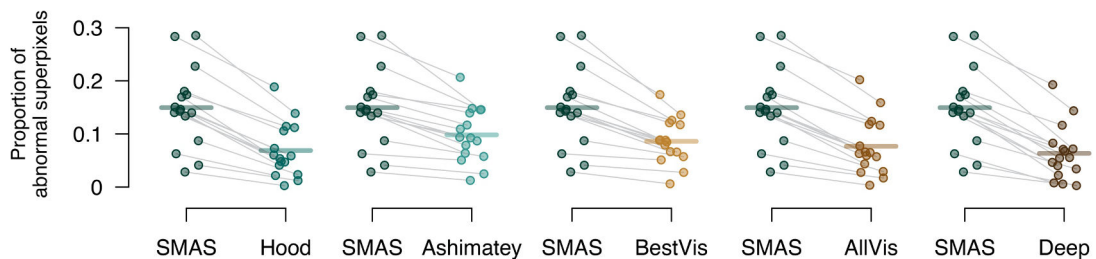
Figure 4.8 Differences of proportion of abnormal superpixels between each single slab method and SMAS. Different methods are colour coded as per Figure 4.7. Grey lines link data from individual images, horizontal lines indicate means. All differences p<0.0001.

The median distance of abnormal superpixels from the ONH changed significantly between different slab methods ($\chi^2_{(5)} = 50.0$, p<0.0001). As shown in Figure 4.9, the *Ashimatey* and *BestVis slabs* were more likely to find defects further away from the ONH compared to methods either assessing the whole RNFL or its deepest portions (*SMAS* and *Deep slab*). Indeed, pairwise differences showed the *Ashimatey slab* and the *BestVis slab* to identify abnormal superpixels significantly further away from the ONH compared to SMAS (differences: 3.5, p=0.003 and 4.6, p<0.0001, respectively). The *Deep slab* found defects closer to the ONH than all other slab methods (all p<0.05) with exception of SMAS, in which distances were smaller but statistically similar (Deep-SMAS: -1.6, p=0.47).
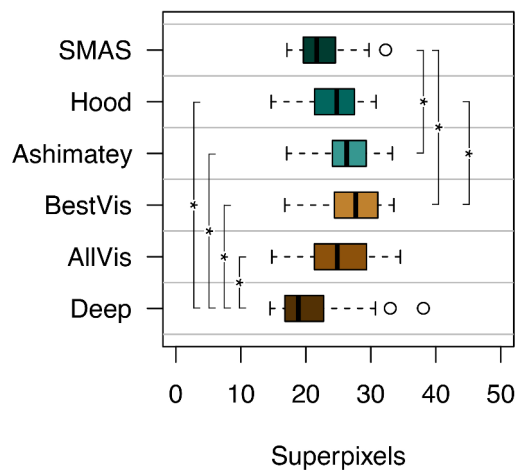
Figure 4.9 Boxplot showing the median distance from the centre of the ONH of abnormal superpixels in eyes with glaucoma by slab method. Data are colour coded according to different slab methods as per previous figures. Slab methods showing statistically significant differences (p<0.05) of defect distance from the ONH are flagged with (*).

As a surrogate measure of the false positive rate we computed the rate of abnormal superpixels in controls. However, none of the slab methods showed superpixels with intensity below the 1st percentile of control eyes, which was the level of significance considered for defining defects in glaucoma group.

Scatterplots showing the relationship between mean and standardised differences in proportion of abnormal superpixels between SMAS and each of the other slab methods are shown in Figure 4.10. The *Hood*, *AllVis* and *Deep Slabs* showed a negative relationship between differences in detection of reflectance abnormalities and average reflectance loss (slopes p<0.01, R^2 range 0.58 to 0.66). This suggests that the benefit of SMAS over these alternatives is greater for earlier defects. The same was not true for the

Ashimatey and *BestVis* slabs*,* whose performance compared to SMAS was relatively consistent across the range of reflectance loss in this sample.
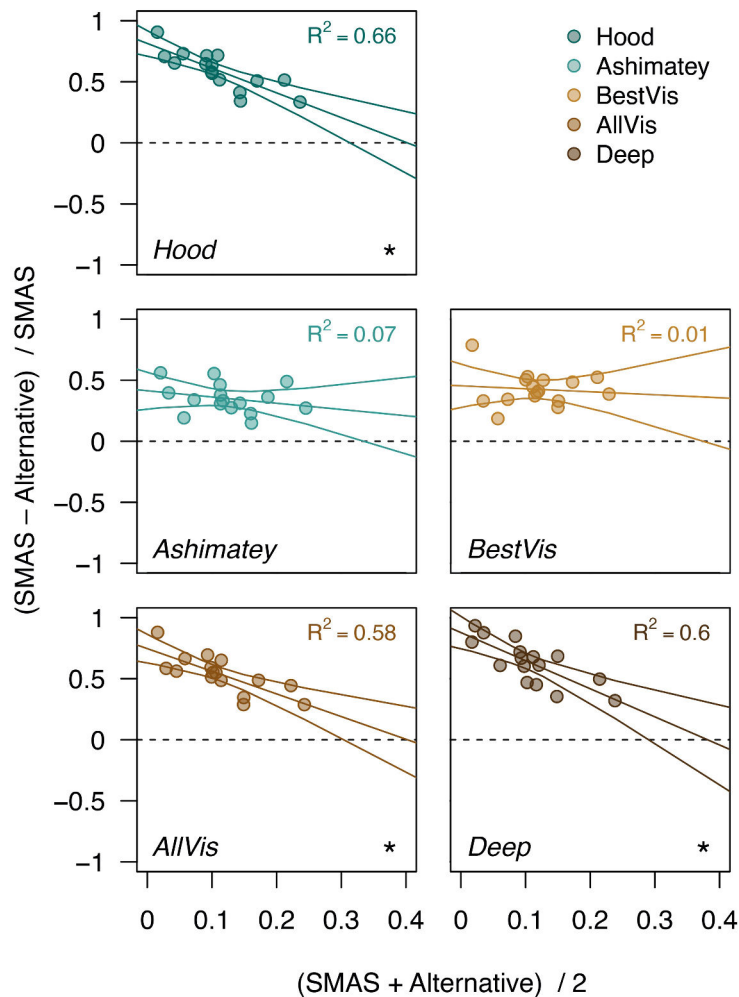
Figure 4.10 Bland-Altman-like scatter plots of the relationship between the mean and standardised difference of proportion of abnormal superpixels between SMAS and alternative slab methods. Standardised difference was calculated as the difference between proportion of abnormal superpixels of SMAS and the alternative slab method, divided by the proportion of abnormal superpixels for SMAS. The best linear fit to the data (with 95% CI) and corresponding R^2 are also reported. Cases where the linear regression line had a slope significantly different from 0 ($p<0.05$) are flagged with (*). Different slab methods are colour-coded as per previous figures.

4.5 Discussion

The construction of en face slab images can be accomplished in several different ways, and there is currently limited evidence on optimal methods to construct slab images for detection of glaucomatous defects. The parameters to be selected include the slab composition, that is the depth below the ILM to be combined in the slab, and the method used to control for anatomical variability. Each of these parameters likely affects slab final appearance and, more importantly, the related ability to detect glaucomatous changes. In addition to slab construction, there is also a lack of strategies for automated, objective definition of defects. These are critical to facilitate a consistent evaluation of reflectance loss as well as the automated use of en face findings

for seeding other investigations, such as custom perimetry (Chong et al., 2014, Denniss et al., 2018, Alluwimi et al., 2018a, Alluwimi et al., 2018b).

In this study we introduced SMAS, a novel approach for the construction of en face slab images and the automated, objective detection of glaucomatous defects in en face images. With SMAS, we aimed to introduce improvements over existing methods, such as i) examination of all depths containing visible RNFBs, by including reflectivity information up to 116µm below the ILM around the ONH; ii) greater consideration of the varying configuration of RNFBs across the retina, by averaging together thinner retinal sections and focussing on areas with presumed RNFBs in healthy eyes; iii) greater consideration of inter-individual anatomical variability, by adopting a superpixel grid of individualised dimensions and arranged to follow the specific raphe-fovea-ONH alignment; and iv) consideration of a wider area of the retina to include all regions containing visible RNFBs in healthy eyes except the temporal raphe.

Compared to other methods, SMAS performed better at identifying glaucomatous defects, with a detected proportion of abnormal superpixels 0.05 (5%) to 0.09 (9%) greater than alternatives (Figure 4.8). Although SMAS was the only method with the key feature of considering multiple summary slabs, analysis of surrogate false positive rate showed no abnormal superpixels in control eyes at the level of significance used to extract defects in glaucoma eyes (1st percentile). This suggests that SMAS detects more defects in eyes with glaucoma without increasing the number of false positives in healthy eyes.

While considering potential reasons for greater capability in detecting glaucoma defects, it should first be noted that methods mainly differed in their slab composition. The greater retinal area of examination and novel treatment of anatomical variability was applied to all slab methods in this study, even if not strictly replicating original settings, where individual anatomy was not or only partially considered (Sakamoto et al., 2019, Hood et al., 2015, Ashimatey et al., 2018a). A more detailed adjustment for the individual anatomy could be expected to reduce variability of measurements, which would otherwise limit

retinal regions in which reflectance modifications could be assessed (Ashimatey et al., 2018a). Increased detection of abnormal superpixels in glaucoma with SMAS could be attributable to the consideration of multiple slabs through the range of depths containing visible RNFBs in healthy eyes. In addition, SMAS aimed to evaluate changes of reflectivity at all retinal depths presenting RNFBs. This might have accounted for the better performance compared to the *Ashimatey* and the *BestVis slabs*, both generating relatively thin slab images circumscribed to the first 20-50µm below the ILM. Accordingly, these two approaches might be as effective as SMAS in the detection of reflectance changes in retinal regions with thin RNFL, with a progressive loss of capability for defects localised further away from the ILM. This may be the case in the glaucoma eye reported in Figure 4.5 and 4.6, where an inferior arcuate defect was distinctly visible starting from 39µm below the ILM. *BestVis* and *Ashimatey slabs* completely or partially failed to detect this defect, whose origin from the ONH was instead well captured by the *Deep slab* (Figure 4.6). Findings from the median distance of abnormal superpixels from ONH seems to conform with this interpretation. In fact, *BestVis* and *Ashimatey slabs* detected, on average, abnormal superpixels further away from the ONH, where the RNFL is thinner and RNFBs are also present at smaller range of depths below the ILM (Cheloni and Denniss, 2021, Varma et al., 1996, Kim et al., 2021). As compared to these two approaches, *Deep slab* examined depths presenting RNFBs only in the nasal retina and unsurprisingly found defects significantly closer to the ONH by 5.1 and 6.2 superpixels, respectively (both p<0.0001).

As an additional reason for the observed better performance, the settings adopted by SMAS might have contributed to a more consistent slab composition. Slab images were generated by depth-averaging thin sections of the retina, likely with more homogeneous composition compared to the slabs with greater depth-averaging. Specifically, depth-averaging of only 4 pixels (~16 µm) per individual slab was selected to contain all bundles in the temporal retina and to minimise mixture with deeper bundle-free layers. Conversely, the combination of thicker sections of the retina (e.g. *Hood* and *AllVis slabs*) might be more prone to inter-individual variability and also facilitate the inclusion of

retinal layers deeper than the RNFL and not presenting RNFBs. This would likely lead to more variable normative data of reflectivity, ultimately impacting the ability to detect glaucomatous changes. This seems to be supported by the better performance of SMAS compared to the *AllVis slab* (0.015 vs 0.077, p<0.0001). Though the two approaches assessed approximately the same total retinal depths, their performance in detection of glaucoma defects differed significantly.

To our knowledge, this is the first attempt to quantify the ability of different methods for en face slab construction to detect changes of reflectivity in glaucoma. Among the few studies available, authors have usually justified their slab parameters with findings from pilot testing, without published data (Hood et al., 2015, Ashimatey et al., 2018a, Sakamoto et al., 2019, Iikawa et al., 2020). Comparison with previous work is further complicated by different aims and settings. Most studies performed subjective evaluation of reflectance abnormalities (Iikawa et al., 2020, Sakamoto et al., 2019, Hood et al., 2015), and the only analysis including objective extraction of glaucoma en face defects is the work of Ashimatey et al. (2018a). Additionally, settings between studies would hardly overlap in terms of retinal areas examined, instruments used and metrics considered to evaluate reflectance. For instance, Hood et al. considered a smaller region centred on the ONH (Hood et al., 2015), whereas other investigators mainly focussed on the macula (Sakamoto et al., 2019, Iikawa et al., 2020). Different target regions would result in different configurations of the RNFBs in the area tested, justifying selection of different slab parameters. Notwithstanding the difficulties, our results appear in agreement with those of previous reports. Indeed, in earlier work (see chapter 2) we showed that RNFB configuration changes through the retina, suggesting that slab parameters need to be adjusted in order to consistently detect defects across the retina (Cheloni and Denniss, 2021). This was confirmed in this study, showing that slab methods do affect the capability to identify defects and a better performance is achievable when refining such aspects. In the study from Ashimatey et al, en face reflectance abnormalities were evaluated with a slab image with the depth averaging characteristics detailed above. These authors noted the inability of that setup to identify all reflectance losses,

and the requirement to extend the analysis further below 52µm to retrieve all existing defects (Ashimatey et al., 2018a). Additional work from the same lab is the only attempt to analyse the effect of different slab parameters to detect glaucoma defects (Cheung and Swanson, 2019). Therein, the authors considered the average reflectance of small circular regions (30 pixel diameter) placed around the ONH, and different ranges of depths were combined together in several slabs. The greatest ability in the detection of glaucoma was achieved by averaging reflectivity from 36 to 60µm below the ILM, as compared to slabs including 0-52µm, 24-52µm and 24-36µm (Cheung and Swanson, 2019). These results seem to confirm the importance of considering greater depths with present RNFBs in order to retrieve glaucoma defects. However, the inclusion of greater depths should not be achieved by averaging together larger sections of retina, but rather with alternative approaches able to preserve consistency of slab composition.

This study has limitations. Although we included processing strategies to adjust for uneven illumination of scans from different retinal locations and computed attenuation coefficients to minimise the impact of artefacts, the final images were still affected by these issues. Indeed, a non-negligible proportion of eyes had to be excluded due to a substantial impact of artefacts, from either activated glial artefacts, uneven illumination and/or low quality B-scans. This has to be considered, especially for the purpose of applying similar approaches in clinical practice. Accordingly, additional analysis, more sophisticated image processing and/or improved image capture could be evaluated to further minimise the impact of artefacts in future studies. For instance, attenuation coefficients may be further refined (Ghafaryasl et al., 2020), and the consideration of the varying incident light beam angle would likely improve en face OCT analyses, especially when relying on wide-field imaging (Huang et al., 2016, Knighton and Huang, 1999, Tan et al., 2021). Furthermore, the small sample size did not allow exploration of the impact of different parameters on the observed reflectance, such as age, eye laterality or ethnicity (Thepass et al., 2017, Wagner et al., 2020, Mauschitz et al., 2018). Larger studies could allow development of normative data adjusted for

covariates with clinically significant impact on reflectivity, ultimately leading to further refinement of the defect extraction method.

One additional constraint might affect the study design. Cross-sectional analysis of ability to detect reflectance changes lacked an appropriate reference standard for defining whether superpixels flagged as defective were truly defective or not. We deliberately avoided the use of structural measures from alternative OCT measures (e.g. cpRNFL thickness) to reduce bias. Accordingly, performance of each slab method could not be evaluated with conventional indices of classification accuracy, and we focussed on the proportion of abnormal super-pixels and a surrogate measure of false positive. Lastly, we assessed the performance of SMAS in a sample with established glaucoma. It should, however, be noted that a key goal of en face imaging is the improvement of early glaucoma detection, when conventional OCT metrics have been showed to be imperfect (Michelessi et al., 2020, Virgili et al., 2018, Stagg and Medeiros, 2020). As such, further evaluation in glaucoma groups with only the earliest signs of glaucoma would be highly relevant. Nonetheless, it could be speculated that to detect these earliest changes, the examination of greatest depths below the ILM would become even more essential than in our sample. More subtle defects could be found deeper in the RNFL, possibly making SMAS more advantageous than its alternatives to detect reflectance modifications at the earliest stages of glaucoma, as supported by Figure 4.10.

4.6 Conclusions

In summary, we developed and presented a novel method for the construction and analysis of slab en face images. The method considers the individual anatomy and all the depths and regions presenting RNFBs, apart from the temporal raphe. SMAS was able to automatically and objectively detect glaucomatous changes of RNFB reflectance and in our sample of participants with glaucoma, this method performed better compared to the other available approaches. Further assessments in eyes with the earliest stages of glaucoma is warranted.

5. Concordance of RNFB reflectance loss in en face OCT images with conventional structural and functional glaucoma changes

5.1 Abstract

Background: We have previously introduced a novel method for en face OCT image construction and objective detection of glaucomatous defects of RNFBs reflectance. As further validation, we assessed how identified abnormalities relate to conventional measures of glaucoma damage, including cpRNFL thickness and VF.

Methods: 16 participants with early glaucoma and 29 controls underwent wide-field SD-OCT, conventional optic disc scans and VF examination. En face images were generated by averaging reflectivity over multiple 16μm thick slabs, up to 116μm below the ILM, according to SMAS methods. Superpixels in glaucoma eyes were considered abnormal when below the 1st percentile of control data. En face defects were deemed concordant with cpRNFL thickness when they had at least one cpRNFL thickness point with p<1%, within ±15° of the predicted insertion on the optic disc. Proportion of concordant en face defects was measured in each eye. VF was examined with SITA Standard 24-2 (SAP) and a dense supra-threshold custom perimetry test. 24-2 locations showing total deviation with p<2% were considered defects. For any VF location, corresponding reflectance was deemed abnormal if at least one en face superpixel lying within ±1° was abnormal. The overall, positive and negative agreement were measured in each participant.

Results: Most en face abnormalities had concordant cpRNFL thickness defects at the mapped sector (median proportion of concordance: 0.85, IQR 0.74 to 0.95). In glaucoma eyes a median of 8.1% (range 2.4% to 23.7%) and 14.9% (range 3.5% to 29.1%) of locations showed corresponding en face and VF defects for SAP and custom perimetry, respectively. Both VF strategies had moderate-good raw agreement with en face analysis (0.66-0.68), with stronger agreement on normal findings compared to defects (0.77-0.78 & 0.4-0.44, respectively).

Conclusions: Objectively extracted reflectance defects with our novel method showed strong concordance with conventional cpRNFL thickness damage, and good agreement with VF. Further minimisation of en face artefacts could enhance the structure-function relationship.

5.2 Introduction

En face OCT imaging is a relatively new approach to assess structural modifications in glaucoma. As detailed in previous chapters, defects within this examination method appear as hypo-reflective regions of the retinal area assessed, likely from a combination of primary loss of RNFB reflectance and thinning of the RNFL (Hood et al., 2015, Ashimatey et al., 2018a). Analyses of reflectivity have shown the potential for early detection of glaucoma (Huang et al., 2011, Liu et al., 2014, Fortune et al., 2013), which remains imperfect with conventional OCT metrics (Michelessi et al., 2020). Also, the ability to provide a direct topographical relationship with functional testing enables assessment of the structure-function relationship independently from VF to ONH maps (see section 1.5b), removing one source of error (Denniss et al., 2018).

As reported in chapter 3 and in concordance with previous studies on en face OCT imaging, changes of RNFB reflectance have a strong relationship with conventional cpRNFL thickness analyses (Ashimatey et al., 2018a, Cheloni et al., 2021b). Good correspondence with VF loss has also been reported (Alluwimi et al., 2018a, Sakamoto et al., 2019, Iikawa et al., 2020, Ashimatey et al., 2021). However, reflectance changes in most of the studies above have been typically evaluated subjectively by clinicians, as has the relationship between en face abnormalities and other tests such as cpRNFL and VF. Indeed, the novelty of en face OCT analysis in glaucoma means there is a lack of widely accepted objective methods for defining defects.

In chapter 4 we introduced SMAS, a novel method for automated and objective extraction of RNFB reflectance defects in en face images (Cheloni et al., 2021a). To further validate this method, this study aimed to assess how extracted en face findings relate to more conventional measures of glaucoma such as structural changes of cpRNFL thickness and VF loss measured by perimetry.

5.3 Methods

Settings and participants

This case-control study included glaucoma participants with early to moderate vision loss, as defined by a SAP MD better than -8dB, and age-similar healthy controls. Participants underwent eye examination including refraction, slit-lamp examination, GAT, and SAP (Humphrey VF, 24-2 SITA Standard). Conventional OCT imaging was performed with Spectralis OCT and included analysis of the ONH and macula, enabling the automated localisation of the fovea and the centre of the ONH. cpRNFL thickness data along a 3.5mm diameter circle scan (768 A-scan points) were also extracted. Beside conventional analyses, participants also underwent dense OCT imaging of the central retina, allowing us to extract en face images, and a custom perimetry test, whose details are provided below.

Details of eligibility criteria have been described in detail earlier (section 2.3). In brief, glaucoma participants required a diagnosis of open angle glaucoma from a consultant ophthalmologist and evidence of structural damage, defined as at least 1 abnormal sector (p<1%) on Spectralis cpRNFL analysis. No criteria based on VF defect were required for inclusion in the glaucoma group, as we aimed to include the earliest defects. Age-similar healthy controls were included if presenting no disease affecting their visual system and all participants were required to have clear ocular media or with or without history of uncomplicated cataract surgery.

En face defects of RNFB reflectivity

Details of OCT imaging and en face slab extraction have been presented earlier (see chapters 2, 3, 4). Briefly, multiple dense OCT scans were collected over the central ±25° of the retina and reflectivity data were extracted as single pixel slabs (3.87µm) containing depth-resolved attenuation coefficients (Vermeer et al., 2013). Single slabs were montaged and combined together according to the SMAS method, as described in section 4.3 (Cheloni et al., 2021a). For each participant, this modality of en face image construction results in 7 slabs with a 16µm thickness, overall encompassing the retinal depths from 8 to 116µm below the ILM. Reflectivity is evaluated on a

superpixel grid, by comparing individual data with normative data from age-similar healthy eyes, only at regions with expected RNFBs. Defects were considered when superpixel intensity fell below the 1st percentile (p<1%) of normative intensity, when occurring in any slab. To account for the individual anatomy, all slab images were adjusted for the participant's fovea-disc and fovea-raphe angles by using geometric transformations (double-shear transformation). Further, superpixel dimensions were individually adjusted to have a fixed number of superpixels (20) separating the fovea and the centre of the ONH across all eyes tested.

For the specific analysis of this study, deviation maps generated with SMAS, were firstly back-transformed to the original retinal configuration of each individual eye, by using the same parameters previously adopted in the double-shear transformation (fovea-disc and fovea-raphe angles). The geometric transformation resulted in a shift of the y axes computed as:

$$x_1 = x_0 \qquad y_1 = y_0 + |x_0| * tan(alpha)$$

where x_1, y_1 and x_0, y_0 refer to coordinates of the transformed and original deviation maps respectively, alpha corresponds to the fovea-raphe angle in the temporal retina and to the fovea-disc angle in the nasal retina. The use of the absolute value was required to account for negative x coordinates in temporal retina. An example is shown in Figure 5.1. Subsequently, the dimension of deviation maps was converted from superpixels to degrees, by using the estimated ONH location in degrees as extracted from Spectralis software. This distance is computed by the software from optic disc and macular scans, making use of the automated identification of the fovea and the ONH centre, as defined by the Bruch's membrane opening. The horizontal distance of the ONH centre from the fovea was divided by the number of pixels separating the two structures as subjectively identified in en face images. The distance in pixels rather than superpixels was considered because superpixel dimension was previously rounded, to achieve a fixed number of 20 superpixels between the fovea and the centre of the ONH (see section 4.3). Eventually, the dimension in degrees of a single pixel was used to identify superpixel diameter, by multiplying by the individual-specific number of pixels

in each superpixel. Superpixels had a median diameter of 0.7° (range 0.67° to 0.85°). Conversion to degrees was set to correspond to the centre of each superpixel.
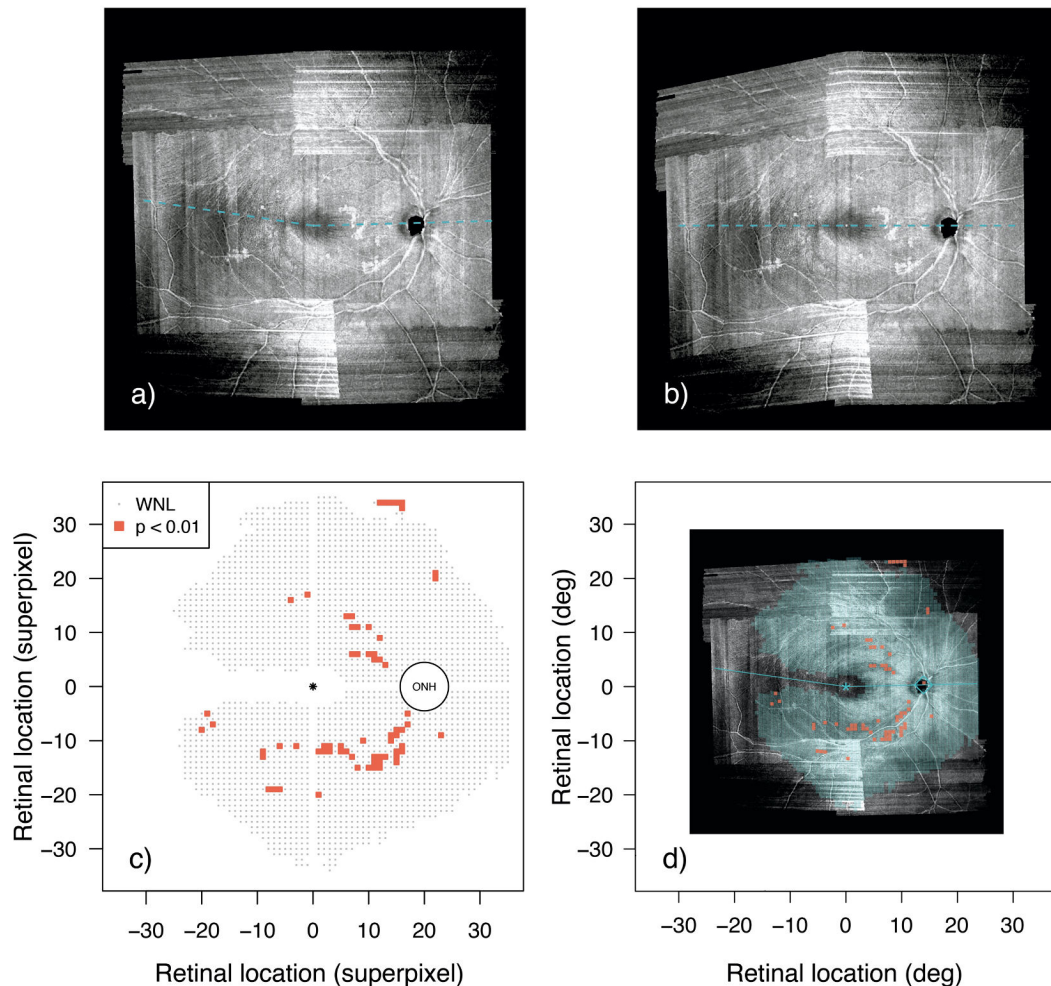
Figure 5.1 Example of the double shear transformation applied within SMAS and conversion from superpixels to degrees for one eye with glaucoma. In (a), the original en face image (8-23μm below ILM) is reported, blue dashed lines are drawn along the fovea-raphe and fovea-disc angles. In (b), the image transformed according to SMAS is showed, and raphe, fovea and ONH are aligned on a common horizontal axis. In (c) the SMAS deviation map is shown, whereas (d) shows the map converted from superpixels to degrees and transformed back in the original retinal configuration (1 superpixel = 0.68°).

Circumpapillary RNFL thickness defects

To identify abnormalities of cpRNFL thickness, deviation maps were generated by comparing measurements of each glaucoma eye with age-adjusted normative data from healthy controls. cpRNFL thickness measurements, automatically segmented by the device software, were extracted from Spectralis OCT after verification of segmentation by one of the authors (RC). Firstly, cpRNFL thickness measurements of all participants were

adjusted for the individual fovea-disc angle as automatically measured by Spectralis, allowing us to set the 0° in correspondence with the fovea-disc axis. Thickness measures at different angular locations in healthy eyes were consistent with normal distributions, as evaluated by visualisation of histograms and Shapiro-Wilk tests (all p>0.05). While determining normative cpRNFL thickness data, age was the only covariate for which data were adjusted, as per current clinical practice (Realini et al., 2015). Since our sample size for linear regression was limited (Green, 1991), age-related changes of cpRNFL thickness were extracted from a German population-based study evaluating a large cohort of healthy eyes with the same OCT device and scan pattern adopted here (Wang et al., 2017). That study provides point-wise estimates of cpRNFL thickness changes induced by age (µm/year). The difference between individual participant's age and the average age of controls was firstly computed and used to identify correcting factors at each of 768 cpRNFL thickness points. These values were then used to correct normative cpRNFL thickness from our controls at all angular locations. The 1st percentile (p<1%) from our controls was used to define the lower limit of normality of cpRNFL thickness. This was computed by subtracting 2.33 SDs from the mean thickness of healthy controls at each of 768 points, after age adjustment.

Eventually, to facilitate the analysis of correspondence with en face defects and to support the required mapping arrangements (see below), individual deviation maps were rotated by their fovea-disc angle. This allowed us to set the 0° temporally, consistently with the structure-function map used to project en face defects at the ONH.

VF examination

On the same day or within maximum 30 days from OCT imaging, participants underwent SAP (24-2, SITA Standard) and a dense supra-threshold custom perimetry test, implemented via an Octopus 900 perimeter (Haag-Streit AG, Koeniz, Switzerland). This platform allows a direct interaction with the VF strategy via computer programming, through the Open Perimetry Interface (Turpin et al., 2012).

Custom perimetry was run over the 24-2 grid limits with 6 extra locations in the nasal step area (i.e. [-27, ±5]; [-27, ±7]; [-25, ±7], considering a right eye) and adopted a varying density, overall greater than SAP (Figure 5.2). As suggested recently (Numata et al., 2017), a 1° density was used in the central 10°, whereas a slightly coarser presentation pattern (2°) was used in the mid-periphery. In addition, a dense (1°) peri-papillary annulus, located in proximity of the estimated position of the blind spot, was included in the grid (Figure 5.2). A white on white paradigm was used, presenting size III Goldmann stimuli (0.43°) on a 10cd/m^2 intensity background with duration of 200ms and allowing up to 1500ms for the observer's response. Presentations were made in the Humphrey dB scale, achieved by setting the 0dB stimulus of the perimeter (maximum intensity) to 10000asb. The Octopus 900 perimeter uses a movable projector to present stimuli at varying locations. By default, the projector is moved immediately prior to the stimulus presentation, with the potential for a cueing effect from the projector's noise. To minimise this risk, a stimulus falling outside the individual perceivable range (46dB) was presented immediately before the actual stimulus at the same location. The 'invisible' stimulus had a duration of 200ms and a variable response window randomly sampled from a set of values between 1 and 1000ms, where shorter periods of time were more likely than longer ones (Appendix A, Supplementary Material 5.1). Hence, a variable amount of time was interposed between the projector's noise and the actual presentation of the stimulus, most frequently (~60%) being below 500msec.

Custom suprathreshold perimetry used a 1-2 pass/fail criteria, presenting a second stimulus at locations where the first one was missed, and considering a defect only in locations where neither of the two produced a response. This criteria might provide similar specificity levels with fewer presentations compared to a 1-3 criteria (Turpin et al., 2016). Stimulus intensity was selected according to age-adjusted and location-specific normative sensitivities, derived from a previous study testing healthy participants with the Humphrey VF's Full Threshold 30-2 test (Heijl et al., 1987). Estimates of retinal sensitivity at locations falling outside the conventional 24-2 grid were computed with spatial interpolation (universal kriging), as proposed elsewhere (Denniss and

Astle, 2016). This technique allowed us to fit surfaces to normative perimetric data, resulting in estimates for locations not tested in the original data set. To establish the age-adjusted normative values as well as the variability of the response at each location, three surfaces were fitted adopting the *MASS* and *Spatial* libraries in *R* (Venables and Ripley, 2002): i) the mean normal sensitivity for 50 year old observers; ii) the lower boundary of normality (2 SDs below the 50 year old mean); and iii) the average age-related loss (dB/year) at each location. For each participant and at each location, stimuli 2dB below the age-adjusted lower bound of normality were presented, i.e. 2 SDs plus 2dB below the age-adjusted normative sensitivity. This approach aimed to increase specificity in spite of a potentially reduced sensitivity (Henson et al., 1999, Henson and Artes, 2002), which is an encouraged practice in glaucoma diagnosis oriented to minimise the false positive rate (Saunders et al., 2015, Trevethan, 2017, Fallon et al., 2017, Mwanza et al., 2018a).
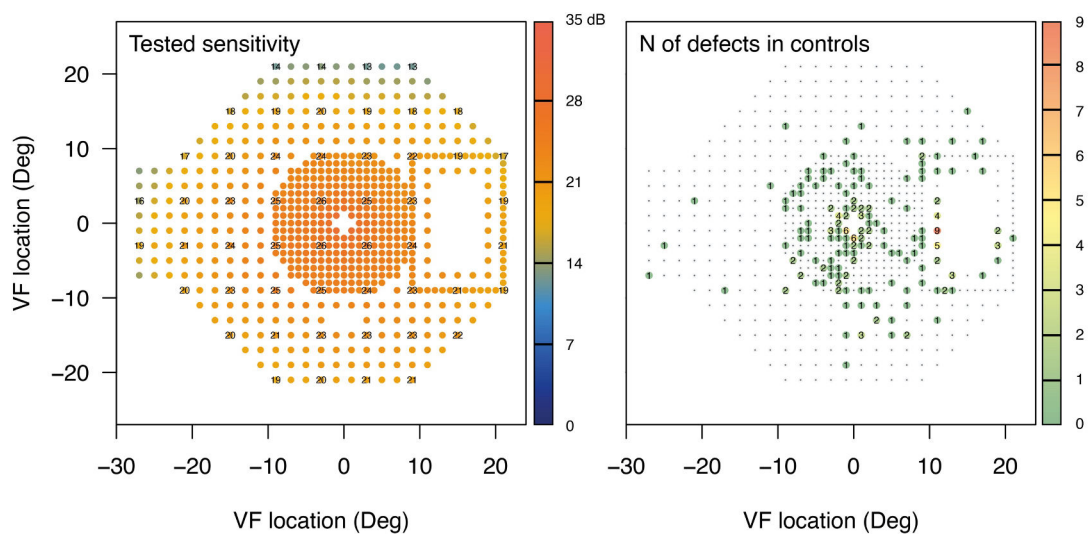
Figure 5.2 Details of the custom perimetry strategy used in this study, for a right eye. The left-hand side plot shows the average sensitivities (dB) tested at each location in all controls (n=29). Points are color-coded according to sensitivity (dB) and the value is reported only at locations corresponding to 24-2 grid. Consistently with the narrow age range of this sample, tested sensitivities showed small variability (mean range across all 643 locations: 1.5dB, ranging from 1 to 3dB). The raw number of VF locations found abnormal in all controls is reported in the right-hand side plot, and colour coded accordingly. The mean number of defects across all locations was 0.3 (range 0 to 9), corresponding to a frequency of 1% (range 0% to 31%). Each individual control presented a mean of 7.2 defects (range 0 to 69; removing one outlier: mean 5 defects, range 0 to 21).

Overall, 643 locations were tested with custom perimetry. The test was accomplished in a single day, and broken down into 8 different sub-grids (~80 locations each), executed in a random order. Each sub-grid roughly required 3-6 minutes to be completed, resulting in an overall 40-70 minutes to

accomplish the whole test, including pauses. To minimise fatigue effects, short breaks between each sub-grid were allowed, as well as 2 to 4 longer breaks, as required. Nevertheless, we expect any fatigue and learning effects to be evenly distributed across the VFs of participants since grids were randomised. Although most of the participants were familiar with VF testing, and underwent SAP examination first, one training session on a random sub-grid of custom perimetry was performed to minimise a potential learning effect with a different instrument.

For both perimetric procedures only reliable tests were included, firstly assessed through subjective observation of fixation stability from the internal fixation monitor (Patel et al., 2015). In addition, reliability indices (false positives, false negatives and fixation losses) were required to be below 20% in all SAP tests. For custom perimetry, stimuli outside the visible range presented throughout the whole procedure served as a surrogate measure of the false positive rate. Only tests with rate below 20% were included.

Custom perimetry provides a dichotomous classification of any VF location as normal or defect. Accordingly, a threshold level was applied to SAP results to obtain a similar dichotomous outcome. For consistency with en face analysis and custom perimetry, SAP findings were also considered by evaluating total deviation values. The level of significance selected to establish VF defects in custom perimetry (2 SDs and 2dB below the mean) approximately corresponds to a probability of 2.3%. Accordingly, SAP locations were considered to be abnormal when showing a total deviation with p<2% or less.

Structure-structure and structure-function maps

To map reflectance defects in en face images to corresponding angular locations around the ONH, a computational model for structure-function mapping was used (Denniss et al., 2012, Denniss et al., 2014a, Denniss et al., 2014c, McKendrick et al., 2017, Turpin and McKendrick, 2021). As detailed in section 1.5b, this produced mapping specific for the individual anatomy of the eye assessed and builds on axonal growth models rather than on hand-tracing of RNFBs. As required by our experiment, this model allowed us to map any location in the retina, and not only those corresponding to a specific VF grid,

to specific angles on the ONH (Denniss et al., 2012). The mapping model is fed by several parameters, some of which were participant-specific, such as the ONH location, measured in degrees from the fovea, and the disc-fovea-raphe angle, measured in en face images (details provided in section 4.3). Additional parameters feeding the model were passed as default values, as suggested by the authors (Denniss et al., 2018, Denniss et al., 2014a). These included the ONH vertical and horizontal diameters (5.35° and 6°), total RGC count (1400000) and fovea radius (0.67°). Data were extracted considering a right eye in retina view and setting angular locations at the ONH as 0° temporally, 90° superiorly, 180° nasally and 270° inferiorly. Individualised maps were extracted specific to each participant's en face deviation map, identifying the corresponding angle of insertion on the cpRNFL profile for each abnormal superpixel (Figure 5.3).
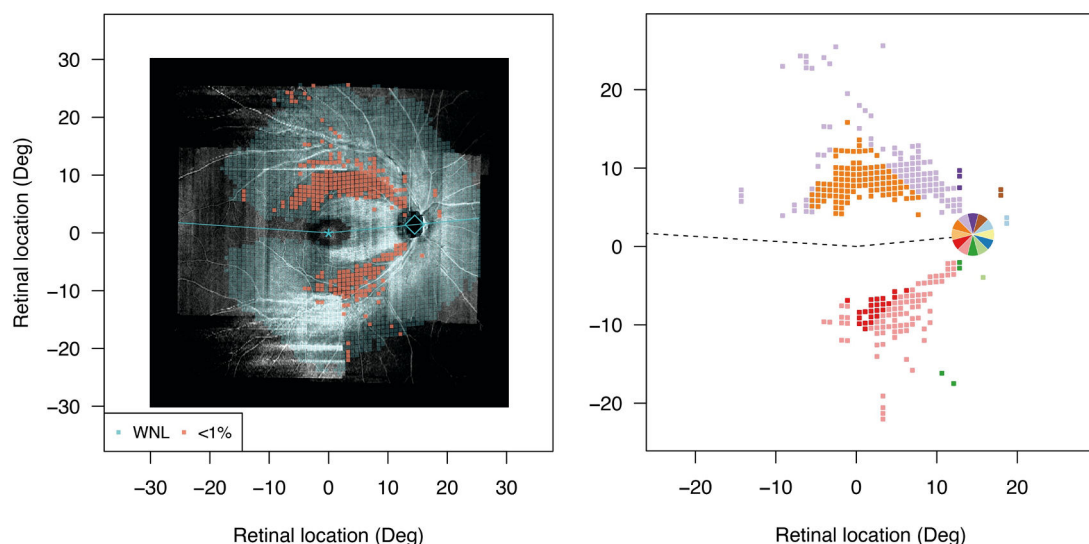
Figure 5.3 Example of the mapping model used in this study to relate en face reflectance defects to angular locations around the optic disc, in a glaucoma participant. The left-most plot shows the SMAS deviation map overlaid on an en face image of the corresponding eye. In the right-most plot, en face defects are color-coded according to the angular location of insertion around the ONH, in 30° sectors.

As reported previously, one advantage of en face imaging is the opportunity to explore the relationship with visual function by an almost direct overlap between reflectivity and VF findings. However, RNFBs are the main target of en face analysis, and these are known to be displaced from the served receptive field in the macula. Accordingly, VF locations were corrected for RGC displacement, according to the model proposed by Drasdo et al. (2007). Conversely, Drasdo correction was not applied to en face data when seeking

relationship with cpRNFL. Indeed, as reported above RNFBs are likely to be already away from the receptive field (e.g. absent in the fovea, and along the fovea-raphe axis – see Figure 5.1). Additional displacement, yet not the one pertinent to Henle's fibre modelled by Drasdo and colleagues, could be defined by the trajectory of the single axon emerging from an RGC and entering the closest bundle. However, we considered this additional source of displacement to be negligible for our purposes and possibly included within the level of tolerance applied (see below).

Data analysis

Consistently with chapter 4, the impact of artefacts on slab images was examined by two investigators (RC, JD), and in case of substantial influence, either whole images or specific regions were excluded from further analysis. Each case was jointly discussed by the authors until consensus on data exclusion was reached. To facilitate comparison and reporting, all en face deviation maps and VF data were transformed into right eye retina view format. En face defects localised within a 2° radius of the centre of the ONH were excluded from the analysis, as these are likely to fall within the ONH or on its margins.

To explore the relationship between reflectance defects and cpRNFL thickness loss (referred to later as structure-structure) we considered the recommendations of a recently proposed framework to assess concordance between glaucoma modifications in the retina and at the optic disc (Turpin and McKendrick, 2021). Accordingly, it should be considered that healthy locations in the central retina could project to abnormal cpRNFL sectors. Axons linking the peripheral retina might also enter the ONH at similar angular locations (i.e. peripheral damage and intact central retina), ultimately leading to cpRNFL defects. Hence, analysis should focus on damaged retinal locations, which should also constitute the starting point in the examination of this relationship. We considered structure-structure to be concordant when the 30° sector centred on the predicted angle of insertion on the ONH of any en face defect presented at least one abnormal cpRNFL thickness point (i.e. $p < 1\%$). A 30° sector (±15° on predicted angle) was selected according to estimates of

mapping variability, as defined by measurement errors of parameters seeding custom structure-function mapping (McKendrick et al., 2017, Denniss et al., 2014a). For each participant, an overall measure of topographic concordance was computed as the concordance ratio, being the proportion of concordant defects among all en face defects. Boot-strapped 95% CI were also computed (n=1000 resampling).

Secondly, we explored agreement between en face defects of reflectance and corresponding visual function in perimetry (referred to later as structure-function relationship). Consistently with structure-structure analysis, sources of variability within structure-function mapping were considered by allowing some degree of tolerance in the evaluation of concordance. For instance, imperfect fixation during perimetry could affect the spatial mapping of structure to function, leading to inconsistencies between the desired and actually tested locations on the retina. Differences between the fovea and the actual preferred retinal locus for fixation should be considered (i.e. fixation bias), since macular damage could result in non-foveal fixation, also in glaucoma (Montesano et al., 2018). Errors may also arise from fixation instability, which was recently modelled in gaze attractions from peripheral stimuli and random eye displacements (Montesano et al., 2021b). These authors used fundus-tracked perimetry and showed fixation bias in healthy eyes to be close to zero when removing the effect of gaze attractions. This latter component was found to be substantial in their healthy participants, possibly due to normal retinal function (more targets seen) and low experience with perimetry. In glaucoma eyes, age-adjusted projection errors due to eye movements were on average 0.89° (95% quantiles: 0.73° to 1.05°) for all components together, whereas 0.55° (95% quantiles: 0.41° to 0.68°) for random displacements only (Montesano et al., 2021b). One additional source of error in structure-function mapping may be linked to current RGCs displacement models. These indicate displacement to be largest within the macula, with peaks of 2-2.5° at 2-3° of eccentricity from the fovea (Drasdo et al., 2007, Turpin et al., 2015, Denniss et al., 2018), and small effects of individual anatomy (Turpin et al., 2015). As a measure of the variability in structure-function mapping from RGC displacement, the original data from the Drasdo model showed an average displacement of 0.601mm

(SD: 0.088mm) at locations with maximum lengths (Drasdo et al., 2007). Accordingly, 4 SDs of displacement should correspond to 1.2° (1°≈0.29mm), at the regions where displacement is maximum.

Based on the data presented above and the sample characteristics, a ±1° tolerance was used for this analysis. In fact, in view of the early moderate-defect of the included glaucoma participants, we speculate that central fixation was relatively preserved. In addition, participants' previous experience with perimetry and requirements from reliability criteria may have contributed to avoid significant gaze attractions. As such, structure-function agreement was established between any VF location and the en face superpixels lying within 1° (Figure 5.4). Only en face superpixels with nearest displaced VF location within 1° were considered and used to generate a dichotomous prediction of VF status. Similarly, VF locations with no en face superpixels within the tolerance distance were ignored. Function was predicted to be abnormal if at least 1 superpixel was found to be abnormal within the relevant cluster of each VF location, for SAP and custom perimetry.
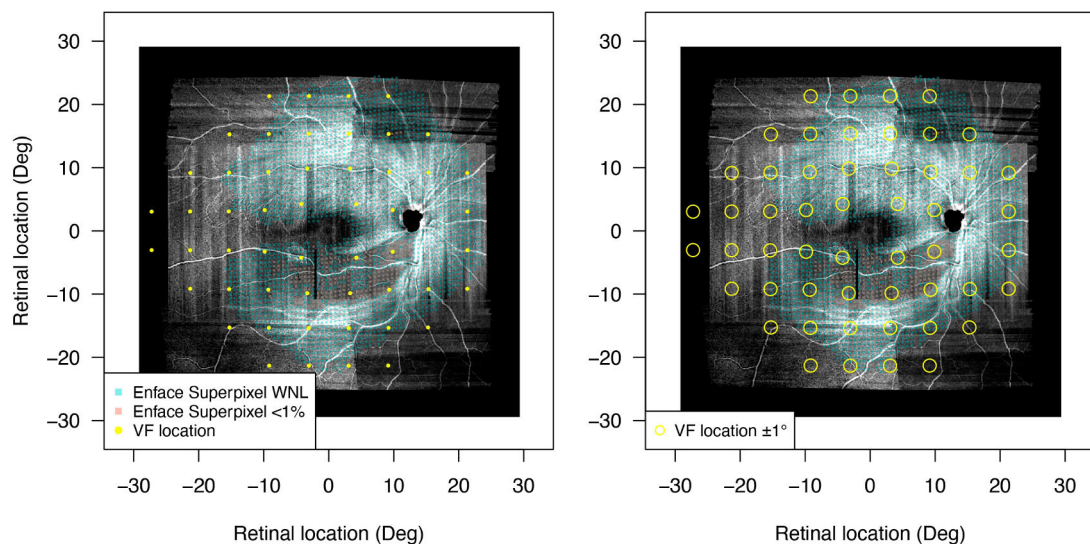
Figure 5.4 Example of tolerance level considered in the analysis of structure-function agreement between en face reflectance and visual function in SAP. The left most image shows displaced VF grids in Goldmann size III stimuli on an en face image and deviation map. The right most image shows the same 24-2 locations, including the ±1° tolerance around VF locations. VF locations (yellow circles) with no en face superpixels underneath were excluded from analysis.

Unlike the evaluation of the structure-structure relationship, analysis of concordance between en face findings and corresponding function considered both defects and healthy regions. As a general measure of concordance we

used overall raw agreement and tetrachoric correlation (Drasgow, 2014, Olsson, 1979), which were computed for each individual eye. Tetrachoric correlation was computed with the *Correlation* package in *R* (Makowski et al., 2020), and provides *rho* coefficients between 2 dichotomous variables. Unlike K statistics, tetrachoric correlation is not affected by the marginal proportions of rating levels (Uebersax, 1987). This method has been proposed to evaluate agreement, and has the advantage of providing a familiar effect size within the field of structure-function relationships in glaucoma. The main assumption of tetrachoric correlation requires underlying distributions of data feeding dichotomous categorisation to follow a continuous distribution, which can be assumed to be held for the testing modalities adopted here.

Although raw agreement (measured as [a + d] / N, as in Table 5.1) represents an intuitive descriptive statistic, it does not provide insights into differential agreement from positive and negative ratings. Therefore, positive and negative conditional agreements were also computed. Positive agreement represents the conditional probability that given a prediction of abnormality from one test, the other test also showed the location to be a defect. Similarly, negative agreement measures the probability of a concordant measure of normality. The two measures of agreement are adjusted for chance and also for uneven distributions of ratings (i.e. prevalence) across different individuals (Cicchetti and Feinstein, 1990). The following equations, referring to Table 5.1, were used (Cicchetti and Feinstein, 1990, Spitzer and Fleiss, 1974): positive agreement: 2a / [2a + b + c]; and negative agreement: 2d / [2d + b + c].

Table 5.1
Data arrangement for dichotomous results of VF and en face analysis, for a given participant. Dealing with dichotomous estimates of normal/abnormal VF and en face reflectance, results were arranged in a 2x2 contingency table.

	Perimetry		
En face OCT	*Defect* (+)	*Normal* (-)	*Total*
Defect (+)	a	b	a + b
Normal (-)	c	d	c + d
Total	a + c	b + d	N = a+b+c+d

For each participant, estimates of overall, positive and negative agreement were computed alongside their boot-strapped 95% CI. Fisher's exact test was

used to test independence of the two test modalities (Uebersax, 1987). Overall measures of correlation were provided as median and range, the former being computed after transformation of correlation coefficients to Z values to measure the median, and then converted back to correlation coefficients.

Lastly, the ability of en face analysis to predict visual function was also evaluated by considering VF results as the reference standard and reporting positive and negative predictive values. These measures provide an estimate of the likelihood of a VF defect given an abnormal en face estimate and a normal VF finding given normal en face measures, respectively. Although positive and negative predictive values are easy to interpret, they are affected by the 'prevalence' of the disease (i.e. percentage of VF defects), with greater positive predictive values for higher prevalence, and vice-versa (Altman and Bland, 1994). Since frequency of defects changed across participants, positive and negative likelihood ratios were also reported, for completeness. These latter measures determine how much the odds of presenting a VF defect (or a normal finding) increase when en face analysis showed a defect (or a normal result). Positive likelihood ratios range from 0 to infinite (the greater the better, 0 being the worst predictor), whereas negative likelihood ratios have the same range, but the smaller the value, the better the predictor, with 0 being the ideal performance (Hajian-Tilaki, 2013, Trevethan, 2017). Diagnostic accuracy indices were computed using the *epiR* package in *R* (Stevenson and Sergeant, 2021). 95% CI for positive and negative likelihood ratios were based on formulae provided by Simel et al. (1991), whereas for predictive values by Collett (2002).

5.4 Results

Overall, 29 controls were included, among which 22 had usable en face images to establish normative levels of RNFB reflectance. Among 21 participants with early-moderate glaucoma imaged with en face OCT, 5 (median age: 69, range 67 to 78; median MD: -4.8dB, range: -6.1 to -1.6dB) were excluded because of substantial artefacts affecting the final images. Also, part of the images of 3 glaucoma participants were censored for similar reasons (refer to section 4.4 for details). Overall, findings from 16 glaucoma participants were included in the analysis, whose demographics are reported

in Table 5.2. All participants with glaucoma but one presented a VF defect according to the criteria used for the inclusion of healthy controls (section 2.3). The remaining eye had three contiguous defective points but one was an edge location.

Table 5.2
Demographics and clinical characteristics of included participants. Continuous data are summarised as median and (interquartile range).

	Controls		Glaucoma
	cpRNFL norms	En face norms	
N	29		16
	29	22	
Age (y)	69 (8.0)	67.5 (4.0)	70 (8.3)
Eye: R/L	15/14	14/8	4/12
Ethnicity: Caucasian/Other	28/1	21/1	16/0
Global cpRNFL thickness (µm)	97 (14)	97.5 (11.5)	68 (14)
SAP MD (dB)	0.5 (1.4)	0.8 (1.4)	-3.3 (2.2)
Axial Length (mm)	23.5 (0.7)	23.3 (0.6)	24.05 (0.95)

Structure-structure concordance

Among glaucoma participants, a median of 322 abnormal superpixels (IQR: 181 to 463) were identified in each eye, out of an average of 2492 superpixels tested (~12.9%).

Objectively extracted reflectance defects showed a strong concordance with conventional cpRNFL thinning, with a median concordance ratio of 0.85 (IQR: 0.74 to 0.95). Data suggest that on average 85% of en face defects in this sample had a cpRNFL location with thickness below the 1st percentile within ±15° of the predicted insertion on the ONH. Figure 5.5 shows details of concordance for each individual eye. The smallest concordance ratio was 0.48 (95%CI: 0.36 to 0.6, ID#1 in Figure 5.6) whereas the eye showing the highest concordance between en face and cpRNFL was ID#13 (0.99, 95%CI: 0.99 to 1.0), where nearly all en face defects were matched by thinned cpRNFL.
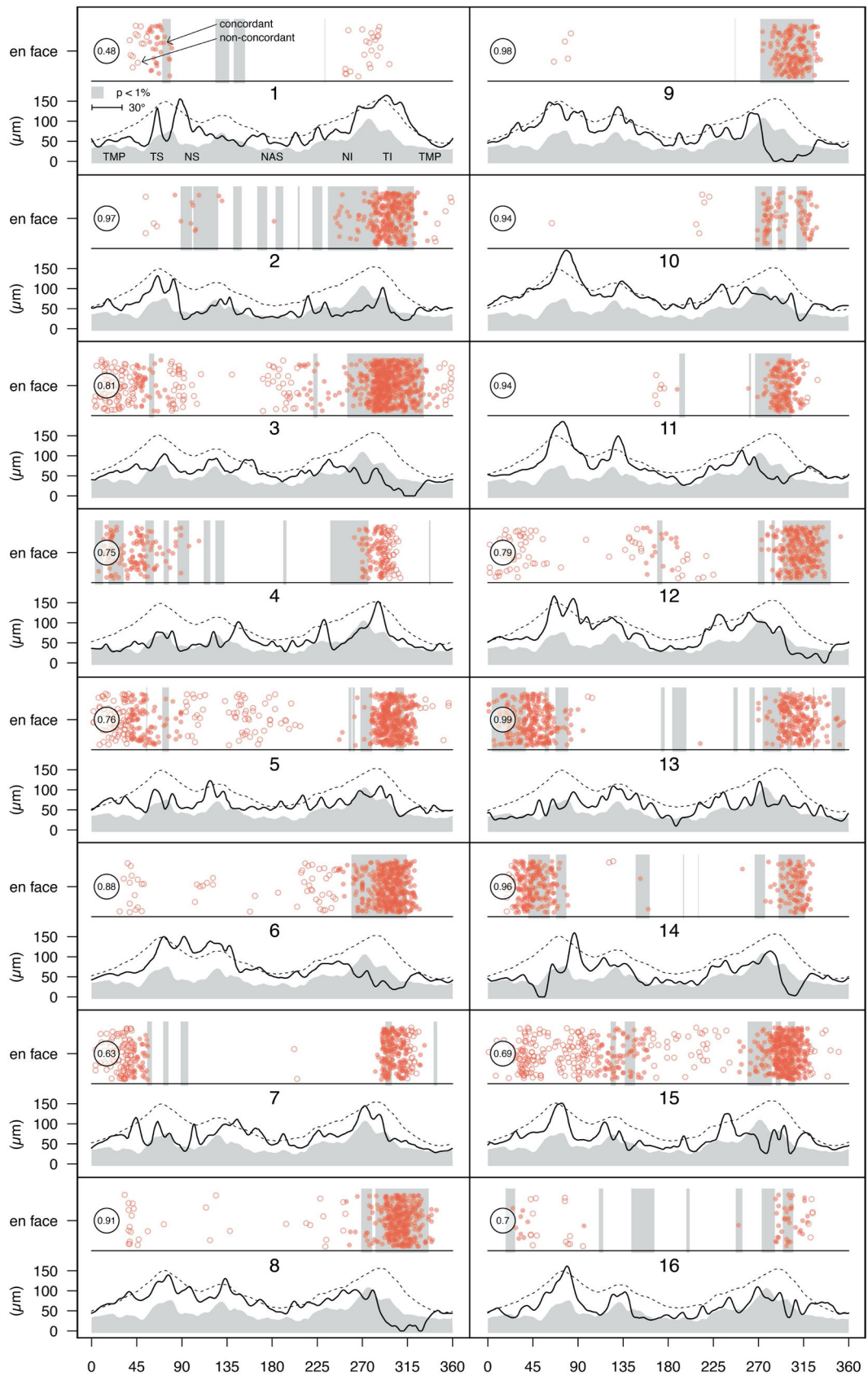
Figure 5.5 Concordance findings for all glaucoma participants between en face defects and cpRNFL thickness. For each participant the bottom of the panel reports the individual cpRNFL thickness (black solid line) compared to age-adjusted normative data (black dashed line) and lower limit of normality (p<1%, grey area). cpRNFL thickness defects (i.e. <1%) are found where cpRNFL thickness falls within

the grey area and are also reported in the top of each panel as grey rectangles. En face defects found in the retina as mapped to the ONH (red circles) are reported at the corresponding angle along the cpRNFL profile. Filled and empty points show concordant and discordant defects, respectively. Concordance ratio for each participant is also reported in the top left corner of each subplot.

Distributions of en face angular projections on the ONH are reported in Figure 5.6, and en face defects most frequently mapped to the superior temporal (40-50°) and inferior temporal (300°) aspects of the ONH.
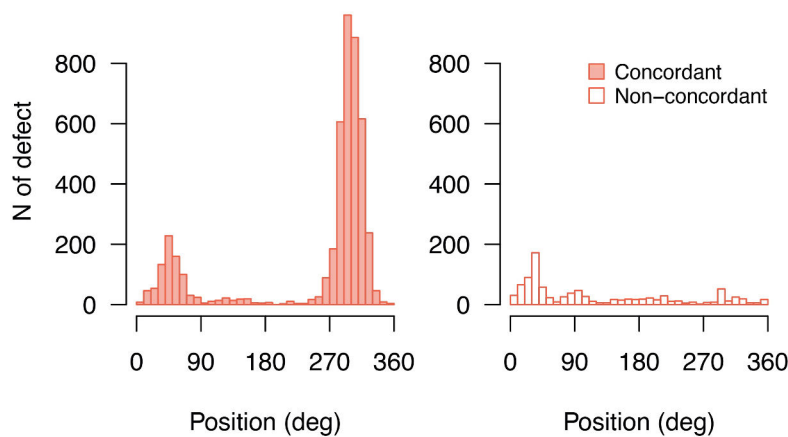
Figure 5.6 Distributions of angular insertions on the ONH as mapped by the model adopted in this study (Denniss et al., 2012). Data are shown separately for concordant en face defects (left-most plot, mode=297°) and abnormal superpixels without corresponding cpRNFL defect (right-most plot, mode=38°).

Structure-function concordance (Custom perimetry & SAP)

Analysis of SAP results was limited to 15 glaucoma eyes, since one of the participants (ID#16) had abnormal locations only in the nasal step area, where no en face predictions were made (see Figure 5.4 for example). As detailed in the methods, SAP locations showing total deviation with p<2% or below were considered abnormal. Dichotomised SAP results were strongly correlated with custom perimetry findings at corresponding locations (tetrachoric correlation, median rho: 0.72, range 0.2 to 0.94). Scaling the significance level of SAP dichotomisation from more liberal (i.e. defect if total deviation with p<5%, or below) to more conservative (i.e. p<1% and p<0.5%) showed a small increase in the strength of correlation: rho 0.76 (range 0.3 to 0.88), 0.79 (range 0.15 to 0.95), and 0.85 (range -0.03 to 0.97), for p<5%, p<1% and p<0.5%, respectively.

The tolerance level selected in the structure-function agreement analysis led to a median of 5.8 (IQR: 5.5 to 6.1) superpixels being considered in each cluster used to make a prediction of visual function from en face data. The two VF modalities were superseded by grids with different spatial densities,

resulting in a different number of VF locations analysed. In SAP, a median of 39 locations per participants (IQR: 37.5 to 40.5) were considered, with the value raising to 543.5 (IQR: 531.4 to 555.6) for custom perimetry.

Findings from agreement analysis are reported in Table 5.3 and summarised in Figure 5.7. Firstly, all glaucoma participants showed at least 1 location where perimetry (both SAP and custom perimetry) and en face analysis were concordant on the presence of a defect. In SAP there were a median of 3 concordant damaged locations (range 1 to 9), whereas in custom perimetry there were on average 82 concordant defects (median 82, range 20 to 159).

Table 5.3
Findings from agreement analysis for SAP and custom perimetry, respectively. Values of overall, positive and negative agreements are reported alongside their bootstrapped 95% CI. *P* values for Fisher exact tests for independence of the two test modalities in the overall agreement analysis are also reported.

ID	SAP (defect if TD with p<2%)				Custom perimetry			
	Overall	*Pos*	*Neg*	*P value*	*Overall*	*Pos*	*Neg*	*P value*
1	0.69 0.58, 0.82	0.36 0.11, 0.62	0.79 0.70, 0.91	0.021	0.76 0.72, 0.79	0.22 0.15, 0.30	0.85 0.83, 0.88	<0.0001
2	0.55 0.40, 0.71	0.24 0.01, 0.48	0.68 0.55, 0.83	0.72	0.61 0.56, 0.65	0.42 0.36, 0.48	0.70 0.66, 0.74	0.004
3	0.58 0.42, 0.74	0.53 0.33, 0.76	0.62 0.46, 0.82	0.15	0.64 0.60, 0.68	0.62 0.57, 0.67	0.66 0.62, 0.71	<0.0001
4	0.51 0.35, 0.68	0.25 0.01, 0.50	0.64 0.50, 0.81	0.71	0.57 0.53, 0.61	0.40 0.33, 0.46	0.66 0.62, 0.70	0.15
5	0.42 0.27, 0.58	0.10 0.01, 0.19	0.58 0.42, 0.77	0.58	0.44 0.40, 0.48	0.20 0.14, 0.26	0.57 0.53, 0.62	0.89
6	**0.85** 0.76, 0.98	0.67 0.45, 1.00	0.91 0.84, 1.00	0.001	0.83 0.80, 0.87	0.72 0.66, 0.78	0.88 0.86, 0.90	<0.0001
7	0.63 0.47, 0.79	0.12 0.01, 0.25	0.77 0.65, 0.91	1	0.58 0.54, 0.63	0.21 0.14, 0.28	0.72 0.68, 0.76	1
8	0.79 0.68, 0.92	0.43 0.14, 0.86	0.87 0.79, 0.98	0.048	0.85 0.82, 0.88	0.73 0.68, 0.79	0.89 0.87, 0.92	<0.0001
9	0.82 0.71, 0.95	0.46 0.15, 0.92	0.89 0.82, 0.98	0.063	**0.92** 0.89, 0.94	0.82 0.78, 0.88	0.94 0.93, 0.96	<0.0001
10	0.76 0.63, 0.90	0.17 0.01, 0.33	0.86 0.78, 0.96	1	0.74 0.70, 0.77	0.39 0.31, 0.47	0.83 0.80, 0.86	<0.0001
11	0.55 0.32, 0.73	0.38 0.05, 0.75	0.64 0.46, 0.89	1	0.68 0.63, 0.73	0.29 0.19, 0.40	0.79 0.75, 0.83	0.009
12	0.77 0.64, 0.90	0.40 0.13, 0.80	0.86 0.77, 0.95	0.024	0.76 0.73, 0.80	0.57 0.50, 0.64	0.84 0.81, 0.87	<0.0001
13	0.63 0.49, 0.78	0.48 0.28, 0.73	0.72 0.60, 0.88	0.30	0.51 0.47, 0.55	0.53 0.48, 0.58	0.49 0.43, 0.54	0.49
14	0.71 0.57, 0.86	0.40 0.13, 0.69	0.81 0.72, 0.94	0.21	0.69 0.65, 0.73	0.54 0.48, 0.60	0.77 0.74, 0.81	<0.0001
15	0.62 0.48, 0.80	0.48 0.28, 0.74	0.71 0.57, 0.87	0.3	0.63 0.59, 0.67	0.47 0.41, 0.53	0.72 0.68, 0.76	<0.0001
16	-	-	-	-	0.72 0.68, 0.76	0.23 0.15, 0.32	0.83 0.80, 0.86	0.074

As shown in Table 5.3 and Figure 5.7, there was a moderate-good overall agreement between SAP results and corresponding estimates of visual function from en face analysis (median: 0.66, IQR: 0.57 to 0.77). Stronger agreement was found between concurrently healthy locations (negative agreement) compared to abnormal areas in both test modalities. In fact, on average, 40% of damaged locations with a given test had a corresponding defect with the other test modality (median positive agreement: 0.40, IQR: 0.24 to 0.47), whereas the likelihood of both tests corresponding on healthy locations was 77% (median negative agreement: 0.77, IQR 0.66 to 0.86). Overall, en face analysis and SAP showed a moderate correlation (median rho: 0.34, range -0.29 to 0.81).

Slightly better results were observed in the agreement between custom perimetry and en face predictions of visual function. On average, 68% of locations had a corresponding prediction of defect or normality (median overall agreement: 0.68, IQR: 0.60 to 0.76). Consistently with SAP analysis, agreement on defects was poorer compared to agreement on healthy locations (median positive agreement: 0.44, IQR 0.27 to 0.58; median negative agreement: 0.78, IQR: 0.69 to 0.84). Correlation between custom perimetry findings and reflectance status across participants was on average stronger compared to SAP analysis (median rho: 0.51, range -0.02 to 0.95).
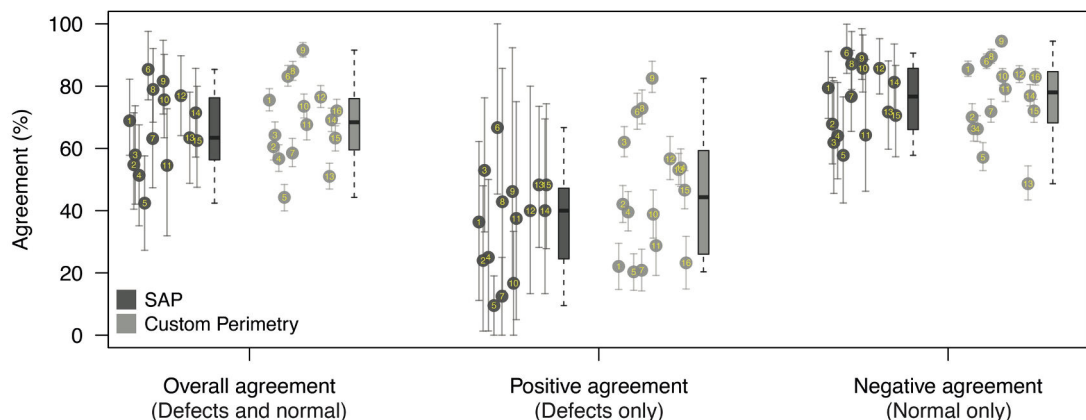
Figure 5.7 Findings from structure-function agreement analysis for SAP and custom perimetry. Overall agreement, positive agreement (both tests showing a defect), and negative agreement (both tests showing a location within normal limits) were computed. Each data point represents a participant (ID number), and error bars show 95%CI. Data are grouped for the type of agreement and colour-coded according to VF test. To summarise measurements, box-plots are reported next to each group of data points, and computed considering the estimate for each participant eye.

Structure-function concordance between SAP and en face analysis for each individual eye is shown in Figure 5.8. Findings for custom perimetry are provided as supplementary material (Appendix A, Supplementary Figure 5.2).
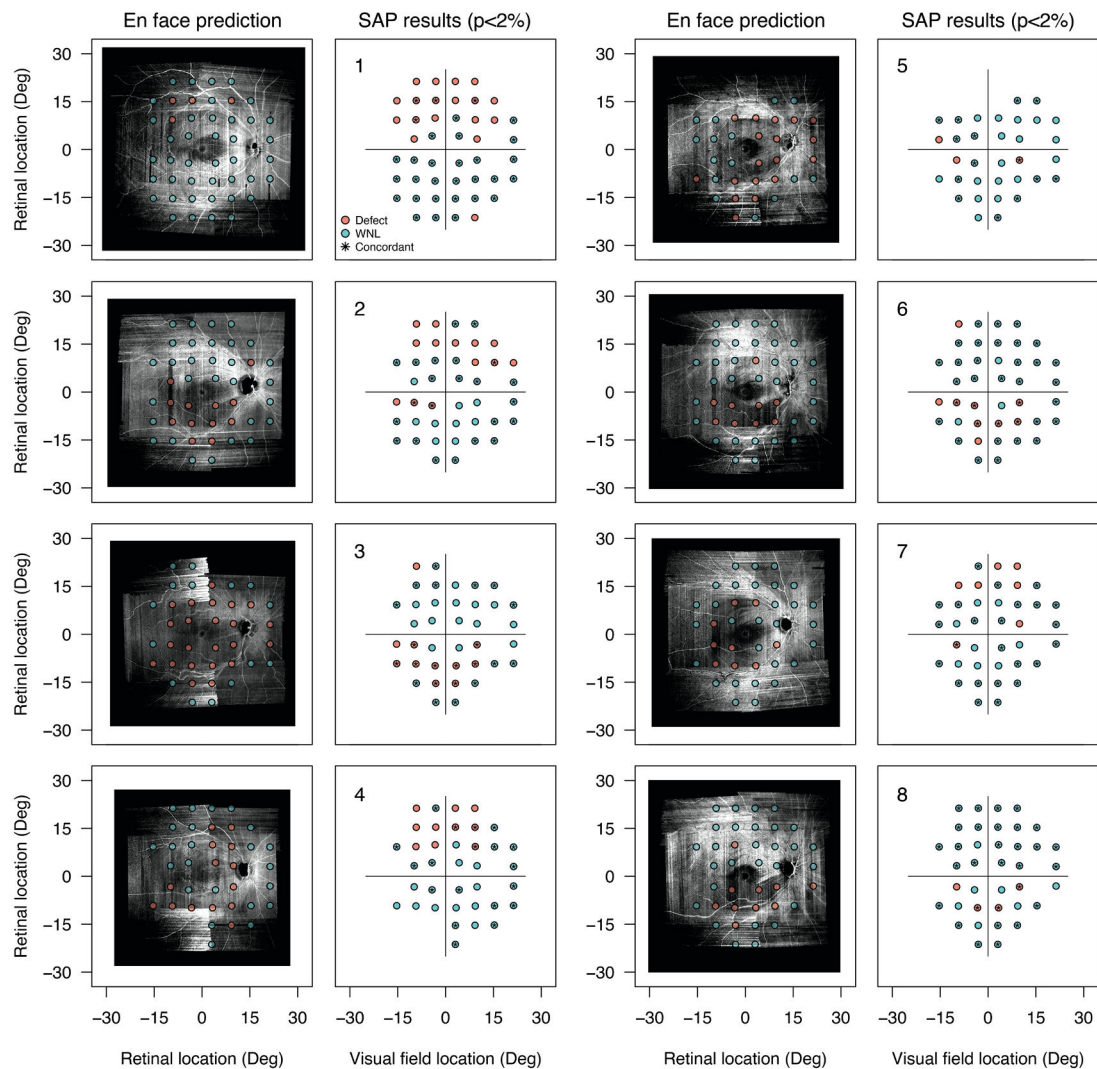
Figure 5.8 Correspondence between en face predictions of visual function and SAP dichotomised results. En face predictions of VF status are overlaid on the en face image of the patient; locations are colour coded as the observed status for each test domain and locations where there was en face-SAP agreement are flagged with asterisks (figure continues on to the next page).
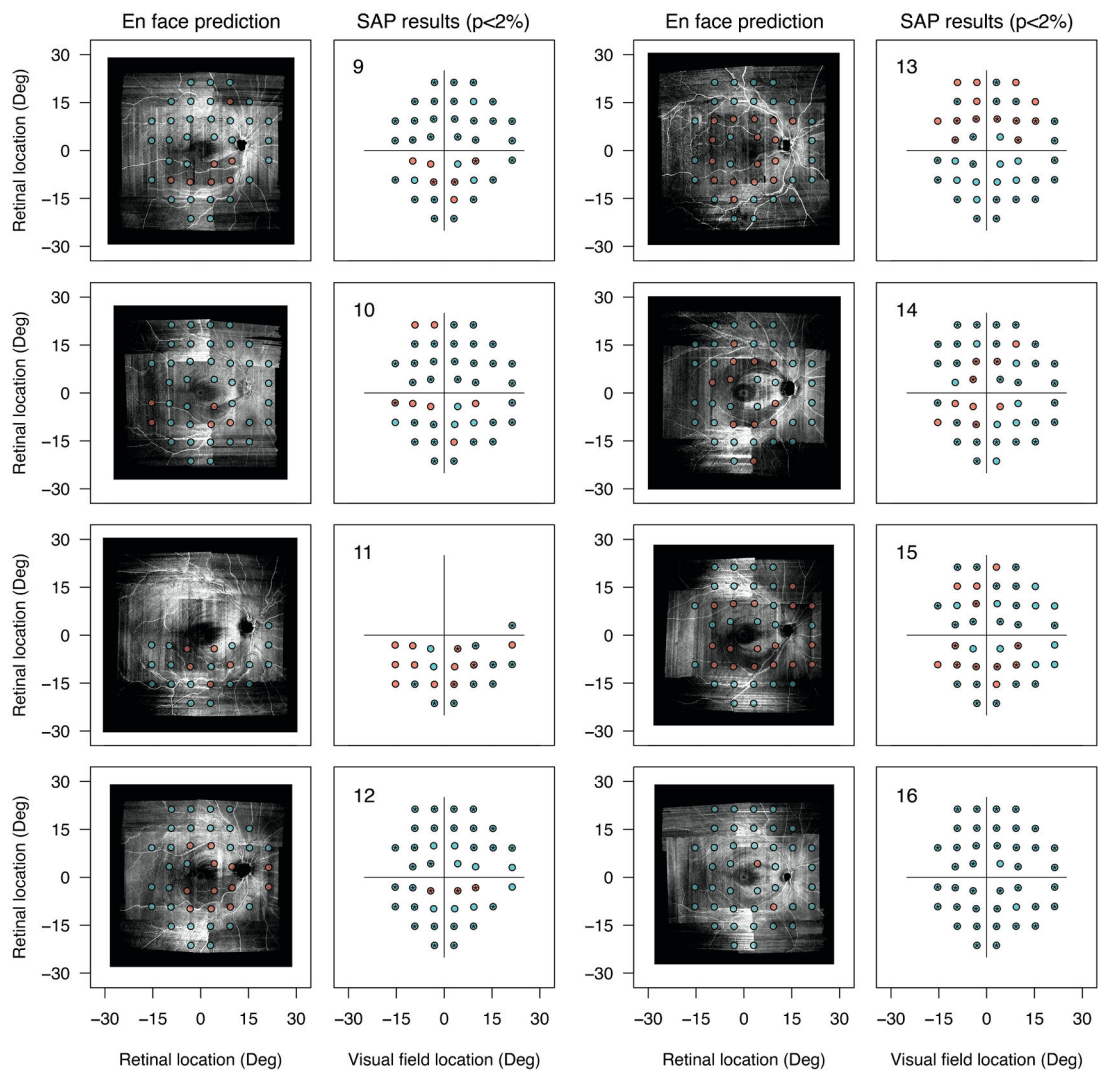
Figure 5.8 Continued.

Findings from predictive values analysis were consistent with the agreement data and are reported in Table 5.4. On average, an en face defect showed a corresponding VF defect in 36% of cases for SAP (median positive predictive value: 0.36, IQR: 0.23 to 0.43) and 47% of cases for custom perimetry (median positive predictive value: 0.47, IQR: 0.32 to 0.65). On the other hand, a normal VF location from a prediction of normal reflectivity was substantially more likely, as showed by high negative predictive values for both SAP and custom perimetry (SAP: median negative predictive value: 0.83, IQR: 0.73 to 0.88; Custom Perimetry: median negative predictive value: 0.86, IQR: 0.75 to 0.93).

Median positive likelihood ratios were 1.63 (IQR: 0.98 to 3.82) and 1.92 (IQR: 1.38 to 4.69) for SAP and custom perimetry, respectively. Median

negative likelihood ratios were 0.71 (IQR: 0.47 to 1.04) and 0.73 (IQR: 0.28 to 0.91) for SAP and custom perimetry, respectively.

Table 5.4
Findings from positive predictive value (PPV) and negative predictive value (NPV) analysis for SAP and custom perimetry. Values are reported alongside their bootstrapped 95% CI. Prevalence, computed as the proportion of VF defects among locations tested, is also reported.

ID	SAP (defect if TD with p<2%)			Custom perimetry		
	PPV	NPV	Prevalence	PPV	NPV	Prevalence
1	1.0 0.4 1.0	0.66 0.49 0.80	0.4	0.71 0.51 0.87	0.76 0.72 0.79	0.27
2	0.25 0.05 0.57	0.67 0.47 0.83	0.31	0.45 0.37 0.52	0.68 0.63 0.73	0.36
3	0.39 0.20 0.61	0.87 0.60 0.98	0.29	0.48 0.43 0.54	0.88 0.83 0.92	0.34
4	0.21 0.05 0.51	0.70 0.47 0.87	0.27	0.45 0.37 0.53	0.62 0.57 0.67	0.4
5	0.06 0.00 0.27	0.87 0.60 0.98	0.09	0.12 0.09 0.17	0.87 0.82 0.91	0.13
6	0.67 0.30 0.93	0.91 0.75 0.98	0.22	0.64 0.57 0.71	0.93 0.90 0.95	0.26
7	0.11 0.00 0.48	0.79 0.60 0.92	0.18	0.14 0.10 0.20	0.86 0.82 0.90	0.14
8	0.30 0.07 0.65	0.96 0.82 1.00	0.11	0.67 0.59 0.74	0.93 0.90 0.95	0.25
9	0.43 0.10 0.82	0.90 0.74 0.98	0.16	0.88 0.80 0.93	0.93 0.90 0.95	0.26
10	0.20 0.01 0.72	0.83 0.67 0.94	0.17	0.67 0.55 0.78	0.74 0.70 0.78	0.31
11	0.60 0.15 0.95	0.53 0.28 0.77	0.5	0.20 0.13 0.29	0.91 0.86 0.95	0.12
12	0.25 0.05 0.57	1.00 0.87 1.00	0.08	0.44 0.37 0.52	0.93 0.90 0.96	0.2
13	0.44 0.20 0.70	0.76 0.55 0.91	0.32	0.60 0.53 0.66	0.43 0.38 0.49	0.58
14	0.36 0.11 0.69	0.84 0.66 0.95	0.21	0.54 0.46 0.61	0.77 0.72 0.81	0.33
15	0.41 0.18 0.67	0.78 0.56 0.93	0.3	0.32 0.27 0.38	0.94 0.90 0.96	0.19
16	-	-	-	0.31 0.21 0.43	0.79 0.75 0.82	0.23

Analysis of structure-function agreement was also performed by location, and findings are reported in Figure 5.9. Agreement and tetrachoric correlation were only evaluated at locations where data were available in more than 50% of glaucoma participants (i.e. locations with data from 8 or more participants). Also, correlation coefficients could not be computed in some of the locations due to lack of perimetry or en face defects, therefore leading to zero variance for one of the variables. Considering SAP and en face predictions of visual function, the median overall agreement from all locations was 0.69 (IQR: 0.57 to 0.80), whereas median correlation was 0.27 (range -0.04 to 0.77). Similar measures were observed between custom perimetry and en face analysis

(median overall agreement: 0.69, IQR: 0.60 to 0.80; median rho: 0.36, range: -0.82 to 0.90).
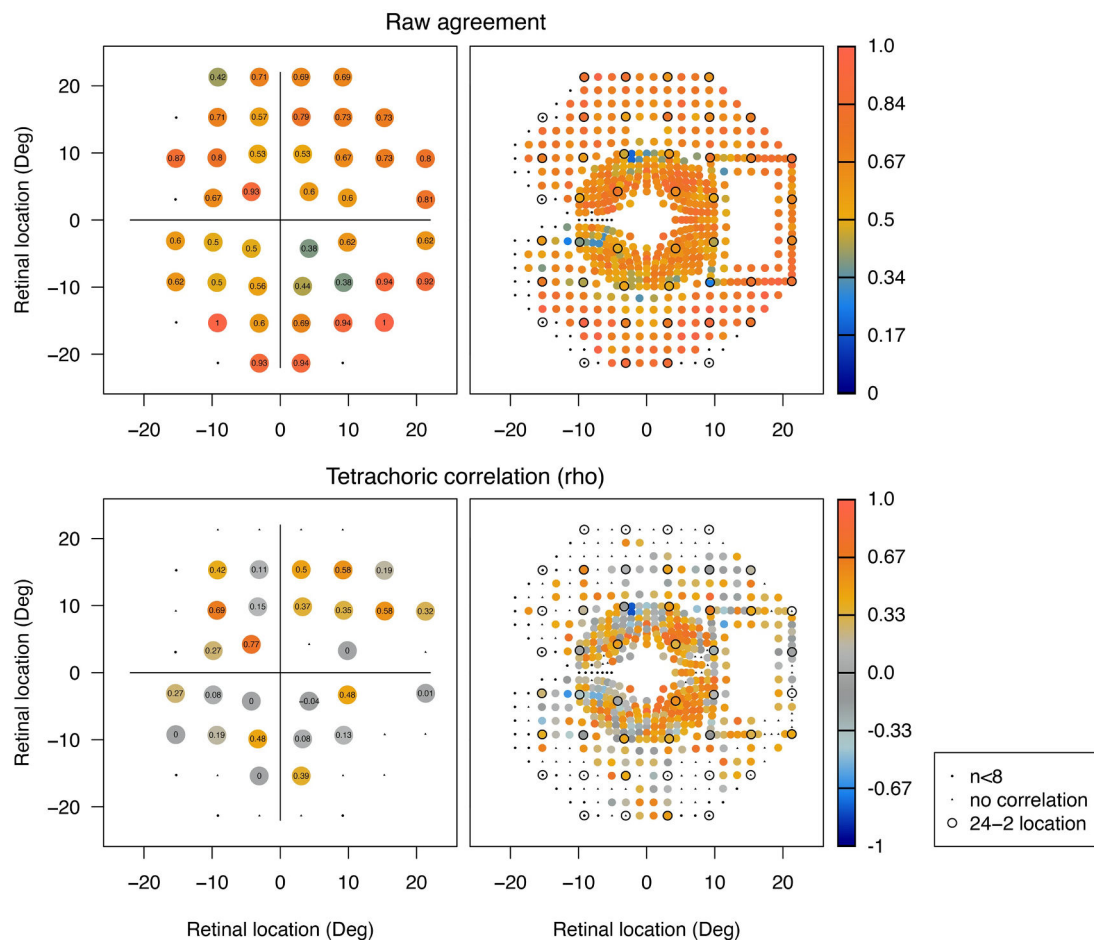
Figure 5.9 Heatmaps of the strength of overall raw agreement (top panel) and tetrachoric correlation (bottom panel) analysed by retinal location. Plots on the left hand side show en face-SAP analysis, whereas relationships between en face and custom perimetry are shown on the right hand side (locations corresponding to 24-2 grid are also marked with a circle). Locations with data from less than 8 participants were censored from analysis (small dots). At locations where either en face or perimetry did not show any defects in the whole sample, correlations could not be computed and were flagged with small triangles, only in the bottom panel.

5.5 Discussion

Evaluation of glaucomatous damage by en face OCT imaging is a clinically appealing and rapidly expanding area, yet optimal methods for objective definition of RNFB reflectance defects remain poorly defined (Cheloni and Denniss, 2021). In Chapter 4, we introduced a method to objectively and automatically extract reflectance defects from OCT en face images, which addressed some of the limitations of currently available methods (Cheloni et al., 2021a). As further validation, in this study we explored how the identified defects relate with conventional measures of glaucoma damage such as

structural changes of cpRNFL thickness and loss of visual function as measured with two different VF strategies. As glaucoma frequently develops in focal patterns (Nicolela and Drance, 1996, Elze et al., 2015), we strived to evaluate structure-structure and structure-function relationships with a topographical and point-wise approach. This was achieved by considering concordance between individual locations of each domain, and the method used adds to the few frameworks available for pointwise comparison across domains (Ballae Ganeshrao et al., 2015, Turpin and McKendrick, 2021, Tsamis et al., 2020, Hirasawa et al., 2020, Mohammadzadeh et al., 2020). The benefit of such an approach is to comply with the typical pattern of onset of glaucoma defects and to minimise conflation of healthy and damaged locations, reported among factors able to bury structure-function relationships (Araie et al., 2014, Chu et al., 2018).

Structure-Structure

Reflectance abnormalities showed a high concordance with conventional cpRNFL thickness loss, with 85% of en face defects having corresponding cpRNFL thinning at the mapped angular location of insertion on the ONH. Concordance ratios among glaucoma participants showed some variance but were generally high, with one outlier standing out from the sample (ID#1, concordance ratio: 0.48). SMAS deviation maps of this participant and the first four slabs feeding SMAS are reported in Figure 5.10. The eye presented anatomical configuration relatively similar to the population average (axial length: 22.7mm; ONH position, x: 16.9°, y: 1.1°; fovea-disc angle: -3.7°; and fovea-disc-raphe angle: 170.3°). Objective analysis of reflectance with SMAS identified a narrow superior arcuate defect, and a number of abnormal superpixels in the inferior-nasal quadrant (23 out of the total 67 abnormal superpixels). Yet, abnormalities found in the inferior retina were likely due to artefacts and this quadrant overall appeared spared by glaucomatous damage in subjective evaluation of en face images. Integrity of the inferior nasal retina was also suggested by the cpRNFL thickness analysis (Figure 5.5, top left panel), where only a superior temporal thinning was observed. A possible explanation for the low concordance ratio found in this participant is a non-substantial loss of reflectance in the superior hemisphere according to SMAS

objective analysis, plus some scattered noise in the inferior hemisphere which had substantial impact on concordance. The small superior-arcuate defect may be under-represented in SMAS, since it appears subjectively more substantial at greater depths (Figure 5.10). For the same participant, a concordance ratio of 0.73 was computed by excluding abnormal locations in the inferior-nasal region, and arguably this could be even higher if considering the shape of the cpRNFL thinning which was close to the 1% lower limit even between the temporal and temporal-superior sector (Figure 5.5). Future improvements in image capture and processing could contribute to reduce the impact on this and similar issues (see limitations).
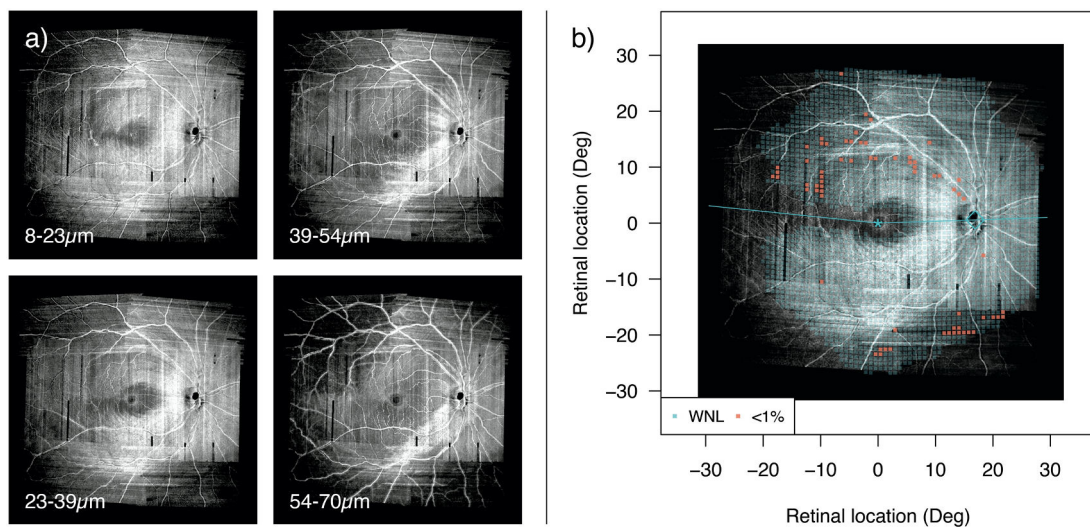
Figure 5.10 En face slab images (first 4 slabs out of the 7 considered in SMAS analysis) are shown in (a). The corresponding depth below the ILM, considered by each single slab is reported in µm. In (b), SMAS deviation map.

Findings of the structure-structure analysis are consistent with the available literature. Although previous works used different methodologies, OCT measures of reflectance loss have been shown to be strongly related to thinning of the RNFL (Ashimatey et al., 2018a, Thepass et al., 2017, Pons et al., 2000, Tan et al., 2021). Our previous work on a largely overlapping sample with this study (see chapter 3) also evaluated this relationship, and found a strong correlation between these two domains. See section 3.5 for further discussion on the evidence available on this relationship and its implications. The work from Ashimatey and colleagues perhaps represents the analysis that most closely relates to the one presented here, since their study objectively evaluated reflectance defects in wide field en face images and related them to

cpRNFL thickness (Ashimatey et al., 2018a). In their work, ratios of en face defects in the superior and inferior retinal hemispheres were strongly related with the depth of cpRNFL thickness defects at corresponding ONH sectors (superior temporal and inferior temporal), with Spearman's rho of 0.74-0.75 and an absolute agreement between en face and cpRNFL abnormalities of ~82% (Ashimatey et al., 2018a). The observed strong concordance is also consistent with the generally high topographical correlation between structural measurements of glaucoma loss at the ONH (cpRNFL) and at the macular RNFL, GCL and IPL (Kim and Park, 2018, Shin et al., 2013, Kim et al., 2014, Wollstein et al., 2004). Overall, these findings suggest that in eyes with established glaucoma, loss of reflectance may be well represented by cpRNFL thickness analysis. This underlines the need to further evaluate the value of RNFL reflectance analyses at the earliest stages of the disease and/or in combination with perimetry.

Structure-function

The relationship between reflectance findings and visual function was explored by relating any VF location (either SAP or custom perimetry) with the cluster of en face superpixels lying within a 2° circle diameter, in each participant. Overall there was moderate-good agreement between en face measures and VF status, more strongly driven by agreement on healthy locations rather than defects. Glaucoma participants showed on average 3 and 82 concordant defects with SAP and custom perimetry, respectively. Similar measures were not computed in controls, since inclusion criteria required normal VF by definition and no abnormal superpixels were detected in healthy eyes at the level of significance considered to establish glaucoma defects (see chapter 4). Although concordance was imperfect, all glaucoma eyes showed some levels of structure-function agreement on presence of abnormalities, and the ability of similar concordance thresholds for glaucoma detection deserves further investigation.

Sources of disagreement between en face and VF findings in this sample were qualitatively explored post-hoc. For simplicity, the analysis was broken down into separate hemispheres, and only considered SAP results. As reported in

Table 5.5, 17 hemispheres showed imperfect agreement and 9 (53%) of them showed en face defects not matched by SAP defects. Yet, in 5 of these cases artefacts contributed at least in part to the abnormal reflectivity, leading to imperfect agreement. The remaining 4 hemispheres could instead support the hypothesis that structural modifications in the form of reflectance loss precede VF changes, yet further validation in larger samples is required. SAP defects with normal corresponding reflectivity was observed in 4 hemispheres (23.5%). However, in 3 out of 4 cases some reflectance abnormalities could be observed subjectively (especially superior wedge defects and peripheral arcuate lesions), and further refinement of SMAS analysis could result in stronger concordance. Lastly, in 4 (23.5%) hemispheres, both domains showed defects, and there was imperfect spatial overlap across domains explaining the disagreement.

Table 5.5
Post-hoc analysis of the sources of imperfect agreement between en face and SAP results. Percentages are computed considering all analysed hemispheres showing imperfect agreement (n=17).
Inf: inferior hemisphere; Sup: superior hemisphere.

Abnormal en face normal SAP		Abnormal SAP normal en face		Both abnormal (spatial disagreement)
Genuine	*Possibly artefacts*	*Genuine*	*Possibly abnormal en face - appears abnormal subjectively*	
ID#2 inf; ID#5 sup ID#7 inf; ID#8 inf	ID#3 sup; ID#4 inf ID#5 inf; ID#12 sup ID#13 inf	ID#13 sup (possibly artefacts)	ID#1 sup; ID#2 sup ID#7 sup	ID#10 inf; ID#11 inf ID#14 inf ID#15 sup (possibly artefacts)
4 (23.5%)	5 (29.4%)	1 (5.9%)	3 (17.6%)	4 (23.5%)
9 (52.9%)		4 (23.5%)		

Beside genuinely different time scales at which en face reflectance and VF could be affected in glaucoma, several reasons might explain disagreement between the two domains. As anticipated in the methods, RGC displacement and eye movements should be considered as a potential confounder in this relationship, and we aimed to control for their impact by allowing a 1° tolerance around VF locations. In addition, VF stimuli could be presented at the border of a scotoma, or the location tested could still have preserved RGCs but not visible/measurable RNFBs with the imaging device used. In these instances,

the two domains could show disagreement. Anomalously high or low hill of vision for the specific participants should also be considered. An additional issue reported in previous studies (Alluwimi et al., 2018b, Alluwimi et al., 2018a) is the possibility of en face defects changing with depth below the ILM when single slabs are considered. Yet, SMAS (see chapter 4) should minimise this risk since it considers multiple depths in the definition of reflectance defects. Besides these potential sources of disagreement, it should also be considered that every structural defect could not necessarily correspond to a VF defect, and vice versa. In fact, the RNFL represents only a minor part of the whole visual pathway, which may fail to capture other consequences of glaucoma (Frezzotti et al., 2014, Frezzotti et al., 2016). Similarly, there may be compensatory mechanisms leading to minimally impacted visual function in the presence of structural glaucoma damage (Bham et al., 2020).

Consistently with our results, studies exploring the relationship between reflectance loss and corresponding visual function have generally shown good agreement (Alluwimi et al., 2018a, Alluwimi et al., 2018b, Iikawa et al., 2020, Sakamoto et al., 2019, Ashimatey et al., 2021, Christopher et al., 2019). Yet, previous work has mostly focussed on the macular region and also differs from the analysis performed here for the subjective assessments of reflectivity and of concordance which were employed. Subjective assessment of reflectance by clinicians is likely to be less prone to image artefacts, yet has limitations in terms of practicality, technique standardisation, and selection bias due to preconceived expectations of RNFB defects. Also, when correspondence with VF is evaluated subjectively, a variable level of tolerance in spatial agreement could be considered at different locations and different eyes by clinicians, hampering consistency. Comparison of previous research with our results is further complicated by different statistical measures of agreement, different slab image extraction and different VF procedures.

Earlier work in this area was conducted by Alluwimi et al., which examined the central retina of 10 glaucoma eyes with en face imaging and the corresponding function with an experimental VF device (Alluwimi et al., 2018b, Alluwimi et al., 2018a). Among the stimuli used, presentations were made with a supra-threshold Goldman size III stimulus (2 levels: 25dB and 28dB). A subjective

approach was used to define abnormalities of en face reflectance and correspondence with visual function. Consistently with our analysis, the authors found good concordance between the two domains, stronger for normal retinal locations compared to damaged areas. In fact, 31-48% of the locations predicted to be abnormal, did not see the VF stimulus, and 74-90% of locations predicted to not have defect, showed normal VF (Alluwimi et al., 2018b). More recent work from the same research group (Ashimatey et al., 2021), used kinetic perimetry to explore visual function along reflectance defects connected to the ONH, as observed in a selected sample of 7 participants. A mixture of stimulus intensities, according to individual isopters, were used, more often being a Goldman size II with ~15dB intensity. The study showed spatial correspondence between en face and functional defects, which, consistently with en face abnormalities, were always connected to the ONH (Ashimatey et al., 2021). Even better agreement was observed by Iikawa and colleagues in a study that superimposed SAP 10-2 results on macular en face images (Iikawa et al., 2020). The presence or absence of RNFBs underneath VF locations was evaluated subjectively by 3 clinicians, and ~85% of locations were correctly predicted to be either defect or normal. Lastly, a different approach was proposed in the work of Sakamoto et al, which focussed on eyes with advanced glaucoma and conducted subjective evaluation of concordance between preserved RNFBs in en face macula images and SAP 10-2 (Sakamoto et al., 2019). The authors concluded that preserved bundles and function had high agreement, however, little detail was provided on specific instruction and/or criteria adopted by clinicians to establish preserved reflectance and agreement itself.

As described above, a common feature of available studies is the subjective definition of en face defects or preserved RNFB. A potential methodological improvement introduced in our work is the objective identification of reflectance abnormalities and the use of an objective framework to establish pointwise agreement independently from clinician judgement. Such advancements should be considered in the perspective of wider clinical usage of en face analysis and greater standardisation of the technique. Although the level of structure-function agreement observed in our sample was perhaps

slightly poorer compared to some of the findings above, overall results appear to be well aligned. Data available so far indicate considerable concordance between en face and visual function, which seems stronger on prediction of normal locations compared to defects. Differences in performance shown by our analysis could be explained by unmitigated impact of artefacts on en face images, different perimetric strategies and criteria for establishing defects, and earlier stages of glaucoma in our sample.

Analyses of structure-function relationships in glaucoma have conventionally considered measures of individual retinal layer thickness, such as RNFL, GCL and IPL, or a combination of these together. Work in this area has adopted a wide variety of approaches to undertake this analysis (Malik et al., 2012, Denniss et al., 2018), with differences ranging from the clinical measures adopted, statistical modelling (from linear correlation to artificial intelligence), up to the elements related. The latter has considered either global summary metrics (e.g. SAP MD and average cpRNFL thickness), or hemispherical and sectorial data (Lee et al., 2017, Hirooka et al., 2016, Sato et al., 2013). Pointwise relationships at individual locations have seldom been considered. There are a few studies representing a more relevant comparison with our data, evaluating the structure-function relationship by overlapping VF data to a structural measure, therefore avoiding the use of ONH mapping. Some of these works considered only the macula, by relating OCT scans of this region with corresponding VF status measured by SAP 10-2 (Hirasawa et al., 2020, Ohkubo et al., 2014, Hashimoto et al., 2021). Results showed an overall strong structure-function correlation, yet imperfect and highly variable across different locations (range 0.3 to 0.77). There are a limited number of studies that considered wider regions of the retina and the relationship with VF data. Highly relevant work compared to our analysis was produced by a study group at the University of Iowa, which performed wide-field SD-OCT in eyes with glaucoma to estimate VF sensitivities from the average GCC (i.e. RNFL, GCL, and IPL tohgether), across a 54 location grid designed to resemble the SAP 24-2 grid (Bogunovic et al., 2014). When considering pointwise structural defects, the analysis included the thinning affecting the RNFL along the potential bundle's pathway from the studied location to the ONH, as defined by a widely used

structure-function map (Garway-Heath et al., 2000b). Subsequently, a machine learning algorithm was trained to predict VF sensitivities from thickness measures. The authors observed strong correlation between estimated and measured sensitivities, with mean Pearson's r of 0.68 (Bogunovic et al., 2014). Subsequently, a slightly updated model was tested on an independent cohort of participants (Guo et al., 2017), and major modifications consisted the use of variable patterns of NFL adjunct to specific locations. Best performance (average Pearson's r: 0.74) was achieved with a NFL adjunct pattern developed through computational methods, which however included regions temporal to the area assessed in some locations (i.e. away from ONH), which is unlikely to be anatomically grounded. In the replication cohort, an average r of 0.49 (range 0.08 to 0.74) was observed when computing only the thickness at each VF location, a value significantly lower than the optimal performance which, however, is more comparable to our analysis (Guo et al., 2017).

Lastly, work from Hood and colleagues (Hood et al., 2019, Tsamis et al., 2020) explored agreement between structural defects in wide-field OCT and VF results from SAP 10-2 and 24-2. Here, the authors proposed a method to objectively and automatically relate structure and function to facilitate glaucoma assessment. A mixture of retinal layers was considered in OCT, being the GCIPL in the macula, and the RNFL in a wide field 9x12mm scan. The analysis focussed on 1° diameter superpixels (chosen 'to be slightly larger than the perimetric stimulus') centred on each VF location, which were considered abnormal if found with p<0.1 of normative data. VF locations were instead considered defects when showing p<0.05 on pattern deviation plots. Agreement between abnormal structure and abnormal function was defined if an eye had at least 2 matched defects, from the 10-2 and 24-2 combined grid. Accordingly, 47 out of 53 defined glaucoma participants showed at least 2 concordant defects (Hood et al., 2019). A relative inconsistency could be noted in defining abnormalities of structure and function, which differs from our approach which aimed to keep criteria compatible across domains. OCT data considered deviations from age-adjusted norms (similar to Total Deviation) with a threshold of abnormality at p<0.1. VF data were instead evaluated by

considering a measure of deviation from individual sensitivity (i.e. Pattern Deviation) and a threshold of significance of p<0.05. As reported in later work, defect criteria and thresholds were derived from analysis on the same sample, aiming to achieve highest agreement in glaucoma eyes and minimal to no agreement in healthy controls (Tsamis et al., 2020). Criteria identified (i.e. OCT <10%, VF, PD <5%, threshold of 2 locations with agreement) were also validated in a different dataset, with 94.7% of glaucoma participants showing matched defects.

When relating our results to the findings above, it must be considered that Hood et al. used a grid with more locations (~96) compared to our SAP analysis, and fewer locations compared to our custom perimetry analysis. The absolute threshold of 2 concordant locations to call concordance employed above roughly corresponds to a 2.1% frequency of locations tested showing a defect in both domains. With a similar criterion all glaucoma participants in our sample but ID#16 (94%) showed concordance for SAP analysis, with a median rate of defects matched in both domains of 8.1% (range 2.4 to 23.7%). Considering custom perimetry analysis, all participants (100%) presented more than 2.1% concordant defects in en face and perimetry (median: 14.9%, range 3.5 to 29.1%). It should also be noted, that a similar level of concordance in our sample was observed by using more conservative levels of significance to consider glaucoma abnormalities (p<0.01 and p<0.02 in en face and perimetry, respectively). This feature might suggest preserved sensitivity with higher specificity, and represent an additional feature peculiar to en face analysis worth further investigation. For instance, the percentage of glaucoma participants with 2 locations showing abnormal structure and function in the work from Tsamis et al. (2020) dropped to 60.4% when considering more conservative criteria, closer to those used in our analysis (i.e. VF: total deviation <2%; OCT: <5%; Table 2 in Tsamis et al., 2020).

Overall, evidence from structure-function analyses seeded by thickness measures appears consistent with en face studies. Although there remains some uncertainty, structural and functional estimates of glaucoma damage show a sizeable relationship and good spatial agreement, which is promising for the use of both tests in conjunction.

This study has some limitations. The lack of a solid reference standard and the cross-sectional design of this study did not allow us to identify true and false defects and the analysis considered raw agreement between different domains and strength of correlation. The use of correlation coefficients to infer agreement is sometimes misleading when dealing with continuous variables, since strong correlations could coexist with poor agreement in case of significant bias or systemic errors (McAlinden et al., 2011). Yet, the dichotomous nature of the measures adopted here should prevent this risk. Among the many measures of agreement between dichotomous variables we choose to report overall, positive and negative agreement, rather than K statistics. The selected measures have the benefit of providing direct meaning in concordance analysis and could be broken down into differences between agreement on defects and normal findings. On the contrary, interpretation of K results is less immediate and requires comparison to arbitrary thresholds (Uebersax, 1987). In addition, agreement was mainly computed within-participants, and since K depends on trait prevalence, measures could not have been compared across eyes, due to the variable frequency of defects. Although the measures of positive and negative agreement used are adjusted for uneven distributions of ratings (Cicchetti and Feinstein, 1990), many statistical measures also reported here are likely affected by the prevalence of defects and caution should be exercised before generalising these results to other samples with different severity of glaucoma.

Additional limitations could affect the mapping of structural data to VF results. The conversion of en face images from superpixels to degrees relied on different scans (Spectralis Glaucoma Module vs. wide field montage) and different methods to identify fovea-ONH distance (automated by Spectralis software vs. subjective in en face analysis). Inaccuracies in detection of anatomical structures and changes in eye position, rotation and scaling across scans could have led to imprecise mapping. Yet, we expect this source of error to be small and controlled for by the tolerances used in the structure-structure and structure-function analyses. The VF was examined without fundus-tracking, and a ±1° tolerance around VF locations was considered to account for mapping errors due to RGC displacement and eye movements in

perimetry. We acknowledge that this represents an over-simplification and these sources of error could have still impacted the structure-function relationship.

Criteria to establish defects in the various domains were selected to be highly specific and also consistent across different tests. The effect on concordance of varying thresholds was not explored in this analysis, but it is possible that levels could be fine-tuned to increase the agreement between structure and function in glaucoma eyes (Tsamis et al., 2020, Turpin and McKendrick, 2021).

Lastly, issues related to en face imaging and the impact of artefacts on analysis of reflectance were discussed earlier (see section 4.5). Nonetheless, artefacts had considerable impact on the agreement between en face analysis and VF. Future improvements of en face processing should focus on several areas, including: i) Minimisation of artefacts: superior and inferior scans appeared particularly burdened, and better strategies to correct for uneven intensities of different scans to achieve seamless montage could be considered; ii) Enhanced ability to detect reflectance defects, especially in the superior and inferior regions of the wide field image (e.g. wedge shaped defects). Also, SMAS showed limited capacity to identify reflectance abnormalities at the nasal step area and in the temporal retina. It is likely that a reduction in variability of reflectance data in healthy eyes, achieved by addressing en face artefacts, would result in improved capability to detect defects. Yet, specific strategies such as adjusting for covariates with clinically significant impact on reflectivity (Mauschitz et al., 2018, Wagner et al., 2020, Thepass et al., 2017), or performing asymmetry analyses between hemispheres could also be evaluated.

It should also be considered that en face image extraction and analysis in SMAS focussed on detection of RNFB defects, targeting evaluation of changes in early glaucoma. However, considering the stronger structure-function agreement on healthy locations, SMAS may accurately predict areas of preserved visual function in perimetry and a paradigm shift from defects if abnormal superpixels were found at any depth to alternatives prioritising

detection of preserved bundles in eyes with advance glaucoma may warrant further investigation.

5.6 Conclusions

In summary, this study further validated SMAS analysis of en face images by establishing correspondence between reflectance abnormalities and conventional measures of glaucoma damage. A framework to objectively assess pointwise relationships between structure-structure and structure-function was introduced. The majority of retinal locations with abnormal reflectance presented a matched cpRNFL thickness defect at the estimated insertion on the ONH. Agreement between en face analysis and VF data was moderate to good and further minimisation of artefacts in en face imaging could lead to stronger concordance. All participants showed a number of locations with abnormal reflectivity and function, and the diagnostic capability of similar criteria to detect early glaucoma deserves dedicated assessment. Stronger agreement was found between regions with normal reflectivity and preserved visual function and this property could be exploited to drive custom perimetry strategies in eyes with advanced glaucoma, aiming to monitor disease progression in areas able to provide more reliable functional data. Further evaluation in representative samples with relevant stages of glaucoma is warranted.

6. Discussion & conclusions

6.1 Discussion

Glaucoma defines a group of optic neuropathies causing irreversible damage to the visual system, among which the open angle subtype does so without associated symptoms at the early stages. Once diagnosed, lifelong treatment and follow-up is essential to minimise risk of any further vision loss at the cost of considerable social and economic burden. Early and accurate diagnosis is, therefore, essential.

At moderate-advanced stages, glaucomatous changes are likely to be easily identified by clinicians, enabling accurate detection with conventional ophthalmoscopy and VF examination. As such, an unresolved challenge in clinic is the differentiation of eyes at risk of glaucoma from those presenting the earliest changes, perhaps in the absence of conventionally defined VF defects (i.e. pre-perimetric). It can be argued that the use of advanced diagnostic imaging, such as OCT, is now essential in the examination of these such cases, where conventional examinations do not resolve the diagnostic dilemma. OCT provides considerable support to clinicians in this regard, yet diagnostic accuracy for manifest glaucoma remains imperfect, and is increasingly poor at earlier stages (Michelessi et al., 2020). Clinicians currently address the shortcoming by looking for changes of suspicious lesions in follow-up examinations, yet this approach is time and resource consuming, besides leading to increased risk of developing vision loss.

Monitoring progression becomes the key clinical task once glaucoma is detected and treatment is commenced. To this end, VF is often considered the reference standard technique, and especially so at moderate-advanced stages of the disease, when OCT measures are reported to suffer from a floor-effect. However, several reports have challenged the credibility of VF thresholds at damaged locations (<20dB), as any measured change of sensitivity could be purely due to test-retest variability. Authors have proposed that clinically significant reduction of variability at damaged locations could not be realistically achieved within current test paradigms (i.e. SAP). Such an aim

would rather derive from a revision of either the test stimuli and pattern or paradigms adopted (Turpin and McKendrick, 2011). One approach may be to preferentially target spatial information of scotoma, rather than sensitivity information, to monitor enlargement of defects, in place of deepening.

En face OCT imaging of RNFB reflectance retains considerable interest for early glaucoma detection. Evidence from animal models, yet to be replicated in clinical settings, indicates that reflectance may be affected earlier than RNFL thickness. En face images may also facilitate custom perimetry procedures which make use of structural information, via a direct overlay of VF test locations on OCT results. Similar strategies could be employed at different stages of glaucoma, either targeting early detection or monitoring of disease progression. There are, therefore, encouraging premises behind en face OCT, however, its relative novelty in glaucoma led to limited usability of the technique. There was limited understanding of RNFB configuration in healthy and glaucoma eyes. Also, no accepted methods for slab construction and objective extraction of reflectance changes were available. This project aimed to fill these gaps of knowledge, by contributing to develop a better understanding of characteristics of RNFBs and a method for objective extraction of defects in this domain. Such information is essential for consistent exploitation of en face OCT in the clinical examination of patients with glaucoma.

The study described in chapter 2 developed preliminary understanding of how visible presence of RNFBs changes across regions of the retina and in different eyes. Consistently with the known configuration of the RNFL, results clearly indicated that both retinal region and individual eye significantly affect visible presence of bundles. This suggested that, rather than the common choice of fixed 50µm thickness, slab construction should consider both these sources of variability to consistently sample RNFBs. Also, depths greater than 50µm below the ILM should be examined to collect all reflectance information, since RNFBs remain present posterior to this this limit.

The second experiment (chapter 3) represents a small deviation from the project's main trajectory. The experiment was born from the observation of a

substantial difference in depths of visible presence of RNFBs between healthy and glaucoma eyes during data extraction. The forced pause to data collection imposed by the Covid-19 pandemic provided the extra time needed to formally conduct this analysis. The results showed that a simple subjective evaluation of the depths at which RNFBs were present could be valuable in the detection of glaucoma. Notably, the results also provided useful insights on the relationship between RNFL reflectance and thickness in clinical settings. The strong correlation observed in our sample suggested that any potential discordance between the two measures may be present earlier in glaucoma pathogenesis, or should be sought via more sensitive automated analyses.

In chapter 4, we presented SMAS, a novel method to assess reflectance defects intended to address some of the limitations of previous approaches. These included better adaptation to individual anatomy and consideration of different layer morphology across the retina. The results showed that methods of slab construction do affect the ability to detect defects, and SMAS found more defects compared to alternatives. This was an indication of the efficacy of the modifications introduced in SMAS, yet, a limitation of this study was the lack of a solid reference standard to conduct conventional diagnostic accuracy analysis. As such, further evaluation may provide additional insights on the capability of SMAS in glaucoma detection.

The final experiment (chapter 5) sought additional validation of defects extracted with SMAS, by exploring their relationship with conventional measures of glaucoma damage. This analysis of cross-domain concordance attempted to respect the localised nature of early glaucomatous damage, by considering pointwise changes of structure and function. The results showed strong concordance between RNFB reflectance loss and cpRNFL thickness defects and moderate-good agreement with VF. In both cases, addressing artefacts of en face analysis may improve concordance.

Taken together, the experiments presented in chapters 2 and 4 provided essential understanding on the configuration of RNFBs across the retina. The results helped to confirm the underlying hypothesis that slab construction affects the capability to detect RNFL reflectance loss in en face OCT images.

Within these two studies, a method that more strictly considers the arrangement of RNFBs across the retina was developed and, although further refinement is required, this enabled objective extraction of reflectance defects and could be trialled in the detection of glaucoma in isolation or in combination with customised perimetric procedures.

Chapters 3 and 5 together provided further insights on the relationship between changes in RNFB reflectance and changes in RNFL thickness. Evidence from animal models indicates that reflectivity is affected earlier than thickness, but still has not been clearly replicated in clinical settings with human subjects (Huang et al., 2011, Fortune et al., 2013), and our work added to the few studies available. Overall, we found a strong correlation in chapter 3 (r = 0.73 to 0.80) and high concordance in chapter 5 (median concordance ratio 0.85, range 0.48 to 0.99). These findings suggest that in eyes with established glaucoma, loss of reflectance may be well represented by cpRNFL thickness analysis. This outcome underlines the importance of further evaluating the value of RNFL reflectance analyses at the earliest stages of glaucoma, or by means of more sensitive analysis (e.g. artificial intelligence, objective analysis), or in combination with perimetry.

The final experiment (chapter 5), also provided insights on the agreement between en face reflectance and visual function. This was similar to previous studies, that, however, employed subjective evaluations of reflectance and concordance (Iikawa et al., 2020, Sakamoto et al., 2019, Alluwimi et al., 2018b). Also consistently with previous work, agreement was stronger on undamaged locations, intimating the potential of en face analysis for monitoring progression in advanced glaucoma: En face imaging could be used to identify retinal regions with preserved RNFBs, and presumably relatively spared visual function. Custom perimetry strategies could then be targeted to such regions, while ignoring or minimally exploring locations with advanced damage. Such a paradigm would aim to tackle the high test-retest variability of perimetry, by focussing on spared VF locations able to provide more reliable measurements. A prompt detection of progression could therefore be achieved by privileging spatial information (e.g. scotoma enlargement, or development of new scotoma), rather than changes in depth of existing

scotoma. This perspective was, in fact, among initially planned analyses, but was aborted because of the inability of recruiting a sufficient number of participants with advanced glaucoma due to the Covid-19 pandemic.

Chapter 5 also suggested that eyes with established glaucoma may show a number of locations where both en face analysis and perimetry present concordant abnormal findings. Notably such a correspondence would consider also abnormal single VF locations, not classifying as a VF defect according to conventional definition of (e.g.) 3 contiguous VF points (Anderson and Patella, 1992). Compared to similar pointwise agreement analyses using retinal layer thickness (Hood et al., 2019, Tsamis et al., 2020), similar or better levels of concordance were achieved with more conservative thresholds to define defects in each specific domain. This finding may suggest higher sensitivity at matched specificity of en face-function analyses and deserves further investigation to explore any enhanced diagnostic capability for detection of early glaucoma in cross-sectional examinations (Karvonen et al., 2021).

An overall consideration is required before generalising results from our studies to different settings. As reported previously (section 2.3), data collection was part of a larger project on the structure-function relationship in glaucoma, aiming to create a large dataset usable to address different research queries. As data from eligible participants became available, these were used to answer research questions relevant to this project while data collection was still ongoing. As such, an increasingly larger number of participants could be included in consecutive studies, and collected data on individual eyes fed multiple analyses with considerable overlap between samples of different studies. As a result, our findings apply to the sample studied, and further work in larger and more diverse populations, in terms of age, ethnicity and anatomical configuration, is required to fully understand how our results can be generalised to different settings.

To further clarify the level of participant overlap between different studies, Table 6.1 shows how different control and glaucoma eyes were included in different analyses, as imposed by data availability and specific inclusion criteria (e.g. limitation to early glaucoma in chapter 5).

Table 6.1
Participants included in different experiments and sample overlap between studies. Where data were usable for the specific participants and participant was included in the analysis this is reported as (**O**). The number in the first column (n) does not correspond with participant's ID reported in individual studies.

n	Chapter 2	Chapter 3		Chapter 4		Chapter 5		
	Control	Control	POAG	Control	POAG	Control En face	Control cpRNFL	POAG
1	O	O	O	O	O	O	O	O
2	O	O	O	O	O	O	O	O
3		O	O				O	
4	O	O	O	O	O	O	O	O
5		O	O	O	O	O	O	
6	O	O	O	O	O	O	O	O
7	O	O	O	O	O	O	O	O
8	O	O	O	O		O	O	
9	O	O	O	O	O	O	O	O
10	O	O	O	O	O	O	O	O
11		O	O	O	O	O	O	O
12	O	O	O	O	O	O	O	O
13		O	O	O	O	O	O	O
14		O	O				O	
15	O	O	O	O		O	O	
16		O	O	O	O	O	O	O
17		O	O	O		O	O	
18		O	O	O	O	O	O	O
19		O	O	O	O	O	O	O
20		O	O	O	O	O	O	O
21				O	O	O	O	O
22						O	O	O
23						O	O	
24						O	O	
25							O	
26							O	
27							O	
28							O	
29							O	
Tot	10	20	20	19	16	22	29	16

Future work in this area should target further refinement of image capture and processing, aiming to reduce artefacts of en face imaging, to make the technique widely usable in primary and secondary care. The impact of artefacts on analysis of reflectance was sizeable, and future improvements of

image processing should focus on several areas, including better strategies to correct for uneven intensity due to varying beam incidence angle and different scans. Work on refinement of en face reflectance analysis should also aim to evaluate larger and more diverse samples in order to explore the effects of covariates, such as age, or ethnicity on measured reflectance and provide adjusted normative values for improved diagnostic capability.

As stressed in the discussion of chapter 3 (section 3.5), it is likely that a full appreciation of the value of en face reflectance for early glaucoma diagnosis requires examination of samples of glaucoma suspects or eyes with the earliest damage. It must be considered that for these cases there is a lack of an accepted reference standard to establish diagnosis, and a considerable risk of bias is introduced by using one test compared to another in favouring index tests from a corresponding domain. This makes longitudinal analysis, as proposed elsewhere (Azuara-Blanco et al., 2016, Virgili et al., 2018, Stagg and Medeiros, 2020), a method of choice for a solid and more independent adjudication of true cases of glaucoma. Such a study, conducted on glaucoma suspects or participants at higher risk of developing the disease, would provide cogent information on the value of several strategies for earlier and accurate detection of glaucoma in cross-sectional examinations. Questions such as, "*Can analysis of en face reflectance increase accuracy of early glaucoma detection in a single examination?",* and *"Can analysis of pointwise structure-function concordance of defects achieve better performance than best OCT measures?*", could be answered by such a research design.

6.2 Conclusions

This thesis showed that RNFB reflectance changes observed in en face OCT images can be examined automatically and objectively. Assessing RNFB reflectance changes is enhanced by considering individual anatomy and varying configuration of bundles across the retina and between individuals. In eyes with established glaucoma, changes of en face reflectance appear to be well represented in conventional RNFL thickness analysis, and strong evidence of reflectance loss without matched thinning was not found. The agreement between reflectance and visual function was moderate to good, being stronger on areas spared by glaucoma. Notably, both testing domains

presented concordant abnormalities in all tested eyes. Following further minimisation of artefacts in en face images, a direct use of reflectance analysis or its combination with perimetry to run OCT-driven perimetric strategies appear viable and with significant potential for the clinical examination of glaucoma.

References

Abrams, L. S., Scott, I. U., Spaeth, G. L., Quigley, H. A. & Varma, R. 1994. Agreement among optometrists, ophthalmologists, and residents in evaluating the optic disc for glaucoma. *Ophthalmology,* 101**,** 1662-7.

Alluwimi, M. S., Swanson, W. H., Malinovsky, V. E. & King, B. J. 2018a. A basis for customising perimetric locations within the macula in glaucoma. *Ophthalmic Physiol Opt,* 38**,** 164-173.

Alluwimi, M. S., Swanson, W. H., Malinovsky, V. E. & King, B. J. 2018b. Customizing Perimetric Locations Based on En Face Images of Retinal Nerve Fiber Bundles With Glaucomatous Damage. *Transl Vis Sci Technol,* 7**,** 5.

Altman, D. G. & Bland, J. M. 1994. Diagnostic tests 2: Predictive values. *Bmj,* 309**,** 102.

Amerasinghe, N. 2018. *Open-Angle Glaucoma - BMJ Best Practice* [Online]. British Medical Journal. Available: https://bestpractice.bmj.com/topics/en-gb/373 [Accessed November 2018].

Anderson, D. R. & Patella, V. M. 1992. Automated static perimetry.

Aoyama, Y., Murata, H., Tahara, M., Yanagisawa, M., Hirasawa, K., Mayama, C. & Asaoka, R. 2014. A method to measure visual field sensitivity at the edges of glaucomatous scotomata. *Invest Ophthalmol Vis Sci,* 55**,** 2584-91.

Araie, M., Saito, H., Tomidokoro, A., Murata, H. & Iwase, A. 2014. Relationship between macular inner retinal layer thickness and corresponding retinal sensitivity in normal eyes. *Invest Ophthalmol Vis Sci,* 55**,** 7199-205.

Aref, A. A. & Budenz, D. L. 2017. Detecting Visual Field Progression. *Ophthalmology,* 124**,** S51-s56.

Artes, P. H., Hutchison, D. M., Nicolela, M. T., Leblanc, R. P. & Chauhan, B. C. 2005. Threshold and variability properties of matrix frequency-doubling technology and standard automated perimetry in glaucoma. *Invest Ophthalmol Vis Sci,* 46**,** 2451-7.

Artes, P. H., Iwase, A., Ohno, Y., Kitazawa, Y. & Chauhan, B. C. 2002. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. *Invest Ophthalmol Vis Sci,* 43**,** 2654-9.

Ashimatey, B. S., King, B. J., Burns, S. A. & Swanson, W. H. 2018a. Evaluating glaucomatous abnormality in peripapillary optical coherence tomography enface visualisation of the retinal nerve fibre layer reflectance. *Ophthalmic Physiol Opt,* 38**,** 376-388.

Ashimatey, B. S., King, B. J. & Swanson, W. H. 2018b. Retinal putative glial alterations: implication for glaucoma care. *Ophthalmic Physiol Opt,* 38**,** 56-65.

Ashimatey, B. S., King, B. J. & Swanson, W. H. 2021. Functional characteristics of glaucoma related arcuate defects seen on OCT en face visualisation of the retinal nerve fibre layer. *Ophthalmic and Physiological Optics,* 41**,** 437-446.

Azuara-Blanco, A., Banister, K., Boachie, C., Mcmeekin, P., Gray, J., Burr, J., Bourne, R., Garway-Heath, D., Batterbury, M., Hernández, R., Mcpherson, G., Ramsay, C. & Cook, J. 2016. Automated imaging technologies for the diagnosis of glaucoma: a comparative diagnostic study for the evaluation of the diagnostic accuracy, performance as triage tests and cost-effectiveness (GATE study). *Health Technol Assess,* 20**,** 1-168.

Bakdash, J. Z. & Marusich, L. R. 2017. Repeated Measures Correlation. *Frontiers in Psychology,* 8.

Ballae Ganeshrao, S., Turpin, A., Denniss, J. & Mckendrick, A. M. 2015. Enhancing Structure-Function Correlations in Glaucoma with Customized Spatial Mapping. *Ophthalmology,* 122**,** 1695-705.

Ballae Ganeshrao, S., Turpin, A. & Mckendrick, A. M. 2018. Sampling the Visual Field Based on Individual Retinal Nerve Fiber Layer Thickness Profile. *Invest Ophthalmol Vis Sci,* 59**,** 1066-1074.

Bates, D., Mächler, M., Bolker, B. & Walker, S. J. a. P. A. 2014. Fitting linear mixed-effects models using lme4.

Bedggood, P., Nguyen, B., Lakkis, G., Turpin, A. & Mckendrick, A. M. 2017. Orientation of the Temporal Nerve Fiber Raphe in Healthy and in Glaucomatous Eyes. *Invest Ophthalmol Vis Sci,* 58**,** 4211-4217.

Bedggood, P., Tanabe, F., Mckendrick, A. M. & Turpin, A. 2016. Automatic identification of the temporal retinal nerve fiber raphe from macular cube data. *Biomedical optics express,* 7**,** 4043-4053.

Belovay, G. W. & Goldberg, I. 2018. The thick and thin of the central corneal thickness in glaucoma. *Eye (Lond),* 32**,** 915-923.

Bengtsson, B. 1981. The prevalence of glaucoma. *Br J Ophthalmol,* 65**,** 46-9.

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**,** 165-9.

Bham, H. A., Dewsbery, S. D. & Denniss, J. 2020. Unaltered Perception of Suprathreshold Contrast in Early Glaucoma Despite Sensitivity Loss. *Investigative Ophthalmology & Visual Science,* 61**,** 23-23.

Bland, J. M. & Altman, D. G. 1995a. Calculating correlation coefficients with repeated observations: Part 1--Correlation within subjects. *BMJ (Clinical research ed.),* 310**,** 446-446.

Bland, J. M. & Altman, D. G. 1995b. Calculating correlation coefficients with repeated observations: Part 2--Correlation between subjects. *Bmj,* 310**,** 633.

Blumberg, D., Skaat, A. & Liebmann, J. M. 2015. Emerging risk factors for glaucoma onset and progression. *Prog Brain Res,* 221**,** 81-101.

Boden, C., Blumenthal, E. Z., Pascual, J., Mcewan, G., Weinreb, R. N., Medeiros, F. & Sample, P. A. 2004. Patterns of glaucomatous visual field progression identified by three progression criteria. *Am J Ophthalmol,* 138**,** 1029-36.

Bogunovic, H., Kwon, Y. H., Rashid, A., Lee, K., Critser, D. B., Garvin, M. K., Sonka, M. & Abramoff, M. D. 2014. Relationships of retinal structure and humphrey 24-2 visual field thresholds in patients with glaucoma. *Invest Ophthalmol Vis Sci,* 56**,** 259-71.

Boling, W., Wudunn, D., Cantor, L. B., Hoop, J., James, M. & Nukala, V. 2012. Correlation between macular thickness and glaucomatous visual fields. *J Glaucoma,* 21**,** 505-9.

Boodhna, T. & Crabb, D. P. 2015. Disease severity in newly diagnosed glaucoma patients with visual field loss: trends from more than a decade of data. *Ophthalmic Physiol Opt,* 35**,** 225-30.

Bourne, R. R. A., Jonas, J. B., Bron, A. M., Cicinelli, M. V., Das, A., Flaxman, S. R., Friedman, D. S., Keeffe, J. E., Kempen, J. H., Leasher, J., Limburg, H., Naidoo, K., Pesudovs, K., Peto, T., Saadine, J., Silvester, A. J., Tahhan, N., Taylor, H. R., Varma, R., Wong, T. Y. & Resnikoff, S. 2018. Prevalence and causes of vision loss in high-income countries and in Eastern and Central Europe in 2015: magnitude, temporal trends and projections. *Br J Ophthalmol,* 102**,** 575-585.

Bowd, C., Zangwill, L. M., Weinreb, R. N., Medeiros, F. A. & Belghith, A. 2017. Estimating Optical Coherence Tomography Structural Measurement Floors to Improve Detection of Progression in Advanced Glaucoma. *Am J Ophthalmol,* 175**,** 37-44.

Bussel, Ii, Wollstein, G. & Schuman, J. S. 2014. OCT for glaucoma diagnosis, screening and detection of glaucoma progression. *Br J Ophthalmol,* 98 Suppl 2**,** ii15-9.

Camp, A. S. & Weinreb, R. N. 2017. Will Perimetry Be Performed to Monitor Glaucoma in 2025? *Ophthalmology,* 124**,** S71-s75.

Cedrone, C., Mancino, R., Cerulli, A., Cesareo, M. & Nucci, C. 2008. Epidemiology of primary glaucoma: prevalence, incidence, and blinding effects. *Prog Brain Res,* 173**,** 3-14.

Chakravarti, T., Moghimi, S., De Moraes, C. G. & Weinreb, R. N. 2021. Central-most Visual Field Defects in Early Glaucoma. *J Glaucoma,* 30**,** e68-e75.

Champely, S. 2020. *pwr: Basic Functions for Power Analysis* [Online]. Available: https://cran.r-project.org/package=pwr [Accessed].

Chan, M. P. Y., Broadway, D. C., Khawaja, A. P., Yip, J. L. Y., Garway-Heath, D. F., Burr, J. M., Luben, R., Hayat, S., Dalzell, N., Khaw, K. T. & Foster, P. J. 2017. Glaucoma and intraocular pressure in EPIC-Norfolk Eye Study: cross sectional study. *Bmj,* 358**,** j3889.

Chang, S. & Bowden, A. K. 2019. Review of methods and applications of attenuation coefficient measurements with optical coherence tomography. *J Biomed Opt,* 24**,** 1-17.

Cheloni, R. & Denniss, J. 2021. Depth-resolved variations in visibility of retinal nerve fibre bundles across the retina in enface OCT images of healthy eyes. *Ophthalmic and Physiological Optics,* 41**,** 179-191.

Cheloni, R., Dewsbery, S. D. & Denniss, J. 2021a. Enhanced Objective Detection of Retinal Nerve Fiber Bundle Defects in Glaucoma With a Novel Method for En Face OCT Slab Image Construction and Analysis. *Translational Vision Science & Technology,* 10**,** 1-1.

Cheloni, R., Dewsbery, S. D. & Denniss, J. 2021b. A Simple Subjective Evaluation of Enface OCT Reflectance Images Distinguishes Glaucoma From Healthy Eyes. *Translational Vision Science & Technology,* 10**,** 31-31.

Chen, B., Gao, E., Chen, H., Yang, J., Shi, F., Zheng, C., Zhu, W., Xiang, D., Chen, X. & Zhang, M. 2016. Profile and Determinants of Retinal Optical Intensity in Normal Eyes with Spectral Domain Optical Coherence Tomography. *PLoS One,* 11, e0148183.

Chen, T. C., Cense, B., Pierce, M. C., Nassif, N., Park, B. H., Yun, S. H., White, B. R., Bouma, B. E., Tearney, G. J. & De Boer, J. F. 2005. Spectral domain optical coherence tomography: ultra-high speed, ultra-high resolution ophthalmic imaging. *Arch Ophthalmol,* 123, 1715-20.

Chen, T. C., Hoguet, A., Junk, A. K., Nouri-Mahdavi, K., Radhakrishnan, S., Takusagawa, H. L. & Chen, P. P. 2018. Spectral-Domain OCT: Helping the Clinician Diagnose Glaucoma: A Report by the American Academy of Ophthalmology. *Ophthalmology,* 125, 1817-1827.

Chen, X., Hou, P., Jin, C., Zhu, W., Luo, X., Shi, F., Sonka, M. & Chen, H. 2013. Quantitative analysis of retinal layer optical intensities on three-dimensional optical coherence tomography. *Invest Ophthalmol Vis Sci,* 54, 6846-51.

Cheung, H. & Swanson, W. H. 2019. Identifying retinal nerve fiber layer defects using attenuation coefficients. *Investigative Ophthalmology & Visual Science,* 60, 5608-5608.

Chong, G. T. & Lee, R. K. 2012. Glaucoma versus red disease: imaging and glaucoma diagnosis. *Curr Opin Ophthalmol,* 23, 79-88.

Chong, L. X., Mckendrick, A. M., Ganeshrao, S. B. & Turpin, A. 2014. Customized, automated stimulus location choice for assessment of visual field defects. *Invest Ophthalmol Vis Sci,* 55, 3265-74.

Chong, L. X., Turpin, A. & Mckendrick, A. M. 2015. Targeted spatial sampling using GOANNA improves detection of visual field progression. *Ophthalmic Physiol Opt,* 35, 155-69.

Christopher, M., Bowd, C., Belghith, A., Goldbaum, M. H., Weinreb, R. N., Fazio, M. A., Girkin, C. A., Liebmann, J. M. & Zangwill, L. M. 2019. Deep Learning Approaches Predict Glaucomatous Visual Field Damage from OCT Optic Nerve Head En Face Images and Retinal Nerve Fiber Layer Thickness Maps. *Ophthalmology*.

Chu, F. I., Marin-Franch, I., Ramezani, K. & Racette, L. 2018. Associations between structure and function are different in healthy and glaucomatous eyes. *PLoS One,* 13, e0196814.

Chylack, L. T., Jr., Wolfe, J. K., Singer, D. M., Leske, M. C., Bullimore, M. A., Bailey, I. L., Friend, J., Mccarthy, D. & Wu, S. Y. 1993. The Lens Opacities Classification System III. The Longitudinal Study of Cataract Study Group. *Arch Ophthalmol,* 111, 831-6.

Cicchetti, D. V. & Feinstein, A. R. 1990. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol,* 43, 551-8.

Coleman, A. L. & Miglior, S. 2008. Risk factors for glaucoma onset and progression. *Surv Ophthalmol,* 53 Suppl1, S3-10.

Collett, D. 2002. *Modelling binary data*, CRC press.

Conlon, R., Saheb, H. & Ahmed, Ii 2017. Glaucoma treatment trends: a review. *Can J Ophthalmol,* 52, 114-124.

Crabb, D. P., Saunders, L. J. & Edwards, L. A. 2017. Cases of advanced visual field loss at referral to glaucoma clinics - more men than women? *Ophthalmic Physiol Opt,* 37, 82-87.

Ctori, I. & Huntjens, B. 2015. Repeatability of Foveal Measurements Using Spectralis Optical Coherence Tomography Segmentation Software. *PLoS One,* 10**,** e0129005.

Cull, G. A., Reynaud, J., Wang, L., Cioffi, G. A., Burgoyne, C. F. & Fortune, B. 2012. Relationship between orbital optic nerve axon counts and retinal nerve fiber layer thickness measured by spectral domain optical coherence tomography. *Invest Ophthalmol Vis Sci,* 53**,** 7766-73.

Curcio, C. A. & Allen, K. A. 1990. Topography of ganglion cells in human retina. *J Comp Neurol,* 300**,** 5-25.

De Moraes, C. G., Liebmann, J. M., Medeiros, F. A. & Weinreb, R. N. 2016. Management of advanced glaucoma: Characterization and monitoring. *Surv Ophthalmol,* 61**,** 597-615.

Delgado, M. F., Nguyen, N. T., Cox, T. A., Singh, K., Lee, D. A., Dueker, D. K., Fechtner, R. D., Juzych, M. S., Lin, S. C., Netland, P. A., Pastor, S. A., Schuman, J. S. & Samples, J. R. 2002. Automated perimetry: a report by the American Academy of Ophthalmology. *Ophthalmology,* 109**,** 2362-74.

Delong, E. R., Delong, D. M. & Clarke-Pearson, D. L. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics,* 44**,** 837-45.

Denniss, J. & Astle, A. T. 2016. Spatial Interpolation Enables Normative Data Comparison in Gaze-Contingent Microperimetry. *Invest Ophthalmol Vis Sci,* 57**,** 5449-5456.

Denniss, J., Mckendrick, A. M. & Turpin, A. 2012. An anatomically customizable computational model relating the visual field to the optic nerve head in individual eyes. *Invest Ophthalmol Vis Sci,* 53**,** 6981-90.

Denniss, J., Mckendrick, A. M. & Turpin, A. 2013. Towards Patient-Tailored Perimetry: Automated Perimetry Can Be Improved by Seeding Procedures With Patient-Specific Structural Information. *Transl Vis Sci Technol,* 2**,** 3.

Denniss, J., Turpin, A. & Mckendrick, A. M. 2014a. Individualized structure-function mapping for glaucoma: practical constraints on map resolution for clinical and research applications. *Invest Ophthalmol Vis Sci,* 55**,** 1985-93.

Denniss, J., Turpin, A. & Mckendrick, A. M. 2014b. Visual Contrast Detection Cannot Be Predicted From Surrogate Measures of Retinal Ganglion Cell Number and Sampling Density in Healthy Young Adults. *Investigative Ophthalmology & Visual Science,* 55**,** 7804-7813.

Denniss, J., Turpin, A. & Mckendrick, A. M. 2018. Relating optical coherence tomography to visual fields in glaucoma: structure-function mapping, limitations and future applications. *Clin Exp Optom*.

Denniss, J., Turpin, A., Tanabe, F., Matsumoto, C. & Mckendrick, A. M. 2014c. Structure-function mapping: variability and conviction in tracing retinal nerve fiber bundles and comparison to a computational model. *Invest Ophthalmol Vis Sci,* 55**,** 728-36.

Douglas, G. R., Drance, S. M. & Schulzer, M. 1974. A correlation of fields and discs in open angle glaucoma. *Can J Ophthalmol,* 9**,** 391-8.

Drance, S., Anderson, D. R. & Schulzer, M. 2001. Risk factors for progression of visual field abnormalities in normal-tension glaucoma. *Am J Ophthalmol,* 131**,** 699-708.

Drasdo, N., Millican, C. L., Katholi, C. R. & Curcio, C. A. 2007. The length of Henle fibers in the human retina and a model of ganglion receptive field density in the visual field. *Vision Res,* 47**,** 2901-11.

Drasgow, F. 2014. Polychoric and polyserial correlations. *Wiley StatsRef: Statistics Reference Online.*

Dwelle, J., Liu, S., Wang, B., Mcelroy, A., Ho, D., Markey, M. K., Milner, T. & Rylander, H. G., 3rd 2012. Thickness, phase retardation, birefringence, and reflectance of the retinal nerve fiber layer in normal and glaucomatous non-human primates. *Invest Ophthalmol Vis Sci,* 53**,** 4380-95.

Elze, T., Pasquale, L. R., Shen, L. Q., Chen, T. C., Wiggs, J. L. & Bex, P. J. 2015. Patterns of functional vision loss in glaucoma determined with archetypal analysis. *J R Soc Interface,* 12.

Fallon, M., Valero, O., Pazos, M. & Anton, A. 2017. Diagnostic accuracy of imaging devices in glaucoma: A meta-analysis. *Surv Ophthalmol,* 62**,** 446-461.

Fawcett, T. 2006. An introduction to ROC analysis. *Pattern recognition letters,* 27**,** 861-874.

Finn, R. H. 1970. A note on estimating the reliability of categorical data. *Educational Psychological Measurement,* 30**,** 71-76.

Flaxman, S. R., Bourne, R. R. A., Resnikoff, S., Ackland, P., Braithwaite, T., Cicinelli, M. V., Das, A., Jonas, J. B., Keeffe, J., Kempen, J. H., Leasher, J., Limburg, H., Naidoo, K., Pesudovs, K., Silvester, A., Stevens, G. A., Tahhan, N., Wong, T. Y. & Taylor, H. R. 2017. Global causes of blindness and distance vision impairment 1990-2020: a systematic review and meta-analysis. *Lancet Glob Health,* 5**,** e1221-e1234.

Fortune, B. 2015. In vivo imaging methods to assess glaucomatous optic neuropathy. *Exp Eye Res,* 141**,** 139-53.

Fortune, B. 2018. Optical coherence tomography evaluation of the optic nerve head neuro-retinal rim in glaucoma. *Clin Exp Optom.*

Fortune, B., Burgoyne, C. F., Cull, G., Reynaud, J. & Wang, L. 2013. Onset and progression of peripapillary retinal nerve fiber layer (RNFL) retardance changes occur earlier than RNFL thickness changes in experimental glaucoma. *Invest Ophthalmol Vis Sci,* 54**,** 5653-61.

Fortune, B., Burgoyne, C. F., Cull, G. A., Reynaud, J. & Wang, L. 2012. Structural and Functional Abnormalities of Retinal Ganglion Cells Measured In Vivo at the Onset of Optic Nerve Head Surface Change in Experimental Glaucoma. *Investigative Ophthalmology & Visual Science,* 53**,** 3939-3950.

Fortune, B., Cull, G., Reynaud, J., Wang, L. & Burgoyne, C. F. 2015. Relating Retinal Ganglion Cell Function and Retinal Nerve Fiber Layer (RNFL) Retardance to Progressive Loss of RNFL Thickness and Optic Nerve Axons in Experimental Glaucoma. *Investigative Ophthalmology & Visual Science,* 56**,** 3936-3944.

Foster, P. J., Buhrmann, R., Quigley, H. A. & Johnson, G. J. 2002. The definition and classification of glaucoma in prevalence surveys. *Br J Ophthalmol,* 86**,** 238-42.

Founti, P., Bunce, C., Khawaja, A. P., Doré, C. J., Mohamed-Noriega, J. & Garway-Heath, D. F. 2020. Risk Factors for Visual Field Deterioration in the United Kingdom Glaucoma Treatment Study. *Ophthalmology,* 127**,** 1642-1651.

Founti, P., Coleman, A. L., Wilson, M. R., Yu, F., Anastasopoulos, E., Harris, A., Pappas, T., Koskosas, A., Kilintzis, V., Salonikiou, A., Raptou, A. & Topouzis, F. 2018. Overdiagnosis of open-angle glaucoma in the general population: the Thessaloniki Eye Study. *Acta Ophthalmol,* 96**,** e859-e864.

Frezzotti, P., Giorgio, A., Motolese, I., De Leucio, A., Iester, M., Motolese, E., Federico, A. & De Stefano, N. 2014. Structural and functional brain changes beyond visual system in patients with advanced glaucoma. *PLoS One,* 9**,** e105931.

Frezzotti, P., Giorgio, A., Toto, F., De Leucio, A. & De Stefano, N. 2016. Early changes of brain connectivity in primary open angle glaucoma. *Hum Brain Mapp,* 37**,** 4581-4596.

Fry, L. E., Fahy, E., Chrysostomou, V., Hui, F., Tang, J., Van Wijngaarden, P., Petrou, S. & Crowston, J. G. 2018. The coma in glaucoma: Retinal ganglion cell dysfunction and recovery. *Prog Retin Eye Res,* 65**,** 77-92.

Ganeshrao, S. B., Mckendrick, A. M., Denniss, J. & Turpin, A. 2015. A perimetric test procedure that uses structural information. *Optom Vis Sci,* 92**,** 70-82.

Gardiner, S. K., Demirel, S., Reynaud, J. & Fortune, B. 2016. Changes in Retinal Nerve Fiber Layer Reflectance Intensity as a Predictor of Functional Progression in Glaucoma. *Invest Ophthalmol Vis Sci,* 57**,** 1221-7.

Gardiner, S. K., Swanson, W. H., Goren, D., Mansberger, S. L. & Demirel, S. 2014. Assessment of the reliability of standard automated perimetry in regions of glaucomatous damage. *Ophthalmology,* 121**,** 1359-69.

Garway-Heath, D. F., Caprioli, J., Fitzke, F. W. & Hitchings, R. A. 2000a. Scaling the hill of vision: the physiological relationship between light sensitivity and ganglion cell numbers. *Invest Ophthalmol Vis Sci,* 41**,** 1774-82.

Garway-Heath, D. F., Crabb, D. P., Bunce, C., Lascaratos, G., Amalfitano, F., Anand, N., Azuara-Blanco, A., Bourne, R. R., Broadway, D. C., Cunliffe, I. A., Diamond, J. P., Fraser, S. G., Ho, T. A., Martin, K. R., Mcnaught, A. I., Negi, A., Patel, K., Russell, R. A., Shah, A., Spry, P. G., Suzuki, K., White, E. T., Wormald, R. P., Xing, W. & Zeyen, T. G. 2015. Latanoprost for open-angle glaucoma (UKGTS): a randomised, multicentre, placebo-controlled trial. *Lancet,* 385**,** 1295-304.

Garway-Heath, D. F., Poinoosawmy, D., Fitzke, F. W. & Hitchings, R. A. 2000b. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology,* 107**,** 1809-15.

Gazzard, G., Konstantakopoulou, E., Garway-Heath, D., Garg, A., Vickerstaff, V., Hunter, R., Ambler, G., Bunce, C., Wormald, R., Nathwani, N., Barton, K., Rubin, G. & Buszewicz, M. 2019. Selective laser trabeculoplasty versus eye drops for first-line treatment of ocular hypertension and glaucoma (LiGHT): a multicentre randomised controlled trial. *Lancet,* 393**,** 1505-1516.

Ghafaryasl, B., Vermeer, K. A., Kalkman, J., Callewaert, T., De Boer, J. F. & Van Vliet, L. J. 2020. Analysis of attenuation coefficient estimation in Fourier-domain OCT of semi-infinite media. *Biomedical Optics Express,* 11**,** 6093-6107.

Gmeiner, J. M., Schrems, W. A., Mardin, C. Y., Laemmer, R., Kruse, F. E. & Schrems-Hoesl, L. M. 2016. Comparison of Bruch's Membrane Opening Minimum Rim Width and Peripapillary Retinal Nerve Fiber Layer Thickness in Early Glaucoma Assessment. *Invest Ophthalmol Vis Sci,* 57**,** Oct575-84.

Gonzalez, R. C., Woods, R. E. & Eddins, S. L. 2004. *Digital image processing using MATLAB*, Pearson Education India.

Gordon, M. O., Beiser, J. A., Brandt, J. D., Heuer, D. K., Higginbotham, E. J., Johnson, C. A., Keltner, J. L., Miller, J. P., Parrish, R. K., 2nd, Wilson, M. R. & Kass, M. A. 2002. The Ocular Hypertension Treatment Study: baseline factors that predict the onset of primary open-angle glaucoma. *Arch Ophthalmol,* 120**,** 714-20; discussion 829-30.

Goren, D., Demirel, S., Fortune, B. & Gardiner, S. K. 2013. Correlating perimetric indices with three nerve fiber layer thickness measures. *Optom Vis Sci,* 90**,** 1353-60.

Green, P. & Macleod, C. J. 2016. SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution,* 7**,** 493-498.

Green, S. B. 1991. How Many Subjects Does It Take To Do A Regression Analysis. *Multivariate Behavioral Research,* 26**,** 499-510.

Grzybowski, A., Och, M., Kanclerz, P., Leffler, C. & Moraes, C. G. 2020. Primary Open Angle Glaucoma and Vascular Risk Factors: A Review of Population Based Studies from 1990 to 2019. *J Clin Med,* 9.

Guo, Z., Kwon, Y. H., Lee, K., Wang, K., Wahle, A., Alward, W. L. M., Fingert, J. H., Bettis, D. I., Johnson, C. A., Garvin, M. K., Sonka, M. & Abramoff, M. D. 2017. Optical Coherence Tomography Analysis Based Prediction of Humphrey 24-2 Visual Field Thresholds in Patients With Glaucoma. *Invest Ophthalmol Vis Sci,* 58**,** 3975-3985.

Ha, A., Kim, Y. K., Kim, J.-S., Jeoung, J. W. & Park, K. H. 2020. Temporal Raphe Sign in Elderly Patients With Large Optic Disc Cupping: Its Evaluation as a Predictive Factor for Glaucoma Conversion. *American Journal of Ophthalmology,* 219**,** 205-214.

Hadwin, S. E., Redmond, T., Garway-Heath, D. F., Lemij, H. G., Reus, N. J., Ward, G. & Anderson, R. S. 2013. Assessment of optic disc photographs for glaucoma by UK optometrists: the Moorfields Optic Disc Assessment Study (MODAS). *Ophthalmic Physiol Opt,* 33**,** 618-24.

Hajian-Tilaki, K. 2013. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian J Intern Med,* 4**,** 627-35.

Hammel, N., Belghith, A., Weinreb, R. N., Medeiros, F. A., Mendoza, N. & Zangwill, L. M. 2017. Comparing the Rates of Retinal Nerve Fiber Layer and Ganglion Cell-Inner Plexiform Layer Loss in Healthy Eyes and in Glaucoma Eyes. *Am J Ophthalmol,* 178**,** 38-50.

Hanley, J. A. & Mcneil, B. J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology,* 143**,** 29-36.

Harwerth, R. S., Carter-Dawson, L., Shen, F., Smith, E. L., 3rd & Crawford, M. L. 1999. Ganglion cell losses underlying visual field defects from experimental glaucoma. *Invest Ophthalmol Vis Sci,* 40, 2242-50.

Harwerth, R. S., Carter-Dawson, L., Smith, E. L., 3rd, Barnes, G., Holt, W. F. & Crawford, M. L. 2004. Neural losses correlated with visual losses in clinical perimetry. *Invest Ophthalmol Vis Sci,* 45, 3152-60.

Harwerth, R. S. & Quigley, H. A. 2006. Visual field defects and retinal ganglion cell losses in patients with glaucoma. *Arch Ophthalmol,* 124**,** 853-9.

Hashimoto, Y., Asaoka, R., Kiwaki, T., Sugiura, H., Asano, S., Murata, H., Fujino, Y., Matsuura, M., Miki, A., Mori, K., Ikeda, Y., Kanamoto, T., Yamagami, J., Inoue, K., Tanito, M. & Yamanishi, K. 2021. Deep learning model to predict visual field in central 10° from optical coherence tomography measurement in glaucoma. *Br J Ophthalmol,* 105**,** 507-513.

Heijl, A., Lindgren, G. & Olsson, J. 1987. Normal variability of static perimetric threshold values across the central visual field. *Arch Ophthalmol,* 105**,** 1544-9.

Heijl, A., Patella, V. M., Chong, L. X., Iwase, A., Leung, C. K., Tuulonen, A., Lee, G. C., Callan, T. & Bengtsson, B. 2019. A New SITA Perimetric Threshold Testing Algorithm: Construction and a Multicenter Clinical Study. *Am J Ophthalmol,* 198**,** 154-165.

Henson, D. B., Artes, P., Chaudry, S. J. & Chauhan, B. J. T. V. 1999. Suprathreshold perimetry: establishing the test intensity. 243-252.

Henson, D. B. & Artes, P. H. 2002. New developments in supra-threshold perimetry. *Ophthalmic Physiol Opt,* 22**,** 463-8.

Hirasawa, K., Matsuura, M., Fujino, Y., Yanagisawa, M., Kanamoto, T., Inoue, K., Nagumo, M., Yamagami, J., Yamashita, T., Murata, H. & Asaoka, R. 2020. Comparing Structure-Function Relationships Based on Drasdo's and Sjöstrand's Retinal Ganglion Cell Displacement Models. *Invest Ophthalmol Vis Sci,* 61**,** 10.

Hirooka, K., Misaki, K., Nitta, E., Ukegawa, K., Sato, S. & Tsujikawa, A. 2016. Comparison of Macular Integrity Assessment (MAIA), MP-3, and the Humphrey Field Analyzer in the Evaluation of the Relationship between the Structure and Function of the Macula. *PLoS One,* 11**,** e0151000.

Hodapp, E., Parrish, R. K. & Anderson, D. R. 1993. *Clinical decisions in glaucoma*, Mosby Inc.

Hohn, R., Nickels, S., Schuster, A. K., Wild, P. S., Munzel, T., Lackner, K. J., Schmidtmann, I., Beutel, M. & Pfeiffer, N. 2018. Prevalence of glaucoma in Germany: results from the Gutenberg Health Study. *Graefes Arch Clin Exp Ophthalmol,* 256**,** 1695-1702.

Hollands, H., Johnson, D., Hollands, S., Simel, D. L., Jinapriya, D. & Sharma, S. 2013. Do findings on routine examination identify patients at risk for primary open-angle glaucoma? The rational clinical examination systematic review. *Jama,* 309**,** 2035-42.

Hollo, G. 2017. Comparison of structure-function relationship between corresponding retinal nerve fibre layer thickness and Octopus visual field cluster defect values determined by normal and tendency-oriented strategies. *Br J Ophthalmol,* 101**,** 150-154.

Hondur, G., Goktas, E., Al-Aswad, L. & Tezel, G. 2018. Age-related changes in the peripheral retinal nerve fiber layer thickness. *Clin Ophthalmol,* 12**,** 401-409.

Hood, D. C. 2017. Improving our understanding, and detection, of glaucomatous damage: An approach based upon optical coherence tomography (OCT). *Prog Retin Eye Res,* 57**,** 46-75.

Hood, D. C. 2019. Does Retinal Ganglion Cell Loss Precede Visual Field Loss in Glaucoma? *Journal of glaucoma,* 28**,** 945-951.

Hood, D. C., De Cuir, N., Blumberg, D. M., Liebmann, J. M., Jarukasetphon, R., Ritch, R. & De Moraes, C. G. 2016. A Single Wide-Field OCT Protocol Can Provide Compelling Information for the Diagnosis of Early Glaucoma. *Transl Vis Sci Technol,* 5**,** 4.

Hood, D. C. & De Moraes, C. G. 2018a. Challenges to the Common Clinical Paradigm for Diagnosis of Glaucomatous Damage With OCT and Visual Fields. *Invest Ophthalmol Vis Sci,* 59**,** 788-791.

Hood, D. C. & De Moraes, C. G. 2018b. Four Questions for Every Clinician Diagnosing and Monitoring Glaucoma. *J Glaucoma,* 27**,** 657-664.

Hood, D. C., Fortune, B., Arthur, S. N., Xing, D., Salant, J. A., Ritch, R. & Liebmann, J. M. 2008. Blood vessel contributions to retinal nerve fiber layer thickness profiles measured with optical coherence tomography. *J Glaucoma,* 17**,** 519-28.

Hood, D. C., Fortune, B., Mavrommatis, M. A., Reynaud, J., Ramachandran, R., Ritch, R., Rosen, R. B., Muhammad, H., Dubra, A. & Chui, T. Y. 2015. Details of Glaucomatous Damage Are Better Seen on OCT En Face Images Than on OCT Retinal Nerve Fiber Layer Thickness Maps. *Invest Ophthalmol Vis Sci,* 56**,** 6208-16.

Hood, D. C. & Kardon, R. H. 2007. A framework for comparing structural and functional measures of glaucomatous damage. *Prog Retin Eye Res,* 26**,** 688-710.

Hood, D. C. & Raza, A. S. 2011. Method for comparing visual field defects to local RNFL and RGC damage seen on frequency domain OCT in patients with glaucoma. *Biomed Opt Express,* 2**,** 1097-105.

Hood, D. C., Raza, A. S., De Moraes, C. G., Liebmann, J. M. & Ritch, R. 2013. Glaucomatous damage of the macula. *Prog Retin Eye Res,* 32**,** 1-21.

Hood, D. C., Tsamis, E., Bommakanti, N. K., Joiner, D. B., Al-Aswad, L. A., Blumberg, D. M., Cioffi, G. A., Liebmann, J. M. & De Moraes, C. G. 2019. Structure-Function Agreement Is Better Than Commonly Thought in Eyes With Early Glaucoma. *Investigative Ophthalmology & Visual Science,* 60**,** 4241-4248.

Horn, F. K., Tornow, R. P., Junemann, A. G., Laemmer, R. & Kremers, J. 2014. Perimetric measurements with flicker-defined form stimulation in comparison with conventional perimetry and retinal nerve fiber measurements. *Invest Ophthalmol Vis Sci,* 55**,** 2317-23.

Huang, D., Swanson, E. A., Lin, C. P., Schuman, J. S., Stinson, W. G., Chang, W., Hee, M. R., Flotte, T., Gregory, K., Puliafito, C. A. & Et Al. 1991. Optical coherence tomography. *Science,* 254**,** 1178-81.

Huang, X. R., Knighton, R. W. & Cavuoto, L. N. 2006. Microtubule contribution to the reflectance of the retinal nerve fiber layer. *Invest Ophthalmol Vis Sci,* 47**,** 5363-7.

Huang, X. R., Knighton, R. W., Feuer, W. J. & Qiao, J. 2016. Retinal nerve fiber layer reflectometry must consider directional reflectance. *Biomed Opt Express,* 7**,** 22-33.

Huang, X. R., Knighton, R. W., Spector, Y. Z. & Feuer, W. J. 2017. Cytoskeletal Alteration and Change of Retinal Nerve Fiber Layer Birefringence in Hypertensive Retina. *Curr Eye Res,* 42**,** 936-947.

Huang, X. R., Zhou, Y., Kong, W. & Knighton, R. W. 2011. Reflectance decreases before thickness changes in the retinal nerve fiber layer in glaucomatous retinas. *Invest Ophthalmol Vis Sci,* 52**,** 6737-42.

Hudson, C., Wild, J. M. & O'neill, E. C. 1994. Fatigue effects during a single session of automated static threshold perimetry. *Invest Ophthalmol Vis Sci,* 35**,** 268-80.

Iikawa, R., Togano, T., Sakaue, Y., Suetake, A., Igarashi, R., Miyamoto, D., Yaoeda, K., Seki, M. & Fukuchi, T. 2020. Estimation of the central 10-degree visual field using en-face images obtained by optical coherence tomography. *PLOS ONE,* 15**,** e0229867.

Jain, V., Jain, M., Abdull, M. M. & Bastawrous, A. 2017. The association between cigarette smoking and primary open-angle glaucoma: a systematic review. *Int Ophthalmol,* 37**,** 291-301.

Jampel, H. D., 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.e1.

Jampel, H. D., Singh, K., Lin, S. C., Chen, T. C., Francis, B. A., Hodapp, E., Samples, J. R. & Smith, S. D. 2011. Assessment of visual function in glaucoma: a report by the American Academy of Ophthalmology. *Ophthalmology,* 118**,** 986-1002.

Jansonius, N. M., Nevalainen, J., Selig, B., Zangwill, L. M., Sample, P. A., Budde, W. M., Jonas, J. B., Lagreze, W. A., Airaksinen, P. J., Vonthein, R., Levin, L. A., Paetzold, J. & Schiefer, U. 2009. A mathematical description of nerve fiber bundle trajectories and their variability in the human retina. *Vision Res,* 49**,** 2157-63.

Jansonius, N. M., Schiefer, J., Nevalainen, J., Paetzold, J. & Schiefer, U. 2012. A mathematical model for describing the retinal nerve fiber bundle trajectories in the human eye: average course, variability, and influence of refraction, optic disc size and optic disc position. *Exp Eye Res,* 105**,** 70-8.

Ji, M. J., Park, J. H., Yoo, C. & Kim, Y. Y. 2020. Comparison of the Progression of Localized Retinal Nerve Fiber Layer Defects in Red-free Fundus Photograph, En Face Structural Image, and OCT Angiography Image. *J Glaucoma,* 29**,** 698-703.

Jindal, A., Ctori, I., Fidalgo, B., Dabasia, P., Balaskas, K. & Lawrenson, J. G. 2019. Impact of optical coherence tomography on diagnostic decision-making by UK community optometrists: a clinical vignette study. *Ophthalmic Physiol Opt,* 39**,** 205-215.

Jonas, J. B., Aung, T., Bourne, R. R., Bron, A. M., Ritch, R. & Panda-Jonas, S. 2017. Glaucoma. *Lancet,* 390**,** 2183-2193.

Jonas, J. B., Budde, W. M. & Panda-Jonas, S. 1999. Ophthalmoscopic evaluation of the optic nerve head. *Surv Ophthalmol,* 43**,** 293-320.

Jones, P. R., Philippin, H., Makupa, W. U., Burton, M. J. & Crabb, D. P. 2020. Severity of Visual Field Loss at First Presentation to Glaucoma Clinics in England and Tanzania. *Ophthalmic Epidemiol,* 27**,** 10-18.

Jung, J. H., Park, J. H., Yoo, C. & Kim, Y. Y. 2018. Localized Retinal Nerve Fiber Layer Defects in Red-free Photographs Versus En Face Structural Optical Coherence Tomography Images. *J Glaucoma,* 27**,** 269-274.

Junoy Montolio, F. G., Wesselink, C. & Jansonius, N. M. 2012. Persistence, Spatial Distribution and Implications for Progression Detection of Blind Parts of the Visual Field in Glaucoma: A Clinical Cohort Study. *PLOS ONE,* 7**,** e41211.

Kansal, V., Armstrong, J. J., Pintwala, R. & Hutnik, C. 2018. Optical coherence tomography for glaucoma diagnosis: An evidence based meta-analysis. *PLoS One,* 13**,** e0190621.

Kapetanakis, V. V., Chan, M. P., Foster, P. J., Cook, D. G., Owen, C. G. & Rudnicka, A. R. 2016. Global variations and time trends in the prevalence of primary open angle glaucoma (POAG): a systematic review and meta-analysis. *Br J Ophthalmol,* 100**,** 86-93.

Karvonen, E., Stoor, K., Luodonpaa, M., Hagg, P., Kuoppala, J., Lintonen, T., Ohtonen, P., Tuulonen, A. & Saarela, V. 2018. Prevalence of glaucoma in the Northern Finland Birth Cohort Eye Study. *Acta Ophthalmol*.

Karvonen, E., Stoor, K., Luodonpää, M., Hägg, P., Leiviskä, I., Liinamaa, J., Tuulonen, A. & Saarela, V. 2021. Combined structure-function analysis in glaucoma screening. *Br J Ophthalmol*.

Karvonen, E., Stoor, K., Luodonpää, M., Hägg, P., Lintonen, T., Liinamaa, J., Tuulonen, A. & Saarela, V. 2020. Diagnostic performance of modern imaging instruments in glaucoma screening. *Br J Ophthalmol,* 104**,** 1399-1405.

Katz, J., Tielsch, J. M., Quigley, H. A., Javitt, J., Witt, K. & Sommer, A. 1993. Automated suprathreshold screening for glaucoma: the Baltimore Eye Survey. *Invest Ophthalmol Vis Sci,* 34**,** 3271-7.

Kerrigan-Baumrind, L. A., Quigley, H. A., Pease, M. E., Kerrigan, D. F. & Mitchell, R. S. 2000. Number of ganglion cells in glaucoma eyes compared with threshold visual field tests in the same persons. *Invest Ophthalmol Vis Sci,* 41**,** 741-8.

Khawaja, A. P. & Viswanathan, A. C. 2018. Are we ready for genetic testing for primary open-angle glaucoma? *Eye (Lond),* 32**,** 877-883.

Killer, H. E. & Pircher, A. 2018. Normal tension glaucoma: review of current understanding and mechanisms of the pathogenesis. *Eye (Lond),* 32**,** 924-930.

Kim, H., Lee, J. S., Park, H. M., Cho, H., Lim, H. W., Seong, M., Park, J. & Lee, W. J. 2021. A Wide-Field Optical Coherence Tomography Normative Database Considering the Fovea-Disc Relationship for Glaucoma Detection. *Translational Vision Science & Technology,* 10**,** 7-7.

Kim, K. E., Oh, S., Baek, S. U., Ahn, S. J., Park, K. H. & Jeoung, J. W. 2020. Ocular Perfusion Pressure and the Risk of Open-Angle Glaucoma: Systematic Review and Meta-analysis. *Sci Rep,* 10**,** 10056.

Kim, K. E. & Park, K. H. 2018. Macular imaging by optical coherence tomography in the diagnosis and management of glaucoma. *Br J Ophthalmol,* 102**,** 718-724.

Kim, K. E., Park, K. H., Yoo, B. W., Jeoung, J. W., Kim, D. M. & Kim, H. C. 2014. Topographic Localization of Macular Retinal Ganglion Cell Loss Associated With Localized Peripapillary Retinal Nerve Fiber Layer Defect. *Investigative Ophthalmology & Visual Science,* 55**,** 3501-3508.

Kim, K. E., Yoo, B. W., Jeoung, J. W. & Park, K. H. 2015. Long-Term Reproducibility of Macular Ganglion Cell Analysis in Clinically Stable Glaucoma Patients. *Invest Ophthalmol Vis Sci,* 56**,** 4857-64.

Knighton, R. W. & Huang, X. R. 1999. Directional and spectral reflectance of the rat retinal nerve fiber layer. *Investigative Ophthalmology & Visual Science,* 40**,** 639-647.

Ko, F., Papadopoulos, M. & Khaw, P. T. 2015. Primary congenital glaucoma. *Prog Brain Res,* 221**,** 177-89.

Koh, V., Tham, Y. C., Tan, N. Y. Q., Yu, M., Thakur, S., Teo, Z. L., Mani, B., Wong, T. Y., Aung, T. & Cheng, C. Y. 2021. Six-Year Incidence and Risk Factors of Primary Glaucoma in the Singapore Indian Eye Study. *Ophthalmol Glaucoma,* 4**,** 201-208.

Kuang, T. M., Zhang, C., Zangwill, L. M., Weinreb, R. N. & Medeiros, F. A. 2015. Estimating Lead Time Gained by Optical Coherence Tomography in Detecting Glaucoma before Development of Visual Field Defects. *Ophthalmology,* 122**,** 2002-9.

Kumar, R. & Indrayan, A. 2011. Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatr,* 48**,** 277-87.

La Bruna, S., Tsamis, E., Zemborain, Z. Z., Wu, Z., De Moraes, C. G., Ritch, R. & Hood, D. C. 2020. A Topographic Comparison of OCT Minimum Rim Width (BMO-MRW) and Circumpapillary Retinal Nerve Fiber Layer (cRNFL) Thickness Measures in Eyes With or Suspected Glaucoma. *J Glaucoma,* 29**,** 671-680.

Lamparter, J., Russell, R. A., Zhu, H., Asaoka, R., Yamashita, T., Ho, T. & Garway-Heath, D. F. 2013. The influence of intersubject variability in ocular anatomical variables on the mapping of retinal locations to the retinal nerve fiber layer and optic nerve head. *Invest Ophthalmol Vis Sci,* 54**,** 6074-82.

Lavinsky, F., Wollstein, G., Tauber, J. & Schuman, J. S. 2017. The Future of Imaging in Detecting Glaucoma Progression. *Ophthalmology,* 124**,** S76-s82.

Lee, J. W., Morales, E., Sharifipour, F., Amini, N., Yu, F., Afifi, A. A., Coleman, A. L., Caprioli, J. & Nouri-Mahdavi, K. 2017. The relationship between central visual field sensitivity and macular ganglion cell/inner plexiform layer thickness in glaucoma. *Br J Ophthalmol,* 101**,** 1052-1058.

Leite, M. T., Zangwill, L. M., Weinreb, R. N., Rao, H. L., Alencar, L. M., Sample, P. A. & Medeiros, F. A. 2010. Effect of disease severity on the performance of Cirrus spectral-domain OCT for glaucoma diagnosis. *Investigative ophthalmology & visual science,* 51**,** 4104-4109.

Leitgeb, R. A. 2019. En face optical coherence tomography: a technology review [Invited]. *Biomed Opt Express,* 10**,** 2177-2201.

Leske, M. C. 2007. Open-angle glaucoma -- an epidemiologic overview. *Ophthalmic Epidemiol,* 14**,** 166-72.

Leske, M. C., Heijl, A., Hyman, L., Bengtsson, B. & Komaroff, E. 2004. Factors for progression and glaucoma treatment: the Early Manifest Glaucoma Trial. *Curr Opin Ophthalmol,* 15**,** 102-6.

Leung, C. K., Cheung, C. Y., Weinreb, R. N., Qiu, Q., Liu, S., Li, H., Xu, G., Fan, N., Huang, L., Pang, C. P. & Lam, D. S. 2009. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a variability and diagnostic performance study. *Ophthalmology,* 116**,** 1257-63, 1263.e1-2.

Leung, C. K., Yu, M., Weinreb, R. N., Ye, C., Liu, S., Lai, G. & Lam, D. S. 2012. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a prospective analysis of age-related loss. *Ophthalmology,* 119**,** 731-7.

Lim, A. B., Park, J.-H., Jung, J. H., Yoo, C. & Kim, Y. Y. 2020. Characteristics of diffuse retinal nerve fiber layer defects in red-free photographs as observed in optical coherence tomography en face images. *BMC Ophthalmology,* 20**,** 16.

Linnet, K., Bossuyt, P. M., Moons, K. G. & Reitsma, J. B. 2012. Quantifying the accuracy of a diagnostic test or marker. *Clin Chem,* 58**,** 1292-301.

Lisboa, R., Mansouri, K., Zangwill, L. M., Weinreb, R. N. & Medeiros, F. A. 2013. Likelihood ratios for glaucoma diagnosis using spectral-domain optical coherence tomography. *Am J Ophthalmol,* 156**,** 918-926.e2.

Liu, C. H., Chang, S. H. L. & Wu, S. C. 2017. Regional Relationship between Macular Retinal Thickness and Corresponding Central Visual Field Sensitivity in Glaucoma Patients. *J Ophthalmol,* 2017**,** 3720157.

Liu, S., Lam, S., Weinreb, R. N., Ye, C., Cheung, C. Y., Lai, G., Lam, D. S. & Leung, C. K. 2011. Comparison of standard automated perimetry, frequency-doubling technology perimetry, and short-wavelength automated perimetry for detection of glaucoma. *Invest Ophthalmol Vis Sci,* 52**,** 7325-31.

Liu, S., Wang, B., Yin, B., Milner, T. E., Markey, M. K., Mckinnon, S. J. & Rylander, H. G., 3rd 2014. Retinal nerve fiber layer reflectance for early glaucoma diagnosis. *J Glaucoma,* 23**,** e45-52.

Lucy, K. A. & Wollstein, G. 2016. Structural and Functional Evaluations for the Early Detection of Glaucoma. *Expert Rev Ophthalmol,* 11**,** 367-376.

Makowski, D., Ben-Shachar, M. S., Patil, I. & Lüdecke, D. 2020. Methods and algorithms for correlation analysis in R. *Journal of Open Source Software,* 5**,** 2306.

Malik, R., Swanson, W. H. & Garway-Heath, D. F. 2012. 'Structure-function relationship' in glaucoma: past thinking and current concepts. *Clin Exp Ophthalmol,* 40**,** 369-80.

Mallett, S., Halligan, S., Thompson, M., Collins, G. S. & Altman, D. G. 2012. Interpreting diagnostic accuracy studies for patient care. *Bmj,* 345**,** e3999.

Mauschitz, M. M., Bonnemaijer, P. W. M., Diers, K., Rauscher, F. G., Elze, T., Engel, C., Loeffler, M., Colijn, J. M., Ikram, M. A., Vingerling, J. R., Williams, K. M., Hammond, C. J., Creuzot-Garcher, C., Bron, A. M., Silva, R., Nunes, S., Delcourt, C., Cougnard-Grégoire, A., Holz, F. G., Klaver, C. C. W., Breteler, M. M. B. & Finger, R. P. 2018. Systemic and Ocular Determinants of Peripapillary Retinal Nerve Fiber Layer Thickness Measurements in the European Eye Epidemiology (E3) Population. *Ophthalmology,* 125**,** 1526-1536.

Mavrommatis, M. A., De Cuir, N., Reynaud, J., De Moraes, C. G., Xin, D., Rajshekhar, R., Liebmann, J. M., Ritch, R., Fortune, B. & Hood, D. C. 2019. An Examination of the Frequency of Paravascular Defects and Epiretinal Membranes in Eyes With Early Glaucoma Using En-face Slab OCT Images. *J Glaucoma,* 28**,** 265-269.

Mcalinden, C., Khadka, J. & Pesudovs, K. 2011. Statistical methods for conducting agreement (comparison of clinical tests) and precision (repeatability or reproducibility) studies in optometry and ophthalmology. *Ophthalmic and Physiological Optics,* 31**,** 330-338.

Mccann, P., Hogg, R., Wright, D. M., Pose-Bazarra, S., Chakravarthy, U., Peto, T., Cruise, S., Mcguinness, B., Young, I. S., Kee, F. & Azuara-Blanco, A. 2020. Glaucoma in the Northern Ireland Cohort for the Longitudinal Study of Ageing (NICOLA): cohort profile, prevalence, awareness and associations. *Br J Ophthalmol,* 104**,** 1492-1499.

Mckendrick, A. M. 2005. Recent developments in perimetry: test stimuli and procedures. *Clin Exp Optom,* 88**,** 73-80.

Mckendrick, A. M., Denniss, J., Wang, Y. X., Jonas, J. B. & Turpin, A. 2017. The Proportion of Individuals Likely to Benefit from Customized Optic Nerve Head Structure-Function Mapping. *Ophthalmology,* 124**,** 554-561.

Mckendrick, A. M. & Turpin, A. 2005. Combining perimetric suprathreshold and threshold procedures to reduce measurement variability in areas of visual field loss. *Optom Vis Sci,* 82**,** 43-51.

Mcmonnies, C. W. 2017. Glaucoma history and risk factors. *J Optom,* 10**,** 71-78.

Medeiros, F. A., Alencar, L. M., Zangwill, L. M., Bowd, C., Sample, P. A. & Weinreb, R. N. 2009. Prediction of functional loss in glaucoma from progressive optic disc damage. *Arch Ophthalmol,* 127**,** 1250-6.

Medeiros, F. A., Gracitelli, C. P., Boer, E. R., Weinreb, R. N., Zangwill, L. M. & Rosen, P. N. 2015. Longitudinal changes in quality of life and rates of progressive visual field loss in glaucoma patients. *Ophthalmology,* 122**,** 293-301.

Mehta, N., Lavinsky, F., Gattoussi, S., Seiler, M., Wald, K. J., Ishikawa, H., Wollstein, G., Schuman, J., Freund, K. B., Singh, R. & Modi, Y. 2018. Increased Inner Retinal Layer Reflectivity in Eyes With Acute CRVO Correlates With Worse Visual Outcomes at 12 Months. *Invest Ophthalmol Vis Sci,* 59**,** 3503-3510.

Michelessi, M., Li, T., Miele, A., Azuara-Blanco, A., Qureshi, R. & Virgili, G. 2020. Accuracy of optical coherence tomography for diagnosing glaucoma: an overview of systematic reviews. *British Journal of Ophthalmology***,** bjophthalmol-2020-316152.

Michelessi, M., Lucenteforte, E., Oddone, F., Brazzelli, M., Parravano, M., Franchi, S., Ng, S. M. & Virgili, G. 2015. Optic nerve head and fibre layer imaging for diagnosing glaucoma. *Cochrane Database Syst Rev***,** Cd008803.

Miki, A., Medeiros, F. A., Weinreb, R. N., Jain, S., He, F., Sharpsten, L., Khachatryan, N., Hammel, N., Liebmann, J. M., Girkin, C. A., Sample, P. A. & Zangwill, L. M. 2014. Rates of retinal nerve fiber layer thinning in glaucoma suspect eyes. *Ophthalmology,* 121**,** 1350-8.

Miura, N., Omodaka, K., Kimura, K., Matsumoto, A., Kikawa, T., Takahashi, S., Takada, N., Takahashi, H., Maruyama, K., Akiba, M., Yuasa, T. & Nakazawa, T. 2017. Evaluation of retinal nerve fiber layer defect using wide-field en-face swept-source OCT images by applying the inner limiting membrane flattening. *PLoS One,* 12**,** e0185573.

Mohammadzadeh, V., Rabiolo, A., Fu, Q., Morales, E., Coleman, A. L., Law, S. K., Caprioli, J. & Nouri-Mahdavi, K. 2020. Longitudinal Macular Structure-Function Relationships in Glaucoma. *Ophthalmology,* 127**,** 888-900.

Montesano, G., Mckendrick, A. M., Turpin, A., Brusini, P., Oddone, F., Fogagnolo, P., Perdicchi, A., Johnson, C. A., Lanzetta, P., Rossetti, L. M., Garway-Heath, D. F. & Crabb, D. P. 2021a. Do Additional Testing Locations Improve the Detection of Macular Perimetric Defects in Glaucoma? *Ophthalmology*.

Montesano, G., Rossetti, L. M., Allegrini, D., Romano, M. R. & Crabb, D. P. 2018. Improving Visual Field Examination of the Macula Using Structural Information. *Transl Vis Sci Technol,* 7**,** 36.

Montesano, G., Rossetti, L. M., Allegrini, D., Romano, M. R., Garway-Heath, D. F. & Crabb, D. P. 2021b. Systematic and Random Mapping Errors in Structure – Function Analysis of the Macula. *Translational Vision Science & Technology,* 10**,** 21-21.

Moreno-Montanes, J., Anton, V., Anton, A., Larrosa, J. M., Martinez-De-La-Casa, J. M., Rebolleda, G., Ussa, F. & Garcia-Granero, M. 2017. Intraobserver and Interobserver Agreement of Structural and Functional Software Programs for Measuring Glaucoma Progression. *JAMA Ophthalmol,* 135**,** 313-319.

Mota, M., Vaz, F. T., Ramalho, M., Pedrosa, C., Lisboa, M., Kaku, P. & Esperancinha, F. 2016. Macular Thickness Assessment in Patients with Glaucoma and Its Correlation with Visual Fields. *J Curr Glaucoma Pract,* 10**,** 85-90.

Muthusamy, V., Turpin, A., Walland, M. J., Nguyen, B. N. & Mckendrick, A. M. 2020. Increasing the Spatial Resolution of Visual Field Tests Without Increasing Test Duration: An Evaluation of ARREST. *Transl Vis Sci Technol,* 9**,** 24.

Mwanza, J. C. & Budenz, D. L. 2018. New developments in optical coherence tomography imaging for glaucoma. *Curr Opin Ophthalmol,* 29**,** 121-129.

Mwanza, J. C., Chang, R. T., Budenz, D. L., Durbin, M. K., Gendy, M. G., Shi, W. & Feuer, W. J. 2010. Reproducibility of peripapillary retinal nerve fiber layer thickness and optic nerve head parameters measured with cirrus HD-OCT in glaucomatous eyes. *Invest Ophthalmol Vis Sci,* 51**,** 5724-30.

Mwanza, J. C., Kim, H. Y., Budenz, D. L., Warren, J. L., Margolis, M., Lawrence, S. D., Jani, P. D., Thompson, G. S. & Lee, R. K. 2015. Residual and Dynamic Range of Retinal Nerve Fiber Layer Thickness in Glaucoma: Comparison of Three OCT Platforms. *Invest Ophthalmol Vis Sci,* 56**,** 6344-51.

Mwanza, J. C., Lee, G., Budenz, D. L., Warren, J. L., Wall, M., Artes, P. H., Callan, T. M. & Flanagan, J. G. 2018a. Validation of the UNC OCT Index for the Diagnosis of Early Glaucoma. *Transl Vis Sci Technol,* 7**,** 16.

Mwanza, J. C., Warren, J. L. & Budenz, D. L. 2018b. Utility of combining spectral domain optical coherence tomography structural parameters for the diagnosis of early Glaucoma: a mini-review. *Eye Vis (Lond),* 5**,** 9.

National Institute of Health and Care Excellence. 2017. *Glaucoma: diagnosis and management* [Online]. Available: https://www.nice.org.uk/guidance/ng81 [Accessed December 2020].

Nevalainen, J., Paetzold, J., Papageorgiou, E., Sample, P. A., Pascual, J. P., Krapp, E., Selig, B., Vonthein, R. & Schiefer, U. 2009. Specification of progression in glaucomatous visual field loss, applying locally condensed stimulus arrangements. *Graefes Arch Clin Exp Ophthalmol,* 247**,** 1659-69.

Nicolela, M. T. & Drance, S. M. 1996. Various glaucomatous optic nerve appearances: clinical correlations. *Ophthalmology,* 103**,** 640-9.

Nouri-Mahdavi, K. 2014. Selecting visual field tests and assessing visual field deterioration in glaucoma. *Can J Ophthalmol,* 49**,** 497-505.

Nouri-Mahdavi, K. & Caprioli, J. 2015. Measuring rates of structural and functional change in glaucoma. *Br J Ophthalmol,* 99**,** 893-8.

Nouri-Mahdavi, K., Nassiri, N., Giangiacomo, A. & Caprioli, J. 2011. Detection of visual field progression in glaucoma with standard achromatic perimetry: a review and practical implications. *Graefes Arch Clin Exp Ophthalmol,* 249**,** 1593-616.

Numata, T., Maddess, T., Matsumoto, C., Okuyama, S., Hashimoto, S., Nomoto, H. & Shimomura, Y. 2017. Exploring Test-Retest Variability Using High-Resolution Perimetry. *Transl Vis Sci Technol,* 6**,** 8.

Numata, T., Matsumoto, C., Okuyama, S., Tanabe, F., Hashimoto, S., Nomoto, H. & Shimomura, Y. 2016. Detectability of Visual Field Defects in Glaucoma With High-resolution Perimetry. *J Glaucoma,* 25**,** 847-853.

Oddone, F., Lucenteforte, E., Michelessi, M., Rizzo, S., Donati, S., Parravano, M. & Virgili, G. 2016. Macular versus Retinal Nerve Fiber Layer Parameters for Diagnosing Manifest Glaucoma: A Systematic Review of Diagnostic Accuracy Studies. *Ophthalmology,* 123**,** 939-49.

Ogden, T. E. 1983. Nerve fiber layer of the macaque retina: retinotopic organization. *Invest Ophthalmol Vis Sci,* 24**,** 85-98.

Ohkubo, S., Higashide, T., Udagawa, S., Sugiyama, K., Hangai, M., Yoshimura, N., Mayama, C., Tomidokoro, A., Araie, M., Iwase, A. & Fujimura, T. 2014. Focal relationship between structure and function within the central 10 degrees in glaucoma. *Invest Ophthalmol Vis Sci,* 55**,** 5269-77.

Olsson, U. 1979. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika,* 44**,** 443-460.

Patel, D. E., Cumberland, P. M., Walters, B. C., Russell-Eggitt, I., Rahi, J. S. & Group, O. S. 2015. Study of Optimal Perimetric Testing in Children (OPTIC): Feasibility, Reliability and Repeatability of Perimetry in Children. *PLOS ONE,* 10**,** e0130895.

Patel, N. B., Luo, X., Wheat, J. L. & Harwerth, R. S. 2011. Retinal Nerve Fiber Layer Assessment: Area versus Thickness Measurements from Elliptical Scans Centered on the Optic Nerve. *Investigative Ophthalmology & Visual Science,* 52**,** 2477-2489.

Pathak, M., Demirel, S. & Gardiner, S. K. 2017. Reducing Variability of Perimetric Global Indices from Eyes with Progressive Glaucoma by Censoring Unreliable Sensitivity Data. *Transl Vis Sci Technol,* 6**,** 11.

Phu, J., Khuu, S. K., Yapp, M., Assaad, N., Hennessy, M. P. & Kalloniatis, M. 2017. The value of visual field testing in the era of advanced imaging: clinical and psychophysical perspectives. *Clin Exp Optom,* 100**,** 313-332.

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**,** 5912-20.

Pinto, L. M., Costa, E. F., Melo, L. A., Jr., Gross, P. B., Sato, E. T., Almeida, A. P., Maia, A. & Paranhos, A., Jr. 2014. Structure-function correlations in glaucoma using matrix and standard automated perimetry versus time-domain and spectral-domain OCT devices. *Invest Ophthalmol Vis Sci,* 55**,** 3074-80.

Pons, M. E., Ishikawa, H., Gurses-Ozden, R., Liebmann, J. M., Dou, H. L. & Ritch, R. 2000. Assessment of retinal nerve fiber layer internal reflectivity in eyes with and without glaucoma using optical coherence tomography. *Arch Ophthalmol,* 118**,** 1044-7.

Porciatti, V. & Ventura, L. M. 2012. Retinal ganglion cell functional plasticity and optic neuropathy: a comprehensive model. *J Neuroophthalmol,* 32**,** 354-8.

Prum, B. E., Jr., Rosenberg, L. F., Gedde, S. J., Mansberger, S. L., Stein, J. D., Moroi, S. E., Herndon, L. W., Jr., Lim, M. C. & Williams, R. D. 2016. Primary Open-Angle Glaucoma Preferred Practice Pattern((R)) Guidelines. *Ophthalmology,* 123**,** P41-p111.

Qassim, A., Mullany, S., Abedi, F., Marshall, H., Hassall, M. M., Kolovos, A., Knight, L. S. W., Nguyen, T., Awadalla, M. S., Chappell, A., Schulz, A. M., Galanopoulos, A., Agar, A., Healey, P. R., Hewitt, A. W., Graham, S. L., Landers, J., Casson, R. J., Siggs, O. M. & Craig, J. E. 2020. Corneal Stiffness Parameters Are Predictive of Structural and Functional Progression in Glaucoma Suspect Eyes. *Ophthalmology*.

Quigley, H. A. 2018. 21st century glaucoma care. *Eye (Lond)*.

Quigley, H. A., Dunkelberger, G. R. & Green, W. R. 1989. Retinal ganglion cell atrophy correlated with automated perimetry in human eyes with glaucoma. *Am J Ophthalmol,* 107**,** 453-64.

Quigley, H. A., Katz, J., Derick, R. J., Gilbert, D. & Sommer, A. 1992. An evaluation of optic disc and nerve fiber layer examinations in monitoring progression of early glaucoma damage. *Ophthalmology,* 99**,** 19-28.

Qureshi, R., Azuara-Blanco, A., Michelessi, M., Virgili, G., Barbosa Breda, J., Cutolo, C. A., Pazos, M., Katsanos, A., Garhöfer, G., Kolko, M., Prokosch-Willing, V., Al Rajhi, A. A., Lum, F., Musch, D., Gedde, S. & Li, T. 2021. What Do We Really Know about the Effectiveness of Glaucoma Interventions?: An Overview of Systematic Reviews. *Ophthalmology Glaucoma*.

R Core Team 2020. R: A language and environment for statistical computing. R version 3.6.3 (2020-02-29) ed.: R Foundation for Statistical Computing, Vienna, Austria.

Ransohoff, D. F. & Feinstein, A. R. 1978. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med,* 299**,** 926-30.

Rao, H. L., Kumbar, T., Addepalli, U. K., Bharti, N., Senthil, S., Choudhari, N. S. & Garudadri, C. S. 2012. Effect of spectrum bias on the diagnostic accuracy of spectral-domain optical coherence tomography in glaucoma. *Invest Ophthalmol Vis Sci,* 53**,** 1058-65.

Rao, H. L., Leite, M. T., Weinreb, R. N., Zangwill, L. M., Alencar, L. M., Sample, P. A. & Medeiros, F. A. 2011. Effect of disease severity and optic disc size on diagnostic accuracy of RTVue spectral domain optical coherence tomograph in glaucoma. *Investigative ophthalmology & visual science,* 52**,** 1290-1296.

Rao, H. L., Qasim, M., Hussain, R. S., Januwada, M., Pillutla, L. N., Begum, V. U., Chaitanya, A., Senthil, S. & Garudadri, C. S. 2015. Structure-Function Relationship in Glaucoma Using Ganglion Cell-Inner Plexiform Layer Thickness Measurements. *Invest Ophthalmol Vis Sci,* 56**,** 3883-8.

Raza, A. S., Cho, J., De Moraes, C. G., Wang, M., Zhang, X., Kardon, R. H., Liebmann, J. M., Ritch, R. & Hood, D. C. 2011. Retinal ganglion cell layer thickness and local visual field sensitivity in glaucoma. *Arch Ophthalmol,* 129**,** 1529-36.

Raza, A. S., Zhang, X., De Moraes, C. G., Reisman, C. A., Liebmann, J. M., Ritch, R. & Hood, D. C. 2014. Improving glaucoma detection using spatially correspondent clusters of damage and by combining standard automated perimetry and optical coherence tomography. *Invest Ophthalmol Vis Sci,* 55**,** 612-24.

Read, R. M. & Spaeth, G. L. 1974. The practical clinical appraisal of the optic disc in glaucoma: the natural history of cup progression and some specific disc-field correlations. *Trans Am Acad Ophthalmol Otolaryngol,* 78**,** Op255-74.

Realini, T., Zangwill, L. M., Flanagan, J. G., Garway-Heath, D., Patella, V. M., Johnson, C. A., Artes, P. H., Gaddie, I. B. & Fingeret, M. 2015. Normative Databases for Imaging Instrumentation. *J Glaucoma,* 24**,** 480-3.

Reis, A. S. C., Zangalli, C. E. S., Abe, R. Y., Silva, A. L., Vianna, J. R., Vasconcellos, J. P. C. & Costa, V. P. 2017. Intra- and interobserver reproducibility of Bruch's membrane opening minimum rim width measurements with spectral domain optical coherence tomography. *Acta Ophthalmol,* 95**,** e548-e555.

Revelle, W. 2018. *psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA* [Online]. [Accessed].

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C. & Müller, M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics,* 12**,** 77.

Rolle, T., Manerba, L., Lanzafame, P. & Grignolo, F. M. 2016. Diagnostic Power of Macular Retinal Thickness Analysis and Structure-Function Relationship in Glaucoma Diagnosis Using SPECTRALIS OCT. *Curr Eye Res,* 41**,** 667-75.

Rudnicka, A. R., Mt-Isa, S., Owen, C. G., Cook, D. G. & Ashby, D. 2006. Variations in primary open-angle glaucoma prevalence by age, gender, and race: a Bayesian meta-analysis. *Invest Ophthalmol Vis Sci,* 47**,** 4254-61.

Sakamoto, M., Mori, S., Ueda, K., Kurimoto, T., Kusuhara, S., Yamada-Nakanishi, Y. & Nakamura, M. 2019. En Face Slab Images Visualize Nerve Fibers With Residual Visual Sensitivity in Significantly Thinned Macular Areas of Advanced Glaucomatous Eyes. *Invest Ophthalmol Vis Sci,* 60**,** 2811-2821.

Sample, P. A., Medeiros, F. A., Racette, L., Pascual, J. P., Boden, C., Zangwill, L. M., Bowd, C. & Weinreb, R. N. 2006. Identifying glaucomatous vision loss with visual-function-specific perimetry in the diagnostic innovations in glaucoma study. *Invest Ophthalmol Vis Sci,* 47**,** 3381-9.

Sato, S., Hirooka, K., Baba, T., Tenkumo, K., Nitta, E. & Shiraga, F. 2013. Correlation between the ganglion cell-inner plexiform layer thickness measured with cirrus HD-OCT and macular visual field sensitivity measured with microperimetry. *Invest Ophthalmol Vis Sci,* 54**,** 3046-51.

Saunders, L. J., Zhu, H., Bunce, C., Doré, C. J., Freemantle, N. & Crabb, D. P. 2015. Ophthalmic statistics note 5: diagnostic tests—sensitivity and specificity. *British Journal of Ophthalmology,* 99**,** 1168-1170.

Sayed, M. S., Margolis, M. & Lee, R. K. 2017. Green disease in optical coherence tomography diagnosis of glaucoma. *Curr Opin Ophthalmol,* 28**,** 139-153.

Schiefer, U., Flad, M., Stumpp, F., Malsam, A., Paetzold, J., Vonthein, R., Denk, P. O. & Sample, P. A. 2003. Increased detection rate of glaucomatous visual field damage with locally condensed grids: a comparison between fundus-oriented perimetry and conventional visual field examination. *Arch Ophthalmol,* 121**,** 458-65.

Schiefer, U., Patzold, J. & Dannheim, F. 2005. [Conventional perimetry I: introduction--basics]. *Ophthalmologe,* 102**,** 627-44; quiz 645-6.

Schulzer, M., Wallace Alward, F Feldman & Nen, A. 1998. Comparison of glaucomatous progression between untreated patients with normal-tension glaucoma and patients with therapeutically reduced intraocular pressures. Collaborative Normal-Tension Glaucoma Study Group. *Am J Ophthalmol,* 126**,** 487-97.

Scott, D. W. 1992. Kernel Density Estimators. *Multivariate Density Estimation.*

Shigueoka, L. S., Vasconcellos, J. P. C., Schimiti, R. B., Reis, A. S. C., Oliveira, G. O., Gomi, E. S., Vianna, J. a. R., Lisboa, R., Medeiros, F. A. & Costa, V. P. 2018. Automated algorithms combining structure and function outperform general ophthalmologists in diagnosing glaucoma. *PLoS One,* 13**,** e0207784.

Shin, H. Y., Park, H. Y., Jung, K. I. & Park, C. K. 2013. Comparative study of macular ganglion cell-inner plexiform layer and peripapillary retinal nerve fiber layer measurement: structure-function analysis. *Invest Ophthalmol Vis Sci,* 54**,** 7344-53.

Shirato, S., Inoue, R., Fukushima, K. & Suzuki, Y. 1999. Clinical evaluation of SITA: a new family of perimetric testing strategies. *Graefes Arch Clin Exp Ophthalmol,* 237**,** 29-34.

Shrout, P. E. & Fleiss, J. L. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull,* 86**,** 420-8.

Simel, D. L., Samsa, G. P. & Matchar, D. B. 1991. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol,* 44**,** 763-70.

Sjöstrand, J., Popovic, Z., Conradi, N. & Marshall, J. 1999. Morphometric study of the displacement of retinal ganglion cells subserving cones within the human fovea. *Graefes Arch Clin Exp Ophthalmol,* 237**,** 1014-23.

Soh, Z. D., Yu, M., Betzler, B. K., Majithia, S., Thakur, S., Tham, Y. C., Wong, T. Y., Aung, T., Friedman, D. S. & Cheng, C.-Y. 2021. The global extent of undetected glaucoma in adults: a systematic review and meta-analysis. *Ophthalmology*.

Spitzer, R. L. & Fleiss, J. L. 1974. A re-analysis of the reliability of psychiatric diagnosis. *Br J Psychiatry,* 125**,** 341-7.

Stagg, B. C. & Medeiros, F. A. 2020. A Comparison of OCT Parameters in Identifying Glaucoma Damage in Eyes Suspected of Having Glaucoma. *Ophthalmology Glaucoma,* 3**,** 90-96.

Steinmetz, J. D., Bourne, R. R. A., Briant, P. S., Flaxman, S. R., Taylor, H. R. B., Jonas, J. B., Collaborators, G. B. a. V. I. & Study. 2021. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. *Lancet Glob Health,* 9**,** e144-e160.

Stevenson, M. & Sergeant, E. 2021. *epiR: Tools for the Analysis of Epidemiological Data. Version 2.0.26* [Online]. Available: https://cran.r-project.org/package=epiR [Accessed].

Strouthidis, N. G., Vinciotti, V., Tucker, A. J., Gardiner, S. K., Crabb, D. P. & Garway-Heath, D. F. 2006. Structure and function in glaucoma: The relationship between a functional visual field map and an anatomic retinal map. *Invest Ophthalmol Vis Sci,* 47**,** 5356-62.

Strouthidis, N. G., Yang, H., Fortune, B., Downs, J. C. & Burgoyne, C. F. 2009. Detection of optic nerve head neural canal opening within histomorphometric and spectral domain optical coherence tomography data sets. *Invest Ophthalmol Vis Sci,* 50**,** 214-23.

Sung, M. S., Heo, H. & Park, S. W. 2019. Structure-function Relationship in Advanced Glaucoma After Reaching the RNFL Floor. *J Glaucoma,* 28**,** 1006-1011.

Susanna, B. N., Ogata, N. G., Jammal, A. A., Susanna, C. N., Berchuck, S. I. & Medeiros, F. A. 2019. Corneal Biomechanics and Visual Field Progression in Eyes with Seemingly Well-Controlled Intraocular Pressure. *Ophthalmology,* 126**,** 1640-1646.

Susanna, R., Jr., Nicolela, M. T., Soriano, D. S. & Carvalho, C. 1994. Automated perimetry: a study of the glaucoma hemifield test for the detection of early glaucomatous visual field loss. *J Glaucoma,* 3**,** 12-6.

Tan, O., Liu, L., You, Q., Wang, J., Chen, A., Ing, E., Morrison, J. C., Jia, Y. & Huang, D. 2021. Focal Loss Analysis of Nerve Fiber Layer Reflectance for Glaucoma Diagnosis. *Translational Vision Science & Technology,* 10**,** 9-9.

Tan, O., Liu, L., You, Q., Wang, J., Jia, Y. & Huang, D. 2020. Focal Loss Analysis of Nerve Fiber Layer Reflectance for Glaucoma Diagnosis. *Investigative Ophthalmology & Visual Science,* 61**,** 5194-5194.

Tappeiner, C., Barthelmes, D., Abegg, M. H., Wolf, S. & Fleischhauer, J. C. 2008. Impact of optic media opacities and image compression on quantitative analysis of optical coherence tomography. *Invest Ophthalmol Vis Sci,* 49**,** 1609-14.

Tatham, A. J. & Medeiros, F. A. 2017. Detecting Structural Progression in Glaucoma with Optical Coherence Tomography. *Ophthalmology,* 124**,** S57-s65.

Tatham, A. J., Medeiros, F. A., Zangwill, L. M. & Weinreb, R. N. 2015. Strategies to improve early diagnosis in glaucoma. *Prog Brain Res,* 221**,** 103-33.

Tatham, A. J., Weinreb, R. N. & Medeiros, F. A. 2014. Strategies for improving early detection of glaucoma: the combined structure-function index. *Clin Ophthalmol,* 8**,** 611-21.

Tham, Y. C., Li, X., Wong, T. Y., Quigley, H. A., Aung, T. & Cheng, C. Y. 2014. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology,* 121**,** 2081-90.

Tharwat, A. 2018. Classification assessment methods. *Applied Computing and Informatics*.

The College of Optometrists. 2018. *Glaucoma (primary open angle) (POAG)* [Online]. Available: https://www.college-optometrists.org/guidance/clinical-management-guidelines/glaucoma-primary-open-angle-poag-.html [Accessed November 2018].

Thepass, G., Lemij, H. G. & Vermeer, K. A. 2017. Attenuation Coefficients From SD-OCT Data: Structural Information Beyond Morphology on RNFL Integrity in Glaucoma. *J Glaucoma,* 26**,** 1001-1009.

Ting, D. S. W., Pasquale, L. R., Peng, L., Campbell, J. P., Lee, A. Y., Raman, R., Tan, G. S. W., Schmetterer, L., Keane, P. A. & Wong, T. Y. 2019. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol,* 103**,** 167-175.

Traquair, H. M. 1939. Clinical Detection of Early Changes in the Visual Field. *Trans Am Ophthalmol Soc,* 37**,** 158-79.

Trevethan, R. 2017. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. *Front Public Health,* 5**,** 307.

Tsamis, E., Bommakanti, N. K., Sun, A., Thakoor, K. A., De Moraes, C. G. & Hood, D. C. 2020. An Automated Method for Assessing Topographical Structure–Function Agreement in Abnormal Glaucomatous Regions. *Translational Vision Science & Technology,* 9**,** 14-14.

Turalba, A. V. & Grosskreutz, C. 2010. A review of current technology used in evaluating visual function in glaucoma. *Semin Ophthalmol,* 25**,** 309-16.

Turpin, A., Artes, P. H. & Mckendrick, A. M. 2012. The Open Perimetry Interface: an enabling tool for clinical visual psychophysics. *J Vis,* 12.

Turpin, A., Chen, S., Sepulveda, J. A. & Mckendrick, A. M. 2015. Customizing Structure-Function Displacements in the Macula for Individual Differences. *Invest Ophthalmol Vis Sci,* 56**,** 5984-9.

Turpin, A. & Mckendrick, A. M. 2011. What reduction in standard automated perimetry variability would improve the detection of visual field progression? *Invest Ophthalmol Vis Sci,* 52**,** 3237-45.

Turpin, A. & Mckendrick, A. M. 2021. Improving Personalized Structure to Function Mapping From Optic Nerve Head to Visual Field. *Translational Vision Science & Technology,* 10**,** 19-19.

Turpin, A., Morgan, W. H. & Mckendrick, A. M. 2018. Improving Spatial Resolution and Test Times of Visual Field Testing Using ARREST. *Transl Vis Sci Technol,* 7**,** 35.

Turpin, A., Myers, J. S. & Mckendrick, A. M. 2016. Development of Visual Field Screening Procedures: A Case Study of the Octopus Perimeter. *Transl Vis Sci Technol,* 5**,** 3.

Tuulonen, A. & Airaksinen, P. J. 1991. Initial glaucomatous optic disk and retinal nerve fiber layer abnormalities and their progression. *American journal of ophthalmology,* 111**,** 485-490.

Uebersax, J. S. 1987. Diversity of decision-making models and the measurement of interrater agreement. *Psychological bulletin,* 101**,**140.

Van Der Schoot, J., Reus, N. J., Garway-Heath, D. F., Saarela, V., Anton, A., Bron, A. M., Faschinger, C., Hollo, G., Iester, M., Jonas, J. B., Topouzis, F., Zeyen, T. G. & Lemij, H. G. 2013. Accuracy of matching optic discs with visual fields: the European Structure and Function Assessment Trial (ESAFAT). *Ophthalmology,* 120**,** 2470-2475.

Van Der Schoot, J., Vermeer, K. A., De Boer, J. F. & Lemij, H. G. 2012. The effect of glaucoma on the optical attenuation coefficient of the retinal nerve fiber layer in spectral domain optical coherence tomography images. *Invest Ophthalmol Vis Sci,* 53**,** 2424-30.

Varma, R., Skaf, M. & Barron, E. 1996. Retinal nerve fiber layer thickness in normal human eyes. *Ophthalmology,* 103**,** 2114-9.

Venables, W. & Ripley, B. 2002. Modern Applied Statistics with S. New York: Springer.

Vermeer, K. A., Mo, J., Weda, J. J., Lemij, H. G. & De Boer, J. F. 2013. Depth-resolved model-based reconstruction of attenuation coefficients in optical coherence tomography. *Biomed Opt Express,* 5**,** 322-37.

Vermeer, K. A., Schoot, J. V. D., Lemij, H. G. & Boer, J. F. D. 2011. OCT-derived Retinal Nerve Fiber Layer Reflectivity Maps for Glaucoma Assessment. *Investigative Ophthalmology & Visual Science,* 52**,** 3666-3666.

Vermeer, K. A., Van Der Schoot, J., Lemij, H. G., De Boer, J. F. J. I. O. & Science, V. 2012. RPE-normalized RNFL attenuation coefficient maps derived from volumetric OCT imaging for glaucoma assessment. 53**,** 6102-6108.

Verticchio Vercellin, A. C., Jassim, F., Poon, L. Y., Tsikata, E., Braaf, B., Shah, S., Ben-David, G., Shieh, E., Lee, R., Simavli, H., Que, C. J., Papadogeorgou, G., Guo, R., Vakoc, B. J., Bouma, B. E., De Boer, J. F. & Chen, T. C. 2018. Diagnostic Capability of Three-Dimensional Macular Parameters for Glaucoma Using Optical Coherence Tomography Volume Scans. *Invest Ophthalmol Vis Sci,* 59**,** 4998-5010.

Vianna, J. R. & Chauhan, B. C. 2015. How to detect progression in glaucoma. *Prog Brain Res,* 221**,** 135-58.

Virgili, G., Michelessi, M., Cook, J., Boachie, C., Burr, J., Banister, K., Garway-Heath, D. F., Bourne, R. R. A., Asorey Garcia, A., Ramsay, C. R. & Azuara-Blanco, A. 2018. Diagnostic accuracy of optical coherence tomography for diagnosing glaucoma: secondary analyses of the GATE study. *Br J Ophthalmol,* 102**,** 604-610.

Wagner, F. M., Hoffmann, E. M., Nickels, S., Fiess, A., Münzel, T., Wild, P. S., Beutel, M. E., Schmidtmann, I., Lackner, K. J., Pfeiffer, N. & Schuster, A. K.-G. 2020. Peripapillary Retinal Nerve Fiber Layer Profile in Relation to Refractive Error and Axial Length: Results From the Gutenberg Health Study. *Translational Vision Science & Technology,* 9**,** 35-35.

Wall, M., Woodward, K. R., Doyle, C. K. & Zamba, G. 2010. The effective dynamic ranges of standard automated perimetry sizes III and V and motion and matrix perimetry. *Arch Ophthalmol,* 128**,** 570-6.

Wang, M., Elze, T., Li, D., Baniasadi, N., Wirkner, K., Kirsten, T., Thiery, J., Löffler, M., Engel, C. & Rauscher, F. G. 2017. Age, ocular magnification, and circumpapillary retinal nerve fiber layer thickness. *Journal of biomedical optics,* 22**,** 121718.

Weinreb, R. N., Aung, T. & Medeiros, F. A. 2014. The pathophysiology and treatment of glaucoma: a review. *Jama,* 311**,** 1901-11.

Weinreb, R. N., Leung, C. K., Crowston, J. G., Medeiros, F. A., Friedman, D. S., Wiggs, J. L. & Martin, K. R. 2016. Primary open-angle glaucoma. *Nat Rev Dis Primers,* 2**,** 16067.

White, B., Pierce, M., Nassif, N., Cense, B., Park, B., Tearney, G., Bouma, B., Chen, T. & De Boer, J. 2003. In vivo dynamic human retinal blood flow imaging using ultra-high-speed spectral domain optical coherence tomography. *Opt Express,* 11**,** 3490-7.

Wild, J. M., Pacey, I. E., Hancock, S. A. & Cunliffe, I. A. 1999. Between-algorithm, between-individual differences in normal perimetric sensitivity: full threshold, FASTPAC, and SITA. Swedish Interactive Threshold algorithm. *Investigative Ophthalmology & Visual Science,* 40**,** 1152-1161.

Wild, J. M., Searle, A. E., Dengler-Harles, M. & O'neill, E. C. 1991. Long-term follow-up of baseline learning and fatigue effects in the automated perimetry of glaucoma and ocular hypertensive patients. *Acta Ophthalmol (Copenh),* 69**,** 210-6.

Wollstein, G., Schuman, J. S., Price, L. L., Aydin, A., Beaton, S. A., Stark, P. C., Fujimoto, J. G. & Ishikawa, H. 2004. Optical coherence tomography (OCT) macular and peripapillary retinal nerve fiber layer measurements and automated visual fields. *Am J Ophthalmol,* 138**,** 218-25.

World Glaucoma Association 2016. *10th Consensus Meeting: Diagnosis of Primary Open Angle Glaucoma.*

Wu, Z., Cimetta, R., Caruso, E. & Guymer, R. H. 2019. Performance of a Defect-Mapping Microperimetry Approach for Characterizing Progressive Changes in Deep Scotomas. *Transl Vis Sci Technol,* 8**,** 16.

Wu, Z. & Medeiros, F. A. 2018. Recent developments in visual field testing for glaucoma. *Curr Opin Ophthalmol,* 29**,** 141-146.

Wu, Z., Medeiros, F. A., Weinreb, R. N., Girkin, C. A. & Zangwill, L. M. 2020a. Specificity of various cluster criteria used for the detection of glaucomatous visual field abnormalities. *Br J Ophthalmol,* 104**,** 822-826.

Wu, Z., Saunders, L. J., Zangwill, L. M., Daga, F. B., Crowston, J. G. & Medeiros, F. A. 2017. Impact of Normal Aging and Progression Definitions on the Specificity of Detecting Retinal Nerve Fiber Layer Thinning. *Am J Ophthalmol,* 181**,** 106-113.

Wu, Z., Vianna, J. R., Reis, A. S. C., Zemborain, Z. Z., Lee, S. H., Thenappan, A., Weng, D. S. D., Tsamis, E., Joiner, D. B., Ritch, R., De Moraes, C. G. V. & Hood, D. C. 2020b. Qualitative evaluation of neuroretinal rim and retinal nerve fibre layer on optical coherence tomography to detect glaucomatous damage. *Br J Ophthalmol,* 104**,** 980-984.

Wudunn, D., Takusagawa, H. L., Sit, A. J., Rosdahl, J. A., Radhakrishnan, S., Hoguet, A., Han, Y. & Chen, T. C. 2021. OCT Angiography for the Diagnosis of Glaucoma: A Report by the American Academy of Ophthalmology. *Ophthalmology.*

Xiong, Y. & Pulli, K. 2010. Color matching for high-quality panoramic images on mobile phones. *IEEE Transactions on Consumer Electronics,* 56**,** 2592-2600.

Xu, G., Weinreb, R. N. & Leung, C. K. S. 2013. Retinal nerve fiber layer progression in glaucoma: a comparison between retinal nerve fiber layer thickness and retardance. *Ophthalmology,* 120**,** 2493-2500.

Yamamoto, S., Sawaguchi, S., Iwase, A., Yamamoto, T., Abe, H., Tomita, G., Tomidokoro, A. & Araie, M. 2014. Primary open-angle glaucoma in a population associated with high prevalence of primary angle-closure glaucoma: the Kumejima Study. *Ophthalmology,* 121**,** 1558-65.

Yang, Q., Reisman, C. A., Wang, Z., Fukuma, Y., Hangai, M., Yoshimura, N., Tomidokoro, A., Araie, M., Raza, A. S., Hood, D. C. & Chan, K. 2010. Automated layer segmentation of macular OCT images using dual-scale gradient information. *Opt Express,* 18**,** 21293-307.

Yohannan, J. & Boland, M. V. 2017. The Evolving Role of the Relationship between Optic Nerve Structure and Function in Glaucoma. *Ophthalmology,* 124**,** S66-s70.

Yu, M., Lin, C., Weinreb, R. N., Lai, G., Chiu, V. & Leung, C. K. 2016. Risk of Visual Field Progression in Glaucoma Patients with Progressive Retinal Nerve Fiber Layer Thinning: A 5-Year Prospective Study. *Ophthalmology,* 123**,** 1201-10.

Zhao, D., Cho, J., Kim, M. H., Friedman, D. S. & Guallar, E. 2015. Diabetes, fasting glucose, and the risk of glaucoma: a meta-analysis. *Ophthalmology,* 122**,** 72-8.

Appendix A – Supplementary material

Supplementary File 2.1

Animation showing an example of the grading task, including the superimposed grid as per Figure 2.4. The animation cycles through one of the presentations provided to the clinicians performing the grading task and shows how visibility of retinal nerve fibre bundles (RNFBs) changes with depth from the inner limiting membrane (ILM). The video of the task can be accessed via the supporting information section in Cheloni and Denniss (2021) paper, at the bottom of the page at this link:

https://onlinelibrary.wiley.com/doi/full/10.1111/opo.12756

Supplementary File 2.2

Written instruction provided to each clinician before the grading task (next page).

Optimisation of enface OCT images for visualisation of retinal nerve fibre bundles

UNIVERSITY of BRADFORD

Principal Investigator: Jonathan Denniss; *PhD Student*: Riccardo Cheloni

Thank you for participating in our study. Enface visualisation of retinal nerve fibre bundles (RNFBs) in glaucoma has been recently enabled from OCT developments. However, the optimum depth at which to visualise RNFBs across the retina and across different patients is unknown. The objective of this study is to identify the optimal depth/s for the observation of RNFBs in healthy eyes in different regions of the retina.

Images

OCT enface images showing the central retina (roughly 45°) of 10 healthy adults will be shown. For every participant we will present 50 slab images, of fixed thickness (≈4μm), corresponding to an increasing depth from the inner limiting membrane (ILM) up to 190μm below it. Enface images are in greyscale, where RNFBs are observable as hyper-reflective bundles originating from the optic nerve head (ONH).

RNFBs' visibility throughout the retina changes with depth, being observable across almost the whole central retina at lower depths (fig below - 19μm) whereas only preserved around the ONH at greater depths
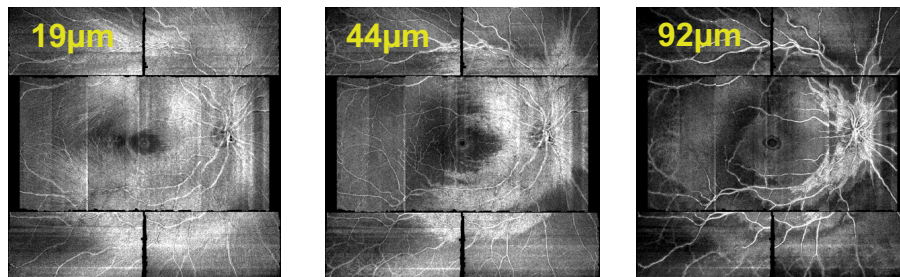
Hyper-reflective patches at superficial depths (i.e., immediately below the ILM) might be ascribable to gliosis retinal alterations. Although similar in term of intensity, these artefacts do not represent RNFBs (figure below left panel: some of the glial alterations are boxed in yellow). Similarly, reaching greater depth, the retinal pigment epithelium is encountered too, resulting in a significant increase of reflectivity.
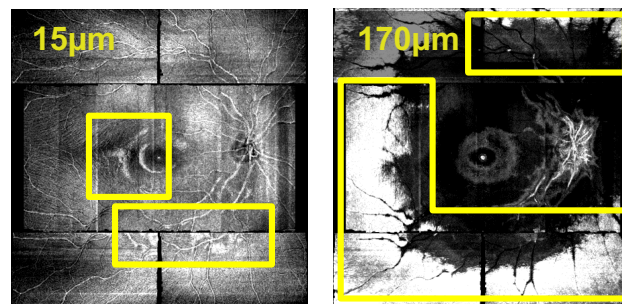
All the images will be presented in a powerpoint-like file in order to facilitate sequential observation. RNFBs at different depths can be easily seen moving from one slide to another with the upper arrow (reducing depth below the ILM) and lower arrow (increasing depth). A number – describing the corresponding depth (μm) – and an identifier of different eyes will be reported on the side of each image.

A grid (10 squares, roughly 15°x15° each) delimitating the retinal regions of interest will be superimposed to each image. This identifies 10 areas of the retina (within blue dashed line) along which your judgment will be required (see image below) Areas of overlap between regions are boxed with red dashed line and shaded in grey - please ignore these regions in performing your judgment.
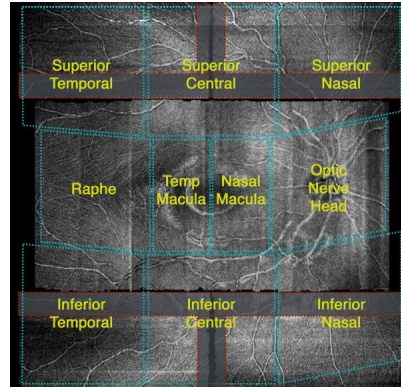
Task

0. Please answer the 6 questions presented in the first page (all fields required) of the spread sheet given and use it to report your data in every relevant cell for the different analysed eyes.

Analyse one eye at time (following the indicated order) and scroll – unlimited times as required – across different depths to observe RNFBs. Express your subjective judgment about RNFBs' visibility across every retinal sector for every of the 10 healthy eyes, answering the questions:

1. *What are the boundaries of visible RNFBs?*
   Superior limit (µm): first depth below the ILM at which RNFBs become visible in that specific sector; **Inferior limit** (µm): last depth below the ILM at which RNFBs are last visible (bundles are no more visible at the consecutive depth) in that specific sector. In doing such judgements consider to be RNFBs to be present/absent when they will occupy ¼ (25%) of the sector's area.

2. *At which depth are RNFBs most visible in each sector?*
   Report the **depth of greatest visibility** (µm) or '**equal**' in the case of equal visibility of RNFBs across the boundaries in which bundles were visible. In doing such judgments ignore the proportion of the sector occupied from RNFBs, focussing instead on the intensity and sharpness of RNFBs.
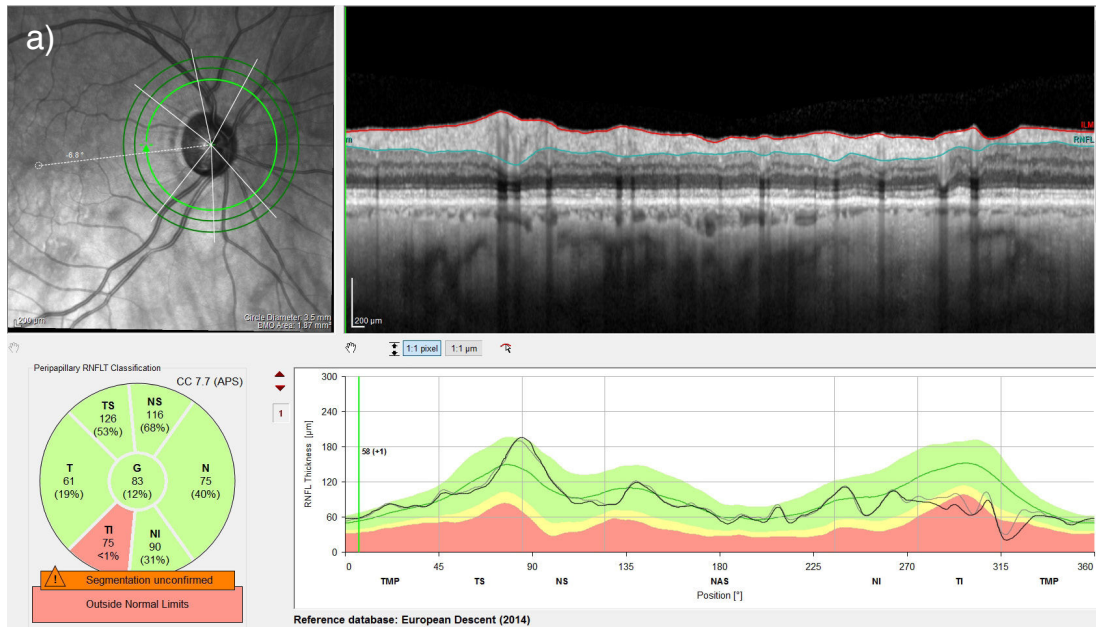
Supplementary File 3.1

Animated version of Figure 3.1, showing how visible presence of RNFBs changes at different depths in healthy and glaucoma eyes. The video can be accessed via the "supplements" section in Cheloni et al. (2021) paper, at the top of the page at this link:
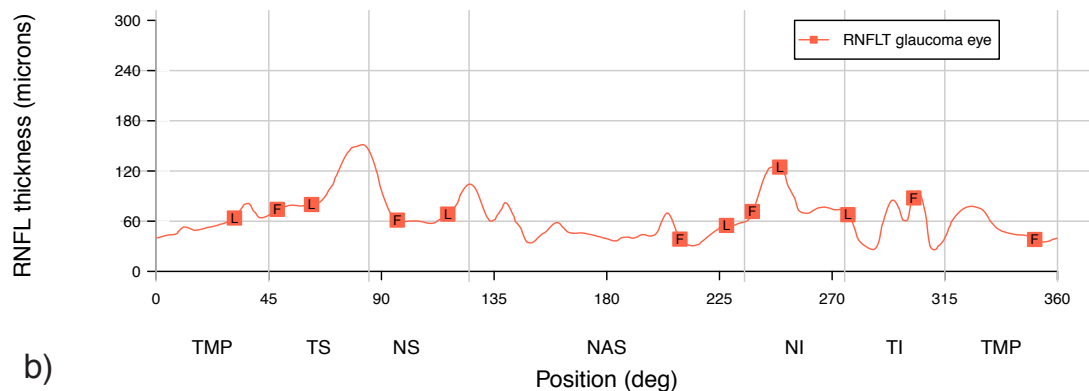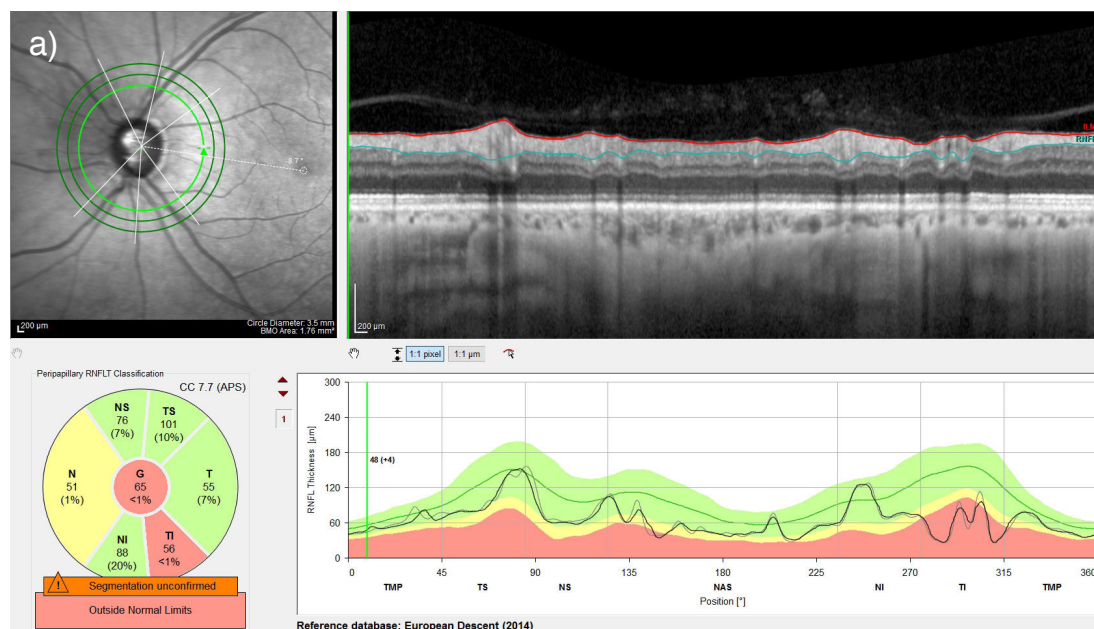
https://tvst.arvojournals.org/article.aspx?articleid=2772641

Supplementary Figure 3.2

Details of an eye with glaucoma reported as an outlier in Figure 3.6, presenting a first gap of visible bundle in the temporal inferior sector at 0µm below ILM and apparent preserved RNFL thickness at corresponding angles. (a) shows the conventional cpRNFL thickness report from Spectralis, whereas (b) shows RNFL thickness at corresponding angles of en face parameters, as per Figure 3.3. This specific participant was 77 years old, had SAP MD of -3.3dB, and was diagnosed with glaucoma 12 years before testing. Substantial RNFL thinning was present at the border between the temporal inferior and temporal sectors, and we speculate that angular incongruences rather than genuine differences might have caused the observed differences (next page).

a)

Peripapillary RNFLT Classification

CC 7.7 (APS)

TS 126 (53%)	NS 116 (68%)
T 61 (19%)	G 83 (12%)	N 75 (40%)
TI 75 <1%	NI 90 (31%)

⚠ Segmentation unconfirmed

Outside Normal Limits

Reference database: European Descent (2014)

b)
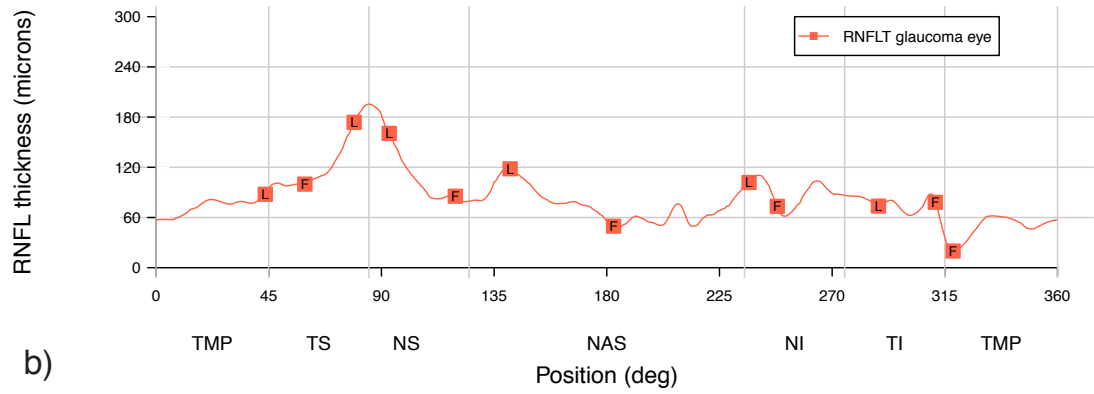
Supplementary Figure 3.3

Details of an eye with glaucoma reported as an outlier in Figure 3.6, presenting a first gap of visible bundle in the temporal inferior sector at 0µm below the ILM and apparent preserved RNFL thickness at corresponding angles. Panels arrangement as per Supplementary Figure 3.2. This specific participant was 61 years old, had SAP MD of -4.0dB, and was diagnosed with glaucoma 1 year before testing. Substantial RNFL thinning was present in the temporal inferior sector, and as per Supplementary Figure 3.2 we speculate that angular incongruences rather than genuine differences might have caused the observed differences.

Supplementary Figure 3.4

Details of an eye with glaucoma reported as an outlier in Figure 3.6, presenting a first gap of visible bundle in the temporal inferior sector at 0µm below the ILM and apparent preserved RNFL thickness at corresponding angles. Panels arrangement as per Supplementary Figure 3.2. This specific participant was 67 years old, had SAP MD of -5.9dB, and was diagnosed with glaucoma 10 years before testing. At the temporal inferior sector, RNFL remains as thin as 50µm at the most abnormal location.
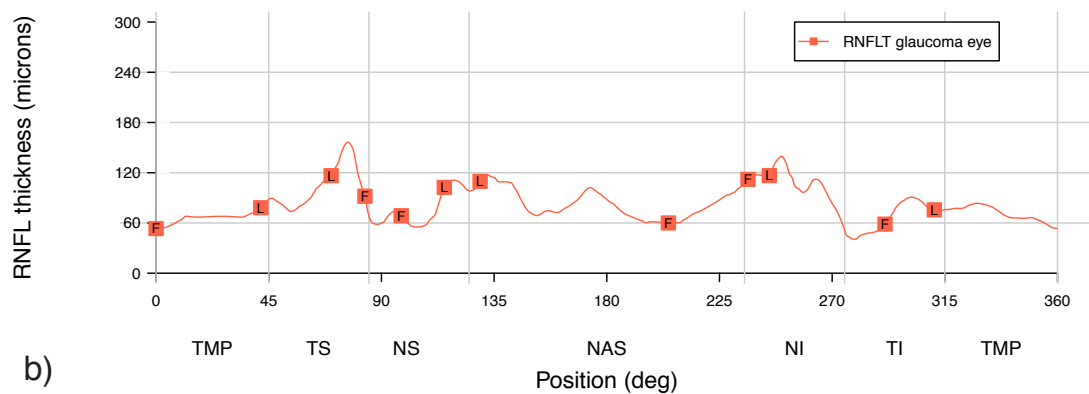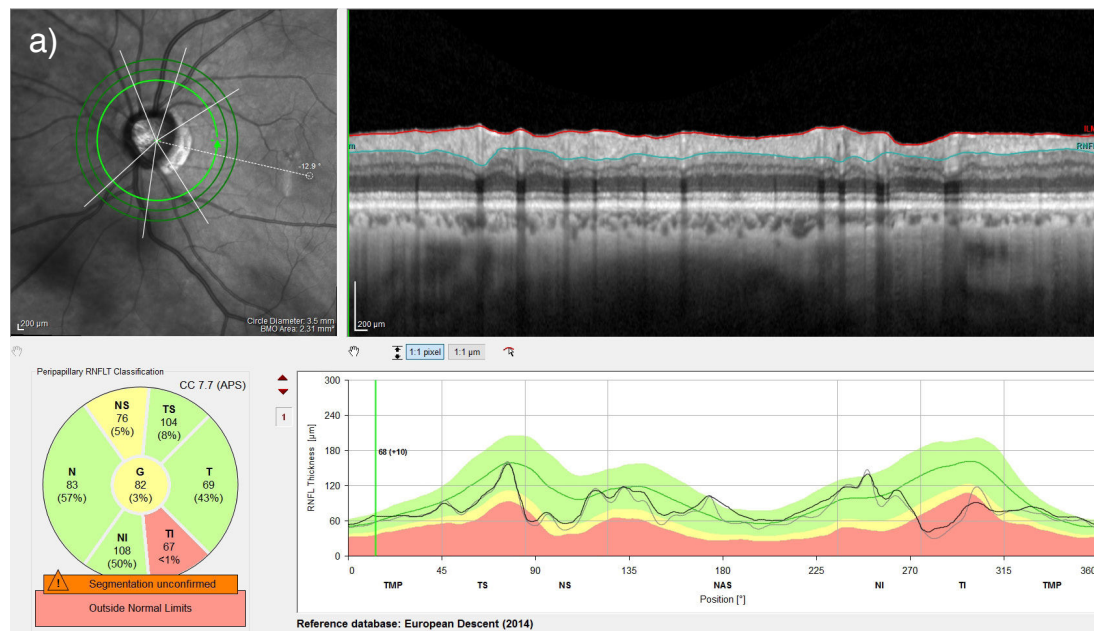
Supplementary Figure 4.1

Estimated frequency distribution of superpixel en face intensities from all control data at different retinal locations for the first five SMAS slabs. Distributions on different levels correspond to different SMAS slabs (e.g. slab #1, 8 to 23µm below ILM, see Table 4.1) and are colour-coded accordingly.
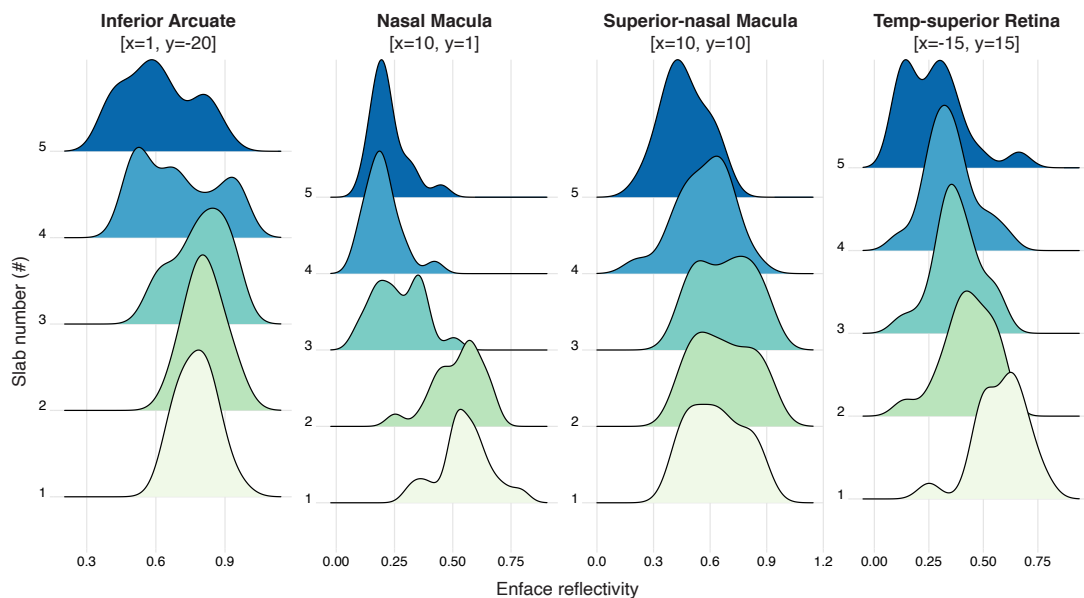
Supplementary Material 5.1

R code snippet used in custom perimetry (OPI) to select the variable response window of the 'invisible' stimulus in custom perimetry:

responseWindow = sample(c(1,1,1,1,400,400,500,700,800,1000),1)

The script above resulted in data consistent with a Poisson distribution, mode: 1 ms (median: 400 ms, IQR: 699 ms).

Supplementary Figure 5.2

Correspondence between en face predictions of visual function and custom perimetry results. Formatting is consistent with Figure 5.8. Locations were colour coded per the observed status for each test domain (blue: undamaged location; red: defect) and locations where there was agreement are flagged with asterisks (continues on next page).
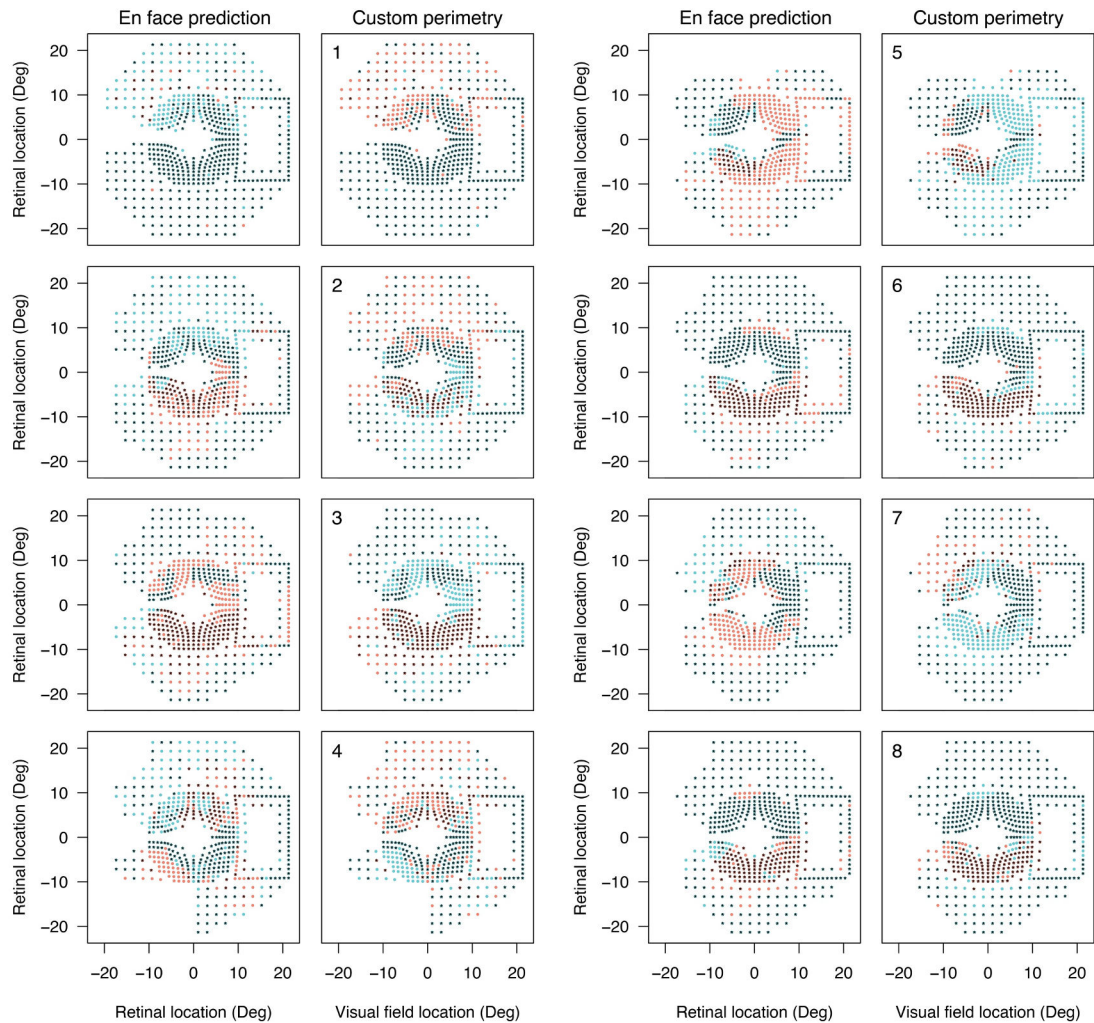
Supplementary Figure 5.2

Continued.
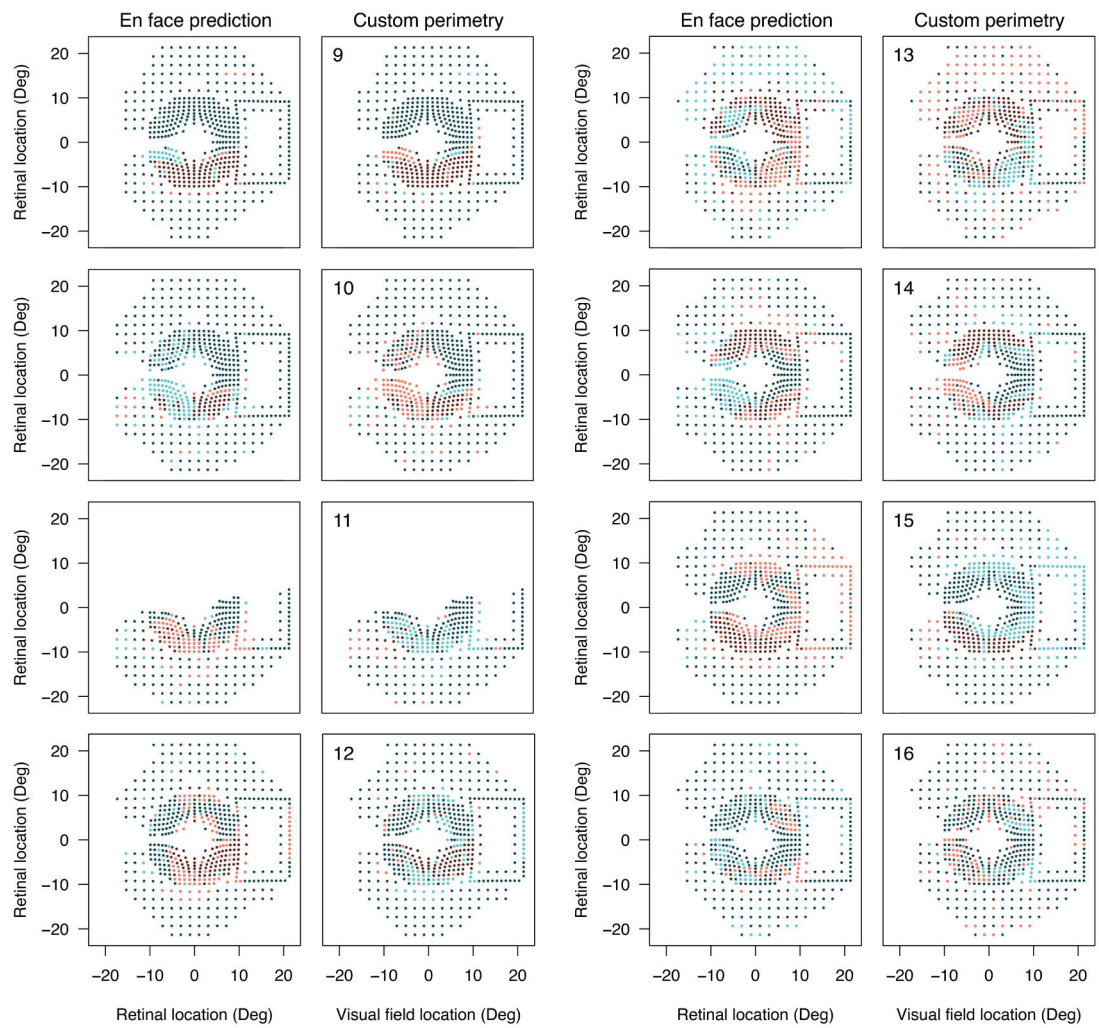
Appendix B – Conference presentations

Cheloni R & Denniss J (2019) *Optimisation of enface OCT images for visualisation of retinal nerve fibre bundles: A pilot study.* British Congress of Optometry and Vision Science, Manchester. (Poster)

Cheloni R & Denniss J (2019) *Optimisation of enface OCT images for visualisation of retinal nerve fibre bundles: A pilot study.* Faculty of Life Science PGR Symposium, University of Bradford. (Paper)

Cheloni R & Denniss J (2020) *Simple subjective evaluation of enface OCT reflectance images can be used to reliably identify RNFL loss in glaucoma.* British Congress of Optometry and Vision Science, held remotely. (Poster)

Cheloni R & Denniss J (2020) *Simple subjective evaluation of enface OCT reflectance images can be used to reliably identify RNFL loss in glaucoma.* Faculty of Life Science PGR Symposium, University of Bradford, held remotely. (Paper)

Cheloni R & Denniss J (2021) *Ability of different enface OCT slab methods to detect RNFL reflectance defects in glaucoma*. ARVO annual conference, held remotely. (Poster)

Cheloni R & Denniss J (2021) *Concordance of retinal nerve fibre reflectance loss in enface OCT images with conventional structural and functional measures of glaucoma damage*. British Congress of Optometry and Vision Science, held remotely. (Paper)

Cheloni R & Denniss J (2021) *Concordance of retinal nerve fibre reflectance loss in enface OCT images with conventional structural and functional measures of glaucoma damage.* Faculty of Life Science PGR Symposium, University of Bradford, UK. (Paper)

Appendix C – List of peer-reviewed publications

Cheloni R & Denniss J (2020) Use of optical coherence tomography in the detection and diagnosis of glaucoma. Optometry in Practice. 21, 1-12.

Cheloni R & Denniss J (2020) Depth-resolved variations in visibility of retinal nerve fibre bundles across the retina in enface OCT images of healthy eyes. Ophthalmic Physiol Opt. 41: 179–191.
Doi: https://doi.org/10.1111/opo.12756

Cheloni, R., Dewsbery, S. D., & Denniss, J (2021) A simple subjective evaluation of enface OCT reflectance images distinguishes glaucoma from healthy eyes. Translational Vision Science & Technology. 10(6): 31.
Doi: https://doi.org/10.1167/tvst.10.6.31

Cheloni, R., Dewsbery, S. D., & Denniss, J (2021) Enhanced objective detection of retinal nerve fibre bundle defects in glaucoma with a novel method for en face OCT slab image construction and analysis. Translational Vision Science & Technology. 10(12): 1.
Doi: https://doi.org/10.1167/tvst.10.12.1

THE COLLEGE OF OPTOMETRISTS

Professional Excellence in Eye Health

Optometry in Practice

Use of optical coherence tomography in the detection and diagnosis of glaucoma

Riccardo Cheloni BSc
Jonathan Denniss PhD MCOptom

School of Optometry and Vision Science, University of Bradford

General Optical Council
Approved CET
For Optometrists

STANDARDS OF PRACTICE   COMMUNICATION   OCULAR EXAMINATION   OCULAR DISEASE

Abstract

Optical coherence tomography (OCT) enables more objective assessment of the retinal structures relevant to glaucoma than conventional ophthalmoscopic techniques, and is rapidly becoming a cornerstone in the detection and diagnosis of glaucoma. Concurrently, OCT devices are becoming increasingly available in optometric practices, thus offering potential for enhanced accuracy in case identification. In this review, we discuss recent findings on the uses and limitations of OCT in glaucoma detection and diagnosis. Evidence suggests that OCT is still imperfect in the detection of early cases following a single examination, though OCT indices are likely to outperform clinicians' subjective optic nerve head assessments in detecting glaucoma. A few strategies, such as acquiring longitudinal data to gain evidence of progressive structural damage, and examining multiple retinal areas to confirm corresponding damage across several structures, can improve confidence in diagnosis and prediction of future functional loss. Even though it remains essential to combine information from a battery of clinical tests in the assessment of patients with or at risk of glaucoma, the liberal use of OCT, where available, is advocated.

Introduction

The term glaucoma encompasses a heterogeneous group of progressive optic neuropathies, all associated with characteristic structural changes and loss of visual sensitivity due to damage to the retinal ganglion cells (RGCs).[1,2] Primary open-angle glaucoma (POAG) is the most common type of glaucoma in Caucasian populations,[3] affecting around 2% of people in Europe[4] but with much higher prevalence in older age groups and some other ethnicities.[5] Despite advances in diagnosis and treatment, glaucoma remains one of the leading causes of irreversible blindness globally.[6,7]

Although the diagnosis of moderate to advanced cases of glaucoma can be straightforward,[8-10] accurate clinical identification of early POAG is difficult, resulting in both under- and overdiagnosis. The early stages of POAG are asymptomatic, resulting in many people with the condition being unaware and thus failing to seek eye care.[11] Further, clinical tests for POAG are imperfect in detecting early disease, meaning that it may not be identified in those who do seek eye care. As a result, one in five glaucoma patients referred to the UK Hospital Eye Service present with already advanced visual field loss in at least one eye,[12] as defined by a stage 4 or worse of the Enhanced Glaucoma Severity Staging[13] typically achieved with either a mean deviation or a pattern standard deviation worse than −15 dB and 14 dB respectively. Imprecise case identification combined with the low overall prevalence of the disease among those seeking primary eye care also results in frequent over-referral from primary care and overdiagnosis of glaucoma at all stages of care, producing considerable unnecessary burden on both patients and healthcare services.[14,15]

Optical coherence tomography (OCT) is increasingly available to optometrists, with 15% of UK practices estimated to have access to OCT in 2014,[16] a figure that is projected to rise.[17] In parallel with the increasing availability of OCT, its role in the detection and diagnosis of glaucoma is also increasing.[18,19]

Signs of glaucomatous damage at the optic nerve head (ONH) and surrounding retinal nerve fibre layer (RNFL), traditionally assessed subjectively by fundoscopy techniques, may be more objectively assessed and recorded by OCT imaging, making diagnosis and monitoring of glaucoma more accurate.[8] Preliminary data on the benefit of OCT use by UK community optometrists have shown that the additional information collected can enhance case identification, possibly resulting in a lower rate of false-positive referral to the Hospital Eye Service.[17] Further, older devices imaging similar ocular structures have already been shown to outperform general ophthalmologists in identifying manifest glaucoma.[20] Although such evidence has not yet been replicated with OCT, it seems reasonable to assume that OCT would perform at least as well with its higher resolution, improved image quality and refined image-processing techniques.[10]

In this article we present recent findings from the scientific literature on the uses and limitations of OCT in glaucoma diagnosis.

Spectral-domain OCT

As a result of increased versatility, improved resolution and reproducibility[21] of measurements, improved identification of anatomical structures[22] and decreased image acquisition time,[23] spectral-domain OCT (SD-OCT) has largely superseded previous imaging technologies used in the diagnosis of glaucoma, such as scanning laser polarimetry (e.g. GDx, Carl Zeiss Meditec, Dublin, CA, USA), confocal scanning laser ophthalmoscopy (e.g. Heidelberg retina tomograph (HRT), Heidelberg Engineering, Heidelberg, Germany) and time-domain OCT (e.g. Stratus OCT, Carl Zeiss Meditec).[24-26] Modern SD-OCTs are able to rapidly perform three-dimensional imaging of the retina with axial resolution of 4–7 μm.[23,27] Furthermore, advancements in image processing have enabled automatic segmentation and analysis of single retinal layers relevant to glaucoma in both the circumpapillary and macular areas,[28] as well as the accurate identification of Bruch's membrane opening in the ONH.[29]

Relevant structures to be examined with OCT

The use and implementation of SD-OCT in glaucoma clinics and research have focused on the examination of three main retinal structures: the circumpapillary RNFL (cpRNFL), the ONH and the macula.[8]

Circumpapillary retinal nerve fibre layer thickness

Capturing the thickness of the RNFL in close proximity to the ONH allows for the sampling of RGC axons projecting from

Figure 1. Circumpapillary retinal nerve fibre layer (cpRNFL) analysis of a 65-year-old glaucoma patient (left panel) and an age-matched healthy patient (right panel). The B-scan and the corresponding cpRNFL profile are shown in (a, b) respectively. (b, c) show the cpRNFL thickness profile and the optic nerve head sectoral analysis, colour-coded according to the probability level compared to the healthy reference population: green, within normal limits; yellow, below the fifth percentile of the reference population; and red, below the first percentile of the reference population. (d) indicates where the circle scan is acquired. In the left panel a typical glaucomatous focal defect (thinning) of the inferior temporal sector is shown in (a, b, c). (Courtesy of Habiba Bham, University of Bradford.)
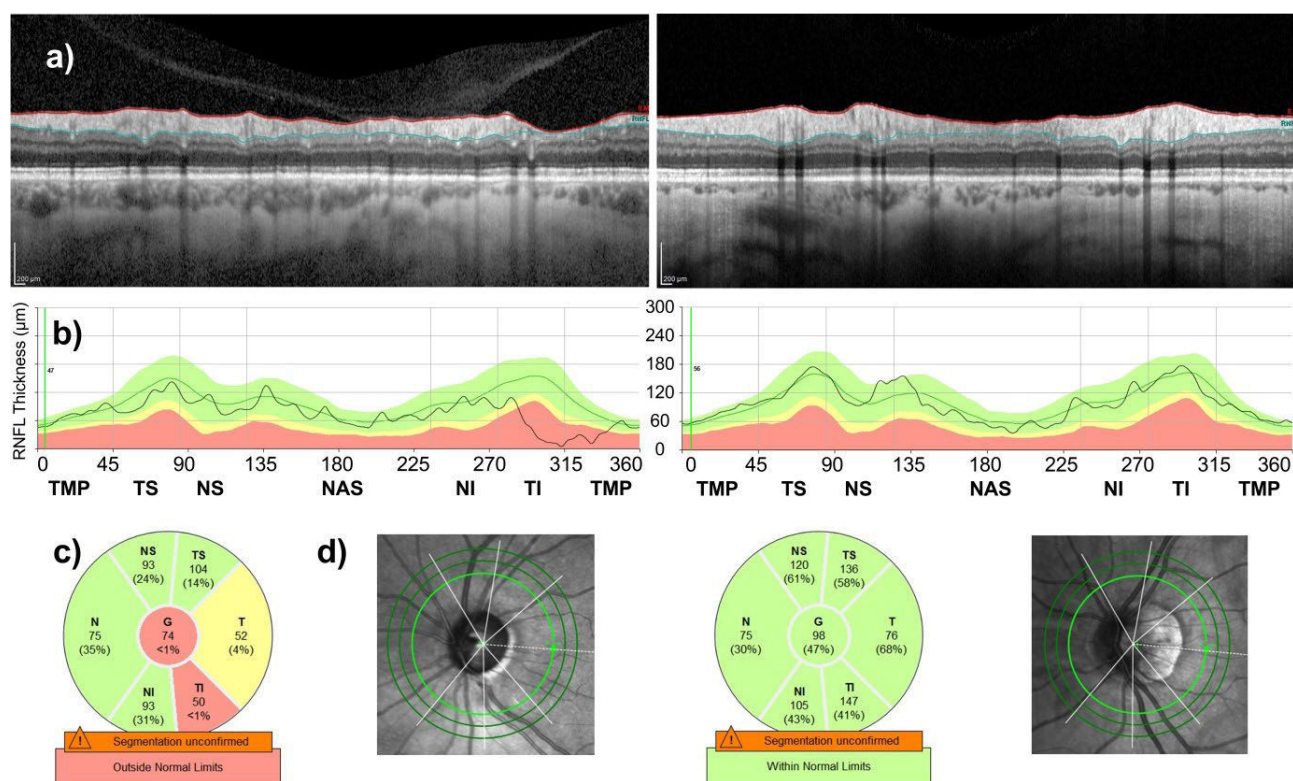
the entire retina (Figure 1). The analysis of cpRNFL thickness is, therefore, key to the current assessment of glaucoma patients and is likely to be the most employed type of scan in this context.[8,9,30,31] In a recent systematic review,[8] among the 59 eligible studies using SD-OCT in glaucoma, 56 included the cpRNFL, while macular and ONH parameters were collected in 36 and 23 studies respectively.

Measurement of cpRNFL uses different scan patterns according to the instrument manufacturer. In different instruments, cpRNFL thickness is determined following a series of concentric circular scans around the ONH (2.5–4.9 mm diameter), radial scans across the ONH or a volumetric scan of the same area (6×6 mm^2 or 7×7 mm^2).[8] RNFL thickness maps are then generated at varying distance from the ONH. In a study comparing different circle sizes of the circular cpRNFL scans using Spectralis OCT (Heidelberg Engineering), diagnosis was typically more accurate when using the smallest-diameter circle among those available (12°, or 3.5 mm, for this instrument).[32]

Different indices can then be determined following the measurement, including global (e.g. average cpRNFL) and localised indices. Careful examination for localised RNFL loss, particularly in the superior- and inferior-temporal cpRNFL sectors, is clinically important as these regions may be affected earlier and more frequently in glaucoma.[18,26]

Optic nerve head

ONH volumetric parameters, including measurements of neuroretinal rim and optic cup based on conventional fundoscopy techniques and those provided by earlier instruments such as the HRT, are widely available from SD-OCT instruments.[8] Notably, as with the HRT, these parameters are determined after setting an arbitrary reference plane at a set distance from the retinal pigment epithelium that varies between instruments.[8] The performance of ONH volumetric measurements in identifying glaucoma has been reported to be inferior to that of cpRNFL thickness and macular retinal layer thickness measurements.[10,18,21,33]

Recent SD-OCT devices have enabled precise localisation of the limits of Bruch's membrane opening (BMO), a more anatomically valid landmark for delimiting neural tissue in the ONH.[29,34] Such instruments provide measurements of the neuroretinal rim based on the minimum distance between the BMO and the inner limiting membrane around the ONH, eliminating the need for use of an arbitrary reference plane.[8] BMO–minimum rim width (BMO–MRW: Figure 2) is calculated by instrument software following radial scans through the ONH and identification of the limits of BMO in each acquired B-scan around the ONH.[8,29] It has been suggested that the BMO–MRW relies on more solid geometrical principles than similar biomarkers, since the rim thickness is measured along the minimum distance from the scleral canal to the nearest point of the optic disc surface (the inner limiting membrane). At this point thickness of the rim is measured perpendicularly to the orientation of its bundles, rather than obliquely as in previously available measurements that may overestimate rim thickness as a result.[18,32] Overall, BMO-

Figure 2. Bruch's membrane opening–minimum rim width analysis from the same patients shown in Figure 1 (left: glaucoma; right: healthy). Panels (a, b, c and d) are analogous to those shown in Figure 1. The B-scan images (top panels) show the limits of Bruch's membrane (red dots) and the minimum distance between these points and the inner limiting membrane (blue arrow). The limits of Bruch's membrane are also shown as red dots on the en face scanning laser ophthalmoscope image (panel d and bottom right). (Courtesy of Habiba Bham, University of Bradford.)
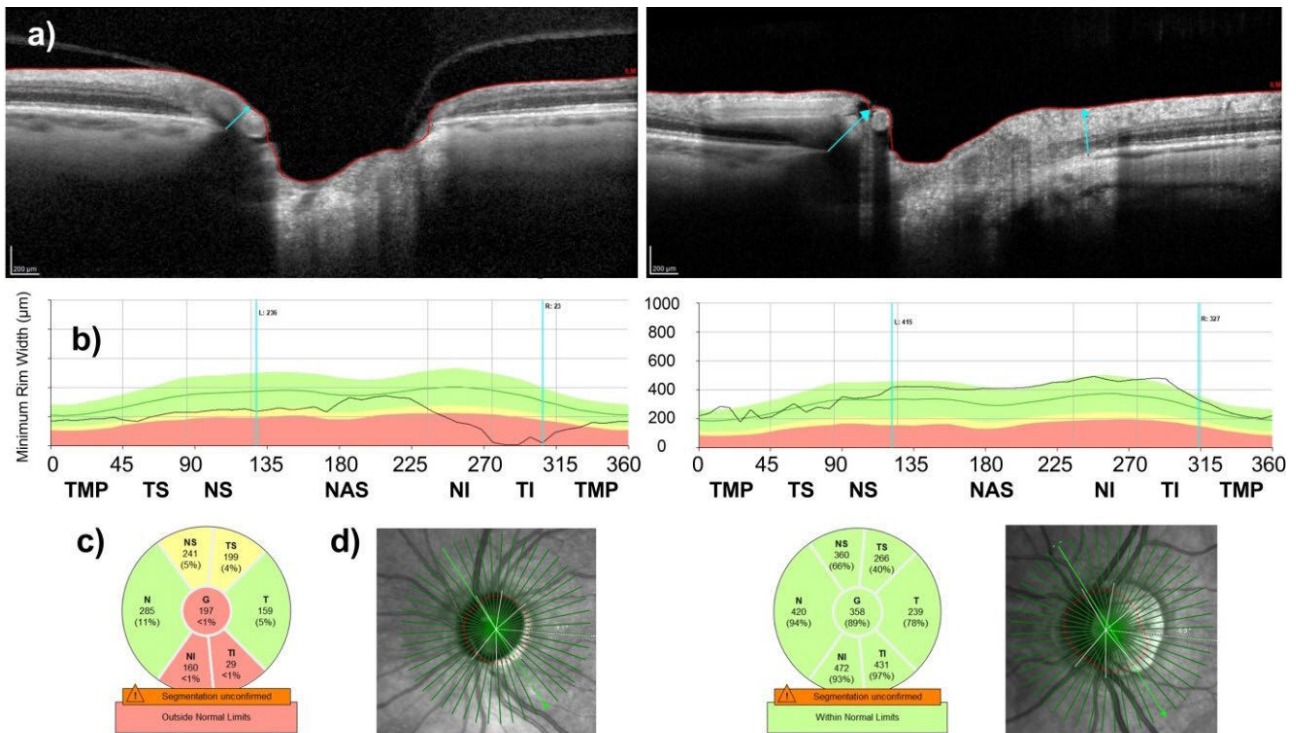
based measurements may have greater validity, reliability and diagnostic accuracy compared to reference plane-based volumetric ONH analysis.[22,35]

Similarly to BMO–MRW, the minimum distance band is a three-dimensional neuroretinal measure, derived from high-density scans of the ONH[36,37] that has yet to translate into commercially available instruments. The index can be considered as a 360° annulus enclosed between the disc margin and the superior limit of the optic cup.[8] A further difference from the BMO–MRW is the use of the retinal pigment epithelium–BMO complex as a reference for the measurement instead of the BMO alone, which is often thinner than the minimum resolvable by SD-OCT. This is reported to yield to a more robust disc margin identification, resulting in a more precise segmentation.[36] A denser scanning protocol combined with a more consistent reference point is suggested to result in a more effective parameter.[8,36]

Macula

The analysis of RGCs in the macula is increasingly common in glaucoma assessment, either to complement or as an alternative to cpRNFL analysis,[28] and is now available in many
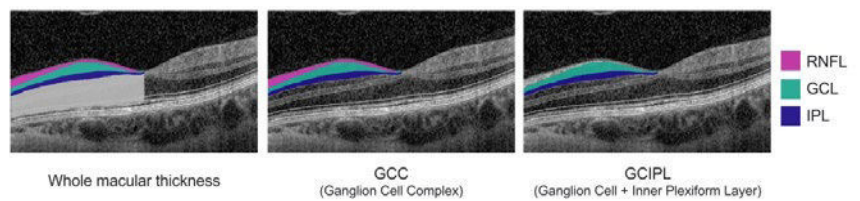
Figure 3. Different approaches adopted in macular segmentation of optical coherence tomography analysis according to different platforms. Segmentation algorithms might consider either all retinal layers (left) or a combination of the retinal nerve fibre layer (RNFL), ganglion cell layer (GCL) and inner plexiform layer (IPL)
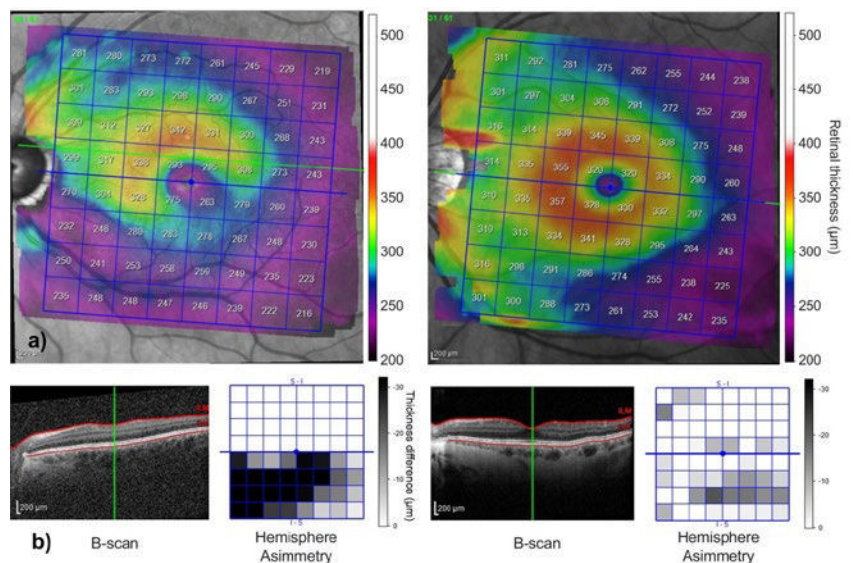
Figure 4. Macular thickness analysis as performed by Spectralis optical coherence tomography, showing images from the same patients as in Figures 1 and 2 (left: glaucoma; right: healthy). (a) The colour-coded macular thickness analysis, with the measurement grid located according to the disc–fovea angle. Thinning relating to the inferior temporal region of the optic nerve head is apparent, in agreement with (b) reporting the B-scan corresponding to the foveal zone besides the asymmetry analysis between the superior and inferior hemispheres. (Courtesy of Habiba Bham, University of Bradford.)

commercial devices. The macula contains approximately 50% of the population of RGCs in a multilayered fashion. RGC bodies in the macula are larger than elsewhere in the retina and up to 20 times greater in diameter than their axons.[38] It has been suggested, therefore, that early glaucomatous damage might be readily identifiable in the macula.[18,28,30] Assessment of this area in glaucoma is supported by growing evidence of early involvement in the disease,[18,31,39] and has been enabled by significant improvements in OCT retinal layer segmentation (Figure 3), allowing quantification of neural tissue thinning caused by glaucoma in the macula. Further potential advantages of assessing this area compared to the cpRNFL include reduced between-individual variability, less interference by non-neural structures (e.g. vasculature, peripapillary atrophy) and the lack of error from placement of the measurement annulus around the ONH.[31,39]

SD-OCT assessment of the macula in glaucoma focuses on the inner retinal layers, but how this is done varies across instrument manufacturers.[31] For instance, in the ganglion cell analysis performed by the Cirrus HD SD-OCT (Carl Zeiss Meditec), either the combination of the ganglion cell and the inner plexiform layer (GCIPL) or the summation of these two layers with the macular RNFL (ganglion cell complex) can be reported.[31] In the latter case the axons, bodies and dendrites of RGCs are all included in the analysis, since the RNFL, the ganglion cell layer and the inner plexiform layer (the retinal layers housing these structures) are respectively segmented and computed together.[8] Until recently, in the Spectralis OCT (Heidelberg Engineering), segmentation of macular layers was not reported in the glaucoma module. Instead, the whole macular thickness could be assessed, and an analysis of posterior pole asymmetry conducted by comparing superior and inferior hemispheres of the same eye (Figure 4).

However, the same manufacturer has recently made an additional analysis of single macular layers (i.e. the macular RNFL, ganglion cell layer and inner plexiform layer) available. With this new analysis glaucomatous thinning can be either assessed by raw values or in a deviation map, comparing thicknesses to a provided reference database (Figure 5).

Strong correlation and topographic correspondence exist between cpRNFL and macular ganglion cell analysis parameters, allowing for them to be considered together.[31] Although there are many advantages to macular analysis in glaucoma, it is limited by the prevalence of other conditions causing macular changes in older adults such as macular degeneration and diabetic retinopathy,[18,31] and analysis tools are currently less well developed than those for cpRNFL thickness.

Figure 5. Macular single-layer thickness analysis as performed by Spectralis optical coherence tomography, including the macular retinal nerve fibre layer (RNFL: top panels), ganglion cell layer (GCL: middle panels) and the inner plexiform layer (IPL: lower panels). Images reported are from the same patients as in Figures 1, 2 and 4 (left: glaucoma; right: healthy). The left panel in (a) shows the colour-coded thickness analysis for the three isolated layers. A reduction of thickness of the inferior or inferior-temporal regions is obvious in all macular sublayers. The right panel in (a) shows the deviation maps; retinal areas are flagged in red when thickness falls below the first percentile of the normative population. (b) The same analysis for a healthy eye.
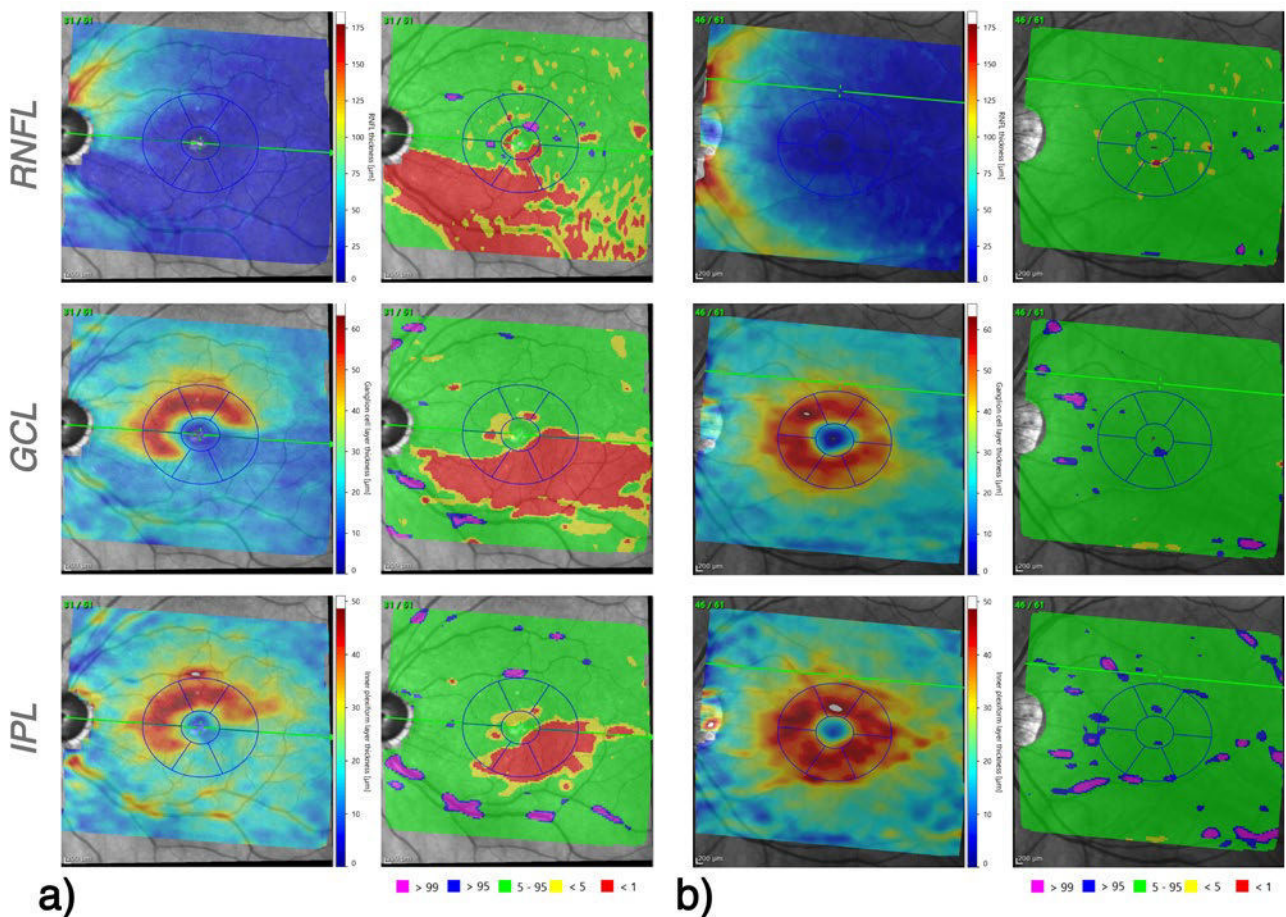
Repeatability of OCT measurements in glaucoma

A variety of studies have assessed the short-term repeatability of OCT, which is generally reported to be excellent for both macular and cpRNFL parameters.[40-45] When assessing global metrics such as the average cpRNFL or macular GCIPL, test–retest repeatability of the latest SD-OCT devices is in the order of 4–5 µm.[41,44,45] Repeatability in localised sectors is worse, in the order of 8–12 µm, though this is not consistent across instruments.[40,43]

It is worth noting that studies on OCT reproducibility have generally been conducted in settings likely to overestimate precision. Typically, healthy participants or glaucoma patients without comorbidities are tested, and only high-quality scans without segmentation or positioning errors are included, which may not mimic many clinical scenarios.[40,46] Furthermore, it is usually intrasession variability that is measured rather than long-term variability, which may be more relevant in clinics. To improve reproducibility, it is recommended to collect multiple initial acquisitions to form a more robust baseline and to use devices with an automated placement of scans, reducing operator dependability.[46] Even small errors in centration of circle scans can result in significant errors in cpRNFL thickness.[47]

## Comparison of measurements to normative databases

In addition to the displayed B-scan images, raw OCT outputs are typically converted by instrument software to measurements of thickness and volume of specific retinal structures. These measurements are commonly presented alongside comparison to percentiles of reference data from a healthy population[18] with classifications such as 'within normal limits', 'borderline' or 'outside normal limits', as can be seen in Figures 1 and 2. This approach can be useful to the clinician, but it has several limitations. For instance, the ability to classify an individual correctly strongly depends on the representativeness of the reference database used.[18] Such databases vary considerably between manufacturers in their size, eligibility criteria and ethnicity included.[48] Criteria for inclusion in normative databases tend to be strict, with databases typically including individuals with European ancestry only and excluding subjects with high refractive error and ocular or systemic comorbidities,[46,48] making comparisons in patients who do not fit into these standards potentially invalid.[18] As a result the diagnostic value of OCT in glaucoma clinics might be hampered in people of non-European ethnicity or with ocular anatomical variants such as high refractive error when solely relying on comparisons to normative data.

Diagnostic ability of OCT

Studies of diagnostic performance of any clinical test are often difficult to interpret for a given clinical context. This is because the reported measurements of diagnostic accuracy only apply to the sample of the study, which is unlikely to match exactly the characteristics of a given clinical population. Indeed, when applied in settings similar to the clinical ones, the diagnostic performance of the same test is likely to be worse than reported in research.[49] Further, inclusion and exclusion criteria vary across different studies, meaning that results from different studies are not directly comparable. More specifically to glaucoma, we might consider that tools like SD-OCT would be most useful in the 'difficult cases', such as those where other clinical test results are equivocal and the diagnosis is therefore uncertain.[10] Unfortunately, it is exactly these cases that are typically excluded from studies of diagnostic performance, leaving only readily identifiable cases of glaucoma and unequivocally healthy controls.[28,50] Nonetheless, diagnostic studies do allow us to directly compare the relative performance of metrics (e.g. cpRNFL vs. BMO–MRW) within a study, and give an indication of which metrics are likely to perform best in the clinic. A further potential bias can arise from the selection of the reference standard by which eyes are grouped (e.g. glaucoma vs. healthy).[10,24,28] For instance, a study may be biased in favour of an ONH-related measurement over a macular measurement when a structural defect of the ONH was originally used to classify eyes as glaucoma or healthy.

Scanning laser polarimetry and confocal scanning laser ophthalmoscopy both preceded OCT in providing objective measurements of RNFL thickness and ONH structure respectively. These instruments had considerably lower resolution than OCT and more assumptions and limitations in their
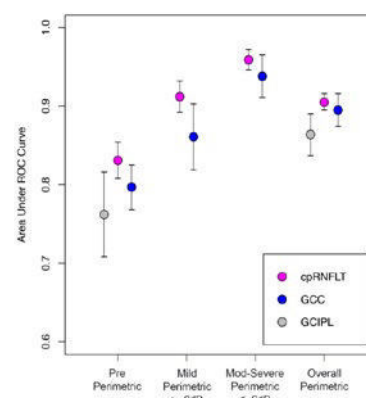
Figure 6. Diagnostic accuracy of optical coherence tomography parameters at different stages of glaucoma, plotted from data in Kansal et al.[9] Diagnosis of pre-perimetric glaucoma was considered when only optic disc appearance was indicative of glaucoma, with no evidence of visual field defects. Error bars report the 95% confidence interval. cpRNFLT, circumpapillary retinal nerve fibre layer thickness; GCC, ganglion cell complex; GCIPL, ganglion cell inner plexiform layer; ROC, receiver operating characteristic

calculation of structural indices analogous to those provided by OCT today. Nevertheless, one important study[18] compared diagnostic performance by early incarnations of those instruments (the GDx VCC, Carl Zeiss Meditec and HRT 1, Heidelberg Engineering) to that of European ophthalmologists viewing stereoscopic ONH photographs of the same eyes. The study found that both of these early instruments already outperformed general ophthalmologists, and performed on a par with UK glaucoma specialist ophthalmologists, despite their limitations compared to contemporary SD-OCT devices.[18] Whilst an equivalent study with SD-OCT has not yet been conducted, it seems safe to assume that SD-OCT would also outperform clinicians' subjective ONH assessment in the detection of glaucoma, and therefore its use, where available, to complement subjective techniques should be strongly advocated.

Overall performance of OCT in glaucoma diagnosis

Recent systematic reviews agree that OCT metrics generally perform well in glaucoma diagnosis, with cpRNFL-related parameters typically performing equally or slightly superiorly to macular parameters.[8,9,24,28] Many studies assessing the diagnostic ability of OCT in glaucoma use the area under the receiver operating characteristic curve (AUC) as an indicator

of overall diagnostic performance.[8] This metric considers both sensitivity (proportion of those with the disease being correctly identified as such) and specificity (proportion of those without the disease being correctly identified as such) over a range of test result cut-off values. For example, in the case of RNFL thickness, a lower cut-off value would yield lower sensitivity and higher specificity than a lower cut-off value. The AUC metric ranges between 0 and 1, with chance performance being 0.5 and perfect performance being 1. It has been suggested that a test can be considered excellent when AUC is 0.9–1 and good when AUC is 0.8–0.9.[8] All current diagnostic tests for glaucoma are imperfect classifiers, requiring a trade-off between minimising false positives or false negatives. Misclassification of healthy and diseased eyes is unlikely to be equally important in clinical scenarios,[51] meaning that the desirability of high sensitivity vs. high specificity varies for different conditions and clinical frameworks.[51,52] Though early diagnosis of glaucoma is important, its low prevalence and high socioeconomic cost of misdiagnosis make minimisation of false positives more desirable than minimising false negatives.[24,53,54] In other words, in the unavoidable compromise that must be made between sensitivity and specificity, high specificity is to be prioritised compared to high sensitivity in glaucoma diagnosis.

Kansal and associates presented a meta-analysis of 150 studies aiming to clarify diagnostic accuracy of different OCT biomarkers at different disease stages.[9] Their results show average AUCs of 0.897 (95% confidence interval (CI), 0.887–0.906), 0.858 (95% CI, 0.835–0.880) and 0.885 (95% CI, 0.869–0.901) respectively for the global cpRNFL, the GCIPL and ganglion cell complex. Although different biomarkers performed similarly, a slight superiority of cpRNFL was reported, with a clear trend toward a reduced diagnostic accuracy in the earlier stages of glaucoma for all metrics. Furthermore, the meta-analysis confirmed that whole macular thickness performed worse than segmented indices.[9] This meta-analysis, however, included studies with wide-ranging reference standards for defining glaucoma and healthy eyes, which might be responsible for considerable heterogeneity in the results.

A slightly older meta-analysis aimed to compare cpRNFL and the macular ganglion cell complex and GCIPL in the diagnosis of manifest glaucoma.[28] Overall, diagnostic performance of the global cpRNFL was excellent and similar to the inferior cpRNFL sectors that performed the best among isolated sectors. Macular parameters performed equal to or only slightly poorer than cpRNFL, with sensitivity values ranging between 0.65 and 0.75 at fixed specificities of 0.90 and 0.95. An overall modest superiority of the cpRNFL was identified which may make this largely adopted parameter preferable when only a single OCT protocol can be performed.[28]

A recent systematic review found only a limited number of studies assessing diagnostic performance of the newer ONH parameters (i.e. BMO–MRW and minimum distance band).[8] According to Chen and colleagues, results regarding BMO–MRW are conflicting, with some earlier evidence of superiority compared to the cpRNFL not confirmed in later studies.[8] On the other hand, the few studies available on the newer minimum distance band are promising, showing superiority to the cpRNFL measurement in detecting glaucoma.[8] Furthermore, in agreement with previous works, the review found a limited number of studies suggesting that macular analysis could outperform cpRNFL for glaucoma diagnosis in highly myopic patients, though further studies are needed in this area where glaucoma diagnosis is notoriously difficult. [8,28,31]

The aforementioned study by Kansal et al.[9] is remarkable since it is the most recent meta-analysis including a subgroup analysis for different stages of the disease (Figure 6). In their perimetric glaucoma group, findings largely mirrored what was observed in the overall sample. Here higher AUCs were computed, approaching 0.9 for cpRNFL or slightly lower for macular indices.[9] However, for the pre-perimetric glaucoma subgroup, a drop in the AUC values was observed (average cpRNFL 0.83, average GCIPL 0.76 and average ganglion cell complex 0.80).[9] This has been subsequently confirmed by Chen and colleagues, with a clear trend for OCT-mediated glaucoma diagnosis to become more accurate when the disease approaches more advanced stages.[8]

Diagnosis of early glaucoma

Combination of multiple indices

More than a dozen glaucoma-related OCT parameters are currently available to clinicians.[55] This can be overwhelming, resulting in difficulty interpreting sometimes conflicting indices.[54] An increasingly adopted approach is the combination of different indices, e.g. cpRNFL with ganglion cell complex analysis, aiming to increase the diagnostic accuracy at earlier stages of the disease.[18] Of course, there are multiple ways in which information can be combined, and which is chosen affects the diagnostic performance of the resulting combined index. For example, an index that requires each of multiple measurements to be abnormal to report overall abnormality would have high specificity but low sensitivity. Alternatively, an index that requires just one of multiple measurements to be abnormal would have high sensitivity but low specificity. As reported above, the former case (high specificity) is generally considered to be preferable for glaucoma.[24,53,54]

In terms of diagnostic performance, the few studies available suggest that combined indices may have advantages over single metrics.[54] The UNC OCT index, developed by researchers at the University of North Carolina (but not yet commercially implemented), includes 16 parameters relating to the GCIPL, ONH and a combination of superior and inferior cpRNFL sectors.[54] After validation, the index was tested in a cohort of early glaucoma with visual field mean deviation no worse than −4.0 dB. The UNC OCT index had AUC 0.96 and 85.4% sensitivity at 95% specificity in this sample, and AUC 0.95 with 81.7% sensitivity at 95% specificity in a subgroup with mean deviation better than −2.0 dB.[54] These values were superior to individual indices, where AUCs were reported between 0.91 and 0.93. Other individual studies of combined OCT indices using a variety of methodologies show high diagnostic accuracy, with AUC values typically above 0.95 in the studied populations. However, the measured improvements over individual indices were often not statistically significant.[55] Though these results are promising, more research is needed to bolster the presently limited evidence to support the use of combined OCT indices in glaucoma diagnosis.

Detection of a progressive defect

Despite advances in OCT image resolution and consequent development of improved metrics for glaucoma diagnosis, it is still often difficult to diagnose glaucoma accurately based on a single examination.[18,54,55] The large variability of ONH appearance between healthy individuals,[46,56] combined with the often minimal structural changes at POAG manifestation,[54,55] contributes to this challenge. Reliance on comparison of snapshot measurements to normative databases risks missing patients who have already lost considerable tissue, but who started from a high baseline and therefore remain within population norms.[57] Consequently, the identification of progressive structural damage across multiple examinations is desirable for a more certain diagnosis and prediction of future functional loss.[18,46]

Where progressive structural damage of the ONH can be identified, the patient's risk of developing a corresponding visual field defect is up to 26 times higher than when only a non-progressive structural defect is identified.[56] The excellent test–retest reproducibility of SD-OCTs makes them suitable for detecting more subtle changes than many other clinical tests.[46] However, it must be noted that measurements are not currently interchangeable between different instruments, possibly because of differences in segmentation algorithms and interaction between devices' optic and light beam with tissue.[40] Therefore, it is strongly recommended that the same instrument must be used for baseline measurement and follow-up.[58]

Progressive thinning noted on SD-OCT must be differentiated from both measurement noise and the normal ageing process in order to be attributed to glaucoma. Unfortunately, OCT data from large-scale longitudinal studies are still scarce.[18] A report on 294 glaucoma suspects showed a significantly faster rate of cpRNFL thickness thinning in those who developed a visual field defect (−2.02 μm/year) compared to participants with no visual field defect at 6-month follow-up (−0.82 μm/year).[59] As underlined by the same study group, it must be noted that, even in the group of participants who did not develop a visual field defect (considered to be healthy, although only after a short follow-up period averaging just above 2 years), thinning was observed and attributed to effects of the normal ageing process.

Both macular and cpRNFL OCT indices have shown faster thinning in glaucomatous eyes compared to healthy eyes, though observed rates are variable.[46] Among the few studies reporting physiological thinning of OCT measurements as a function of age, Leung and colleagues followed a small group of healthy individuals for 30 months and found cpRNFL to decrease by 0.52 μm/year (95% CI, 0.17–0.86 μm) – significantly faster than for the macular GCIPL (−0.32 μm/year).[60] Another study found a thinning rate in healthy subjects of −0.48 μm/year and −0.14 μm/year for cpRNFL and GCIPL respectively, after an average follow-up of 1.7 years.[61] These figures were significantly lower than those in glaucomatous eyes (cpRNFL, −0.98 μm/year, 95% CI, −1.20 to −0.76 μm; GCIPL, −0.57 μm/year, 95% CI, −0.73 μm to −0.41 μm).[61] One additional report followed 45 healthy participants for an average of 3.2 years, finding cpRNFL changes of −0.54±0.23 μm/year.[62]

It is apparent that the above data were collected in limited samples, with variable follow-up durations and using varying devices and protocols. Additionally, effects of other factors in clinical populations such as ethnicity, comorbidity and image quality remain unknown.[18] Overall, agreement is lacking on a precise rate of progression that could confirm diagnosis of glaucoma. Instrument software (e.g. guided progression analysis, Carl Zeiss Meditec) may allow the clinician to identify changes exceeding the expected test–retest variability (event-based analysis) and to observe the rate of change over time (trend-based analysis). This type of analysis can be performed for both ONH and macular parameters.

Emerging techniques

Subjective assessment of OCT images

Hood and De Moraes[30] criticised the current application of OCT in glaucoma as being too reliant on numerical analysis of cpRNFL scans, including summary thickness indices, either global or sectoral, and generated thickness deviation maps. The actual circumpapillary B-scans on OCT reports are often too small to be observed and analysed, or are sometimes omitted completely from printed reports.[26,30] The utility of assessing additional areas besides the cpRNFL has also been clearly identified in the literature.[8,18,39,63] In other areas of ophthalmology, clinicians often look carefully through B-scans and this approach may be underexploited in glaucoma.[30] The assessment of an enlarged image of the cpRNFL could be useful for several reasons, such as appraising the accuracy of automated retinal layer segmentation performed by the software. Furthermore, the approach has the potential to allow clinicians to focus attention on preserved RNFL regions in advanced glaucoma, where other indices might be affected by the 'measurement floor' (a minimum thickness beyond which further thinning cannot be identified due to lack of resolution).[30]
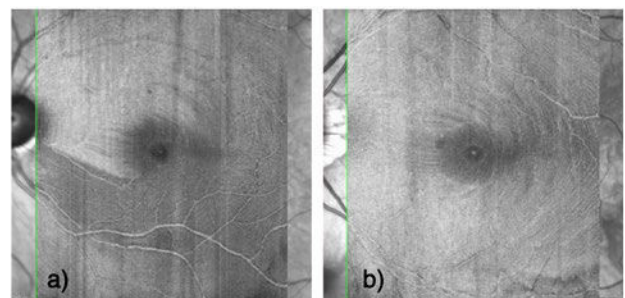
Figure 7. En face (transverse) imaging of the central retina as performed by Spectralis optical coherence tomography, showing images from the same patients as in Figures 1, 2, 4 and 5 (left: glaucoma; right: healthy). In both cases the images reproduce a 40-μm-thick slab, corresponding to retinal depths from 10 to 50 μm below the inner limiting membrane. Whereas hyperreflective bundles are appreciable throughout the imaged retina in the healthy eye (b), an inferior arcuate defect is observable in the glaucoma eye in the form of a hyporeflective patch in the inferior hemisphere (a).

En face visualisation of retinal nerve fibre bundles

According to studies on experimental glaucoma in primates, the cpRNFL thickness may show a measurable thinning after only 10–15% of RGC axons have been lost.[64] Adopting different biomarkers could, therefore, allow for an even earlier detection of structural damage. RGC axons are highly reflective due to the ordered structure of proteins comprising their cytoskeleton, which may be disrupted prior to axon loss.[18] The potential for assessing changes in RNFL reflectance due to cellular disruption in early glaucoma is being investigated, with promising results.[36,65] Animal models showed decreased RNFL reflectance earlier in glaucoma than measurable RNFL thinning.[66] Similarly, altered reflectance of the RNFL may predict more rapid reduction of visual field sensitivity.[67] A recent wide-field OCT study assessed reflectivity defects as observed in en face images of the RNFL and showed a strong correlation with cpRNFL defects.[68] En face visualisation of the RNFL is now available in several commercial OCT instruments (Figure 7) but at present automated analysis tools are not provided, therefore only subjective assessment of these images is possible.

Conclusions

OCT is rapidly reported to be essential in the diagnosis and management of glaucoma. Instruments are becoming more widely available in community optometric practices, and the provided software is becoming increasingly adept at detecting structural damage due to glaucoma. OCT indices are likely to outperform clinicians' subjective ONH/RNFL assessments in detecting glaucoma and therefore should be fully utilised in clinical practice where available. Clinicians should be aware of the strengths and limitations of OCT when assessing patients for glaucoma. Indeed, despite great technological improvements, the detection of early glaucoma remains a significant clinical challenge requiring input from a battery of clinical tests. As such OCT should continue to be used in conjunction with visual field testing, subjective assessment of the ONH, measurement of intraocular pressure and assessment of the anterior-chamber drainage angle.

Relevance to clinical practice

OCT is becoming increasingly available to optometrists, and is concurrently becoming increasingly important in the detection and diagnosis of glaucoma. This article discusses the strengths and limitations of using OCT for the detection and diagnosis of glaucoma, and gives practical advice on how OCT can best be employed in optometric case-finding situations.

Practical tips for using OCT in practice

1. Typically, cpRNFL thickness scans offer the best diagnostic performance of a single scan, but macular scans offer complementary information and may be particularly useful in highly myopic patients.
2. Check that the retinal layers of interest have been accurately segmented by the OCT software to avoid over- or underestimating the thickness of the target tissue.
3. Look at B-scan images carefully for focal tissue loss, even if software marks the region as a whole as 'within normal limits'.
4. Where a defect is uncertain, it may help to confirm it in multiple regions by scanning the ONH, cpRNFL and macula.
5. Perform baseline scans on patients at future risk of developing glaucoma in order to look for change over time as a more certain indication of early glaucoma than snapshot measurements.

References

1. Jonas JB, Aung T, Bourne RR et al. Glaucoma. *Lancet* 2017;**390**:2183–2193.

2. Weinreb RN, Leung CK, Crowston JG et al. Primary open-angle glaucoma. *Nat Rev Dis Primers* 2016;**2**:16067.

3. Yamamoto S, Sawaguchi S, Iwase A et al. Primary open-angle glaucoma in a population associated with high prevalence of primary angle-closure glaucoma: the Kumejima study. *Ophthalmology* 2014;**121**:1558–1565.

4. Kapetanakis VV, Chan MP, Foster PJ et al. Global variations and time trends in the prevalence of primary open angle glaucoma (POAG): a systematic review and meta-analysis. *Br J Ophthalmol* 2016;**100**:86–93.

5. Rudnicka AR, Mt-Isa S, Owen CG et al. Variations in primary open-angle glaucoma prevalence by age, gender, and race: a Bayesian meta-analysis. *Invest Ophthalmol Vis Sci* 2006;**47**:4254–4261.

6. Bourne RRA, Jonas JB, Bron AM et al. Prevalence and causes of vision loss in high-income countries and in Eastern and Central Europe in 2015: magnitude, temporal trends and projections. *Br J Ophthalmol* 2018;**102**:575–585.

7. Flaxman SR, Bourne RRA, Resnikoff S et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *Lancet Glob Health* 2017;**5**:e1221–e1234.

8. Chen TC, Hoguet A, Junk AK et al. Spectral-domain OCT: helping the clinician diagnose glaucoma: a report by the American Academy of Ophthalmology. *Ophthalmology* 2018;**125**:1817–1827.

9. Kansal V, Armstrong JJ, Pintwala R et al. Optical coherence tomography for glaucoma diagnosis: an evidence based meta-analysis. *PLoS One* 2018;**13**:e0190621.

10. Michelessi M, Lucenteforte E, Oddone F et al. Optic nerve head and fibre layer imaging for diagnosing glaucoma. *Cochrane Database Syst Rev* 2015;Cd008803.

11. Chan MPY, Broadway DC, Khawaja AP et al. Glaucoma and intraocular pressure in EPIC-Norfolk Eye Study: cross sectional study. *BMJ* 2017;**358**:j3889.

12. Crabb DP, Saunders LJ, Edwards LA. Cases of advanced visual field loss at referral to glaucoma clinics – more men than women? *Ophthalmic Physiol Opt* 2017;**37**:82–87.

13. Brusini P, Filacorda S. Enhanced Glaucoma Staging System (GSS 2) for classifying functional damage in glaucoma. *J Glaucoma* 2006;**15**:40–46.

14. Hohn R, Nickels S, Schuster AK et al. Prevalence of glaucoma in Germany: results from the Gutenberg Health Study. *Graefes Arch Clin Exp Ophthalmol* 2018;**256**:1695–1702.

15. Karvonen E, Stoor K, Luodonpaa M et al. Prevalence of glaucoma in the Northern Finland Birth Cohort Eye Study. *Acta Ophthalmol* 2019;**97**:200–207

16. Dabasia PL, Edgar DF, Garway-Heath DF et al. A survey of current and anticipated use of standard and specialist equipment by UK optometrists. *Ophthalmic Physiol Opt* 2014;**34**:592–613.

17. Jindal A, Ctori I, Fidalgo B et al. Impact of optical coherence tomography on diagnostic decision-making by UK community optometrists: a clinical vignette study. *Ophthalmic Physiol Opt* 2019;**39**:205–215.

18. Tatham AJ, Medeiros FA, Zangwill LM et al. Strategies to improve early diagnosis in glaucoma. *Prog Brain Res* 2015;**221**:103–133.

19. World Glaucoma Association. *10th Consensus Meeting: Diagnosis of Primary Open Angle Glaucoma*; 2016.

20. Reus NJ, Lemij HG, Garway-Heath DF et al. Clinical assessment of stereoscopic optic disc photographs for glaucoma: the European Optic Disc Assessment Trial. *Ophthalmology* 2010;**117**:717–723.

21. Bussel II, Wollstein G, Schuman JS. OCT for glaucoma diagnosis, screening and detection of glaucoma progression. *Br J Ophthalmol* 2014;**98**(Suppl 2):ii15–ii19.

22. Fortune B. Optical coherence tomography evaluation of the optic nerve head neuro-retinal rim in glaucoma. *Clin Exp Optom* 2019;**102**:286–290.

23. Chen TC, Cense B, Pierce MC et al. Spectral domain optical coherence tomography: ultra-high speed, ultra-high resolution ophthalmic imaging. *Arch Ophthalmol* 2005;**123**:1715–1720.

24. Fallon M, Valero O, Pazos M et al. Diagnostic accuracy of imaging devices in glaucoma: a meta-analysis. *Surv Ophthalmol* 2017;**62**:446–461.

25. Lavinsky F, Wollstein G, Tauber J et al. The future of imaging in detecting glaucoma progression. *Ophthalmology* 2017;**124**: S76–S82.

26. Hood DC. Improving our understanding, and detection, of glaucomatous damage: an approach based upon optical coherence tomography (OCT). *Prog Retin Eye Res* 2017;**57**:46–75.

27. White B, Pierce M, Nassif N et al. In vivo dynamic human retinal blood flow imaging using ultra-high-speed spectral domain optical coherence tomography. *Opt Express* 2003;**11**:3490–3497.

28. Oddone F, Lucenteforte E, Michelessi M et al. Macular versus retinal nerve fiber layer parameters for diagnosing manifest glaucoma: a systematic review of diagnostic accuracy studies. *Ophthalmology* 2016;**123**:939–949.

29. Chauhan BC, Danthurebandara VM, Sharpe GP et al. Bruch's membrane opening minimum rim width and retinal nerve fiber layer thickness in a normal white population: a multicenter study. *Ophthalmology* 2015;**122**:1786–1794.

30. Hood DC, De Moraes CG. Challenges to the common clinical paradigm for diagnosis of glaucomatous damage with OCT and visual fields. *Invest Ophthalmol Vis Sci* 2018;**59**:788–791.

31. Kim KE, Park KH. Macular imaging by optical coherence tomography in the diagnosis and management of glaucoma. *Br J Ophthalmol* 2018;**102**:718–724.

32. Gmeiner JM, Schrems WA, Mardin CY et al. Comparison of Bruch's membrane opening minimum rim width and peripapillary retinal nerve fiber layer thickness in early glaucoma assessment. *Invest Ophthalmol Vis Sci* 2016;**57**:575–584.

33. Ervin AM, Boland MV, Myrowitz EH et al. Screening for glaucoma: comparative effectiveness. AHRQ comparative effectiveness reviews. In: *Screening for Glaucoma: Comparative Effectiveness*. Rockville, MD: Agency for Healthcare Research and Quality (US); 2012.

34. Chauhan BC, Burgoyne CF. From clinical examination of the optic disc to clinical assessment of the optic nerve head: a paradigm change. *Am J Ophthalmol* 2013;**156**:218–227.e2.

35. Danthurebandara VM, Sharpe GP, Hutchison DM et al. Enhanced structure–function relationship in glaucoma with an anatomically and geometrically accurate neuroretinal rim measurement. *Invest Ophthalmol Vis Sci* 2014;**56**:98–105.

36. Mwanza JC, Budenz DL. New developments in optical coherence tomography imaging for glaucoma. *Curr Opin Ophthalmol* 2018;**29**:121–129.

37. Shieh E, Lee R, Que C et al. Diagnostic performance of a novel three-dimensional neuroretinal rim parameter for glaucoma using high-density volume scans. *Am J Ophthalmol* 2016;**169**:168–178.

38. Curcio CA, Allen KA. Topography of ganglion cells in human retina. *J Comp Neurol* 1990;**300**:5–25.

39. Hood DC, Raza AS, de Moraes CG et al. Glaucomatous damage of the macula. *Prog Retin Eye Res* 2013;**32**:1–21.

40. Pierro L, Gagliardi M, Iuliano L et al. Retinal nerve fiber layer thickness reproducibility using seven different OCT instruments. *Invest Ophthalmol Vis Sci* 2012;**53**:5912–5920.

41. Leung CK, Cheung CY, Weinreb RN et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a variability and diagnostic performance study. *Ophthalmology* 2009;**116**:1257–1263, 1263.e1–2.

42. Wadhwani M, Bali SJ, Satyapal R et al. Test–retest variability of retinal nerve fiber layer thickness and macular ganglion cell-inner plexiform layer thickness measurements using spectral-domain optical coherence tomography. *J Glaucoma* 2015;**24**:e109–e115.

43. Reis ASC, Zangalli CES, Abe RY et al. Intra- and interobserver reproducibility of Bruch's membrane opening minimum rim width measurements with spectral domain optical coherence tomography. *Acta Ophthalmol* 2017;**95**:e548–e555.

44. Kim KE, Yoo BW, Jeoung JW et al. Long-term reproducibility of macular ganglion cell analysis in clinically stable glaucoma patients. *Invest Ophthalmol Vis Sci* 2015;**56**:4857–4864.

45. Mwanza JC, Chang RT, Budenz DL et al. Reproducibility of peripapillary retinal nerve fiber layer thickness and optic nerve head parameters measured with cirrus HD-OCT in glaucomatous eyes. *Invest Ophthalmol Vis Sci* 2010;**51**:5724–5730.

46. Tatham AJ, Medeiros FA. Detecting structural progression in glaucoma with optical coherence tomography. *Ophthalmology* 2017;**124**:S57–S65.

47. Cheung CY, Yiu CK, Weinreb RN et al. Effects of scan circle displacement in optical coherence tomography retinal nerve fibre layer thickness measurement: a RNFL modelling study. *Eye (Lond)* 2009;**23**:1436–1441.

48. Realini T, Zangwill LM, Flanagan JG et al. Normative databases for imaging instrumentation. *J Glaucoma* 2015;**24**:480–483.

49. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;**299**:926–930.

50. Rao HL, Kumbar T, Addepalli UK et al. Effect of spectrum bias on the diagnostic accuracy of spectral-domain optical coherence tomography in glaucoma. *Invest Ophthalmol Vis Sci* 2012;**53**:1058–1065.

51. Mallett S, Halligan S, Thompson M et al. Interpreting diagnostic accuracy studies for patient care. *BMJ* 2012;**345**:e3999.

52. Saunders LJ, Zhu H, Bunce C et al. Ophthalmic statistics note 5: diagnostic tests—sensitivity and specificity. *Br J Ophthalmol* 2015;**99**:1168–1170.

53. Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. *Front Public Health* 2017;**5**:307.

54. Mwanza JC, Lee G, Budenz DL et al. Validation of the UNC OCT index for the diagnosis of early glaucoma. *Transl Vis Sci Technol* 2018;**7**:16.

55. Mwanza JC, Warren JL, Budenz DL. Utility of combining spectral domain optical coherence tomography structural parameters for the diagnosis of early glaucoma: a mini-review. *Eye Vis (Lond)* 2018;**5**:9.

56. Medeiros FA, Alencar LM, Zangwill LM et al. Prediction of functional loss in glaucoma from progressive optic disc damage. *Arch Ophthalmol* 2009;**127**:1250–1256.

57. Sayed MS, Margolis M, Lee RK. Green disease in optical coherence tomography diagnosis of glaucoma. *Curr Opin Ophthalmol* 2017;**28**:139–153.

58. Vianna JR, Chauhan BC. How to detect progression in glaucoma. *Prog Brain Res* 2015;**221**:135–158.

59. Miki A, Medeiros FA, Weinreb RN et al. Rates of retinal nerve fiber layer thinning in glaucoma suspect eyes. *Ophthalmology* 2014;**121**:1350–1358.

60. Leung CK, Yu M, Weinreb RN et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a prospective analysis of age-related loss. *Ophthalmology* 2012;**119**:731–737.

61. Hammel N, Belghith A, Weinreb RN et al. Comparing the rates of retinal nerve fiber layer and ganglion cell-inner plexiform layer loss in healthy eyes and in glaucoma eyes. *Am J Ophthalmol* 2017;**178**:38–50.

62. Wu Z, Saunders LJ, Zangwill LM et al. Impact of normal aging and progression definitions on the specificity of detecting retinal nerve fiber layer thinning. *Am J Ophthalmol* 2017;**181**:106–113.

63. Hood DC, De Cuir N, Blumberg DM et al. A single wide-field OCT protocol can provide compelling information for the diagnosis of early glaucoma. *Transl Vis Sci Technol* 2016;**5**:4.

64. Cull GA, Reynaud J, Wang L et al. Relationship between orbital optic nerve axon counts and retinal nerve fiber layer thickness measured by spectral domain optical coherence tomography. *Invest Ophthalmol Vis Sci* 2012;**53**:7766–7773.

65. Liu S, Wang B, Yin B et al. Retinal nerve fiber layer reflectance for early glaucoma diagnosis. *J Glaucoma* 2014;**23**:e45–e52.

66. Huang XR, Zhou Y, Kong W et al. Reflectance decreases before thickness changes in the retinal nerve fiber layer in glaucomatous retinas. *Invest Ophthalmol Vis Sci* 2011;**52**:6737–6742.

67. Gardiner SK, Demirel S, Reynaud J et al. Changes in retinal nerve fiber layer reflectance intensity as a predictor of functional progression in glaucoma. *Invest Ophthalmol Vis Sci* 2016;**57**:1221–1227.

68. Ashimatey BS, King BJ, Burns SA et al. Evaluating glaucomatous abnormality in peripapillary optical coherence tomography enface visualisation of the retinal nerve fibre layer reflectance. *Ophthalmic Physiol Opt* 2018;**38**:376–388.

CET multiple choice questions

This article has been approved for one non-interactive point under the GOC's Enhanced CET Scheme. The reference and relevant competencies are stated at the head of the article. To gain your point visit the College's website **college-optometrists.org/oip** and complete the multiple choice questions online. The deadline for completion is 30 April 2021. Please note that the answers that you will find online are not presented in the same order as in the questions below, to comply with GOC requirements.

1. Primary open-angle glaucoma affects approximately what percentage of the European population?

- 6%
- 2%
- 0.8%
- 4%

2. Spectral-domain optical coherence tomography:

- Has similar axial resolution in the retina to the latest scanning laser polarimetry instruments
- Has improved transverse resolution but reduced axial resolution in the retina compared to time-domain OCT
- Can measure the whole retinal thickness in the macula but not the thickness of individual retinal layers
- Enables assessment of the optic nerve head using Bruch's membrane opening as an anatomical landmark

3. The assessment of the macula for the diagnosis of early glaucoma:

- Can be conducted by slit-lamp binocular indirect ophthalmoscopy
- Is not advocated, since central vision is typically affected only at later stages of the disease
- Is particularly useful in high myopes
- Offers the best opportunity to detect early disease compared to other scan areas

4. The diagnostic accuracy of OCT for glaucoma:

- Increases as a function of the disc size
- Is greater in pre-perimetric glaucoma compared to glaucoma with manifest visual field loss
- Is greater for macular parameters compared to cpRNFL parameters
- Increases with advancing disease stage

5. A 'within' or 'outside normal limit' result from the comparison of OCT measurements with population normative data:

- Can misclassify patients because of large variability in healthy anatomy
- Might be unreliable in the presence of ocular comorbidities or high refractive error
- Has varying diagnostic accuracy across different patient ethnicities
- All the above

6. The automated combination of different OCT indices for a more accurate diagnosis of glaucoma:

- Should combine only indices from the nasal cpRNFL sector and superior macula
- Still requires refinement before translation into clinics
- Enhances diagnostic accuracy only when more than five indices are combined
- None of the above

CPD exercise

After reading this article, can you identify areas in which your knowledge of the use of optical coherence tomography in the detection and diagnosis of glaucoma has been enhanced?

How do you feel you can use this knowledge to offer better patient advice?

Are there any areas you still feel you need to study and how might you do this?

Which areas outlined in this article would you benefit from reading in more depth, and why?

Depth-resolved variations in visibility of retinal nerve fibre bundles across the retina in enface OCT images of healthy eyes

Riccardo Cheloni [iD] and Jonathan Denniss [iD]

School of Optometry and Vision Science, University of Bradford, Bradford, UK

Abstract

Purpose: Recent developments in optical coherence tomography (OCT) technology enable direct enface visualisation of retinal nerve fibre bundle (RNFB) loss in glaucoma. However, the optimum depth at which to visualise RNFBs across the retina is unknown. We aimed to evaluate the range of depths and optimum depth at which RNFBs can be visualised across the retina in healthy eyes.

Methods: The central $\pm$ 25° retina of 10 healthy eyes from 10 people aged 57--75 years (median 68.5 years) were imaged with spectral domain OCT. Slab images of maximum axial resolution (4 μm) containing depth-resolved attenuation coefficients were extracted from 0 to 193.5 μm below the inner limiting membrane (ILM). Bundle visibility within 10 regions of a superimposed grid was assessed subjectively by trained optometrists ($n = 8$), according to written instructions. Anterior and posterior limits of RNFB visibility and depth of best visibility were identified for each grid sector. Effects of retinal location and individual eye on RNFB visibility were explored using linear mixed modelling with likelihood ratio tests. Intraclass correlation coefficient (ICC) was used to measure overall agreement and repeatability of grading. Spearman's correlation was used to measure correlation between depth range of visible RNFBs and retinal nerve fibre layer thickness (RNFLT).

Results: Retinal location and individual eye affected anterior limit of visibility ($\chi^2_{(9)}$ = 58.6 and 60.5, both $p < 0.0001$), but none of the differences exceeded instrument resolution, making anterior limit consistent across the retina and different eyes. Greater differences were observed in the posterior limit of visibility across retinal areas ($\chi^2_{(9)}$ = 1671.1, $p < 0.0001$) and different eyes ($\chi^2_{(9)}$ = 88.7, $p < 0.0001$). Optimal depth for visualisation of RNFBs was around 20 μm below the ILM in most regions. It varied slightly with retinal location ($\chi^2_{(9)}$ = 58.8, $p < 0.0001$), but it was not affected by individual eye ($\chi^2_{(9)}$ = 10.7, $p = 0.29$). RNFB visibility showed good agreement between graders (ICC 0.89, 95%CI 0.87–0.91), and excellent repeatability (ICC 0.96–0.99). Depth range of visible RNFBs was highly correlated with RNFLT ($\rho = 0.9$, 95%CI: 0.86–0.95).

Conclusions: The range of depths with visible RNFBs varies markedly across the healthy retina, consistently with RNFLT. To extract all RNFB information consistently across the retina, slab properties should account for differences across retinal locations and between individual eyes.

Introduction

Optical coherence tomography (OCT) is now widely available in glaucoma clinics, being used to objectively quantify structural changes to the retina and optic nerve head.[1,2] Nevertheless, glaucoma detection, diagnosis and monitoring are still imperfect.[3–5] As such, glaucoma and its progression must currently be confirmed across multiple tests, resulting in delayed treatment and increased socio-economic burden.[6] There remains, therefore, considerable interest in improved identification of glaucomatous damage, which has led to consideration of retinal nerve fibre layer (RNFL) reflectance information from enface OCT images.[7,8]

Retinal nerve fibre bundles (RNFBs) are highly reflective compared to other retinal structures because of the ordered structure of ganglion cell axons' cytoskeleton.[9] Recent OCT devices allow us to qualitatively explore reflectance in enface images of the RNFL.[10–12] Enface analysis is often accomplished with volumetric scans by collapsing the pixel intensity of each A-scan over a certain range of depths into a 2D image. The result is a single transverse retinal section, or slab, frequently derived from a fixed retinal thickness (often 50 μm) beginning anteriorly from the inner limiting membrane (ILM).[11,12] In these images, glaucomatous defects appear as regions of impaired reflectance, often in typical arcuate or wedge shapes as well as more generalised loss. Defects are likely to be induced from a combination of primary loss of reflectivity, caused by cytoskeletal disruption,[9,13] and thinning of the RNFL which leads to the inclusion of deeper hypo-reflective retinal layers in the slab.[11] Previous studies of enface OCT imaging in glaucoma showed a strong correlation with circumpapillary RNFL defects,[10] and excellent agreement between residual RNFBs and preserved visual function in the macula.[12,14]

It has been suggested that disruption of RNFB reflectivity may be measurable earlier than RNFL thinning in glaucoma,[8,15] making this approach encouraging for earlier identification of defects. Further, clinicians are encouraged to confirm a suspected structural lesion in the functional domain to increase confidence in diagnosis.[16–19] Enface images offer the opportunity to relate structure and function without use of structure-function maps, thought to represent an additional source of noise in this relationship.[20] Indeed, newer perimetric strategies that aim to incorporate structural information for greater efficiency,[21–23] or that assess specific regions of interest in greater detail,[24–26] may be facilitated by enface imaging.

Although the use of enface OCT images is promising, current understanding of the arrangement of visible RNFBs in healthy eyes is lacking. Understanding of the normal appearance of RNFBs in enface OCT images is important in the future development of objective methods for detecting glaucomatous defects in these images. For instance, it is unclear how RNFB visibility in enface OCT images varies across the healthy retina and between individual eyes. It is, therefore, also unclear what depths should be considered to capture a consistent portion of the RNFL through the retina. Using a fixed thickness slab to examine portions of the retina with varying morphology may be expected to lead to an uneven composition whereby some areas contain RNFL only, whilst others also include deeper retinal layers with different reflectivity. This would most likely result in different appearance and detectability of glaucoma defects across the retina.[10] Additionally, since there may be situations where it is desirable to use a single slab thickness across the retina, the axial placement of this slab within the thicker regions of RNFL may affect the detectability of glaucomatous defects. It may therefore be advantageous to know whether there is an optimum depth for visualisation of RNFBs and how this varies across the retina. Looking for glaucomatous defects at this depth may achieve higher specificity than at other depths if it results in a lower chance of misidentifying normal variation as a defect.

In this study we aimed to identify the range of depths and the optimal depth for the subjective visualisation of RNFBs across different retinal regions in eyes of healthy adults. Further, we assessed the agreement between- and repeatability within- clinicians in determining RNFB visibility. Since RNFL thickness varies across the healthy retina,[27] we expected RNFBs to be visible over a range of depths that varies across the retina and between individual eyes. The results contribute to a preliminary understanding of the three-dimensional configuration of RNFBs in healthy eyes as viewed subjectively in enface OCT images, identifying the depths that future automated image analysis methods should consider to consistently assess the RNFL.

Methods

Participants

Healthy volunteers aged over 50 years were recruited for imaging. We targeted older adults to include an age range relevant to primary open angle glaucoma.[28] One eye per participant was included. Participants were eligible if they had visual acuity ≤ 0.20 logMAR (6/9.5 Snellen) measured at 6 m with appropriate refractive correction using an electronic logMAR chart. Further inclusion criteria were clear optical media and normal visual field (Humphrey Field Analyser III SITA standard, www.zeiss.com/meditec), as defined by normal mean deviation ($p > 0.05$), Glaucoma Hemifield Test within normal limits and absence of three contiguous non-edge points with $p < 5\%$ on the pattern deviation plot. Participants were excluded if they had any condition affecting their eyes or visual system, or

intraocular pressure > 21 mmHg in either eye or between eye difference > 4 mmHg by Goldmann applanation tonometry. Written informed consent was given by each participant. Ethical approval was obtained from the National Health Service Research Ethics Service and the study adhered to the Declaration of Helsinki.

OCT imaging and processing

Multiple high density, high-speed scans encompassing the central ± 25° of the retina were acquired with Spectralis OCT (Heidelberg Engineering, www.heidelbergengineering.com). Seven volume scans comprising 30 μm-separated B-scans (9.65 B-scans per degree) were captured in different positions of gaze (Figure 1). Measurement noise was reduced by the use of built-in automated retinal tracking to average 16 B-scans per scan location. The orientation of B-scans was adjusted to yield high image resolution in a shorter time (Figure 1). Pupils were dilated only if inadequate scan quality was achieved in physiological conditions (i.e., below 25db as suggested by the manufacturer and elsewhere[29]). All images were acquired with signal to noise ratio above 20dB as per the manufacturer's instructions. Traditional OCT imaging of the optic nerve head (ONH) was also performed, including a circle B-scan (3.5 mm diameter) around the ONH that provided the mean circumpapillary RNFL (cpRNFL) thickness as automatically computed by Spectralis built-in software.

For each eye, we extracted 50 single-pixel thick slabs containing depth-resolved attenuation coefficients[30] encompassing the volume from 0 to 193.5 μm below the ILM. Each slab represents a 3.87 μm thick slab through the

retina, corresponding to the instrument's digital axial resolution. The greatest depth explored with this method (193.5 μm) was selected in order to include the maximum expected cpRNFL thickness for a 50-year-old healthy individual. As proposed elsewhere,[10] we extracted attenuation coefficients instead of the raw intensity of each pixel using equations 17 and 18 from Vermeer et al.[30] Attenuation coefficients are purported to describe how quickly incident light is attenuated when passing through the retinal portion of interest irrespective of the amount of light received.[31] Although attenuation coefficients were originally developed to minimise shadowing effects of blood vessels on underlying tissue, they should also minimise reflectance artefacts from media opacities and poor quality B-scans.[10] Images for each retinal area and depth were converted from the raw output of the Spectralis to attenuation coefficients using custom software written in R version 3.6.3(www.r-project.org),[32] prior to exporting to Matlab version 9.6.0 for further processing (MathWorks, www.mathworks.com).

At each depth, for each individual eye, we collated the seven arrays together into a single pseudo-montage image (Figure 1b). Because of some extent of overlap between individual scans, this method resulted in partial duplication of information from certain retinal areas that was accounted for at later stages (see below). Then, we applied a smoothing filter (3 × 3 Wiener filter) to all images to reduce noise. Subsequently, a 100 × 50 pixel rectangle was extracted from the raphe area of the composite image 35 μm below the ILM of each participant, and a value 30% below the median pixel value within this rectangle was computed as 'background'. The background value is,
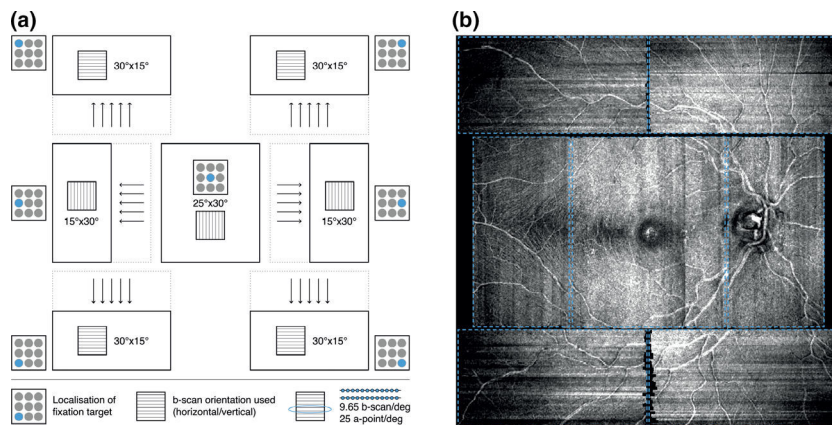
Figure 1. (a) The wide field OCT acquisition protocol comprised of 7 volume scans (black boxes) shown within the acquisition window (dashed boxes) at each scan location. Scan dimensions and orientation are shown for every location. Arrows indicate the movement of scan positions from the manufacturer's default positions. The level of overlap between different scans was considerable between superior scans and inferior scans as well as between superior/inferior and central scans. The overlap between central scans was instead marginal. (b) Example of the resulting attenuation coefficients pseudo-montage for one eye 20 μm below the inner limiting membrane. Individual volume scan positions are shown by the blue boxes.

therefore, computed from an area expected to be lacking any RNFBs. This was used as a reference to reset the lower limit of the image.[10] A two-step thresholding process was then applied. First, the background value was subtracted from all pixels of all slabs, with values falling below 0 being clipped to 0. Then, for each participant the median of the 99th percentiles of all 50 slabs was computed and used to normalise each slab by division, with values above 1 being clipped to 1. This process left all pixel values within a 0 to 1 range, before 8-bit images were extracted for display in the grading task.

Grading task

Subjective visibility of RNFBs was rated over 10 retinal regions of interest, each approximately consisting of a 15 × 15° square with the exception of the macular area that was further split into temporal and nasal regions (*Figure 2*). To account for inter-subject anatomical variability,[33,34] the central row of the grid was tilted to follow each individual's fovea-disc and fovea-raphe angles (*Figure 2*). The number of retinal regions delimited represented a trade-off between resolution and the time required to complete the task within manageable limits. To minimise potential bias from duplication of information, overlapping areas from different scans were obscured and graders were instructed to ignore those areas during the task (*Figure 2*). An example of the presentation file presented to clinicians is available as supporting material (Supplementary Material 1).

Images were displayed in randomised order on a MacBook Pro 13" (2017 version, Apple, www.apple.com) under standardised lighting conditions. Before performing the task, all graders were provided with the same written instructions (Supplementary Material 2), which included information on characteristic appearance of RNFBs, visible artefacts and potential confounders (e.g., putative glial alteration[35]). Subsequently, they were allowed to scroll through the images for a given eye unlimited times and without a time limit. The following subjective judgements were collected for each region of each imaged eye:

1. Boundaries of RNFB visibility: The first depth below the ILM at which RNFBs become visible in that specific region (anterior limit – μm) and the depth below the ILM at which RNFBs are last visible in that specific region (posterior limit – μm). Graders were instructed to consider RNFBs to be present when 25% or more of the region was occupied.

2. Best visibility of RNFBs: The depth of greatest RNFB visibility (μm) according to features such as intensity and sharpness while ignoring the proportion of the region occupied from RNFBs. Since the best visibility value aims to identify a single pixel depth for optimum observation of bundles, in the case of equal visibility of RNFBs across multiple depths, graders were invited to report 'none'.

Eight optometrists including the two authors (median age 29.5 years, range 26–41; median years since qualification 7, range 2–16) completed the grading task. The optometrists had varying levels of post-qualification training
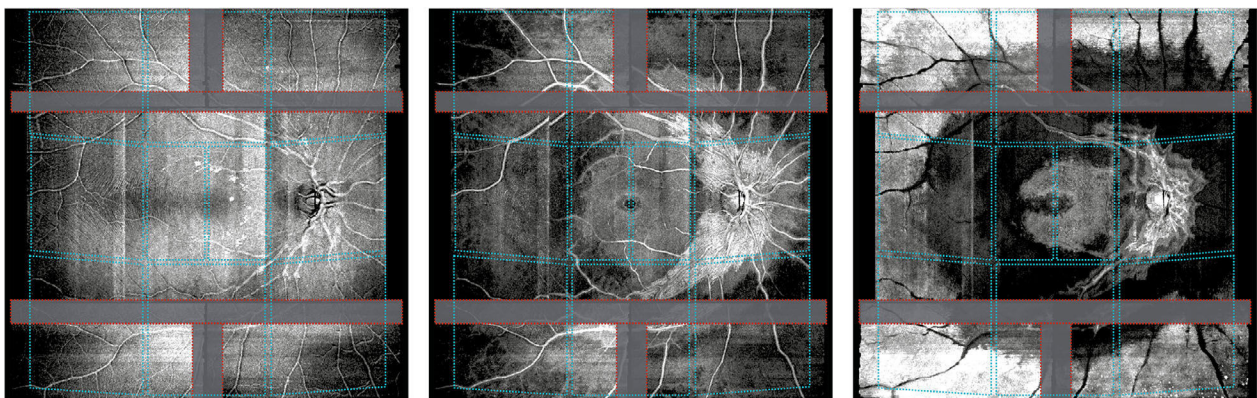
Figure 2. Example of the grid (blue lines) superimposed onto 3 of the 50 pseudo-montaged slab images for one eye, delimiting the regions of interest in the retina. Grey patches within red dashed bounding lines indicate the overlapping areas that graders were instructed to ignore. From left to right, the images correspond to 12 μm, 93 μm and 186 μm below the ILM, showing how the visible presence of RNFBs changes moving away from the ILM. At 12 μm below the ILM, RNFBs are visible throughout almost the whole retina, with exception of the raphe, the fovea and peripheral portions of the inferior temporal, nasal and superior temporal regions. The reduction of reflectivity in these areas can be interpreted as absence of RNFBs. At 93 μm below the ILM, RNFBs are present only in the ONH region, with arcuate wedges of RNFBs in the inferior and superior central retina following the vascular arcades. Lastly, at 186 μm below the ILM, it is possible to appreciate hyper-reflectivity of the retinal pigment epithelium in almost all peripheral regions and the macula, whereas only a few visible bundles are present around the ONH.

(four out of eight had undertaken specific training in glaucoma), but apart from the two authors they were all naïve to enface OCT images showing reflectivity details of healthy retinae. All were pre-presbyopic, with self-reported normal vision, and were invited to perform the task with their habitual refractive correction. To assess repeatability of the gradings, on a subgroup of two graders, the task was repeated three times on three randomly selected eyes and with no information regarding previous grades available. Repeated measures were taken one week apart.

Data analysis

All data were anonymised before analysis in the open-source environment R.[32] The effects of retinal region and individual eye on anterior limit of visibility, posterior limit of visibility and best depth for RNFB visibility were explored with linear mixed modelling using the R package *lme4* (www.r-project.org).[36] Means were computed to summarise ratings from all graders at each region of each eye. For greater robustness of results, best visibility analysis was limited to cases where the proportion of 'none' ratings was below 30%. The first model tested whether retinal region affected boundaries of visible bundles and best visibility, accounting for random effects of eye and grader. This model had the form:

$$y \sim 1 + region + (1|eye) + (1|grader) + \varepsilon \quad (1)$$

where y signifies the measure of interest (e.g., best visibility depth), 1 signifies the intercept and ϵ signifies random error. Second, effects of individual eye on RNFB visibility were tested accounting for different retinal regions and graders using a model of the form:

$$y \sim 1 + eye + (1|region) + (1|grader) + \varepsilon \quad (2)$$

whose symbols are as defined for equation 1. Chi-squared likelihood ratio tests were used to assess statistical significance of variables, with $p < 0.05$ considered significant. Where relevant, post-hoc analysis was performed to test pairwise differences, adjusting for multiple comparisons with the Tukey method.

To explore overall agreement between graders and repeatability within graders, we computed intraclass correlation coefficients (ICC), according to Shrout and Fleiss,[37] using the R package *psyc* (www.r-project.org).[38] ICC classes ICC(2,1) and ICC(3,1) for a single rating were used to estimate reliability and repeatability, respectively. ICC is strongly influenced by the variance of the sample,[39] thereby not allowing for comparison among subgroups of data with inconsistent variances. Hence, to further evaluate variability between graders for different tasks and retinal regions we

also computed the central 90% range width (difference between 5th and 95th percentiles) of ratings at each region of each image. This was done for ratings of anterior limit, posterior limit and best visibility. We henceforth refer to this measure as between-grader variability. As above, linear mixed models were used to explore the effects of retinal region and individual task on agreement. The model testing the effect of retinal region on between-grader variability included task and individual eye as random effects and had the form:

$$variability \sim 1 + region + (1|task) + (1|eye) + \varepsilon \quad (3)$$

where 1 signifies the intercept and ϵ signifies random error. Effects of task were assessed accounting for retinal region and individual eye:

$$variability \sim 1 + task + (1|region) + (1|eye) + \varepsilon \quad (4)$$

where terms are as per equation 3.

We also sought validation that the object of the subjective grading task was actually the visibility of RNFBs, and not some other retinal image feature. We therefore evaluated correlation (Spearman's ρ) between thickness of visible RNFBs, computed as the axial distance (μm) between the subjectively-defined anterior and posterior limits of visible RNFBs, and mean RNFL thickness in the corresponding region. To this end, RNFL thickness data as automatically segmented by Spectralis OCT were extracted from the wide-field scans and the mean thickness computed along the same grid adopted for the grading task. Lastly, correlation between thickness of visible bundles and between-grader variability was tested.

Following a previous study,[40] we calculated that the selected combination of number of graders, images and repetition should produce 95% confidence intervals (CI) within 0.1 for the agreement ICC for any value of ICC, and within 0.1 for repeatability ICC > 0.79.

**Table 1.** Demographics and RNFL parameters for the 10 imaged eyes

Participant	Eye	Age (y)	Mean cpRNFL thickness (μm)	Mean thickness of visible RNFBs (μm)
1	R	57	92	38
2	R	63	108	48
3	R	66	90	43
4	L	72	97	42
5	R	69	111	50
6	R	73	98	43
7	L	67	99	48
8	R	68	96	42
9	L	75	78	38
10	R	73	103	46

Results

Images from 10 eyes from 10 healthy participants were included in the study (5 females, 9 Caucasian, median age 68.5 years, range 57–75), whose demographic and clinical details are reported in *Table 1*. The grading task required a median total of 73 min (range 51–144 min).

The grand mean and (±) standard deviation for anterior limit of visibility was 9.9 ± 0.8 μm (range 8.8 ± 1.5 μm for superior nasal region to 11.3 ± 1.5 μm for ONH region) and the grand mean for posterior limit of visibility was 53.7 ± 30.6 μm (range 28.1 ± 1.2 μm for raphe region

to 131.6 ± 13.7 μm for ONH region). As shown in *Figure 3*, retinal region affected the anterior limit of RNFB visibility ($\chi^2_{(9)} = 58.6$, $p < 0.0001$), and the greatest difference was found between the optic disc region and the superior nasal region (2.5 μm, $p < 0.0001$). Nonetheless, this difference was below the instrument's digital axial resolution, and is therefore not clinically significant. Similarly, there were significant differences between individual eyes in anterior limit of RNFB visibility ($\chi^2_{(9)} = 60.5$, $p < 0.0001$), but again these differences did not exceed digital axial resolution, and so are not clinically significant (greatest pairwise difference: 2.5 μm, $p < 0.0001$). Conversely,
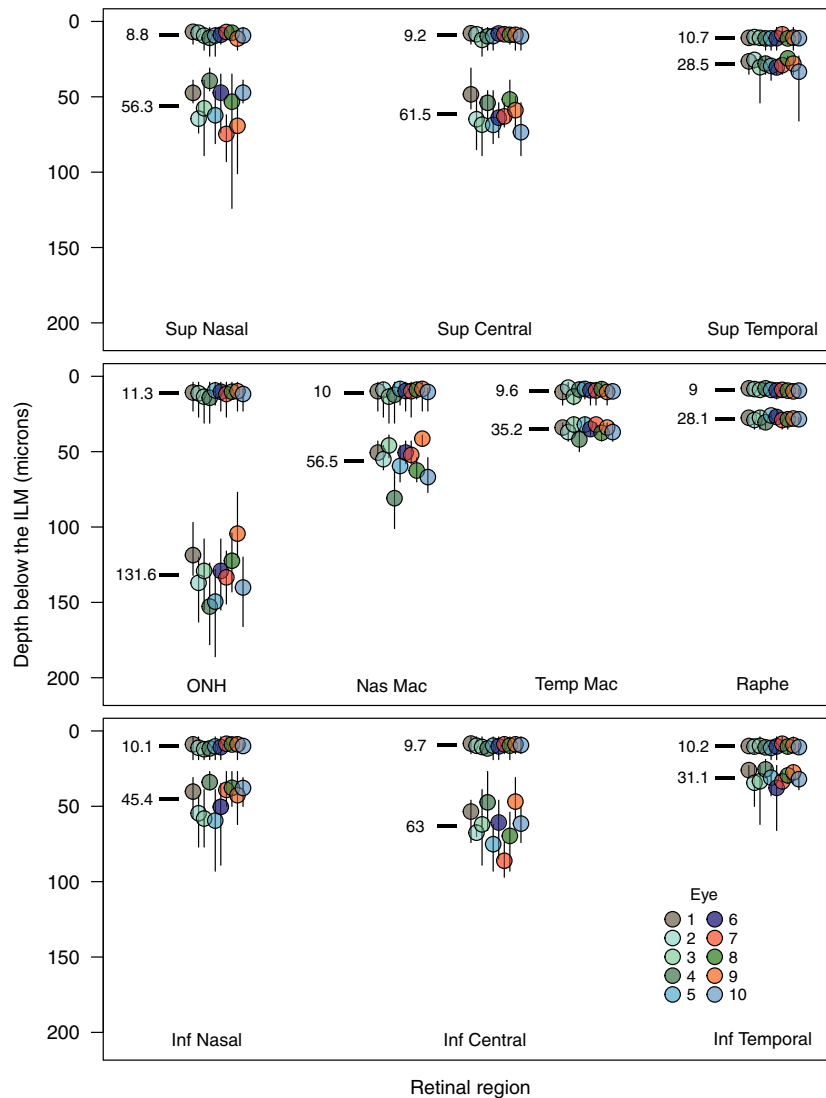
**Figure 3.** Anterior and posterior limits of RNFB visibility (μm) at different retinal regions. The top, middle and lower panels correspond to the superior, central and inferior retina respectively. Points and error bars show respectively the mean and range of all ratings for each of the 10 individual eyes. Black lines and numbers on the left of each cluster of points show the group mean limit of visibility across all eyes. Points are colour-coded according to individual eyes numbered as in *Table 1*. ILM, inner limiting membrane, ONH, optic nerve head; Sup, superior; Inf, inferior; Nas, nasal; Temp, temporal; Mac, macula.

differences in posterior limit of RNFB visibility across both retinal regions ($\chi^2_{(9)} = 1671.1$, $p < 0.0001$) and individual eyes ($\chi^2_{(9)} = 88.7$, $p < 0.0001$) were both statistically and clinically significant (mean difference between retinal regions 31.1 µm, range 0.3–103.5 µm; mean difference between individual eyes 4.6 µm, range 0.2–12 µm). Mean limits of RNFBs visibility for individual regions are reported in *Figure 3*.

As shown in *Figure 4*, the optimum depth for best RNFB visibility was affected by retinal region ($\chi^2_{(9)} = 58.8$, $p < 0.0001$). The rate of ungradable judgements for each

retinal region is shown in *Figure 4*, and it was within 30% for most of retinal regions. In the ONH region, graders could not identify a single depth with best visibility of bundles 40% of the time. For greater robustness, this region was excluded from this analysis. The grand mean for best RNFBs visibility was 20.3 ± 1.9 µm (range 17.4 ± 1.3 µm for raphe region to 22.8 ± 2.0 µm for the nasal macula), whereas mean limits for individual regions are reported in *Figure 4*. Pairwise analysis showed most of the significant differences to be just above the instrument's digital axial resolution (greatest difference between nasal macula and
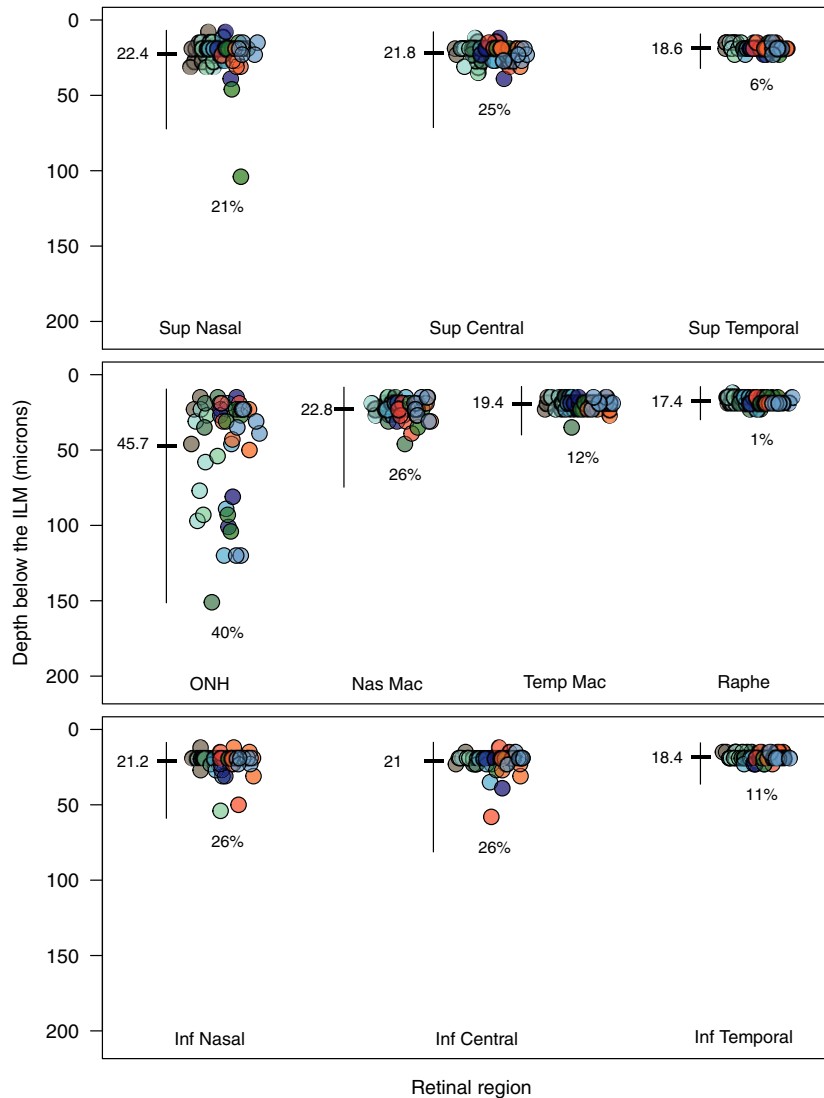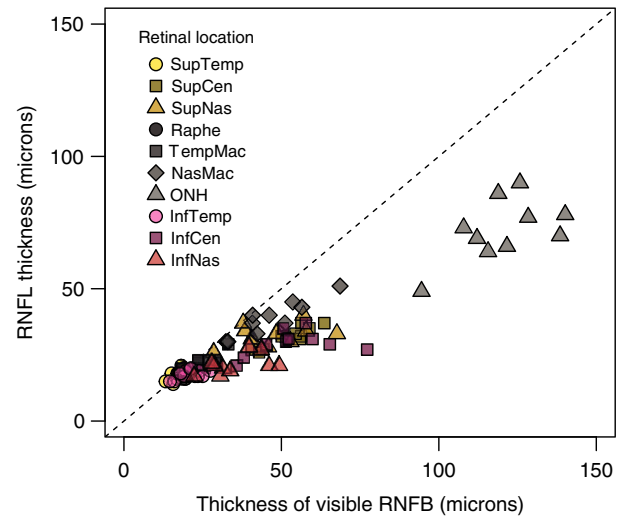

Figure 4. Optimum depth for best RNFB visibility at different retinal regions. Individual points show individual graders' ratings of individual eyes with eyes colour-coded as in *Figure 3*. Horizontal black bars and numbers show mean ratings across all graders and eyes. Vertical bars indicate depths where RNFBs were visible (5th and 95th percentiles of overall anterior and posterior limit of visibility respectively). Percentage of ungradable ratings (no single best depth identified) is shown for each location. Panel arrangement and abbreviations are as in *Figure 3*.
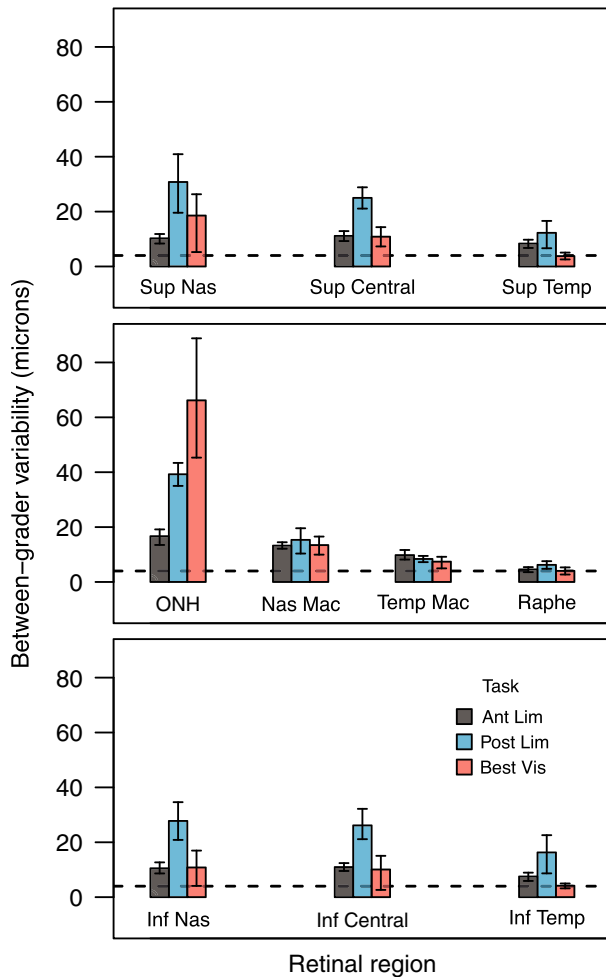
Figure 5. Between-grader variability in ratings for each task and retinal region. Between-grader variability was computed as the central 90% range width (difference between 5th and 95th percentiles) of ratings at each region of each image. Panel arrangement and abbreviations as in *Figure 3*. Error bars represent bootstrap 95% confidence intervals. Horizontal dashed lines indicate the instrument's digital axial resolution.

Figure 6. Thickness of visible retinal nerve fibre bundles (RNFB) vs. mean retinal nerve fibre layer (RNFL) thickness in the corresponding retinal sector. Points are coded according to the retinal location as shown in the key. The dashed line represents a 1:1 relationship.

raphe regions, 5.4 μm, $p < 0.0001$). In contrast, differences in optimum depth for best RNFB visibility between eyes were not significant ($\chi^2_{(9)} = 10.7$, $p = 0.29$).

Overall we found good agreement between graders' ratings of RNFB visibility (ICC(2,1) = 0.89, 95%CI: 0.87–0.91). Similarly, graders' estimates were highly repeatable (grader 1, ICC(3,1) = 0.96, 95%CI: 0.95–0.97; grader 2, ICC(3,1) = 0.99, 95%CI: 0.98–0.99). Between-grader variability is shown in *Figure 5*. Retinal region had a significant effect on between-grader variability ($\chi^2_{(9)} = 158.3$, $p < 0.0001$), with greatest differences found between the ONH region and other retinal regions (max difference: ONH/raphe = 35.8 μm, $p < 0.0001$). Ratings appeared progressively more variable in nasal regions compared to

temporal regions (superior nasal-superior temporal: 11.74 μm, $p = 0.003$; ONH-raphe: 35.76 μm, $p < 0.0001$; and inferior nasal-inferior temporal: 7.04 μm, $p = 0.33$). Similarly, task significantly affected between-grader variability ($\chi^2_{(9)} = 38.5$, $p < 0.0001$), with the identification of the posterior limit of RNFB visibility being the most variable (mean between-grader variability 20.7 μm) compared to anterior limit (10.3 μm) and best visibility (14.9 μm). Differences in between-grader variability among tasks were all statistically significant once adjusted for multiple comparisons (anterior and best: −4.63 μm, $p = 0.01$; anterior and posterior: −10.44 μm, $p < 0.0001$; best and posterior: −5.81 μm, $p = 0.001$).

Figure 6 shows the strong correlation between thickness of visible RNFBs (axial distance between anterior and posterior limits of visibility) and RNFL thickness as measured by the Spectralis OCT (Spearman's ρ: 0.9, 95%CI: 0.85–0.93, $p < 0.0001$). For all tasks, consistency in ratings between clinicians showed inverse correlation with thickness of visible RNFBs (greater consistency was apparent in thinner areas, Spearman's ρ: 0.66, 95%CI: 0.58–0.72, $p < 0.0001$).

Discussion

The detection of RNFB abnormalities in enface OCT images is a promising approach to diagnose and monitor glaucoma, complementary to existing morphological examination.[10–12,41,42] Yet clinical use of this information is limited by coarse understanding of the normal appearance of

RNFBs at different depths and a lack of methods to objectively define defects. For both reliable imaging and defect definition, ideally achieved by objective and quantitative methods, we require adequate understanding of configuration in healthy eyes, including the depths at which RNFBs are visible across the retina. Such information was not previously available and therefore was the objective of this investigation.

Our data suggest that the first depth below the ILM at which bundles become visible is consistent throughout the retina as well as in different eyes. Conversely, retinal region and between-eye factors had statistically and clinically significant effects on the posterior limit of visibility, requiring consideration when extracting enface slab images. Notably, pairwise analysis showed that not all retinal regions had a different posterior limit of visibility, identifying some with a similar range of depths of visible bundles (*Figure 7a*). Optimum depth for best bundle visibility varied with retinal region but was similar across different eyes. The number of ungradable ratings changed across the retina, exceeding our chosen 30% cut-off in the ONH region where optimal visibility was also found at a wider range of depths. Among the remaining retinal regions, best depth was often found at around 20 μm below the ILM (mean 20.3 μm, range 17.4–22.8), and although pairwise differences were statistically significant for a few regions (*Figure 7b*), they were close to the OCT instrument's digital axial resolution (greatest difference: 5.4 μm, $p < 0.0001$), meaning that depth differences were 1–2 pixels at most. This index provides a single depth at which visualisation of RNFBs appeared optimal across the retinae of the studied

eyes, and at which any associated enhanced value in detecting defects could be further explored in future studies of glaucoma.

Overall, we found good repeatability and conformity between graders for identification of RNFB visibility, supporting findings from a recent study where raters consistently reported the presence or absence of bundles.[14] A more detailed analysis of between-grader variability showed greater variability for best visibility and posterior limit of visibility compared to anterior limit of visibility. Further, we found between-grader variability to change with retinal region, becoming progressively greater in the nasal regions, where the RNFL thickness is known to increase.[27,43] It is possible that the presence of visible bundles at an increased range of depths led to this increase in variability between-graders. This hypothesis is also supported by the correlation between between-grader variability and the thickness of visible bundles in the corresponding region. Even though clinicians were more variable in certain retinal regions, the overall high to excellent ICCs support the validity of the visible ranges identified.

No previous studies have characterised RNFB visibility of the central ± 25° of healthy retinae. Additionally, our analysis is the first to consider single slabs of the maximum axial resolution as the unit of assessment of RNFB reflectivity. Most previous work has collapsed a volume of a fixed retinal thickness into a single 2D slab image to explore glaucomatous defects.[10–12,14] Among the earliest, Hood et al[11] assessed loss of reflectivity of the central retina using fixed thickness slabs obtained by averaging intensity values up to 52 μm below the ILM (roughly corresponding to the
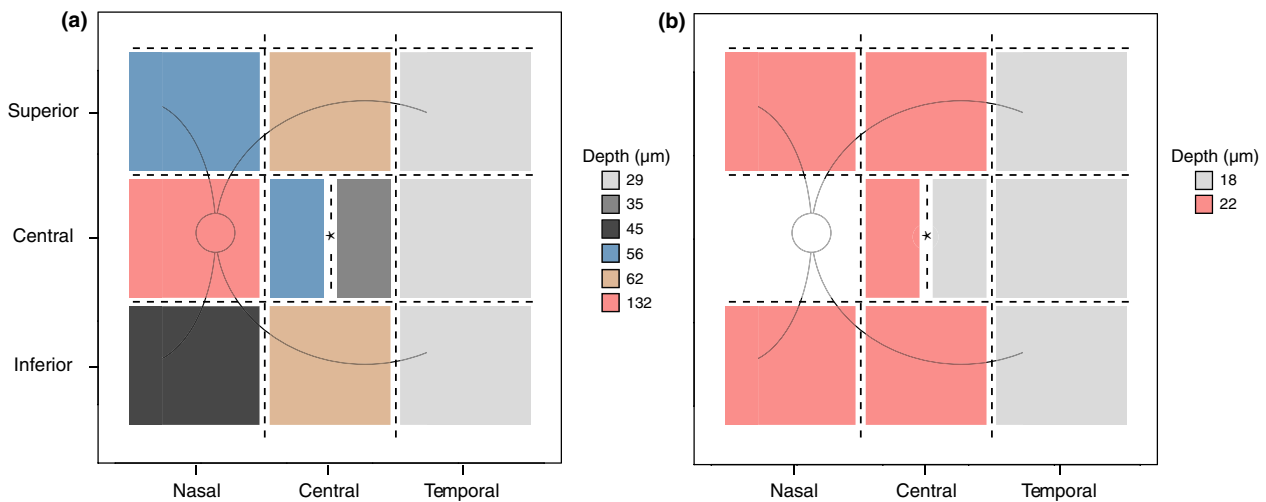
Figure 7. Schematic representations of pairwise differences between retinal regions for: (a) posterior limit of visibility; and (b) the optimum depth of visibility. Retinal regions are coloured differently in cases of a statistically significant difference exceeding the instrument's digital axial resolution (4 μm). This analysis identified 6 and 2 macro-regions for the posterior limit and best visibility respectively. The anterior limit of visibility was omitted since differences between retinal regions were all below 4 μm.

average of 14 slabs in our study). Their axial depth was a compromise between a small thickness – able to show local changes of reflectivity – and a large enough number of pixels to reduce noise. The authors acknowledged that in certain damaged regions the RNFL might have been as thin as 20–30 μm, leading to the inclusion of deeper hypo-reflective layers (i.e., ganglion cell and inner plexiform layers). Their analysis may also have been limited by the incomplete capture of the full depth of RNFBs in some regions. In a larger study, Ashimatey et al[10] extracted single slabs of varying thickness to minimise artefacts from hyper-reflective glial alterations found immediately below the ILM. They averaged pixels from 24 to 52 μm below the ILM in the ONH area, from 24 to 36 μm in the region between ONH and fovea and lastly reduced the sampled depths to 16–24 μm in the temporal macula/raphe area. Although the approach retrieved a large proportion of reflectivity defects, it was possible to identify all RNFB lesions when the inspection was extended beyond 52 μm below the ILM and generalised loss was considered.[10] In a study exploring the structure-function relationship in the macula of eyes with advanced glaucoma, Sakamoto et al[12] used 50 μm fixed-thickness slabs from the ILM. The authors argued that the adopted axial depth would include approximately 20 μm from the ganglion cell layer beneath a thinned RNFL but that this would be unlikely to alter reflectance results due to the hypo-reflectance of the ganglion cell layer.[12]

Using a fixed depth slab to sample regions of the retina with different RNFL thickness may fail to capture all glaucomatous defects. While often included as a limitation, most studies using enface OCT imaging in glaucoma have not accounted for differences between retinal regions, nor between different eyes.[10–12,14] According to our data, the conventional 50 μm thick slab might only be inclusive of all RNFBs in a limited number of retinal regions. In the nasal macula, the superior/inferior central retina and the ONH region, information regarding the status of bundles might be overlooked by this approach (Figure 3), as shown in one previous report.[10] Further, such methods would likely exceed the RNFL in its thinner regions, resulting in the inclusion of deeper, hypo-reflective retinal layers. Consequently, the observed lower intensities of RNFBs could arise not only because of a primary loss of reflectivity but also from the inclusion of deeper layers in the slab. As such, the ability to identify enface defects may vary according to slab thickness and composition. Since RNFBs are visible at different depths in different areas of the retina, slab thickness should also vary across the retina to include a consistent proportion of the RNFL.

It has been suggested that glaucomatous RNFL defects observed as loss of reflectivity might not necessarily be matched by a reduction in thickness.[11,41] Reflectivity defects may, therefore, provide additional information on

RNFL status in glaucoma.[41,42] Disagreement between these two approaches might have several sources, including the method used to generate data. Indeed, thickness analysis relies fully on segmentation which is sometimes inaccurate, especially for the proximal RNFL boundary.[11] Conversely, enface imaging depends only on the vitreous-ILM segmentation, possibly the easiest surface to automatically detect.[44] Additionally, and perhaps more importantly, current enface images provide a combination of RNFL thickness and intensity, which might show evidence of glaucoma lesions on a different timescale.[8,11,15] Despite possible disagreement in glaucoma, RNFL thickness and the range of depths across which bundles are visible were highly correlated in the healthy eyes assessed in this study (Figure 6). This strong correlation provides face validity that the graders' subjective ratings were indeed based on visible RNFBs, supporting the clinical relevance of the technique. Although highly correlated, the slope of the points in the scatterplot is below the 1:1 line, suggesting that the range of depths across which bundles are visible overestimates RNFL thickness. This overestimate is likely to be an artefact of our instructions to graders; bundles were considered to be visible when occupying more than 25% of the region. This could have led to incongruences with the RNFL thickness, calculated as the mean across the region, which may be more likely to represent the depth with visible bundles in 50% of the region.

Several limitations of this study require consideration. Firstly, only a small number of healthy eyes of a single ethnic group (9 of the 10 participants were Caucasian), were imaged. Although this provided adequate power to assess agreement amongst graders, it is possible that some more unusual RNFL configurations or thicknesses may not have been captured in our sample. Similarly, the opportunistic selection of a small number of eyes may limit generalisability to different populations or settings. Nonetheless, the included age group was selected to be similar to that of open angle glaucoma,[28] limiting the impact of healthy ageing on the applicability of our findings.

Previous work considered raw OCT intensities to explore reflectivity defects of the RNFL,[7] yet the overall incident light strength, media quality and age were reported to affect measurements.[42,45–47] To account for the variable amount of incident light, pixel intensity can be normalised to that of a reference layer.[41,42,48] Alternatively, as in our study, depth-resolved attenuation coefficients can be used to determine intrinsic optical properties of retinal tissue independent of the amount of light received as well as of additional segmentation requirements.[30,31,49] Whilst this method is expected to reduce artefacts, one investigation suggested that age and scan quality may still have a significant effect on attenuation coefficients.[49] In this study, these effects were likely to be minimised by the qualitative nature

of the grading task, the inclusion of subjects only within the age range of interest and the inclusion only of scans well above the manufacturer's recommended minimum quality.

Finally, the image processing methods applied in our work have some drawbacks. For instance, as proposed elsewhere[10] we accounted for background intensity of individual images by extracting the value in the raphe region. However, this relies on the assumption of constant background throughout the whole retina which may not necessary hold for all participants. Further, scan montaging software was not available at the time of analysis. In the pseudo-montaged images used, overlapping areas of adjacent images were obscured by greyish masks for display (*Figure 2*) and clinicians were invited to ignore those regions while assessing RNFB visibility. Whilst this method of displaying the images represents a compromise between maintaining positional context of the individual images and avoiding duplication of information and registration errors, the graders were able to overcome the drawbacks to produce highly consistent and repeatable data. As such we do not believe the results of this study would be significantly altered by improved montaging of the individual images.

Due to the moving ILM reference plane with RNFL thinning in glaucoma, further work is needed to assess the slab properties required to fully capture RNFB defects in glaucoma. For example, whilst 50 μm below the ILM may be well within the RNFL of a healthy retina close to the optic disc, it may be within the ganglion cell layer of a glaucomatous retina with significant RNFL thinning. Nevertheless, our data suggest that it is possible that the fixed-slab approaches taken in previous studies may not fully capture all defects in early glaucoma. The data presented here provide preliminary evidence on the range of depths that should be considered by studies using enface OCT imaging in glaucoma.

In conclusion, the range of depths with visible RNFBs varied markedly across the retina of the healthy participants in this study, consistent with the RNFL thickness. The optimal single depth for visualisation of RNFBs across healthy retinae examined was around 20 μm below the ILM. At this depth a cross section of the RNFB pattern was visible in all retinal regions and future work could explore whether these parameters would remain similar in a larger population or in different ethnic groups. To fully extract all RNFB information consistently across the retina, slab properties should account for changes in retinal location, and differences between individual eyes should be considered.

Acknowledgements

Conflict of interest

Author contribution

Riccardo Cheloni: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Resources (equal); Writing-original draft (equal); Writing-review & editing (equal). **Jonathan Denniss:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (lead); Investigation (lead); Methodology (lead); Project administration (lead); Resources (equal); Supervision (lead); Validation (lead); Visualization (equal); Writing-original draft (equal); Writing-review & editing (equal).

References

1. Tatham AJ, Medeiros FA, Zangwill LM & Weinreb RN. Strategies to improve early diagnosis in glaucoma. *Prog Brain Res* 2015; 221: 103–133.

2. Chen TC, Hoguet A, Junk AK *et al*. Spectral-domain OCT: helping the clinician diagnose glaucoma: a report by the American Academy of Ophthalmology. *Ophthalmology* 2018; 125: 1817–1827.

3. Fallon M, Valero O, Pazos M & Anton A. Diagnostic accuracy of imaging devices in glaucoma: a meta-analysis. *Surv Ophthalmol* 2017; 62: 446–461.

4. de Moraes CG, Liebmann JM, Medeiros FA & Weinreb RN. Management of advanced glaucoma: characterization and monitoring. *Surv Ophthalmol* 2016; 61: 597–615.

5. Kansal V, Armstrong JJ, Pintwala R & Hutnik C. Optical coherence tomography for glaucoma diagnosis: an evidence based meta-analysis. *PLoS One* 2018; 13: e0190621.

6. World Glaucoma Association. *10th Consensus Meeting: Diagnosis of Primary Open Angle Glaucoma*. R.N. Weinreb, D.F. Garway-Heath, C. Leung, F.A. Medeiros, Liebmann J, editors, 2016.

7. Vermeer KA, Schoot J, Lemij HG & Boer JF. OCT-derived retinal nerve fiber layer reflectivity maps for glaucoma assessment. *Invest Ophthalmol Vis Sci* 2011;52:3666.

8. Liu S, Wang B, Yin B *et al*. Retinal nerve fiber layer reflectance for early glaucoma diagnosis. *J Glaucoma* 2014; 23: e45–e52.

9. Huang XR, Knighton RW & Cavuoto LN. Microtubule contribution to the reflectance of the retinal nerve fiber layer. *Invest Ophthalmol Vis Sci* 2006; 47: 5363–5367.

10. Ashimatey BS, King BJ, Burns SA & Swanson WH. Evaluating glaucomatous abnormality in peripapillary optical coherence tomography enface visualisation of the retinal nerve fibre layer reflectance. *Ophthalmic Physiol Opt* 2018; 38: 376–388.

11. Hood DC, Fortune B, Mavrommatis MA *et al*. Details of glaucomatous damage are better seen on OCT en face images than on OCT retinal nerve fiber layer thickness maps. *Invest Ophthalmol Vis Sci* 2015; 56: 6208–6216.

12. Sakamoto M, Mori S, Ueda K *et al*. En face slab images visualize nerve fibers with residual visual sensitivity in significantly thinned macular areas of advanced glaucomatous eyes. *Invest Ophthalmol Vis Sci* 2019; 60: 2811–2821.

13. Huang XR, Knighton RW, Spector YZ & Feuer WJ. Cytoskeletal alteration and change of retinal nerve fiber layer birefringence in hypertensive retina. *Curr Eye Res* 2017; 42: 936–947.

14. Iikawa R, Togano T, Sakaue Y *et al*. Estimation of the central 10-degree visual field using en-face images obtained by optical coherence tomography. *PLoS One* 2020; 15: e0229867.

15. Huang XR, Zhou Y, Kong W & Knighton RW. Reflectance decreases before thickness changes in the retinal nerve fiber layer in glaucomatous retinas. *Invest Ophthalmol Vis Sci* 2011; 52: 6737–6742.

16. Malik R, Swanson WH & Garway-Heath DF. 'Structure-function relationship' in glaucoma: past thinking and current concepts. *Clin Exp Ophthalmol* 2012; 40: 369–380.

17. Shigueoka LS, Vasconcellos JPC, Schimiti RB *et al*. Automated algorithms combining structure and function outperform general ophthalmologists in diagnosing glaucoma. *PLoS One* 2018; 13: e0207784.

18. Raza AS, Zhang X, De Moraes CG *et al*. Improving glaucoma detection using spatially correspondent clusters of damage and by combining standard automated perimetry and optical coherence tomography. *Invest Ophthalmol Vis Sci* 2014; 55: 612–624.

19. Hood DC, Tsamis E, Bommakanti NK *et al*. Structure-function agreement is better than commonly thought in eyes with early glaucoma. *Invest Ophthalmol Vis Sci* 2019; 60: 4241–4248.

20. Denniss J, Turpin A & McKendrick AM. Relating optical coherence tomography to visual fields in glaucoma: structure-function mapping, limitations and future applications. *Clin Exp Optomet* 2019; 102: 291–299.

21. Denniss J, McKendrick AM & Turpin A. Towards patient-tailored perimetry: automated perimetry can be improved by seeding procedures with patient-specific structural information. *Transl Vis Sci Technol* 2013; 2: 3.

22. Montesano G, Rossetti LM, Allegrini D, Romano MR & Crabb DP. Improving visual field examination of the macula using structural information. *Transl Vis Sci Technol* 2018; 7: 36.

23. Ganeshrao SB, McKendrick AM, Denniss J & Turpin A. A perimetric test procedure that uses structural information. *Optom Vis Sci* 2015; 92: 70–82.

24. Ballae Ganeshrao S, Turpin A & McKendrick AM. Sampling the visual field based on individual retinal nerve fiber layer thickness profile. *Invest Ophthalmol Vis Sci* 2018; 59: 1066–1074.

25. Alluwimi MS, Swanson WH, Malinovsky VE & King BJ. A basis for customising perimetric locations within the macula in glaucoma. *Ophthalmic Physiol Opt* 2018; 38: 164–173.

26. Alluwimi MS, Swanson WH, Malinovsky VE & King BJ. Customizing perimetric locations based on en face images of retinal nerve fiber bundles with glaucomatous damage. *Transl Vis Sci Technol* 2018; 7: 5.

27. Varma R, Skaf M & Barron E. Retinal nerve fiber layer thickness in normal human eyes. *Ophthalmology* 1996; 103: 2114–2119.

28. Rudnicka AR, Mt-Isa S, Owen CG, Cook DG & Ashby D. Variations in primary open-angle glaucoma prevalence by age, gender, and race: a Bayesian meta-analysis. *Invest Ophthalmol Vis Sci* 2006; 47: 4254–4261.

29. Ctori I & Huntjens B. Repeatability of foveal measurements using spectralis optical coherence tomography segmentation software. *PLoS One* 2015; 10: e0129005.

30. Vermeer KA, Mo J, Weda JJ, Lemij HG & de Boer JF. Depth-resolved model-based reconstruction of attenuation coefficients in optical coherence tomography. *Biomed Optics Express* 2013; 5: 322–337.

31. Chang S & Bowden AK. Review of methods and applications of attenuation coefficient measurements with optical coherence tomography. *J Biomed Optics* 2019; 24: 1–17.

32. R Core Team. *R: A language and environment for statistical computing*. R version 3.6.3 (2020–02-29) ed: R Foundation for Statistical Computing, Vienna, Austria; 2020.

33. Bedggood P, Nguyen B, Lakkis G, Turpin A & McKendrick AM. Orientation of the temporal nerve fiber raphe in healthy and in glaucomatous eyes. *Invest Ophthalmol Vis Sci* 2017; 58: 4211–4217.

34. Denniss J, McKendrick AM & Turpin A. An anatomically customizable computational model relating the visual field to the optic nerve head in individual eyes. *Invest Ophthalmol Vis Sci* 2012; 53: 6981–6990.

35. Ashimatey BS, King BJ & Swanson WH. Retinal putative glial alterations: implication for glaucoma care. *Ophthalmic Physiol Opt* 2018; 38: 56–65.

36. Bates D, Mächler M, Bolker B & Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw* 2015; 67: 1–48.

37. Shrout PE & Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86: 420–428.

38. Revelle W. *psych: Procedures for Personality and Psychological Research*. Evanston, IL: Northwestern University, 2018.

39. Finn RH. A note on estimating the reliability of categorical data. *Educ Psycholog Measure* 1970; 30: 71–76.

40. Denniss J, Turpin A, Tanabe F, Matsumoto C & McKendrick AM. Structure-function mapping: variability and conviction in tracing retinal nerve fiber bundles and comparison to a computational model. *Invest Ophthalmol Vis Sci* 2014; 55: 728–736.

41. Gardiner SK, Demirel S, Reynaud J & Fortune B. Changes in retinal nerve fiber layer reflectance intensity as a predictor of functional progression in glaucoma. *Invest Ophthalmol Vis Sci* 2016; 57: 1221–1227.

42. Vermeer KA, van der Schoot J, Lemij HG & de Boer JF. RPE-normalized RNFL attenuation coefficient maps derived from volumetric OCT imaging for glaucoma assessment. *Invest Ophthalmol Vis Sci* 2012; 53: 6102–6108.

43. Hondur G, Goktas E, Al-Aswad L & Tezel G. Age-related changes in the peripheral retinal nerve fiber layer thickness. *Clin Ophthalmol* 2018; 12: 401–409.

44. Yang Q, Reisman CA, Wang Z *et al.* Automated layer segmentation of macular OCT images using dual-scale gradient information. *Opt Express* 2010; 18: 21293–21307.

45. Chen X, Hou P, Jin C *et al.* Quantitative analysis of retinal layer optical intensities on three-dimensional optical coherence tomography. *Invest Ophthalmol Vis Sci* 2013; 54: 6846–6851.

46. Tappeiner C, Barthelmes D, Abegg MH, Wolf S & Fleischhauer JC. Impact of optic media opacities and image compression on quantitative analysis of optical coherence tomography. *Invest Ophthalmol Vis Sci* 2008; 49: 1609–1614.

47. Chen B, Gao E, Chen H *et al.* Profile and determinants of retinal optical intensity in normal eyes with spectral domain optical coherence tomography. *PLoS One* 2016; 11: e0148183.

48. Mehta N, Lavinsky F, Gattoussi S *et al.* Increased inner retinal layer reflectivity in eyes with acute CRVO correlates with worse visual outcomes at 12 months. *Invest Ophthalmol Vis Sci* 2018; 59: 3503–3510.

49. Thepass G, Lemij HG & Vermeer KA. Attenuation coefficients from SD-OCT data: structural information beyond morphology on RNFL integrity in glaucoma. *J Glaucoma* 2017; 26: 1001–1009.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Supplementary Material 1. Animation showing an example of the grading task, including the superimposed grid as per *Figure 2*. The animation cycles through one of the presentations provided to the clinicians performing the grading task and shows how visibility of retinal nerve fibre bundles (RNFBs) changes with depth from the inner limiting membrane (ILM).

Supplementary Material 2. Written instruction provided to each clinician before the grading task.

Article

A Simple Subjective Evaluation of Enface OCT Reflectance Images Distinguishes Glaucoma From Healthy Eyes

Riccardo Cheloni[1], Simon D. Dewsbery[2], and Jonathan Denniss[1]

[1] School of Optometry and Vision Science, University of Bradford, UK
[2] Ophthalmology Department, Leeds Teaching Hospitals NHS Trust, Leeds, UK

Purpose: We present a subjective approach to detecting glaucomatous defects in enface images and assess its diagnostic performance. We also test the hypothesis that if reflectivity changes precede thickness changes in glaucoma there should be reduced correlation between the modalities in glaucoma compared to controls.

Methods: Twenty glaucoma participants and 20 age-matched controls underwent high-resolution OCT scans of one eye. 4 μm-thick enface slabs were constructed through the retina. Enface indices were depths of *first gap* in visible retinal nerve fiber bundles (RNFBs) and *last visible bundle,* subjectively evaluated in six sectors of a 3.5 mm circle around the optic disc. Retinal nerve fiber layer thickness (RNFLT) along the same circle was extracted at angles corresponding to enface indices. Between-group differences were tested by linear mixed models. Diagnostic performance was measured by partial receiver operating characteristic area (pAUC).

Results: *First gap* and *last visible bundle* were closer to the inner limiting membrane in glaucoma eyes (both $P < 0.0001$). Enface indices showed excellent diagnostic performance (pAUCs 0.63–1.00), similar to RNFLT (pAUCs 0.63–0.95). Correlation between enface and RNFLT parameters was strong in healthy ($r = 0.81$–0.92) and glaucoma eyes ($r = 0.73$–0.80).

Conclusions: This simple subjective method reliably identifies glaucomatous defects in enface images with diagnostic performance at least as good as existing thickness indices. Thickness and reflectivity were similarly related in healthy and glaucoma eyes, providing no strong evidence of reflectivity loss preceding thinning. Objective analyses may realize further potential of enface OCT images in glaucoma.

Translational Relevance: Novel enface OCT indices may aid glaucoma diagnosis.

Introduction

Early diagnosis of glaucoma is desirable to minimize visual impairment,[1,2] but the burden of lifelong treatment demands accurate diagnosis with minimization of false positives.[3,4] Optical coherence tomography (OCT) has become a mainstay of glaucoma assessment,[5,6] however, diagnosis from single examinations remains challenging.[7–9] Among several factors, subtle early changes to ocular structures with a diverse anatomy among healthy eyes[10,11] and suboptimal usage of collected information[12,13] contribute to the imperfect diagnostic capability of OCT.

Conventionally, OCT is employed in glaucoma clinics to evaluate retinal nerve fiber layer (RNFL) thickness. An additional source of structural information is retinal nerve fiber bundle (RNFB) reflectivity, which varies according to integrity and density.[14,15] Hyper-reflectivity of RNFBs is due to their highly ordered structure, and loss of reflectivity occurs when RNFB axon cytoskeleton is disrupted.[16,17] Changes in RNFL reflectance have been assessed in glaucoma clinics long before the introduction of OCT, either by ophthalmoscopy or fundus photography.[18,19] Evidence from animal models suggests that loss of reflectivity may precede measurable reduction of RNFL thickness,[20,21] but this has not been consistently replicated in humans.[14] Recent developments in OCT

| Healthy control | Glaucoma | Glaucoma |
| (73y) | (65y, MD: -2.1dB) | (73y, MD: -5.7dB) |
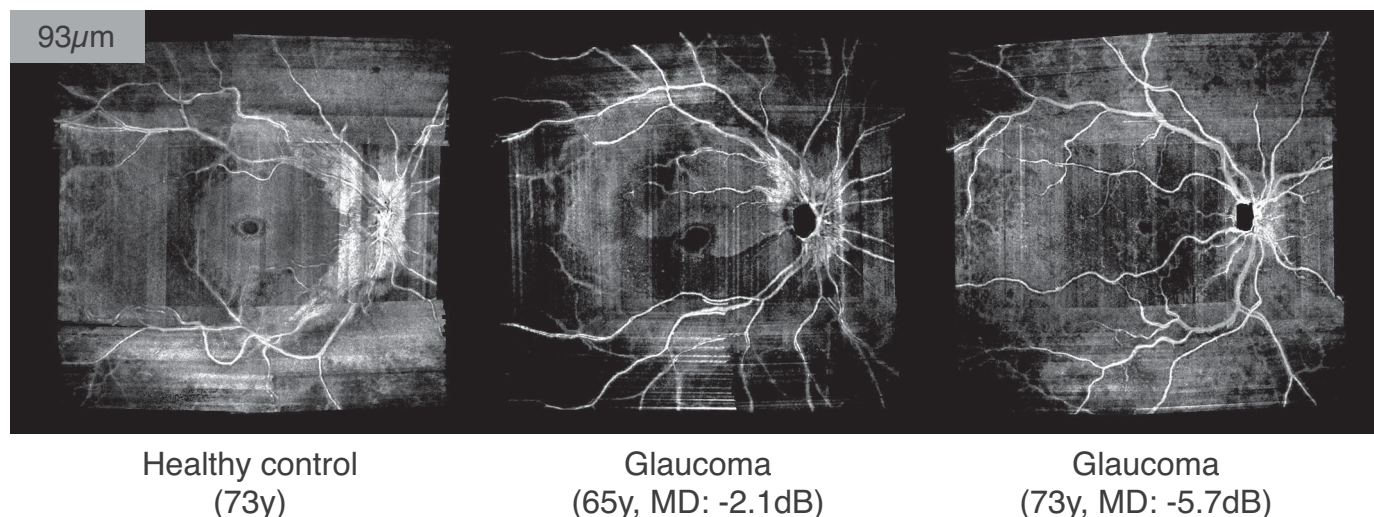
Figure 1. Example of how visible presence of RNFBs (93 µm below ILM) changes in a healthy eye and at different stages of glaucoma. The images are single pixel deep enface OCT images without depth-averaging. At this depth, RNFBs are still visible all around the optic disc in the healthy eye, whereas RNFBs have already disappeared in the rest of the retina where the slab encompasses deeper and hyporeflective retinal layers. In the early glaucoma eye (central panel) a substantial loss of RNFBs can be seen in the temporal and temporal inferior sectors, with no visible presence of RNFBs. In the more advanced glaucoma eye (right panel) no bundles are visible around the optic disc or elsewhere, with the only hyperreflective elements provided by blood vessels. An animated version of this figure, showing a range of depths below the ILM, is provided in Supplementary Figure S1. MD = mean deviation.

technology allow us to generate enface images, enabling visualization and quantification of RNFB reflectance. Following dense volumetric scans of the area of interest, transverse enface slabs can be obtained by averaging the intensity of each A-scan over a certain depth below the inner limiting membrane (ILM), producing a two-dimensional image.[22–26] As such, enface images allow direct observation of RNFBs, that in healthy eyes appear hyper-reflective due to the ordered structure of ganglion cell axon cytoskeletons.[16] According to the above models, clinicians could exploit changes in RNFB reflectance as an additional marker of glaucomatous damage.[23]

Further limitations on diagnostic use of OCT for early glaucoma may arise from data analysis. It has been suggested that the current focus on RNFL thickness indices and red/green classification may limit diagnosis as it does not make full use of available information.[13,27,28] Accordingly, clinicians are recommended to look in greater detail at B-scans for evidence of glaucoma damage missed by RNFL thickness analysis.[12,13,23] Enface images may be one way to observe glaucomatous lesions missed by the conventional RNFL thickness approach.

Though the analysis of enface images is promising, the lack of established methodology currently limits clinical value. No accepted objective criteria to define defects in this domain are available and proposed subjective analyses[23,24] have not been validated, nor have their diagnostic performance been evaluated.

The purpose of this study was to evaluate a simple approach for subjective identification of RNFB reflectance loss in glaucoma (Fig. 1 and Supplementary Fig. S1), and to quantify the diagnostic performance of this approach. Further, since the hypothesis that reflectance loss occurs before thickness changes has been minimally investigated in humans, we additionally aimed to test for discordance between RNFB reflectivity and RNFL thickness changes that may indicate a temporal decoupling between these parameters that could be exploited for glaucoma diagnosis.

Methods

The study adhered to the tenets of the Declaration of Helsinki and received ethical approval from the National Health Service's Research Ethics Service. All participants gave written informed consent and were free to withdraw at any time.

Participants

Participants were recruited through advertisements placed in the university eye clinic, local eye hospitals,

local newspapers and through charities and local interest groups. Recruited participants underwent a detailed eye examination including refraction, Goldmann applanation tonometry, slit lamp examination, spectral domain OCT (Spectralis, Heidelberg Engineering, Heidelberg, Germany) and visual field testing (24-2 SITA-Standard, Humphrey Field Analyzer 3, Carl Zeiss Meditec Inc., Dublin, California). Glaucoma participants were included if older than 40 years and with a confirmed clinical diagnosis of open-angle glaucoma with evidence of both structural and functional (visual field) defects. Structural defects were defined as at least one abnormal sector ($P < 1\%$) from the 3.5 mm diameter circumpapillary RNFL (cpRNFL) thickness OCT scan. Visual field defects were defined as at least three contiguous non-edge points with $P < 5\%$ on the Pattern Deviation plot. Glaucoma participants had no other disease except glaucoma that could affect vision. Age-similar healthy controls were included if presenting with no eye conditions, including ocular hypertension or different intraocular pressure between eyes (>4 mmHg). Healthy participants required normal visual fields (Mean Deviation $P > 5\%$, Glaucoma Hemifield Test within normal limits and no visual field defect as defined for the glaucoma group), but no specific OCT criteria were applied. All participants had best corrected visual acuity ≤ 0.20 logMAR (6/9.5 Snellen), refractive error between ± 6.00DS and less than 3.00DC and clear optical media with or without history of uncomplicated cataract surgery in the included eye.

One eye per participant was included. If both eyes were eligible, the included eye was selected at random in controls, whereas the eye with milder defect (as identified by a less negative Mean Deviation) was included among glaucoma participants.

OCT Imaging & Processing of En Face Images

Details of the OCT imaging procedure have been described previously.[26] This consisted of seven high-density, high-speed OCT scans (9.65 B-scans per degree), collected in different retinal locations. Overall, the central $\pm 25°$ of the retina was covered and all images were acquired with signal-to-noise ratio above 20 dB as per the manufacturer's instructions. Single slab images of the maximum digital axial resolution (3.87 µm), containing depth-resolved attenuation coefficients calculated according to equation 17 from Vermeer et al.,[29] were extracted from 0 to 193.5 µm below the ILM using custom software written in *R* (version 3.6.3).[30] Attenuation coefficients represent an intrinsic optical property of the retinal tissue[31] and

their use has been proposed to reduce the impact of artifacts on enface images.[22] Conventional cpRNFL thickness was also measured at the 3.5 mm diameter circle around the optic nerve head (ONH) as segmented by the device's built-in software.

Attenuation coefficient images were imported into MATLAB (Version 9.6.0, MathWorks Inc., Natick, Massachusetts) for montaging and further processing. Custom software was used to produce a montage of single images using the macular scan as a reference image and choosing the highest intensity pixel in regions of overlap between scans. For each participant, obtained montages underwent further image processing to optimize visualization, the details of which have been published.[26] Briefly, the intensity of an area within the raphe region, 35 µm below the ILM, with no RNFBs was set as background by subtracting this lower limit from all pixels and clipping negative values to zero. Then, the average 99th percentile from all depths was used to normalize the attenuation coefficients arrays to a 0 to 1 range. Figure 1 shows examples of final montages.

Data Extraction

Images of individual eyes were arranged in presentation files allowing observation of sequential slabs with perfect spatial alignment. Visible presence of RNFBs was evaluated around the ONH in six sectors corresponding to those of the Spectralis 3.5 mm diameter circle scan. The sectors adopted[32] were the temporal (90°), nasal (110°), two superior and two inferior sectors, split into temporal and nasal (40° per each of four sectors). A sector grid (Fig. 2) with fixed and standardized dimension was overlaid on the enface images of all participants at each depth. Aiming to adopt the same 3.5 mm circle of the cpRNFL thickness analysis, the corresponding scanning laser ophthalmoscope (SLO) image of Spectralis analysis from one participant was used to drive the construction of the grid by overlapping corresponding retinal structures. The resulting grid was subsequently verified on a second participant. Grid dimensions were then preserved unchanged and applied to all assessed eyes except that the grid was tilted to follow the individual fovea-disc angle as subjectively identified in the enface image. Accordingly, the temporal sector of each individual eye was centered on the fovea-disc axis, mimicking the arrangement used in the Spectralis cpRNFL analysis, facilitating comparison (Fig. 2).

Visible presence of RNFBs was recorded subjectively by one of the authors (RC), viewing images on a MacBook Pro 13" computer (2017 version, Apple Inc., Cupertino, California) under standardized light-

Glaucoma

Control
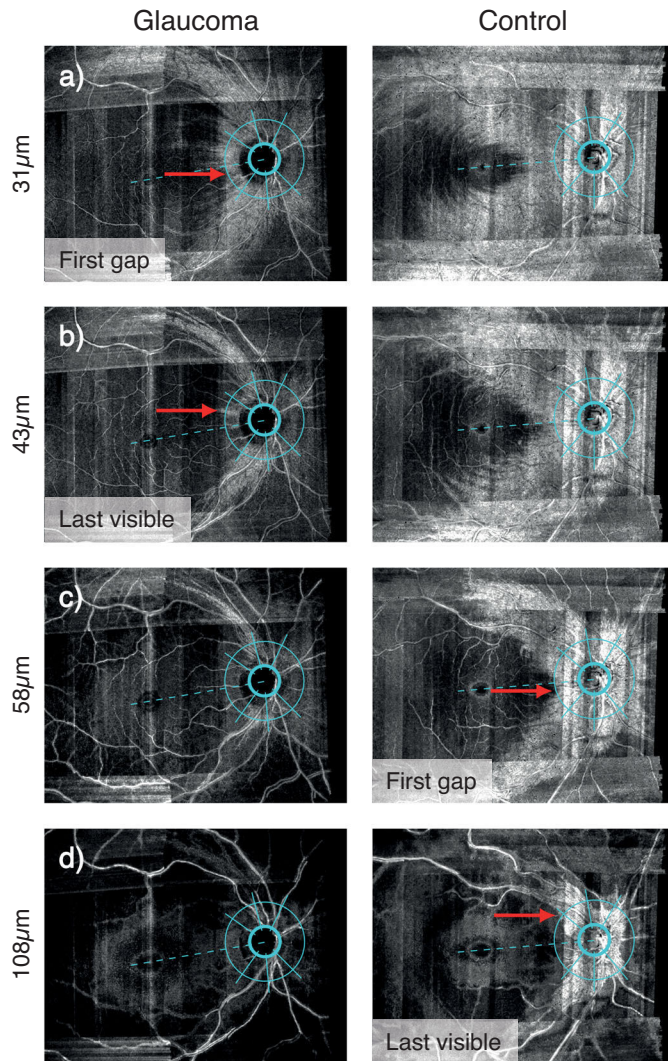
Figure 2. Example of the task in the temporal sector for a glaucoma eye (left panels) and an age-similar healthy participant (right panels). In (**a**) the red arrow shows the first gap for the glaucoma eye at 31 μm below the ILM, whereas the corresponding depth for the healthy eye is reached at 58 μm below the ILM (red arrow in **c**). The depth of last visible bundle (Last visible) was 43 μm (red arrow in **b**) and 108 μm (red arrow in **d**) below the ILM for the glaucoma and healthy eye, respectively.

ing. Both the depth and corresponding angle (with 0° at the fovea-disc axis and angles increasing clockwise for right eyes and anticlockwise for left eyes) of two enface indices were extracted at each ONH sector (Fig. 2). **First gap in visible bundles** (Figs. 2a, c), subsequently referred to as *first gap*, refers to the first (most anterior) depth at which a gap between RNFBs can be seen crossing the 3.5 mm circle in the sector of interest. **Last visible bundle** (Figs. 2b, d) represents the most posterior depth at which one or more visible bundles crosses the 3.5 mm circle in the sector of interest.

To reduce measurement bias, the grader was blinded to the depth of each image during the grading

task. It was not possible to mask the grader to the disease status since typical glaucomatous arcuate defects originating from the ONH were readily observable while viewing the enface images. However, to minimize effects of preconception, the grading task was performed first in eyes with glaucoma, and subsequently in healthy controls. Hyper-reflectivity from blood vessels was ignored in performing the judgment.

In addition to the enface indices, RNFL thickness along the same 3.5 mm circle was extracted at the angles of the enface parameters in each ONH sector (Fig. 3). Conventional cpRNFL parameters including mean sector thickness and global thickness were also extracted. These data served to establish a comparison between the enface parameters (first gap and last visible bundle) and the conventional thickness measures in our sample. Lastly, the average (mean) first gap and last visible bundle were computed for each eye, weighting for the width of each ONH sector by multiplying the indices by the width of the corresponding sector, summing and then dividing by 360°. Weighted average RNFL thickness at first gap and last visible bundle angles were computed the same way.

Statistical Analysis

All analyses were conducted in the open-source environment *R* (version 3.6.3).[30] Linear mixed models[33] and likelihood ratio tests were used to evaluate the overall effect of glaucoma on first gap and last visible bundle, while accounting for repeated-measures from six ONH sectors within each eye. We tested whether the depth of these enface indices was affected by glaucoma (fixed effect), accounting for the individual eye and the ONH sector as random effects. This analysis was limited to data from individual sectors. Models took the form:

$$y \sim 1 + Disease\ Status + (1|eye) + (1|ONH\ sector) + \epsilon \quad (1)$$

where y signifies the measure of interest (first gap or last visible bundle), 1 signifies the intercept and ε signifies random error. A model of the same form was applied to counterpart RNFL thickness data. Post-hoc independent *t*-tests were used to evaluate between-group differences in individual sectors, adjusting for multiple comparisons by Bonferroni correction. Diagnostic capability was quantified with standardized partial receiver operating characteristic area (pAUC).[34] To focus on the highest levels of specificity,[3,4] pAUCs with bootstrap 95% confidence intervals (CI) were calculated at specificity between 90% and 100%, with the trapezoid method.[35] Comparisons between pAUCs of enface indices and corresponding RNFL thickness parameters were made with the DeLong method.[36] An
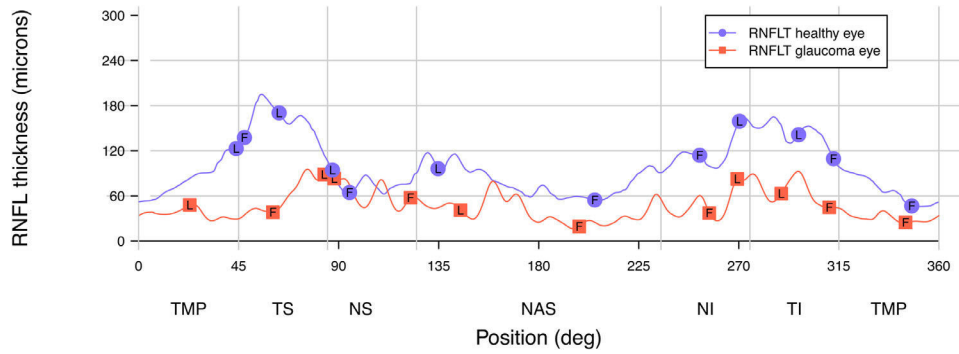
Figure 3. Circumpapillary RNFL thickness (RNFLT) profiles for the same control and glaucoma eye as shown in Figure 2. Points marked by F and L represent the RNFL thickness at the angles corresponding to the enface measures of first gap and last visible bundle, respectively. For both eyes and in every ONH sector, RNFL thicknesses at angle of first gap (F) were smaller than thicknesses at angle of last visible bundle (L), hence appearing lower on the y-axis. TMP = temporal, TS = temporal superior, NS = nasal superior, NAS = nasal, NI = nasal inferior, TI = temporal inferior.

overall measure of the strength of correlation between depth of visible presence of RNFBs and RNFL thickness was estimated by repeated measures correlation, using the *RMcorr R* package,[37] according to Bland & Altman.[38,39] This method allowed us to account for non-independence of data, and provides a measure of strength of common association among individuals (*r*), interpretable as a Pearson correlation coefficient. For consistency, Pearson correlation was also used to assess the strength of individual enface-thickness relationships within each ONH sector. Correlation analyses were further explored according to disease status, testing the hypothesis that if reflectivity loss precedes thinning, correlation would be poorer in glaucoma compared to healthy eyes.

A power calculation suggested that two groups as small as $n = 5$ per group would provide 90% power at $\alpha = 0.05$ to detect between-group differences of the magnitude found in recent data on global cpRNFL thickness in healthy and glaucoma eyes.[40] Data from the overall group ($n = 40$) would provide 90% power ($\alpha = 0.05$) for identification of correlation of at least

$r = 0.48$.[41] On the other hand, when grouping for disease status ($n = 20$), a correlation of at least $r = 0.65$ could be identified, at the same power and alpha.[41]

Results

We included 20 glaucoma participants and 20 age-similar healthy controls, whose demographics are given in Table 1. Seven recruited controls and 12 recruited glaucoma participants were excluded for not meeting inclusion criteria (e.g., pathology, VF defect; four controls, eight glaucoma), for ptosis affecting imaging and perimetry (two controls), unable to obtain reliable VF (one control, three glaucoma) or unwillingness to undergo extended OCT scans (one glaucoma). On average, glaucoma participants had an early-moderate visual field defect, with the majority (17/20) showing a Mean Deviation (MD) better than or equal to -6 dB. The remaining three participants presented MD of -6.13 dB, -8.62 dB, and -14.9 dB, respectively.

Table 1. Demographics of Included Participants. Continuous Data are Summarized as Mean and (Standard Deviation). *P*-Values Are Calculated Using *t*-Tests for Continuous Data or Proportion Tests for Proportions

	Control	Glaucoma	P
N	20	20	–
Male/female	8/12	9/11	1
Ethnicity, caucasian/others	19/1	20/0	1
Age (years)	68.6 (5.0)	69.3 (5.1)	0.66
Mean spherical equivalent (D)	+0.7 (2.0)	+0.0 (1.4)	0.21
SAP Mean Deviation (dB)	0.6 (1.1)	−4.5 (3.1)	<0.0001
Average cpRNFL thickness (μm)	95.1 (9.3)	66.3 (9.4)	<0.0001
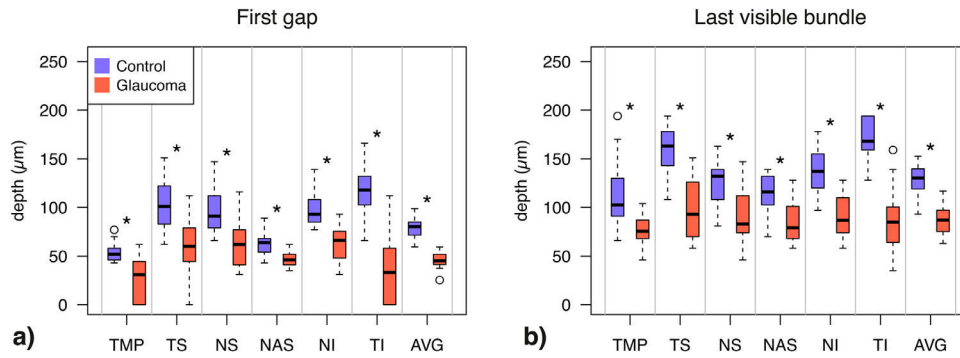
Figure 4. Boxplots showing differences between glaucoma and control eyes for the first gap (**a**) and last visible bundle (**b**) for every ONH sector and the sectors-average. At each ONH sector, control and glaucoma data are reported by the left-most and right-most box, respectively, and color-coded accordingly. After Bonferroni correction (14 comparisons), pairwise differences were considered significant when $P < 0.0036$, and flagged with (*). Boxes report medians and 25th to 75th percentiles. Whiskers represent maximum and minimum values of data within $1.5\times$ interquartile range above or below the limits of the box. Unfilled symbols represent outliers. ONH sectors acronyms as per Figure 3; AVG = average.

Table 2. Diagnostic Performance (Standardized pAUC at Specificity 90–100% with 95% CIs) of Enface Indices, RNFL Thickness (RNFLT) at Corresponding Angles and Conventional cpRNFL Thickness Measurements. ONH Sector Labels as per Figure 3.

| ONH Sector	Enface First Gap	RNFLT at First Gap Angle	Enface Last Visible Bundle	RNFLT at Last Visible Angle	cpRNFL Thickness
TMP	0.86 (0.75, 0.95)	0.82 (0.71, 0.95)	0.63 (0.53, 0.87)	0.67 (0.53, 0.92)	0.70 (0.59, 0.95)
TS	0.79 (0.67, 0.92)	0.79 (0.68, 0.92)	0.80 (0.68, 0.95)	0.76 (0.66, 0.90)	0.74 (0.63, 0.92)
NS	0.82 (0.71, 0.92)	0.67 (0.58, 0.82)	0.76 (0.65, 0.90)	0.65 (0.53, 0.79)	0.72 (0.61, 0.84)
NAS	0.67 (0.55, 0.91)	0.63 (0.53, 0.79)	0.71 (0.57, 0.90)	0.70 (0.59, 0.83)	0.70 (0.55, 0.87)
NI	0.91 (0.83, 0.99)	0.84 (0.71, 0.97)	0.83 (0.71, 0.95)	0.76 (0.66, 0.92)	0.92 (0.82, 1)
TI	0.94 (0.84, 1)	0.88 (0.76, 1)	0.95 (0.87, 1)	0.92 (0.84, 1)	0.95 (0.87, 1)
AVG	1 (1, 1)	0.95 (0.87, 1)	0.90 (0.74, 1)	0.83 (0.72, 0.97)	0.95 (0.87, 1)

Overall, disease status had a significant effect on both enface RNFB indices ($\chi^2_{(1)} = 63.3$ & 51.6, both $P < 0.0001$). Both first gap and last visible bundle were closer to the ILM in glaucomatous eyes (mean difference: 39.1 μm, 95% CI: 33.0 to 45.3 and 48.1 μm, 95% CI: 38.8 to 57.4 μm, respectively, both $P < 0.0001$). Pairwise differences (Fig. 4) showed both enface indices to be smaller in eyes with glaucoma compared to healthy controls across all ONH sectors (all $P < 0.0036$). The greatest separation was found in the temporal inferior sector for both first gap (difference: 80.8 μm, 95% CI: 62.3 to 98.7 μm; $t_{35.2} = 9.1$, $P < 0.0001$) and last visible bundle (82.9 μm, 95% CI: 66.0 to 99.9 μm; $t_{34.2} = 9.9$, $P < 0.0001$). As expected from existing knowledge, linear mixed models showed that RNFL thickness at angles corresponding

to enface first gaps and last visible bundles were also significantly smaller in glaucoma ($\chi^2_{(1)} = 59.6$ & 37.4, both $P < 0.0001$). Pairwise differences among RNFL thickness parameters are shown in Supplementary Figure S2.

Diagnostic performance of enface RNFB indices and RNFL thickness parameters at corresponding angles is reported in Table 2. The performance of conventional cpRNFL thickness analysis in this sample is also reported for comparison.

Several enface indices showed excellent diagnostic capability (pAUCs > 0.9). The enface first gap indices with best diagnostic performance (inferior temporal and sectors-average) performed slightly better than RNFL thickness counterparts, but they were statistically similar ($P = 0.18$ and $P = 0.16$). Similarly,
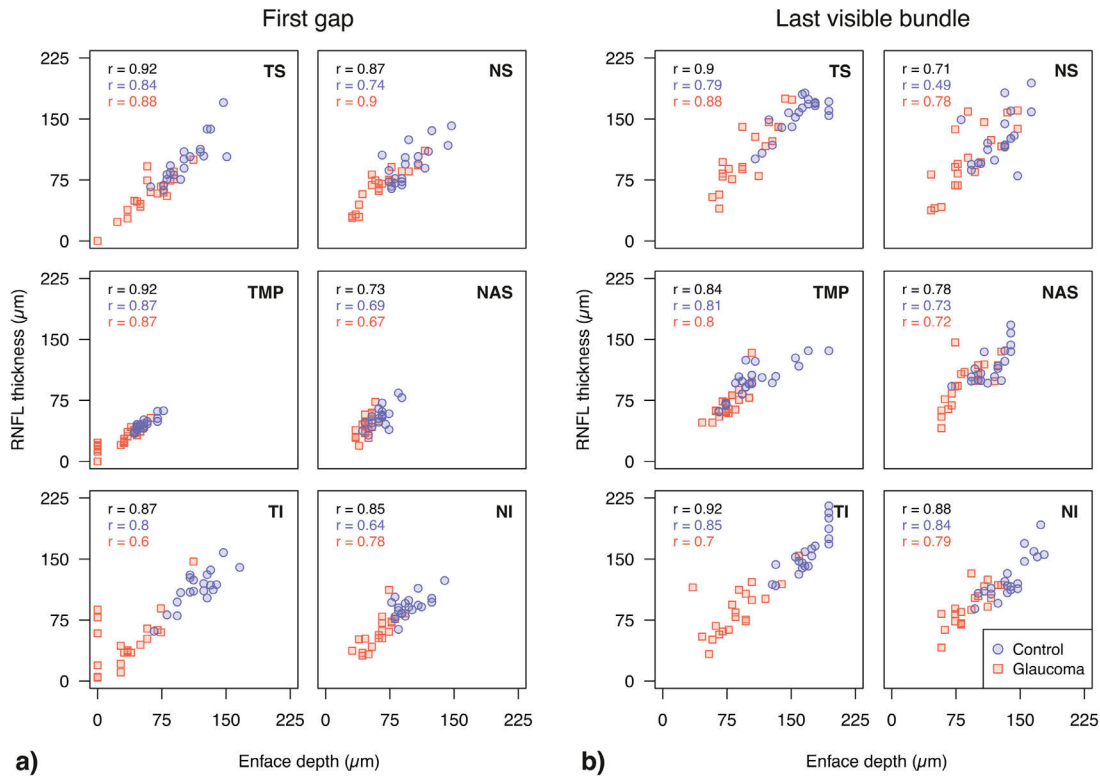
Figure 5. Relationships between (**a**) first gap & (**b**) last visible bundle with RNFL thickness at the corresponding angle in each sector of the ONH. Points are color coded and shaped according to disease status. The Pearson correlation coefficients are computed with data from both glaucoma and healthy eyes combined (black), as well as grouping data according to disease status (color coded accordingly). In the combined group, all correlation coefficients were $P < 0.0001$. All correlation coefficients from data grouped according to disease status were $P < 0.001$, with the exception of last visible bundle at NS and NAS in controls ($P = 0.02$ and 0.002, respectively) and NAS and NI first gap in glaucoma eyes ($P = 0.012$ and 0.005, respectively). ONH sectors are labeled as in Figure 3.

best performing enface last visible bundle indices (inferior temporal and sectors-average) outperformed corresponding RNFL thickness parameters, but differences were not statistically significant ($P = 0.33$ and $P = 0.30$).

Diagnostic accuracy analysis was repeated in a subgroup of early glaucoma participants with MD better than or equal to -4.0 dB ($n = 11$, Supplementary Table S3). Limiting the analysis to patients with glaucoma at earlier stages aimed to remove more advanced cases that are easier to diagnose.[42] Nonetheless, pAUCs in the earlier glaucoma group were similar to the ones identified in the overall sample, suggesting no loss of diagnostic performance. Indeed, the best enface and RNFL thickness parameters (temporal inferior and sectors-average) were similar between the overall group and the early glaucoma subgroup (all $P > 0.05$).

Repeated measures correlation analysis for all eyes showed a strong relationship between first gap and RNFL thickness at the same angle ($r_{df = 199} = 0.87$,

95% CI: 0.83 to 0.90, $P < 0.0001$). Last visible bundle was also strongly correlated with corresponding RNFL thickness for all eyes ($r_{df = 199} = 0.78$, 95% CI: 0.72 to 0.83, $P < 0.0001$). The relationship between enface indices and corresponding RNFL thickness in each ONH sector is shown in Figure 5. For first gap, the strongest correlation was found in the temporal sector ($r = 0.92$, 95% CI: 0.86 to 0.96, $P < 0.0001$), whereas the temporal inferior sector showed the strongest correlation for last visible bundle ($r = 0.92$, 95% CI: 0.85 to 0.96, $P < 0.0001$).

To test the hypothesis that loss of reflectivity might precede thinning of the RNFL, we analyzed the strength of correlation between enface indices and RNFL thickness when grouping data according to disease status. Across all sectors, repeated measures correlation for first gap was stronger in healthy eyes compared to glaucoma eyes ($r_{df = 99} = 0.92$, 95% CI: 0.89 to 0.95, and $r_{df = 99} = 0.80$, 95% CI: 0.73 to 0.87, $P < 0.0001$, respectively). Overall correlation between last visible bundle and corresponding RNFL thick-
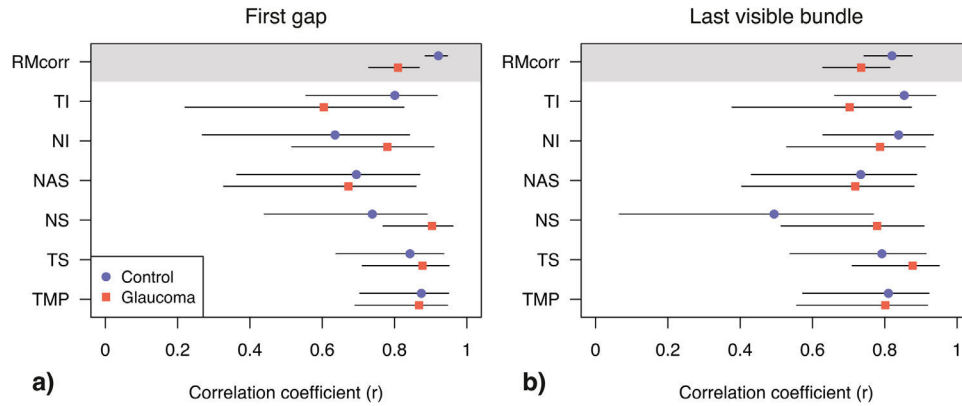
Figure 6. Pearson's correlation coefficients and their 95% CI for first gap (**a**) & last visible bundle (**b**) and corresponding RNFL thickness at each ONH sector. Top panel in each plot reports the overall correlation and its 95% CI limits, computed with repeated measure correlation (RMCorr). ONH sectors are labeled as in Figure 3.

ness was also higher in healthy eyes ($r_{df = 99} = 0.81$, 95% CI: 0.74 to 0.88, $P < 0.0001$) than glaucoma eyes ($r_{df = 99} = 0.73$, 95% CI: 0.62 to 0.81, $P < 0.0001$), though 95% confidence intervals overlapped. Figure 6 shows sector-wise differences in strength of correlation between glaucoma and healthy eyes. Coefficients were similar in many ONH sectors for both enface parameters, sometimes greater in healthy eyes (e.g., TI, first gap) and vice-versa in other ONH sectors (e.g., NS & NI first gap).

Discussion

Exploiting RNFL reflectance for early glaucoma diagnosis has attracted significant research interest,[15,31,43–46] and enface imaging now provides clinicians with a potentially powerful tool to this end.[22,24,47–49] Enface images may facilitate a more detailed clinical approach to OCT in glaucoma than the sole consideration of thickness measurements,[13,23,27,28] and some reports suggest that the technique might show RNFL reflectance changes before measurable thinning.[20,21] Yet, clinical usability of enface imaging remains limited, with most objective and subjective methods for the assessment of reflectivity currently confined to research settings.[15,22,24,31,43,46,50,51] In this study we present a simple and clinically usable method for the evaluation of glaucomatous changes in enface images, which focuses on the subjective assessment of visible presence of RNFBs around the ONH.

Enface parameters were able to identify glaucomatous changes in our sample. Both first gap and last visible bundle were significantly closer to the ILM in glaucoma, with temporal inferior sector and

sectors-average measures showing greatest differences (Fig. 4). This is not surprising since the hallmark RNFL thinning in glaucoma is most easily detected in some ONH sectors including the temporal inferior sector.[11,42,52–54] In case of thinner RNFL, deeper and hyporeflective retinal layers such as the ganglion cell layer and the inner plexiform layer would be encountered at depths closer to the ILM. Further, some preserved RNFBs in glaucoma eyes might show reduced reflectivity, hence mimicking lack of bundles and contributing to smaller enface depths in our study.[23,26] Previous studies focused on quantitative assessment of RNFL reflectance and also found a significant effect of glaucoma, in agreement with our findings. Lower reflectivity of the RNFL was shown with time-domain OCT,[43] and later replicated with more recent technology.[15,31,50] Irrespective of the analysis performed, reflectivity of the RNFL was reduced in eyes with glaucoma compared to healthy eyes, and increasingly so with more severe disease.[15,31,43,50]

Diagnostic performance of enface indices was excellent in many ONH sectors (Table 2), yet statistically similar to corresponding RNFL thickness parameters. The high accuracy of conventional cpRNFL thickness suggests that glaucomatous defects in this sample were already well captured by the typical morphological OCT analysis. This is unsurprising given that our inclusion criteria required a structural defect. Nonetheless, it is notable that enface indices performed similarly to or better than conventional thickness measurements (pAUC higher in 12 of 14 comparisons, though all differences $P > 0.05$). The performance of both enface and conventional indices might not necessarily be representative of clinical settings aiming to diagnose the earliest glaucoma

cases, in which conventional OCT analyses perform more poorly.[55–58] Despite this limitation, some of the enface parameters presented hold promise for early glaucoma detection. For instance, sectors-average first gap discriminated glaucoma perfectly in this sample, warranting further exploration of the use of this parameter in glaucoma diagnosis. We also explored diagnostic accuracy in a "more difficult" subgroup of early glaucoma eyes, with MD better than −4 dB. Diagnostic performance was similar between the early subgroup and the overall sample for all indices (Supplementary Table S3), though it should be noted that these eyes still had a structural defect measurable by conventional cpRNFL thickness.

Among studies employing RNFL reflectivity for glaucoma diagnosis, none have adopted similar approaches to those proposed here. Some reports conducted quantitative analysis of RNFL reflectance for the discrimination of glaucoma with mixed results.[31,44,46] Liu et al. compared cpRNFL thickness and a pigment epithelium normalized reflectance index of the cpRNFL for glaucoma diagnosis.[44] Conventional thickness showed similar accuracy to reflectance indices to detect definite glaucoma, whereas some superiority of reflectivity analysis (0.05 differences in AUCs) was found for detecting glaucoma suspects.[44] A similar normalized reflectance index was employed later by a study aiming to detect glaucoma progression.[45] In that study, reflectance analysis did not outperform cpRNFL thickness in predicting functional progression, but, for a fixed amount of thinning, loss of reflectivity related to more rapid visual field degradation. Recently, Tan and colleagues further refined normalized reflectance indices and tested the related diagnostic capability for glaucoma compared to cpRNFL thickness.[46] In that study, reflectance analysis outperformed thickness in terms of sensitivity at 99% specificity in both glaucoma groups. Lastly, in a case control study by Thepass and colleagues,[31] global cpRNFL thickness outperformed measures of RNFL reflectance from the same OCT scan, both in terms of AUC (0.97 vs. 0.83) and sensitivity at 90% specificity (97% vs. 60%). Our results, therefore, align with the current evidence, which overall suggests that reflectance analysis performs well for glaucoma detection, though not likely to be substantially superior to thickness analysis. Further studies are needed to test whether reflectance information can be combined with thickness measurements to further improve OCT diagnostic accuracy. It is worth noting, though, that there remains scope for improvement in the observation and quantification of defects in enface images, and, given the strong performance of simple approaches such as ours, such improvements may yield greater diagnostic performance.

Although published data were generated with dissimilar approaches to the one used here, previous studies have shown strong correlations between RNFL thickness and reflectance.[22,31,43,46] Our analysis also showed strong relationships between enface parameters and corresponding RNFL thickness (Fig. 5), in concordance with the literature. The reflectance-thickness relationship we found was strong but imperfect, and several reasons for incongruences should be considered. For instance, blood vessels could be expected to have a larger impact on thickness measurement, since these could be distinguished subjectively from RNFBs in enface analysis. An estimate of such effect could be inferred from the slightly poorer correlation in last visible bundle than first gap parameters (≈0.1). This may be attributable to the presence of blood vessels as major blood vessels are usually located in regions with thicker RNFL,[59,60] where RNFBs are also expected to be visible at greater depths. First gap more often coincided with regions of thinner RNFL which are also more likely to be areas free from major blood vessels. Although methods for removal of blood vessels from OCT scans exist,[59,61,62] they are not routinely adopted in clinics, and their usage here would likely result in further improvement of an already strong correlation. Additionally, segmentation inaccuracies of the proximal RNFL boundary could also play a role. This surface is difficult to segment,[63] especially in areas with established damage.[15] Since enface images only depend on the more straightforward vitreous-ILM surface segmentation these inaccuracies only affected thickness measures. The reduced dependence of enface approaches on device software's segmentation and analysis might represent additional advantages of this technique compared to thickness analysis, with the further potential of stronger interdevice comparability of results.

To explore the hypothesis that reflectivity loss precedes thinning of the RNFL, the strength of correlation was evaluated in healthy and glaucoma eyes separately. A weaker overall correlation was found in glaucoma between first gap and corresponding RNFL thickness (Fig. 6a). This was also the case for last visible bundle, though that difference was not statistically significant (Fig. 6b). A more detailed look at individual first gap sector data suggests that correlation was similar between healthy and glaucoma eyes in every sector but the temporal inferior. This was confirmed by similar repeated-measure correlation between the two groups found once censoring the temporal inferior sector from first gap data (glaucoma:

$r_{\mathrm{df}=79} = 0.89$, 95% CI: 0.84 to 0.93, $P < 0.0001$; and controls: $r_{\mathrm{df}=79} = 0.91$, 95% CI: 0.87 to 0.94, $P < 0.0001$). The scatterplot corresponding to the aforementioned relationship (Fig. 5a, bottom left subplot) shows the presence of three outliers in an otherwise strong relationship. Enface images of these participants showed first gap in the temporal inferior sector at 0 μm below the ILM, meaning that in part of this sector there were no visible bundles, albeit still measurable RNFL thickness at corresponding angles. However, all these eyes did present significant thinning of the temporal inferior RNFL thickness, yet slightly apart from angular locations of enface first gap. It is possible that experimental settings and/or effects of blood vessels on RNFL segmentation in significantly thinned regions might have caused these observed differences. Notably, for each of these three participants, the RNFL thickness was markedly outside normal limits, allowing the device's classification system to flag this area as a defect. On the whole, data from this sample did not seem to provide compelling evidence supporting loss of reflectivity without loss of thickness of the RNFL.

Despite considerable research interest, the temporal relationship between changes of reflectivity and thickness of the RNFL is not fully understood. Findings from models of experimental glaucoma[20,21,64–66] suggest that reflectivity deteriorates earlier than a measurable thinning of the RNFL, but evidence for a measurable time delay between reflectance and thickness changes in human glaucoma remains sparse. Overall these results fit well with findings from diagnostic accuracy studies, suggesting that analysis of RNFL reflectivity might be useful at the earliest stages of glaucoma, becoming progressively less valuable in later stages where correlation with thickness measurements is strong.[44,45,66,67] This, however, does not preclude other uses of enface imaging providing additional value. For example, enface imaging may be useful in combination with other test modalities such as visual fields, enabling direct exploitation of the structure-function relationship without the need for spatial structure-function mapping, and in facilitating custom-perimetry based on structural data.[68–72]

Our study has limitations. We presented a novel approach for assessing glaucoma changes in enface images whose translation to practice would require little software adjunction. Clinicians could inspect the suspicious ONH sector in detail for evidence of focal loss of RNFB reflectivity and extract enface parameters. As we showed here, reduced first gap or last visible bundle seemed promising for glaucoma detection. However, additional research is needed on more diverse populations, including earliest glaucoma cases across a wider range of ages and ethnicities, to further characterize enface parameters and identify which one would be best in clinics. Further, subtle reflectivity changes may be overlooked by our subjective method, and more sensitive quantitative methods may be required to fully exploit the value of enface images.[15,22] Second, we studied eyes with established glaucoma, and the hypothesis of discrepancies between RNFL reflectance and thickness at earlier disease stages should be considered. A more thorough longitudinal analysis on reflectivity-thickness relationship is warranted. Cross-sectional approaches such as ours adopt a reference standard for glaucoma diagnosis requiring signs (e.g., defined RNFL thinning) that bias the study in favor of tests used in the inclusion criteria. In our case, all glaucoma participants had measurable cpRNFL defects, therefore, the high performance of cpRNFL is not surprising, though this unfavorable bias does make the similar or better performance of the enface indices more noteworthy. Lastly, there are limitations due to the image processing and the grading task. We did not include correction for beam light incident angle, which is among the determinants of RNFL reflectance[73] and is known to change in circumpapillary scans.[46,74] We speculate that this caveat may be less detrimental to subjective evaluation of RNFB presence compared to quantification of reflectance. The enface parameters considered here could be measured with a single cube scan centered on the ONH, further reducing the impact of beam incident angle. Additionally, enface images and thickness measurements were obtained from different scans that were not mutually registered. To minimize disagreement, data in each domain were adjusted for individual fovea-disc angle, and the strong correlation found suggests that any angular incongruences were small. Concerning the grading task, the order of enface images was not randomized between glaucoma and controls, and the presence of visible glaucomatous changes in many enface images precluded masking of the grader to disease status. Further, the grading task was performed by a single observer on a single occasion. However, previous work showed both consistency between observers and repeatability within observers to be excellent in the assessment of visible presence of RNFBs.[26,48]

In conclusion, our simple method to observe visible presence of RNFBs reliably identified glaucomatous defects in enface OCT images, with diagnostic performance at least as good as existing thickness parameters. No strong evidence of reflectivity loss without corresponding thickness loss was found. Development of more sensitive automated analyses and integration with perimetry may realize further potential of enface OCT images in glaucoma.

Acknowledgments

References

1. Prum BE, Rosenberg LF, Gedde SJ, et al. Primary open-angle glaucoma preferred practice pattern guidelines. *Ophthalmology*. 2016;123(1):p41–p111.
2. National Institute of Health and Care Excellence. Glaucoma: diagnosis and management 2017 [Available from: https://www.nice.org.uk/guidance/ng81].
3. Mwanza JC, Lee G, Budenz DL, et al. Validation of the UNC OCT index for the diagnosis of early glaucoma. *Transl Vis Sci Technol*. 2018;7(2):16.
4. Fallon M, Valero O, Pazos M, Anton A. Diagnostic accuracy of imaging devices in glaucoma: a meta-analysis. *Surv Ophthalmol*. 2017;62(4):446–461.
5. Tatham AJ, Medeiros FA, Zangwill LM, Weinreb RN. Strategies to improve early diagnosis in glaucoma. *Prog Brain Res*. 2015;221:103–133.
6. Jindal A, Ctori I, Fidalgo B, Dabasia P, Balaskas K, Lawrenson JG. Impact of optical coherence tomography on diagnostic decision-making by UK community optometrists: a clinical vignette study. *Ophthalmic Physiol Opt*. 2019;39(3):205–215.
7. Founti P, Coleman AL, Wilson MR, et al. Overdiagnosis of open-angle glaucoma in the general population: the Thessaloniki Eye Study. *Acta Ophthalmol (Copenh)*. 2018;96(7):e859–e864.
8. Hohn R, Nickels S, Schuster AK, et al. Prevalence of glaucoma in Germany: results from the Gutenberg Health Study. *Graefes Arch Clin Exp Ophthalmol*. 2018;256(9):1695–1702.
9. Karvonen E, Stoor K, Luodonpää M, et al. Prevalence of glaucoma in the Northern Finland Birth Cohort Eye Study. *Acta Ophthalmol*. 2019;97(2):200–207.
10. 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(7):1359–1369.
11. Kansal V, Armstrong JJ, Pintwala R, Hutnik C. Optical coherence tomography for glaucoma diagnosis: an evidence based meta-analysis. *PLoS One*. 2018;13(1):e0190621.
12. Hood DC, De Moraes CG. Four questions for every clinician diagnosing and monitoring glaucoma. *J Glaucoma*. 2018;27(8):657–664.
13. Hood DC, De Moraes CG. Challenges to the common clinical paradigm for diagnosis of glaucomatous damage with OCT and visual fields. *Invest Ophthalmol Vis Sci*. 2018;59(2):788–791.
14. Fortune B. In vivo imaging methods to assess glaucomatous optic neuropathy. *Exp Eye Res*. 2015;141:139–153.
15. Vermeer KA, van der Schoot J, Lemij HG, de Boer JF. RPE-normalized RNFL attenuation coefficient maps derived from volumetric OCT imaging for glaucoma assessment. *Invest Ophthalmol Vis Sci*. 2012;53(10):6102–6108.
16. Huang XR, Knighton RW, Cavuoto LN. Microtubule contribution to the reflectance of the retinal nerve fiber layer. *Invest Ophthalmol Vis Sci*. 2006;47(12):5363–5367.
17. Huang XR, Knighton RW, Spector YZ, Feuer WJ. Cytoskeletal alteration and change of retinal nerve fiber layer birefringence in hypertensive retina. *Curr Eye Res*. 2017;42(6):936–947.
18. 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(1):19–28.
19. Tuulonen A, Airaksinen PJ. Initial glaucomatous optic disk and retinal nerve fiber layer abnormalities and their progression. *Am J Ophthalmol*. 1991;111(4):485–490.
20. Huang XR, Zhou Y, Kong W, Knighton RW. Reflectance decreases before thickness changes in the retinal nerve fiber layer in glaucomatous retinas. *Invest Ophthalmol Vis Sci*. 2011;52(9):6737–6742.
21. Fortune B, Burgoyne CF, Cull G, Reynaud J, Wang L. Onset and progression of peripapillary retinal nerve fiber layer (RNFL) retardance changes occur earlier than RNFL thickness

translational vision science & technology

changes in experimental glaucoma. *Invest Ophthalmol Vis Sci*. 2013;54(8):5653–5661.

22. Ashimatey BS, King BJ, Burns SA, Swanson WH. Evaluating glaucomatous abnormality in peripapillary optical coherence tomography enface visualisation of the retinal nerve fibre layer reflectance. *Ophthalmic Physiol Opt*. 2018;38(4):376–388.

23. Hood DC, Fortune B, Mavrommatis MA, et al. Details of glaucomatous damage are better seen on OCT en face images than on OCT retinal nerve fiber layer thickness maps. *Invest Ophthalmol Vis Sci*. 2015;56(11):6208–6216.

24. Sakamoto M, Mori S, Ueda K, et al. En face slab images visualize nerve fibers with residual visual sensitivity in significantly thinned macular areas of advanced glaucomatous eyes. *Invest Ophthalmol Vis Sci*. 2019;60(8):2811–2821.

25. Mavrommatis MA, De Cuir N, Reynaud J, et al. An examination of the frequency of paravascular defects and epiretinal membranes in eyes with early glaucoma using en-face slab OCT images. *J Glaucoma*. 2019;28(3):265–269.

26. Cheloni R, Denniss J. Depth-resolved variations in visibility of retinal nerve fibre bundles across the retina in enface OCT images of healthy eyes. *Ophthalmic Physiol Opt*. 2021;41(1):179–191.

27. Sayed MS, Margolis M, Lee RK. Green disease in optical coherence tomography diagnosis of glaucoma. *Curr Opin Ophthalmol*. 2017;28(2):139–153.

28. Chong GT, Lee RK. Glaucoma versus red disease: imaging and glaucoma diagnosis. *Curr Opin Ophthalmol*. 2012;23(2):79–88.

29. Vermeer KA, Mo J, Weda JJ, Lemij HG, de Boer JF. Depth-resolved model-based reconstruction of attenuation coefficients in optical coherence tomography. *Biomed Opt Express*. 2013;5(1):322–337.

30. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/.

31. Thepass G, Lemij HG, Vermeer KA. Attenuation coefficients from SD-OCT data: structural information beyond morphology on RNFL integrity in glaucoma. *J Glaucoma*. 2017;26(11):1001–1019.

32. Garway-Heath DF, Poinoosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology*. 2000;107(10):1809–1815.

33. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. 2015;67(1), 1–48.

34. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12(1):77.

35. McClish DK. Analyzing a portion of the ROC curve. *Med Decis Making*. 1989;9(3):190–195.

36. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837–845.

37. Bakdash JZ, Marusich LR. Repeated measures correlation. *Front Psychol*. 2017;8:456.

38. Bland JM, Altman DG. Calculating correlation coefficients with repeated observations: Part 1–correlation within subjects. *BMJ (Clinical research ed)*. 1995;310(6977):446–.

39. Bland JM, Altman DG. Calculating correlation coefficients with repeated observations: Part 2–correlation between subjects. *BMJ (Clinical research ed)*. 1995;310(6980):633.

40. Verticchio Vercellin AC, Jassim F, Poon LY, et al. Diagnostic capability of three-dimensional macular parameters for glaucoma using optical coherence tomography volume scans. *Invest Ophthalmol Vis Sci*. 2018;59(12):4998–5010.

41. Champely S. pwr: Basic Functions for Power Analysis 2020 [R package version 1.3-0: Available from: https://cran.r-project.org/package=pwr].

42. Chen TC, Hoguet A, Junk AK, et al. Spectral-domain OCT: helping the clinician diagnose glaucoma: a report by the American Academy of Ophthalmology. *Ophthalmology*. 2018;125(11):1817–1827.

43. Pons ME, Ishikawa H, Gurses-Ozden R, Liebmann JM, Dou HL, Ritch R. Assessment of retinal nerve fiber layer internal reflectivity in eyes with and without glaucoma using optical coherence tomography. *Arch Ophthalmol. (Chicago, Ill: 1960)*. 2000;118(8):1044–1047.

44. Liu S, Wang B, Yin B, et al. Retinal nerve fiber layer reflectance for early glaucoma diagnosis. *J Glaucoma*. 2014;23(1):e45–e52.

45. Gardiner SK, Demirel S, Reynaud J, Fortune B. Changes in retinal nerve fiber layer reflectance intensity as a predictor of functional progression in glaucoma. *Invest Ophthalmol Vis Sci*. 2016;57(3):1221–1227.

46. Tan O, Liu L, You Q, Wang J, Jia Y, Huang D. Focal loss analysis of nerve fiber layer reflectance for glaucoma diagnosis. *Invest Ophthalmol Vis Sci*. 2020;61(7):5194–.

47. Hood DC, De Cuir N, Blumberg DM, et al. A single wide-field OCT protocol can provide

compelling information for the diagnosis of early glaucoma. *Transl Vis Sci Technol*. 2016;5(6):4.

48. Iikawa R, Togano T, Sakaue Y, et al. Estimation of the central 10-degree visual field using en-face images obtained by optical coherence tomography. *PLoS One*. 2020;15(3):e0229867.

49. Leitgeb RA. En face optical coherence tomography: a technology review [Invited]. *Biomed Opt Express*. 2019;10(5):2177–2201.

50. van der Schoot J, Vermeer KA, de Boer JF, Lemij HG. The effect of glaucoma on the optical attenuation coefficient of the retinal nerve fiber layer in spectral domain optical coherence tomography images. *Invest Ophthalmol Vis Sci*. 2012;53(4):2424–2430.

51. Miura N, Omodaka K, Kimura K, et al. Evaluation of retinal nerve fiber layer defect using wide-field en-face swept-source OCT images by applying the inner limiting membrane flattening. *PLoS One*. 2017;12(10):e0185573.

52. Hood DC. Improving our understanding, and detection, of glaucomatous damage: an approach based upon optical coherence tomography (OCT). *Prog Retin Eye Res*. 2017;57:46–75.

53. Weinreb RN, Leung CK, Crowston JG, et al. Primary open-angle glaucoma. *Nat Rev Dis Primers*. 2016;2:16067.

54. Jonas JB, Aung T, Bourne RR, Bron AM, Ritch R, Panda-Jonas S. Glaucoma. *Lancet (London, England)*. 2017;390(10108):2183–2193.

55. Stagg BC, Medeiros FA. A comparison of OCT parameters in identifying glaucoma damage in eyes suspected of having glaucoma. *Ophthalmol Glaucoma*. 2020;3(2):90–96.

56. Virgili G, Michelessi M, Cook J, et al. Diagnostic accuracy of optical coherence tomography for diagnosing glaucoma: secondary analyses of the GATE study. *Br J Ophthalmol*. 2018;102(5):604–610.

57. Karvonen E, Stoor K, Luodonpää M, et al. Diagnostic performance of modern imaging instruments in glaucoma screening. *Br J Ophthalmol*. 2020;104(10):1399–1405.

58. Michelessi M, Li T, Miele A, Azuara-Blanco A, Qureshi R, Virgili G. Accuracy of optical coherence tomography for diagnosing glaucoma: an overview of systematic reviews. *Br J Ophthalmol*. 2021;105(4):490–495

59. Patel NB, Luo X, Wheat JL, Harwerth RS. Retinal nerve fiber layer assessment: area versus thickness measurements from elliptical scans centered on the optic nerve. *Invest Ophthalmol Vis Sci*. 2011;52(5):2477–2489.

60. Hood DC, Fortune B, Arthur SN, et al. Blood vessel contributions to retinal nerve fiber layer thickness profiles measured with optical coherence tomography. *J Glaucoma*. 2008;17(7):519–528.

61. Strouthidis NG, Yang H, Fortune B, Downs JC, Burgoyne CF. Detection of optic nerve head neural canal opening within histomorphometric and spectral domain optical coherence tomography data sets. *Invest Ophthalmol Vis Sci*. 2009;50(1):214–223.

62. Ballae Ganeshrao S, Turpin A, Denniss J, McKendrick AM. Enhancing structure-function correlations in glaucoma with customized spatial mapping. *Ophthalmology*. 2015;122(8):1695–1705.

63. Yang Q, Reisman CA, Wang Z, et al. Automated layer segmentation of macular OCT images using dual-scale gradient information. *Opt Express*. 2010;18(20):21293–21307.

64. Dwelle J, Liu S, Wang B, et al. Thickness, phase retardation, birefringence, and reflectance of the retinal nerve fiber layer in normal and glaucomatous non-human primates. *Invest Ophthalmol Vis Sci*. 2012;53(8):4380–4395.

65. Fortune B, Burgoyne CF, Cull GA, Reynaud J, Wang L. Structural and functional abnormalities of retinal ganglion cells measured in vivo at the onset of optic nerve head surface change in experimental glaucoma. *Invest Ophthalmol Vis Sci*. 2012;53(7):3939–3950.

66. Fortune B, Cull G, Reynaud J, Wang L, Burgoyne CF. Relating retinal ganglion cell function and retinal nerve fiber layer (RNFL) retardance to progressive loss of RNFL thickness and optic nerve axons in experimental glaucoma. *Invest Ophthalmol Vis Sci*. 2015;56(6):3936–3944.

67. Xu G, Weinreb RN, Leung CKS. Retinal nerve fiber layer progression in glaucoma: a comparison between retinal nerve fiber layer thickness and retardance. *Ophthalmology*. 2013;120(12):2493–2500.

68. Denniss J, McKendrick AM, Turpin A. Towards patient-tailored perimetry: automated perimetry can be improved by seeding procedures with patient-specific structural information. *Transl Vis Sci Technol*. 2013;2(4):3.

69. Montesano G, Rossetti LM, Allegrini D, Romano MR, Crabb DP. Improving visual field examination of the macula using structural information. *Transl Vis Sci Technol*. 2018;7(6):36.

70. Alluwimi MS, Swanson WH, Malinovsky VE, King BJ. Customizing perimetric locations based

on en face images of retinal nerve fiber bundles with glaucomatous damage. *Transl Vis Sci Technol.* 2018;7(2):5.

71. Turpin A, Morgan WH, McKendrick AM. Improving spatial resolution and test times of visual field testing using ARREST. *Transl Vis Sci Technol.* 2018;7(5):35.

72. Ganeshrao SB, McKendrick AM, Denniss J, Turpin A. A perimetric test procedure that uses structural information. *Optom Vis Sci.* 2015;92(1):70–82.

73. Huang XR, Knighton RW, Feuer WJ, Qiao J. Retinal nerve fiber layer reflectometry must consider directional reflectance. *Biomed Opt Express.* 2016;7(1):22–33.

74. Knighton RW, Huang XR. Directional and spectral reflectance of the rat retinal nerve fiber layer. *Invest Ophthalmol Vis Sci.* 1999;40(3):639–647.

Article

Enhanced Objective Detection of Retinal Nerve Fiber Bundle Defects in Glaucoma With a Novel Method for En Face OCT Slab Image Construction and Analysis

Riccardo Cheloni[1], Simon D. Dewsbery[2], and Jonathan Denniss[1]

[1] School of Optometry and Vision Science, University of Bradford, Bradford, UK
[2] Department of Ophthalmology, Leeds Teaching Hospitals NHS Trust, Leeds, UK

Correspondence: Jonathan Denniss, School of Optometry and Vision Science, University of Bradford, Richmond Rd, Bradford BD7 1DP, UK. e-mail: j.denniss@bradford.ac.uk

Purpose: To introduce and evaluate the performance in detecting glaucomatous abnormalities of a novel method for extracting en face slab images (SMAS), which considers varying individual anatomy and configuration of retinal nerve fiber bundles.

Methods: Dense central retinal spectral domain optical coherence tomography scans were acquired in 16 participants with glaucoma and 19 age-similar controls. Slab images were generated by averaging reflectivity over different depths below the inner limiting membrane according to several methods. SMAS considered multiple 16 μm thick slabs from 8 to 116 μm below the inner limiting membrane, whereas 5 alternative methods considered single summary slabs of various thicknesses and depths. Superpixels in eyes with glaucoma were considered abnormal if below the first percentile of distributions fitted to control data for each method. The ability to detect glaucoma defects was measured by the proportion of abnormal superpixels. Proportion of superpixels below the fitted first percentile in controls was used as a surrogate false-positive rate. The effects of slab methods on performance measures were evaluated with linear mixed models.

Results: The ability to detect glaucoma defects varied between slab methods, $\chi^2_{(5)} = 120.9$, $P < 0.0001$, with SMAS showing proportion of abnormal superpixels 0.05 to 0.09 greater than alternatives (all $P < 0.0001$). No slab method found abnormal superpixels in controls.

Conclusions: SMAS outperformed alternatives in detecting abnormalities in eyes with glaucoma. SMAS evaluates all depths with potential retinal nerve fiber bundle presence by combining multiple slabs, resulting in greater detection of reflectance abnormalities with no increase in surrogate false positives.

Translational Relevance: SMAS may be used to objectively detect glaucoma defects in en face optical coherence tomography images.

Introduction

Optical coherence tomography (OCT) is increasingly used to assess structural changes of the retina owing to glaucoma.[1–3] Such changes are conventionally evaluated in cross-sectional scans assessing the thickness of either the retinal nerve fiber layer (RNFL) or the ganglion cell and inner plexiform layers.[2,4] En face OCT imaging is a relatively new approach that uses transverse retinal images to assess reflectance properties of retinal nerve fiber bundles (RNFBs).[5] Compared with assessing RNFL reflectance in fundus photographs,[6,7] en face OCT has advantages in better visualization of narrow defects and preserved bundles, the ability to examine below the superficial RNFL, and being less affected by lens opacities and light fundus pigmentation.[8–10] En face OCT analysis of reflectivity has also demonstrated potential for early glaucoma detection[11–13] and is a potential means to facilitate custom perimetry strategies that target specific regions of interest.[14–18] Although direct

1

observation of RNFBs may be beneficial, optimal methods to construct en face slab images are yet to be determined, and automated, objective methods to detect glaucoma defects in this domain are also lacking.[13,19]

En face images are usually generated from dense volumetric scans of the area of interest followed by the projection of pixel intensities from a certain range of depths within each A-scan into a transverse slab image.[20] Here, healthy RNFBs appear hyperreflective because of the ordered structure of their axonal cytoskeletons,[21] and this property may be lost early in glaucoma, decreasing reflectivity.[22] Thinning of the RNFL leads to inclusion of deeper, hyporeflective retinal layers in the slab image, also decreasing reflectivity.[5,23] Damaged bundles, therefore, appear in en face images as loss of reflectivity following typical patterns such as arcuate and wedge-shaped defects.[24]

There are several possible approaches to en face image extraction, including variations in the region of retina imaged and the composition of slabs as defined by the depths below the inner limiting membrane (ILM) over which A-scan pixels are averaged. Furthermore, different arithmetic methods to convert three-dimensional data into transverse images and different approaches to account for individual anatomic variability may be considered. We previously showed that the configuration of RNFBs in healthy eyes varies with retinal location and individual anatomy.[19] Accordingly, the final appearance and diagnostic usefulness of en face slab images is likely to be affected by the methods used for slab construction. These effects have been investigated minimally and therefore choices are currently made based on limited information.

Previous work in this area has often averaged the first 50 μm below the ILM in a single slab image.[5,23,25,26] Consistently with RNFL thickness,[27,28] RNFBs are present at depths of more than 50 μm proximal to the optic nerve head (ONH),[19] and limiting en face analysis to this depth might, therefore, miss early glaucoma defects in some regions.[24] Further, although several authors have recognized a need to adjust slab characteristics to individual anatomy and the varying morphology across the retina,[5,23–25] these adjustments have not been fully considered.

In this study, we introduce summary of multiple anatomically adjusted slabs (SMAS), a novel method for the construction and analysis of slab images. SMAS aims to address current limitations of en face imaging, including (i) adapting to individual anatomy, (ii) considering all depths and regions that contain visible RNFBs in healthy eyes, and (iii) adjusting for different layer morphology across the retina. We also evaluate the ability of several alternative slab construction methods to objectively detect glaucoma defects as compared with SMAS.

Methods

Participants

Twenty-two participants with open angle glaucoma and 19 age-similar healthy controls were recruited for this study. All participants underwent ophthalmic examination including subjective refraction, slit lamp assessment, Goldmann applanation tonometry, retinal OCT (Spectralis, Heidelberg Engineering, Heidelberg, Germany) and standard automated perimetry (24-2 SITA-Standard, Humphrey Field Analyzer III, Carl Zeiss Meditec Inc., Jena, Germany). Participants with glaucoma were only included if older than 40 years and presenting a clinical diagnosis of open angle glaucoma. In addition, inclusion required evidence of structural damage defined as at least one ONH sector with $P < 1\%$ from the Spectralis circumpapillary RNFL thickness analysis. No visual field inclusion criteria were applied to the glaucoma group to include the earliest cases. Participants with refractive error magnitude greater than 6.00 DS or 3.00 DC, evidence of lens opacification,[29] or other eye conditions except glaucoma were excluded. Healthy participants were included if they had no eye disease or history of eye disease and normal visual field as defined by a normal Mean Deviation ($P > 5\%$), glaucoma hemifield test within normal limits and absence of three contiguous non-edge points with $P < 5\%$ on the pattern deviation plot. One eye per participant was included. If both eyes were eligible, the tested eye was selected at random in healthy controls, whereas the one with milder defect (less negative Mean Deviation) was included in participants with glaucoma.

All participants provided written informed consent to participate and were free to withdraw at any time. The study adhered to the tenets of the Declaration of Helsinki and achieved ethical approval from the National Health Service's Research Ethics Service.

OCT Imaging and Image Processing

Seven high-speed dense OCT scans were taken of the central ±25° of the retina (30 μm separation between B-scans) of each participant. The OCT procedure used has been described in detail previously.[13,19] All images were acquired with signal to noise ratio of greater than 20 dB as recommended by the manufacturer.

Single-pixel deep slabs ($n = 50$) of the instrument's maximum digital axial resolution (3.87 μm), ranging from the ILM to 193.5 μm below it, were extracted from individual B-scans. Slab images were converted to depth-resolved attenuation coefficients,[30] which represent an intrinsic optical property of the retinal tissue[31] and have been advocated to minimize the impact of artefacts in en face images.[24] Attenuation coefficient data were imported into MATLAB (Version 9.6.0, The MathWorks Inc., Natick, MA) for montaging and image processing. Before montaging, gamma correction was used to smooth intensity differences between OCT images from different retinal locations by matching the luminance of overlapping regions of neighboring images at each depth using the central macular image as the reference.[32] For this gamma correction, we used the ratio between the median intensities of the individual slab and the macular image in corresponding overlapping regions as the gamma coefficient. We then montaged the images using custom software, again using the macular scan as a reference image.

The highest pixel value was selected from overlapping regions.

Montaged images were processed as described in detail previously.[19] In brief, the intensity of an area with no RNFBs within the raphe region was extracted 35μm below the ILM, and set as background with a threshold transformation. Then, pixel intensity was normalized by dividing by the mean of the 99th percentiles from each depth. Values were clipped to 1, resulting in images with pixel intensities in the range of 0 to 1.

Adjusting for Individual Anatomy

We aimed to minimize the impact of individual anatomy by adjusting en face images to the fovea–disc and fovea–raphe angles. We used geometric image transformations to align the ONH, fovea, and raphe along a common horizontal axis (Fig. 1). Left eyes were flipped to right eye format and the slab image offering best visibility of the foveal pit was used to
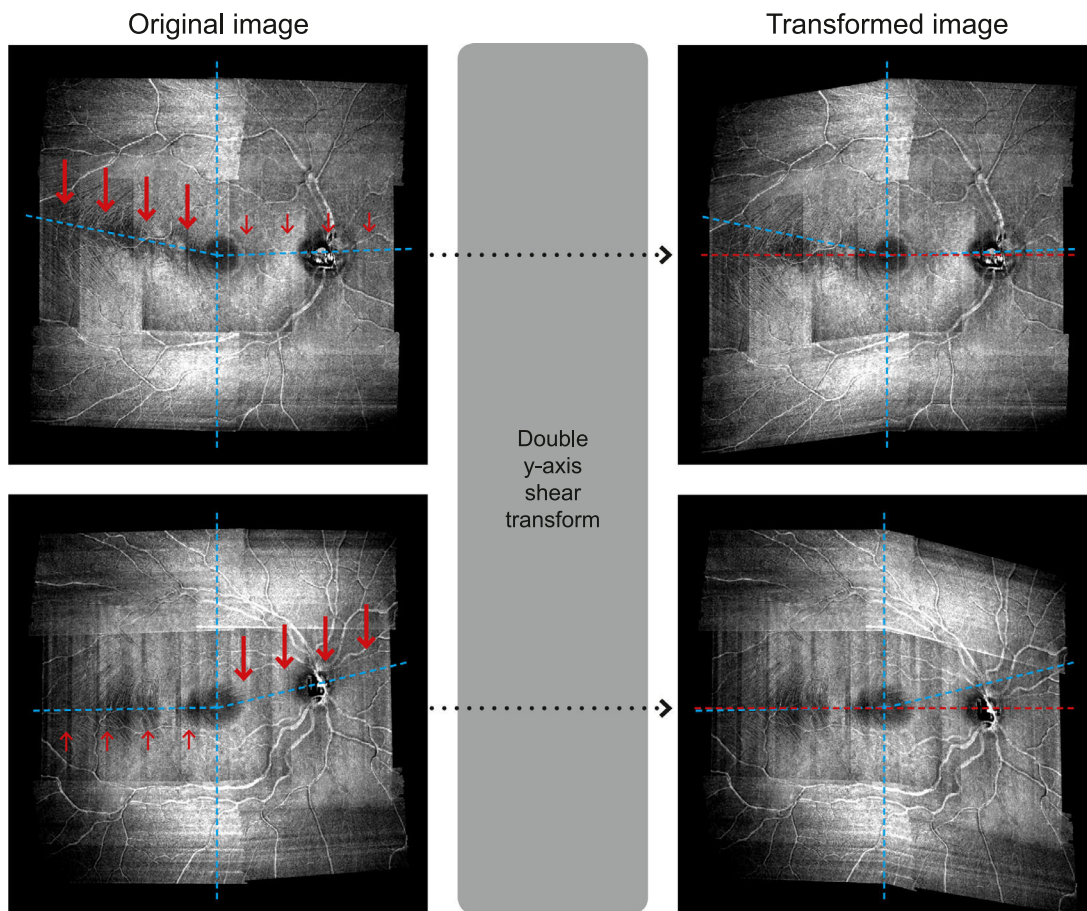
Figure 1. Example of the double vertical shear transformation applied to the en face images of two healthy controls. For the image shown in the *top panels*, the major transformation (*thick red arrows*) was applied to the temporal retina whereas a more minor transformation (*small red arrows*) was applied to the nasal retina. The opposite applies for the image shown in the bottom panels. Irrespectively of the original anatomy (*blue dashed lines*), transformed images (*right column*) align the raphe, fovea, and ONH along a horizontal line (*red dashed lines*).

manually extract the coordinates of the fovea and the center of the ONH. The fovea–disc angle was defined as a straight line between these two points. The orientation of the raphe is known to change with individual anatomy and can be measured in both healthy eyes and eyes with glaucoma.[33] Accordingly, the fovea–raphe angle was extracted following an existing method.[34] Using the single slab image with best visibility of RNFBs in the raphe area (median, 15.5 μm below the ILM; range, 11.6–27.1 μm), the fovea–raphe angle was measured by tracing lines connecting the fovea to five manually selected points in the raphe gap region. We took the average of the five angles identified as the fovea–raphe angle.[34] We then aligned the raphe, fovea and ONH along a horizontal line by applying vertical shear transformations separately to the image regions either side of the fovea (Fig. 1). Shear transformations enable the shift of a single dimension of the image (vertical in this case) by a given angular value, leaving the other (horizontal) dimension unmodified. This approach enabled evaluation of reflectivity over a square array of superpixels (discussed elsewhere in this article), with landmark retinal locations laying on a common horizontal axis (Fig. 1).

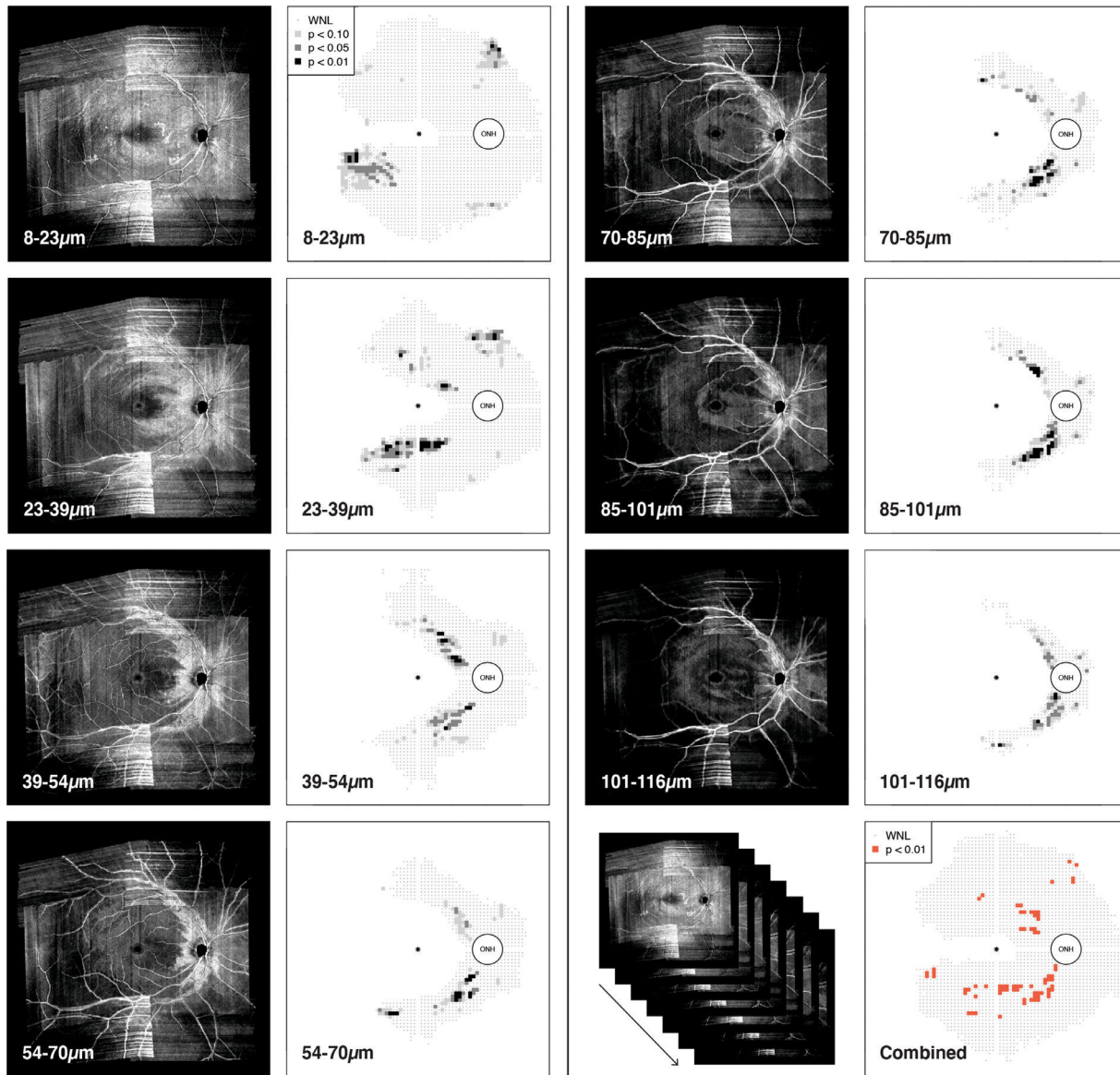## Extraction of Reflectance Abnormalities With SMAS

With the objective of considering all depths with present RNFBs and the differing layer morphology across the retina, we averaged together groups of four single-pixel deep slabs starting from 7.8 μm up to 193.5 μm below the ILM. The first two depths (i.e., <7.8 μm below the ILM) were excluded from slab construction because they do not contain visible

RNFBs in healthy eyes,[19] but are likely to contain vitreous interface and glial artefacts.[24,35] The averaging of groups of four slabs together aimed to combine sufficient single pixel slabs to minimize image noise while also minimizing the mixture of retinal layers (i.e., RNFL with deeper layers such as ganglion cell and inner plexiform layers).[19] RNFBs are visually present at the narrowest range of depths in the temporal retina, and the averaging adopted by SMAS should allow inclusion of all bundles from these regions in the first slab (from 7.8 μm to 23.2 μm below the ILM). This goal should also be achieved in eyes with particularly thin RNFL.[19] This process yielded 12 slabs (each approximately 15.5 μm thick) from 7.8 μm to 193.5 μm below the ILM (Supplementary Table S1).

An analysis of reflectivity was performed on a superpixel grid centered on the fovea, with each superpixel composed of a number of individual pixels in a $n \times n$ pixel configuration. The intensity of each superpixel was the mean of its constituent pixels. Additional mitigation of anatomic variability was achieved by controlling for the varying distance between the fovea and ONH by adjusting the size of superpixels in individual images such that a fixed 20 superpixels separated the fovea and ONH. This number was chosen to target a superpixel dimension of 20 × 20 pixels, previously suggested to represents a suitable compromise between between-subject variability and sufficient resolution to detect wedge shaped defects.[24] Superpixels in individual images contained median 20 × 20 pixels (range, 17 × 17 to 23 × 23) and this system of coordinates maximized consistency between retinal locations among different eyes. See Figures 2 and 3 for examples of the superpixel grid.
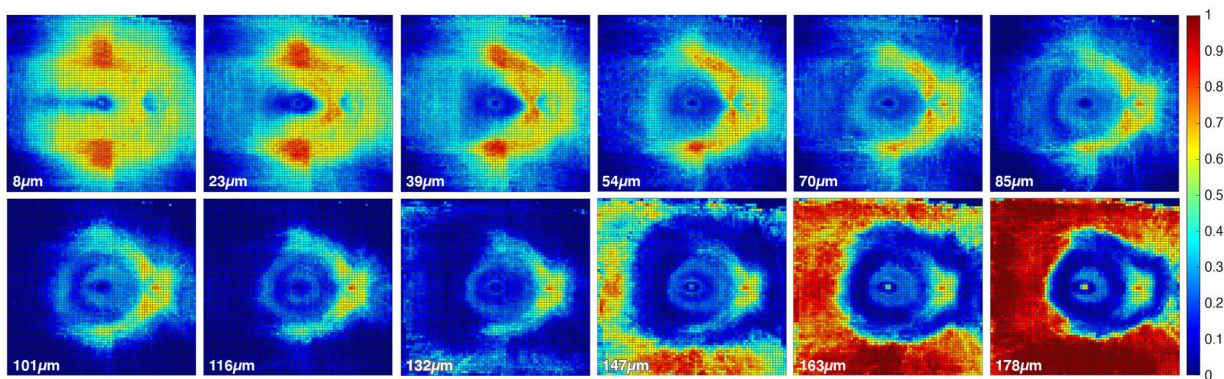
Figure 2. Heat maps of median normative data for the 12 slabs generated in the SMAS method. It is possible to identify the main retinal structures such as the temporal raphe, fovea, and ONH. At greater depths the hyper-reflectivity of the retinal pigment epithelium becomes visible. Depths (in micrometers) shown correspond with the anterior depth at which each of the 15.5-μm-thick slabs commenced. The heat maps also show hyper-reflective artefacts in the superior- and inferior regions (first 3–4 slabs), likely owing to the varying beam light incident angles from wide-field OCT imaging (see Discussion).[37]

Figure 3. Example of the multiple slab images generated with the SMAS method for one participant with glaucoma, as well as the combined deviation map reporting all reflectance abnormalities detected in any slab. At each specific slab depth, whose starting and ending depths are reported in micrometers, the left-most image shows the actual slab image and the right-most image shows the corresponding deviation map. In deviation maps grey points indicate superpixels found within normal limits (WNL), and superpixels below the tenth, fifth, and first percentiles are reported as squares color-coded according to the level of significance. The bottom-right panel shows the combined deviation map with abnormal superpixels (<1%) identified at any depth. In deviation maps, each data point corresponds with 1 superpixel.

The distributions of superpixel intensities at all depths in control eyes were explored visually at different retinal locations and Shapiro–Wilk tests for normality were performed. The majority of distributions were either multimodal or significantly skewed. To account for the observed distributions and the modest sample size, summary statistics and limits of normality were derived from kernel density-estimated frequency distributions rather than the empirical data.[36] We extracted the estimated median (Fig. 2),

tenth, fifth, and first percentiles at all depths of all superpixels.

As reported previously,[19] and as shown in Figure 2, the presence of RNFBs throughout the retina varies with retinal location and depth below the ILM. An evaluation of the reflectance in locations where RNFBs are not expected to be visible even in healthy retinae would have no diagnostic value; therefore, these areas were censored from analysis. Accordingly, analysis was restricted to regions of interest in the first 7 slabs (up

to 116 µm below the ILM). Regions of interest were manually identified as those containing visible RNFBs in the control eyes. For each depth individually, a value 2.5 standard deviations below the grand mean intensity within regions of interest across all control images was set as a threshold, and regions with a lower mean intensity in control eyes were excluded from analysis in all images. This threshold (2.5 standard deviations below the mean) was chosen as the best compromise between maximizing the retinal area evaluated and the adequate exclusion of regions with no visible RNFBs among several cut-offs trialed (Supplementary Fig. S1).

Finally, slab images were extracted in all participants with glaucoma according to the SMAS approach. Superpixel values within the previously defined regions of interest at each depth were compared with corresponding normal limits from controls. Reflectance abnormalities were identified in single depth deviation maps because intensities below the estimated tenth, fifth, and first percentiles of control data (Fig. 3). The seven deviation maps from each depth were then combined into a summary deviation map reporting abnormal superpixels ($<1\%$) identified at any depth (Fig. 3, bottom right).

Extraction of Reflectance Abnormalities With Alternative En Face Slab Methods

En face images for all participants were also generated using several other slab methods based primarily on previous studies.[5,19,24] For each method, transformed single pixel slabs of individual eyes were averaged together over a specific range of depths. Normative data from controls were then extracted as described earlier and reflectance abnormalities were evaluated in eyes with glaucoma using the method described elsewhere in this article. Adjustments for individual anatomy made for SMAS were applied identically for other slab methods, as was the positioning and spacing of the superpixel grid. Therefore, different slab methods differed only in the retinal depths averaged (Fig. 4). All slab methods were evaluated over the same region of the retina tested by SMAS.

Hood Slab

This method was similar to that of Hood et al.,[5] in which the pixel intensity was averaged over a 52-µm deep slab starting from the ILM. We averaged the first 13 single-pixel slabs together, encompassing depths
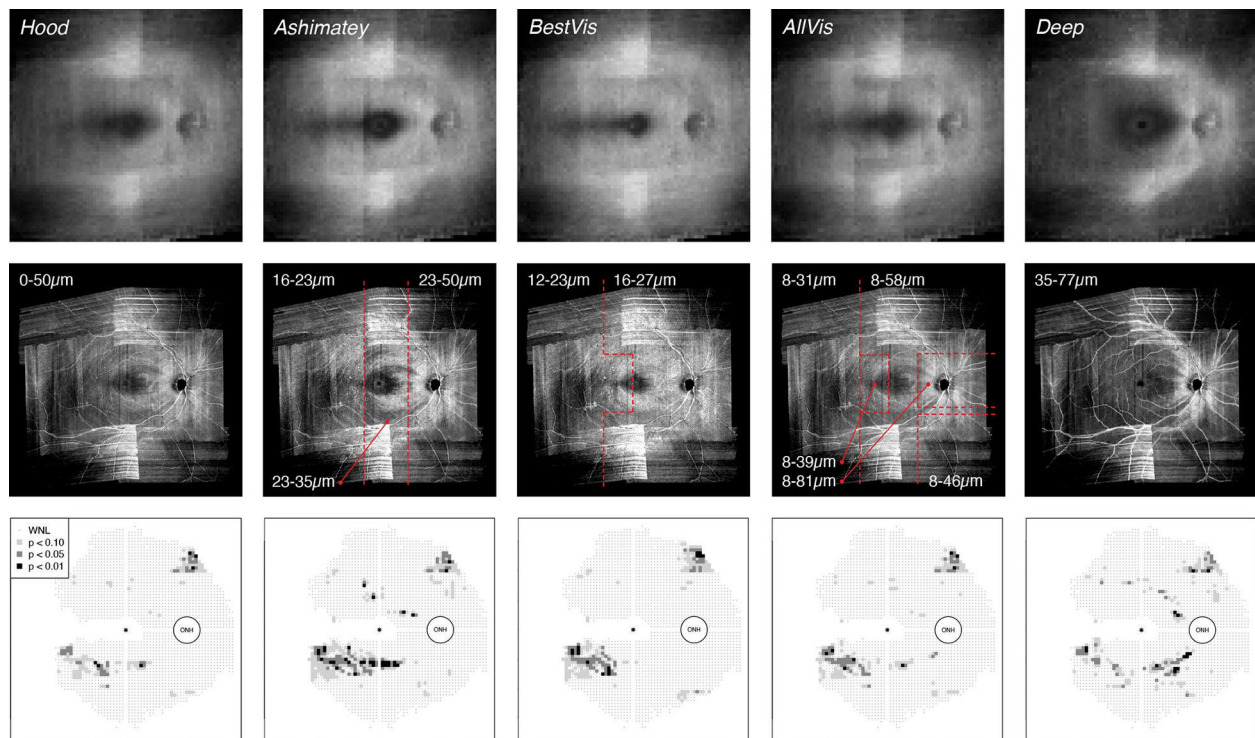
Figure 4. Alternative methods of slab construction explored in this study in addition to SMAS. (*Top row*) Method-specific normative en face slabs derived from control eyes. (*Middle row*) Individual slab images for the same participant with glaucoma as shown in Figure 3. Red dashed lines demarcate different regions of the slab characterized by different depths considered (see text), as labelled in µm. (*Bottom row*) Corresponding deviation maps for different slab methods. Format of deviation maps as per Figure 3.

from the ILM to 50.3 µm below the ILM as the closest possible match to the method of Hood et al.[5]

Ashimatey Slab

Ashimatey et al.[24] used a slab with decreasing thickness from the ONH to the temporal retina. They averaged pixel intensity from 24 to 52 µm below the ILM in the optic disc region, from 24 to 36 µm in the central retina and from 16 to 24 µm in the temporal macula and raphe region. To reproduce a similar slab configuration, we considered three vertically separated regions with different thickness (Fig. 4): the 7th through the 13th pixels in the ONH region (approximately 23.3 to 50.3 µm), the 7th to the 9th pixels in the macular area (approximately 23.3 to 34.8 µm), and the 5th to the 6th pixels in the raphe area (approximately 15.6 to 23.2 µm).

Best Visibility Slab (BestVis)

This slab was centered on the depth of best RNFB visibility across the retina in healthy eyes, which we found previously at an average of 20.3 ± 1.9 µm below the ILM, with slight differences between the temporal and nasal retina.[19] Accordingly, the single pixel slab at the depth of best RNFB visibility was averaged together with the one above and the one below in the BestVis slab (Fig. 4). The fourth through the sixth pixels were included in the raphe and temporal macula (approximately 11.7 to 23.2 µm; best = 18 µm), whereas the fifth through the seventh pixels were averaged in the rest of the retina (15.6 to 27.1 µm; best = 22 µm).

All Visible RNFBs Slab (AllVis)

As per the best visibility slab, this approach considered our previous work[19] and averaged all depths expected to contain visible RNFBs in healthy eyes. Differences in RNFB visibility across the healthy retina were accounted for by averaging varying depths in different regions of the retina (Fig. 4). Hence, the pixel depths included were the 3rd to 8th in the raphe (approximately 7.8 to 31 µm); 3rd to 10th in the temporal macula (approximately 7.8 to 38.7 µm); 3rd through 12th in the inferior nasal quadrant (approximately 7.8 to 46.4 µm); 3rd through 15th in the central and superior-nasal retina (approximately 7.8 to 58.1 µm); and 3rd through 21st in the ONH region (approximately 7.8 to 81.3 µm).

Deep Slab

The deep slab included greater depths below the ILM than considered by most of the methods discussed elsewhere in this article and was included as a control. The deep slab averaged intensity starting from depths close to the posterior limit of the Hood slab and the Ashimatey slab up until the greatest depths at which arcuate regions and the nasal retina around the ONH still present RNFBs in healthy eyes.[19] The slab averaged the 10th to 20th pixels through the whole retina, corresponding with 34.9 to 77.4 µm below the ILM.

Analysis

All extracted slabs and corresponding deviation maps were examined by two authors (RC, JD) for the impact of artefacts. Either whole images or specific regions from participants with substantial effect of artefacts were excluded from further analysis. Artefacts of en face images could arise from low-quality B-scans, floaters, and glial cell alterations.[35] A joint discussion of single cases was performed until a consensus on data exclusion was reached.

The performance of different slab methods was compared by the proportion of abnormal superpixels identified in each participant with glaucoma. This metric was computed as the number of superpixels below the first percentile of the corresponding normative data divided by the number of tested superpixels. For SMAS, the combined deviation map was considered. The differences in the proportion of abnormal superpixels between slab methods were explored with linear mixed models and χ^2 likelihood ratio tests.[38] The slab method was considered as a fixed effect, whereas individual participants were modelled as random effects to account for the repeated-measures design. Statistical significance was considered at $P < 0.5$ and the model had the following form:

$$y \sim 1 + Slab\ Method + (1|Eye) + \varepsilon \qquad (1)$$

where y signifies the outcome of interest (e.g., the proportion of abnormal superpixels), and 1 and ε signify intercept and random error, respectively. Pairwise differences were tested with post hoc t tests, adjusting for multiple comparisons with the Tukey method.

To further characterize individual slab methods, the median distance of abnormal superpixels from the ONH was extracted from all participants with glaucoma for all different approaches. Because specific methods might include different retinal sections across the area examined, the resulting slab composition and therefore the ability to detect abnormalities could also change with retinal location.

False-positive rates of different slab methods could not be evaluated directly owing to the lack of an independent reference standard. As a surrogate measure, we explored the rate of abnormal superpixels

in control eyes at the 1% level of significance (derived from kernel density estimation as described earlier). Linear mixed models of the form above were used to evaluate differences of distance from the ONH and surrogate false-positive rate among different slab methods. Last, we tested whether differences in performance between SMAS and alternatives were related to the severity of reflectance defects. As such, we computed coefficients of determination (R^2) between the mean and standardized difference in proportions of abnormal superpixels between each slab method and SMAS.

As estimated with the *simr R* package,[39] this study had 91% power (95% confidence interval [CI], 89.2–92.8) to detect a 0.02 difference in the proportion of abnormal superpixels at an alpha of 0.05.

Results

Images from six eyes with glaucoma (median age, 69 years; range, 67–78 years; median Mean Deviation, –6.0 dB; range, –1.6 to –12.8 dB) contained significant artefacts and were excluded from the main analysis. Further, part of the en face images of three participants with glaucoma were censored for similar reasons. For one participant, the whole upper hemifield was excluded, whereas a horizontal band in the upper retina and the inferior temporal retina were censored in the remaining two cases. Overall, 19 controls (median age, 68 years; range, 56–75 years) and 16 participants with glaucoma (median age, 70 years; range, 61–77 years) were included. All participants in the glaucoma group but one had a visual field defect according to the definition used for the control group's exclusion criteria. The remaining participant with glaucoma had three contiguous defective points, but one was an edge location. Table reports detailed participant demographics.

Figure 5 shows the proportion of abnormal superpixels identified by different slab methods for all participants with glaucoma. SMAS found a greater proportion of abnormal superpixels than all alternative slab methods in all participants with glaucoma.

There were significant differences between slab methods in the proportion of abnormal superpixels identified in the eyes with glaucoma, $\chi^2_{(5)} = 120.9; P < 0.0001$. Pairwise differences between SMAS and each other slab method are shown in Figure 6. All other methods identified smaller proportions of abnormal superpixels compared with SMAS (all $P < 0.0001$). The smallest difference in proportion of abnormal superpixels to SMAS was found for the Ashimatey slab (–0.051; 95% CI, –0.063 to –0.039; $P < 0.0001$), whereas the deep slab showed the largest difference (–0.086; 95% CI, –0.098 to –0.074; $P < 0.0001$). Among alternative methods, the Ashimatey slab performed significantly better than the Hood, AllVis, and deep slabs by 0.03 (95% CI, 0.018–0.042), 0.022 (95% CI, 0.01–0.034), and 0.035 (95% CI, 0.023–0.047), respectively (all $P < 0.05$), whereas the BestVis slab outperformed the deep slab by 0.023 (95% CI, 0.011–0.035; $P = 0.005$).

The median distance of abnormal superpixels from the ONH differed significantly between different slab methods, $\chi^2_{(5)} = 50.0; P < 0.0001$. As shown in Figure 7, the median distances from the ONH of abnormal superpixels using the Ashimatey and Best Visibility slabs were greater than those for SMAS and the deep slab. The median distance of abnormal superpixels from the ONH for the Ashimatey and Best visibility slabs were significantly greater than for SMAS (differences, 3.5 superpixels [$P = 0.003$] and 4.6 superpixels [$P < 0.0001$], respectively). The median distance of abnormal superpixels found with the deep slab was closer to the ONH than all other slab methods (all $P < 0.05$) apart from SMAS, for which distances were smaller but statistically similar (deep - SMAS, –1.6; $P = 0.47$).

Table. Demographics of Included Participants

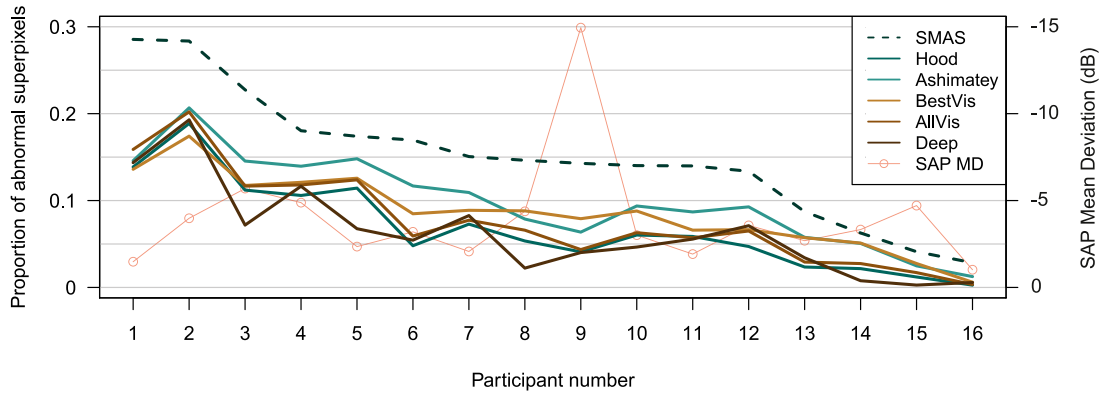|	Control Group	Glaucoma Group
n	19	16
Age (years)	68 (6)	70 (8.25)
Caucasian/other ethnicity	18/1	16/0
Male/female	8/11	8/8
Standard automated perimetry Mean Deviation (dB)	0.8 (1.4)	−3.3 (2.2)
Average cpRNFL thickness (µm)	98 (11.5)	68 (14)
Axial length (mm)	23.26 (0.50)	24.12 (0.96)

cpRNFL, circumpapillary retinal nerve fiber layer.
Continuous data are reported as median and (interquartile range).

Figure 5. Proportion of abnormal superpixels identified by different slab methods in all participants with glaucoma. Visual field mean deviation is also shown. Images from participants 1, 6, and 8 were partially censored owing to artefacts.
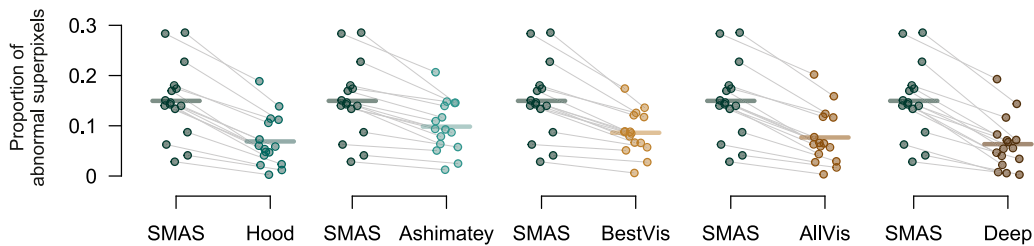
Figure 6. Differences of proportion of abnormal superpixels between each single slab method and SMAS. Different methods are color coded as per Figure 5. Grey lines link data from individual images, horizontal lines indicate means. All differences $P < 0.0001$.
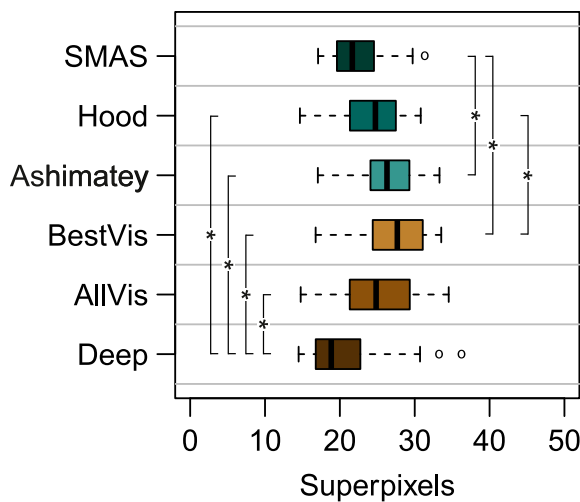
Figure 7. Boxplot showing the median distance from the center of the ONH of abnormal superpixels in eyes with glaucoma by slab method. Data are color coded according to different slab methods as per previous figures. Slab methods showing statistically significant differences ($P < 0.05$) of defect distance from the ONH are flagged with (*).

We computed the rate of abnormal superpixels in controls as a surrogate measure of the false-positive rate. None of the slab methods showed superpixels with intensity below the first percentile of control eyes,

which was the cut-off used to define defects in the glaucoma group.

Scatterplots showing the relationship between the mean and standardized differences in proportion of abnormal superpixels between SMAS and each of the other slab methods are shown in Figure 8. The Hood, AllVis, and deep slabs showed a negative relationship between differences in detection of reflectance abnormalities and average reflectance loss (slopes $P < 0.01$; R^2 0.58–0.66). This finding suggests that the benefit of SMAS over these alternatives is greater for earlier defects. The same was not true for the Ashimatey and BestVis slabs, whose performance compared with SMAS was relatively consistent across the range of reflectance loss in this sample.

Discussion

There are many possible ways to construct en face slabs from OCT images, and there is currently limited evidence on optimal slab construction methods for detection of glaucomatous defects. In addition to slab construction, there is also a lack of strategies
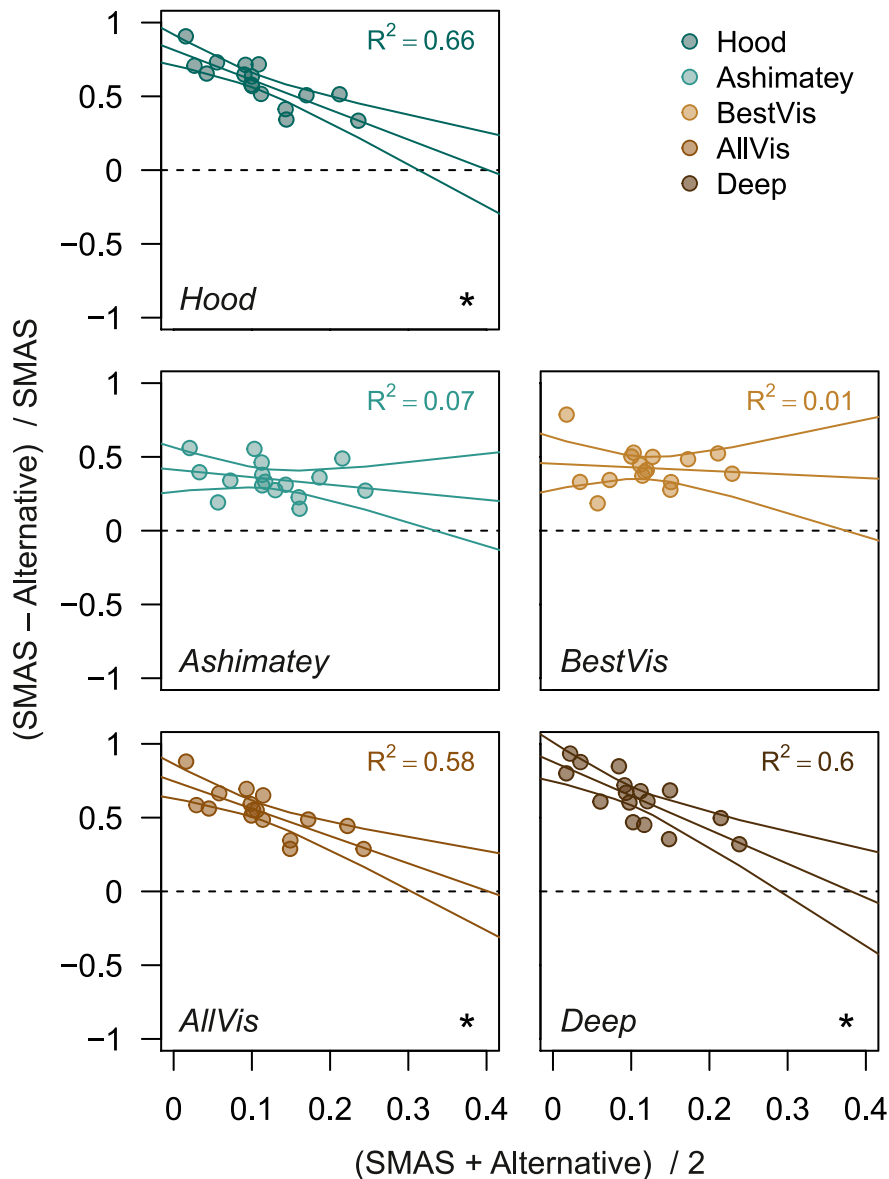
Figure 8. Bland-Altman–like scatter plots assessing the relationship between the mean and standardized difference of proportion of abnormal superpixels between SMAS and alternative slab methods. Standardized difference was calculated as the difference between proportion of abnormal superpixels of SMAS and the alternative slab method, divided by the proportion of abnormal superpixels of SMAS. The best linear fit to the data (including 95% CI) and corresponding R^2 are also reported. Cases where the linear regression line presented a slope significantly different from 0 ($P < 0.05$) are flagged with (*). Different slab methods are color-coded as per previous figures.

for automated, objective definition of defects, which should ideally account for anatomic variation between individuals. Such strategies may facilitate the consistent evaluation of reflectance loss, as well as the use of en face findings for seeding other investigations, such as custom perimetry.[14,18,40,41] In this study, we introduced SMAS, a novel approach for the construction of en face slab images and the automated, objective detection of glaucomatous defects in the en face images. SMAS improves on existing methods in several

ways, such as (i) examination of all depths that contain visible RNFBs in healthy eyes; (ii) greater consideration of the varying configuration of RNFBs across the retina; (iii) greater consideration of interindividual anatomic variability; and (iv) consideration of a wider area of the retina to include all regions containing visible RNFBs in healthy eyes, except the temporal raphe.

Compared with other methods, SMAS detected a greater proportion of abnormal superpixels in eyes

with glaucoma (Fig. 6). Although SMAS was the only method to combine information from multiple separate depth slabs, our analysis of surrogate false positive rate yielded no abnormal superpixels in control eyes at the cutoff used to identify defects in eyes with glaucoma, the same as the other slab methods. This finding suggests that SMAS detects more defects in eyes with glaucoma without increasing the number of false positives in healthy eyes.

The increased detection of abnormal superpixels in eyes with glaucoma with SMAS is attributable to the consideration of multiple slabs through the range of depths containing visible RNFBs in healthy eyes, because the greater retinal area and novel treatment of anatomic variability was applied to all slab methods in this study. The adjustments for anatomic variability were applied to all slab methods, even though individual anatomy was not adjusted for in the same way in the original applications of these methods.[5,23,24] This adjustment for individual anatomy could be expected to decrease the variability of measurements and increase the retinal area available for assessment.[24]

SMAS aimed to evaluate changes of reflectivity at all retinal depths and regions that contain visible RNFBs in healthy eyes. The assessment of the full range of depths accounts for the increased proportion of defects compared with the Ashimatey and BestVis slabs, which both assess relatively thin slabs restricted to the first 20 to 50 μm below the ILM. Accordingly, these two approaches might be as effective as SMAS at detecting reflectance changes in retinal regions with physiologically thin RNFL where they capture the full thickness. Figures 3 and 4 exemplify this. In Figure 3, an inferior arcuate defect is visible starting from 39 μm below the ILM. The BestVis and Ashimatey slabs completely or partially failed to detect this defect, although the defect's origin at the ONH was instead well-depicted in the deep slab (Fig. 4). The median distances of abnormal superpixels from the ONH conform with this interpretation; the BestVis and Ashimatey slabs on average detected abnormal superpixels further away from the ONH, where the RNFL is thinner and RNFBs are present at a smaller range of depths below the ILM.[19,27] Conversely, the deep slab examined depths that only contain RNFBs in the nasal retina and unsurprisingly found defects significantly closer to the ONH by 5.1 and 6.2 superpixels, respectively (both $P < 0.0001$).

In SMAS, the depth-averaging of only 4 pixels (approximately 16 μm) per individual slab minimizes contamination of slab images by deeper retinal layers that do not contain RNFBs, even in the temporal region where the RNFL is thinnest.[19] The

greater depth-averaging of other methods (e.g., the Hood and AllVis slabs) may be more prone to between-individual variability leading to more variable normative reflectivity data, ultimately impacting the ability to detect glaucomatous changes. This hypothesis is supported by the greater identification of defects by SMAS compared with the AllVis slab, which assessed approximately the same retinal depths but with depth-averaging across the whole depth assessed.

To our knowledge, this study is the first attempt to quantify the ability of different en face slab construction methods to detect changes of reflectivity owing to glaucoma. In the few previous studies, authors have usually based their slab construction method on pilot testing only.[5,23–25] Direct comparison of our results with previous work is complicated by differing study aims, methods, OCT devices, retinal area examined, and populations. Further, most studies performed subjective evaluation of reflectance abnormalities[5,23,25] and, to our knowledge, the only previous analysis including an objective extraction of glaucoma en face defects is the work of Ashimatey et al.[24] Hood et al considered a smaller region centered on the ONH,[5] whereas other investigators mainly focused on the macula.[23,25] Different target regions would result in different configurations of the RNFBs in the area tested, justifying the selection of different parameters. Notwithstanding the difficulties in direct comparison, our results are broadly in line with those of previous studies. In our own previous work, we showed that RNFB configuration varies across the retina, suggesting that slab parameters should be adjusted to detect defects consistently across the retina.[19] This finding was confirmed in this study, showing that slab methods do affect the capability to identify defects. Ashimatey et al.[24] noted that the inability of their slab method to identify all reflectance losses and the requirement to extend the analysis further below 52 μm to retrieve all defects. Further work from the same lab is the only previous attempt to analyze the effect of different slab construction parameters on the detection of glaucoma defects.[42] That study considered the average reflectance of small circular regions (30 pixels diameter) placed around the ONH with different ranges of depths combined together in several slabs. The greatest ability to detect glaucoma was achieved by averaging reflectivity from 36 to 60 μm below the ILM, as compared with slabs of 0 to 52 μm, 24 to 52 μm, and 24 to 36 μm.[42] These results confirm the importance of considering greater depths with present RNFBs to retrieve glaucoma defects. However, the inclusion of greater depths should not be achieved by averaging across large depths of retina, but rather with alternative

approaches able to preserve consistency of slab composition.

This study has limitations. Although we included processing strategies to adjust for uneven illumination of scans from different retinal locations and computed attenuation coefficients to minimize the impact of artefacts, the final images were still affected by these issues. Indeed, some eyes had to be excluded owing to a substantial impact of artefacts, from either activated glial artefacts, uneven illumination and/or low quality B-scans. More sophisticated image processing and/or improved image capture may further decrease the impact of such artefacts in the future. The consideration of the varying incident light beam angle at the OCT image acquisition stage would also likely improve en face OCT analyses, especially when imaging the wider retina.[37,43,44] Furthermore, the small sample size did not allow for an exploration of the impact of different parameters on the observed reflectance, such as age, eye laterality or ethnicity.[28,31,45] Larger studies could allow the development of normative data adjusted for covariates with clinically significant impact on reflectivity, ultimately leading to further refinement of the slab extraction method.

An additional constraint on our study design is the lack of an appropriate reference standard for identifying whether superpixels flagged as defective are flagged correctly or not. Accordingly, the performance of each slab method could not be evaluated with conventional indices of classification accuracy, and we focused on the proportion of abnormal superpixels and a surrogate measure of the false-positive rate. Last, we assessed the performance of SMAS in a sample with established glaucoma, while a key goal of en face imaging is early glaucoma detection, when conventional OCT metrics have been showed to be imperfect.[46–48] As such, further evaluation in glaucoma groups with only the earliest signs of glaucoma would be useful. Nonetheless, we speculate that to detect the earliest changes, the examination of greater depths below the ILM would become even more pertinent than in our sample as more subtle defects may be more likely to be found deeper in the RNFL, possibly making SMAS more advantageous over its alternatives in earlier cases of glaucoma as supported by Figure 8.

In conclusion, we developed and presented a novel method for the construction and objective analysis of OCT en face slab images. The method considers all depths and regions containing visible RNFBs in healthy eyes, with the exception of the temporal raphe, as well as the individual anatomy of the eye. With this method, we are able to automatically and objectively detect glaucomatous changes of RNFB reflectance. In our glaucoma sample, this method outperformed other available approaches in detecting defects. Further assessment of this technique is warranted.

Acknowledgments

References

1. Chen TC, Hoguet A, Junk AK, et al. Spectral-domain OCT: helping the clinician diagnose glaucoma: a report by the American Academy of Ophthalmology. *Ophthalmology*. 2018;125(11):1817–1827.
2. Hood DC. Improving our understanding, and detection, of glaucomatous damage: an approach based upon optical coherence tomography (OCT). *Prog Retin Eye Res*. 2017;57:46–75.
3. Dabasia PL, Edgar DF, Garway-Heath DF, Lawrenson JG. A survey of current and anticipated use of standard and specialist equipment by UK optometrists. *Ophthalmic Physiol Opt*. 2014;34(5):592–613.
4. Tatham AJ, Medeiros FA, Zangwill LM, Weinreb RN. Strategies to improve early diagnosis in glaucoma. *Prog Brain Res*. 2015;221:103–133.
5. Hood DC, Fortune B, Mavrommatis MA, et al. details of glaucomatous damage are better seen on OCT en face images than on OCT retinal nerve fiber layer thickness maps. *Invest Ophthalmol Vis Sci*. 2015;56(11):6208–6216.
6. 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(1):19–28.
7. Tuulonen A, Airaksinen PJ. Initial glaucomatous optic disk and retinal nerve fiber layer abnormalities and their progression. *Am J Ophthalmol*. 1991;111(4):485–490.
8. Jung JH, Park JH, Yoo C, Kim YY. Localized retinal nerve fiber layer defects in red-free photographs

translational vision science & technology

versus en face structural optical coherence tomography images. *J Glaucoma*. 2018;27(3):269–274.

9. Ji MJ, Park JH, Yoo C, Kim YY. Comparison of the progression of localized retinal nerve fiber layer defects in red-free fundus photograph, en face structural image, and OCT angiography image. *J Glaucoma*. 2020;29(8):698–703.

10. Lim AB, Park J-H, Jung JH, Yoo C, Kim YY. Characteristics of diffuse retinal nerve fiber layer defects in red-free photographs as observed in optical coherence tomography en face images. *BMC Ophthalmol*. 2020;20(1):16.

11. Huang XR, Zhou Y, Kong W, Knighton RW. Reflectance decreases before thickness changes in the retinal nerve fiber layer in glaucomatous retinas. *Invest Ophthalmol Vis Sci*. 2011;52(9):6737–6742.

12. Fortune B, Burgoyne CF, Cull G, Reynaud J, Wang L. Onset and progression of peripapillary retinal nerve fiber layer (RNFL) retardance changes occur earlier than RNFL thickness changes in experimental glaucoma. *Invest Ophthalmol Vis Sci*. 2013;54(8):5653–5661.

13. Cheloni R, Dewsbery SD, Denniss J. A simple subjective evaluation of enface OCT reflectance images distinguishes glaucoma from healthy eyes. *Transl Vis Sci Technol*. 2021;101:31.

14. Alluwimi MS, Swanson WH, Malinovsky VE, King BJ. Customizing perimetric locations based on en face images of retinal nerve fiber bundles with glaucomatous damage. *Transl Vis Sci Technol*. 2018;7(2):5.

15. Denniss J, McKendrick AM, Turpin A. Towards patient-tailored perimetry: automated perimetry can be improved by seeding procedures with patient-specific structural information. *Transl Vis Sci Technol*. 2013;2(4):3.

16. Ganeshrao SB, McKendrick AM, Denniss J, Turpin A. A perimetric test procedure that uses structural information. *Optom Vis Sci*. 2015;92(1):70–82.

17. Montesano G, Rossetti LM, Allegrini D, Romano MR, Crabb DP. Improving visual field examination of the macula using structural information. *Transl Vis Sci Technol*. 2018;7(6):36.

18. Denniss J, Turpin A, McKendrick AM. Relating optical coherence tomography to visual fields in glaucoma: structure-function mapping, limitations and future applications. *Clin Exp Optom*. 2018;102(3):291–299.

19. Cheloni R, Denniss J. Depth-resolved variations in visibility of retinal nerve fibre bundles across the retina in enface OCT images of healthy eyes. *Ophthalmic Physiol Opt*. 2021;41(1):179–191.

20. Leitgeb RA. En face optical coherence tomography: a technology review [invited]. *Biomed Opt Express*. 2019;10(5):2177–2201.

21. Huang XR, Knighton RW, Cavuoto LN. Microtubule contribution to the reflectance of the retinal nerve fiber layer. *Invest Ophthalmol Vis Sci*. 2006;47(12):5363–5367.

22. Huang XR, Knighton RW, Spector YZ, Feuer WJ. Cytoskeletal alteration and change of retinal nerve fiber layer birefringence in hypertensive retina. *Curr Eye Res*. 2017;42(6):936–947.

23. Sakamoto M, Mori S, Ueda K, et al. En face slab images visualize nerve fibers with residual visual sensitivity in significantly thinned macular areas of advanced glaucomatous eyes. *Invest Ophthalmol Vis Sci*. 2019;60(8):2811–2821.

24. Ashimatey BS, King BJ, Burns SA, Swanson WH. Evaluating glaucomatous abnormality in peripapillary optical coherence tomography enface visualisation of the retinal nerve fibre layer reflectance. *Ophthalmic Physiol Opt*. 2018;38(4):376–388.

25. Iikawa R, Togano T, Sakaue Y, et al. Estimation of the central 10-degree visual field using en-face images obtained by optical coherence tomography. *PloS One*. 2020;15(3):e0229867.

26. Miura N, Omodaka K, Kimura K, et al. Evaluation of retinal nerve fiber layer defect using wide-field en-face swept-source OCT images by applying the inner limiting membrane flattening. *PloS One*. 2017;12(10):e0185573.

27. Varma R, Skaf M, Barron E. Retinal nerve fiber layer thickness in normal human eyes. *Ophthalmology*. 1996;103(12):2114–2119.

28. Mauschitz MM, Bonnemaijer PWM, Diers K, et al. Systemic and ocular determinants of peripapillary retinal nerve fiber layer thickness measurements in the European Eye Epidemiology (E3) population. *Ophthalmology*. 2018;125(10):1526–1536.

29. Chylack LT, Wolfe JK, Singer DM, et al. The Lens Opacities Classification System III. The Longitudinal Study of Cataract Study Group. *Arch Ophthalmol*. 1993;111(6):831–836.

30. Vermeer KA, Mo J, Weda JJ, Lemij HG, de Boer JF. Depth-resolved model-based reconstruction of attenuation coefficients in optical coherence tomography. *Biomed Opt Express*. 2013;5(1):322–337.

31. Thepass G, Lemij HG, Vermeer KA. Attenuation coefficients from SD-OCT data: structural information beyond morphology on RNFL integrity in glaucoma. *J Glaucoma*. 2017;26(11):1001–1009.

translational vision science & technology

32. Xiong Y, Pulli K. Color matching for high-quality panoramic images on mobile phones. *IEEE Trans Consumer Electronics*. 2010;56(4):2592–2600.

33. Bedggood P, Nguyen B, Lakkis G, Turpin A, McKendrick AM. Orientation of the temporal nerve fiber raphe in healthy and in glaucomatous eyes. *Invest Ophthalmol Vis Sci*. 2017;58(10):4211–4217.

34. Bedggood P, Tanabe F, McKendrick AM, Turpin A. Automatic identification of the temporal retinal nerve fiber raphe from macular cube data. *Biomed Opt Express*. 2016;7(10):4043–4053.

35. Ashimatey BS, King BJ, Swanson WH. Retinal putative glial alterations: implication for glaucoma care. *Ophthalmic Physiol Opt*. 2018;38(1):56–65.

36. Scott DW. Kernel density estimators. In: *Multivariate density estimation*. New York: Wiley; 1992. pp. 125–193.

37. Huang XR, Knighton RW, Feuer WJ, Qiao J. Retinal nerve fiber layer reflectometry must consider directional reflectance. *Biomed Opt Express*. 2016;7(1):22–33.

38. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Software*. 2014;67(1):1–48.

39. Green P, MacLeod CJ. SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol Evol*. 2016;7(4):493–498.

40. Chong LX, McKendrick AM, Ganeshrao SB, Customized Turpin A., automated stimulus location choice for assessment of visual field defects. *Invest Ophthalmol Vis Sci*. 2014;55(5):3265–3274.

41. Alluwimi MS, Swanson WH, Malinovsky VE, King BJ. A basis for customising perimetric locations within the macula in glaucoma. *Ophthalmic Physiol Opt*. 2018;38(2):164–173.

42. Cheung H, Swanson WH. Identifying retinal nerve fiber layer defects using attenuation coefficients. *Invest Ophthalmol Vis Sci*. 2019;60(9):5608.

43. Tan O, Liu L, You Q, Wang J, Jia Y, Huang D. Focal loss analysis of nerve fiber layer reflectance for glaucoma diagnosis. *Invest Ophthalmol Vis Sci*. 2020;61(7):5194.

44. Knighton RW, Huang XR. Directional and spectral reflectance of the rat retinal nerve fiber layer. *Invest Ophthalmol Vis Sci*. 1999;40(3):639–647.

45. Wagner FM, Hoffmann EM, Nickels S, et al. Peripapillary retinal nerve fiber layer profile in relation to refractive error and axial length: results from the Gutenberg Health Study. *Transl Vis Sci Technol*. 2020;9(9):35.

46. Michelessi M, Li T, Miele A, Azuara-Blanco A, Qureshi R, Virgili G. Accuracy of optical coherence tomography for diagnosing glaucoma: an overview of systematic reviews. *Br J Ophthalmol*. 2020;105(4):490–495.

47. Virgili G, Michelessi M, Cook J, et al. Diagnostic accuracy of optical coherence tomography for diagnosing glaucoma: secondary analyses of the GATE study. *Br J Ophthalmol*. 2018;102(5):604–610.

48. Stagg BC, Medeiros FA. A comparison of OCT parameters in identifying glaucoma damage in eyes suspected of having glaucoma. *Ophthalmol Glaucoma*. 2020;3(2):90–96.