APPLICATION OF NAÏVE BAYESIAN ARTIFICIAL INTELLIGENCE TO REFERRAL REFINEMENT OF CHRONIC OPEN ANGLE GLAUCOMA

JOHN CHARLES GURNEY

Doctor of Optometry

ASTON UNIVERSITY

October 2016

“Learn from yesterday, live for today, hope for tomorrow. The important thing is not to stop questioning”

Albert Einstein(1879-1955)

Aston University

Application of Naïve Bayesian Artificial Intelligence to Referral Refinement of Chronic Open Angle Glaucoma

John Charles Gurney

Doctor of Optometry October 2016

Abstract

The purpose of this study was to determine whether naïve Bayesian artificial intelligence could accurately predict clinical decisions made during the referral refinement of Chronic open angle glaucoma (COAG) by three specialist independent prescribing optometrists using the highly structured standard operating procedure (SOP) adopted by the Community Ophthalmology Team (COT) of the West Kent Clinical Commissioning Group (CCG). The effectiveness of the COT, in terms of reducing false positive referrals and costs to the National Health Service (NHS), was also explored. This was the first study of its kind.

Treating the study as a clinical audit allowed collection of unconsented fully anonymised data from the worst affected eyes or right eyes of 1006 cases referred into the COT. Each case was classified according to race, sex, age, family history of COAG, reason for referral, intraocular pressure and its inter-ocular asymmetry (Goldmann Applanation Tonometry), several optic nerve head dimensions (vertical size, cup disc ratio and its inter-ocular asymmetry; dilated stereoscopic slit lamp biomicroscopy with Volk lens), central corneal thickness (ultrasound pachymetry) and the severity of any visual field defects (Humphrey Visual Field Analyser, SITA FAST 24-2 testing strategy, Hodapp-Parrish-Anderson classification). Grouping of data into multiple cut-off points was informed by previous research and National Institute for Health and Care Excellence (NICE) guidelines.

Preliminary analyses showed that most cases (79%) were discharged, 7% were followed up and 14% were referred to the NHS hospital eye service. The high discharge rate led to NHS cost savings of over £50 per case. Previous reports of increased intraocular pressure with central corneal thickness and increased cup disc ratios with cup disc size were also confirmed.

Despite a high degree of inter-dependency between clinical tests, which violated the key assumption of naïve Bayesian analyses, the scheme learned rapidly and its weighted accuracy, based on randomised stratified tenfold cross-validation, was high (95%, 2.0% SD). However, false discharge (3.4%, 1.6% SD) and referral rates (3.1%, 1.5% SD) were considered unsafe. Making the analysis cost sensitive led to an 80 fold increase in COT follow-ups that would have reduced cost effectivity. The transferability of likelihood ratios was explored along with their use, compared to Chi-square, to rank clinical tests and explore redundancy in the SOP adopted by the COT.

In summary, high discharge rates were consistent with the level of false positive referrals for suspected COAG reported in the literature and reduced NHS costs. Although use of a structured SOP led to high accuracy, naïve Bayesian artificial intelligence could not safely predict the decisions of COT optometrists as it caused too many false discharges and referrals. More sophisticated forms of machine learning need to be explored.

Dedication

To Claire and our three beautiful boys Josh, Max and Finn.

Without your love, support and belief in me, none of this would have been possible.

Thank You.

Acknowledgments

Firstly I would like to note my sincere thanks and gratitude to my supervisor Dr Mark Dunne for his knowledge, encouragement, enthusiasm and kindness which has made this research not only fulfilling but hugely enjoyable. I hope to work with Mark in the future on more research projects for the benefit of our patients and the optometric and ophthalmology communities.

Many thanks also to Dr Deacon Harle and Mr Niall O'Kane for taking time out of their busy practices and clinics to help with data collection for this project.

Thanks to Prof Ejaz Ansari for continued support and mentoring of the Community Ophthalmology Team allowing optometrists to enhance their clinical skills and expertise whilst facilitating further research.

Thanks to Prof Ian Nabney and Dr Alexis Boukouvalas for their guidance on naïve Bayesian analysis and to Dr Richard Armstrong for his help on statistical matters.

And finally to Claire and our boys for all your love, support and patience in allowing me to pursue my dreams.

Contents

Chapter 3: Preliminary Analyses

Chapter 4: Naïve Bayesian artificial intelligence

Chapter 5: Study findings and limitations with recommendations for further work

List of Figures

List of Tables

Glossary

Abbreviation	Meaning
AEL	Aston Eyetech Limited
ANOVA	Analysis of variance
AOP	Association of Optometrists
AUC	Area under the Receiver Operating Curve
Bayes	Naïve Bayes' theorem
CCG	Clinical Commissioning Group
CCT	Central Corneal Thickness
CMP	Clinical Management Plan
COAG	Chronic Open Angle Glaucoma
COO	College of Optometrists
COT	Community Ophthalmology Team
db	Decibel (unit adopted for VF assessment)
Dip Glauc	Diploma in Glaucoma
Dip OC	Diploma in Ocular Conditions (now defunct)
Dis	Discharge (COT management decision)
FDR	First discharge rates (patient discharged at first visit no disease present)
Fup	Follow-up for suspected POAG (COT management decision)
GAT	Goldmann Applanation Tonometry
GOC	General Optical Council
GON	Glaucomatous Optic Neuropathy
GP	General Practitioner
GRR	Glaucoma Referral Refinement
HAP	Hodapp-Anderson-Parrish grading system (part of VF assessment)
HES	Hospital Eye Service
Higher Cert Glauc	Professional Higher Certificate in Glaucoma
IOP	Intraocular Pressure
LR	Likelihood Ratio
MCC	Matthew's correlation coefficient
MFF	Market Forces Factor
N / LTG	Normal or Low Tension Glaucoma
NCT	Non-Contact Tonometry
NHS	National Health Service
NHSCB	NHS commissioning board
NICE	National Institute for Health and Care Excellence
NPV	Negative Predictive Value
OHT	Ocular Hypertension
OHTS	Ocular Hypertension Treatment Study
ONHA	Optic Nerve Head Assessment
OPwSI	Optometrist with Special Interest in Ophthalmology
OSI	Optometrists with a special interest in glaucoma
POAG	Primary Open Angle Glaucoma
PPV	Positive Predictive Value
Prof Cert Glauc	Professional Certificate in Glaucoma
Refer	Refer to HES for suspected POAG (COT management decision)
RTM	Regression to the Mean
SARI	System Analytics Research Group
SI	Sight Impaired
SITA	Swedish interactive thresholding algorithm
SOP	Standard Operating Procedure
SSI	Severely Sight Impaired
VCDR	Vertical cup-to-disc ratio (part of the ONHA)
VCDR diff	Inter-ocular difference in VCDR (part of the ONHA)
VDS	Vertical disc size (part of the ONHA)
VF	Visual Field
VFA	Humphrey Visual Field Analyser

Chapter 1: Background

1.1 Introduction

Previous research carried out at Aston University had indicated that artificial intelligence based on naïve Bayes' theorem (referred to as Bayes from now on) could replicate diagnoses made by an experienced optometrist with a high level of accuracy as long as a suitable Standard Operating Procedure (SOP) was adopted (1). The aim of the study described in this thesis was to determine whether this was true by applying Bayes to clinical data collected by a Community Ophthalmology Team (COT) which had developed a SOP for the referral refinement of suspected Chronic Open Angle Glaucoma (COAG, also known as Primary Open Angle Glaucoma, POAG). Chapter 1 provides a brief review of Bayes in relation to ophthalmic research before describing how it was applied in the present study. A review of the literature relating to COAG follows prior to describing the current guidelines for its detection and management, issued by the National Institute for Health and Care Excellence (NICE). The review moves on to the referral refinement of COAG by optometrists across the UK before introducing the COT that collected the data for this study. Finally, the study aims are outlined together with an overview of the chapters that follow.

1.2 Bayes in ophthalmic research

The fascinating history of Bayes' theorem has been the subject of a book written by Sharon Bertsch McGrayne and entitled "The theory that would not die. How Bayes' rule cracked the Enigma code, hunted down Russian submarines & emerged triumphant from two centuries of controversy" (2). In the context of optometry, Bayes can be used to calculate the probability of an eye condition being present from its prevalence and any new evidence that arises from diagnostic tests.

Table 1.1 provides a chronological list of studies in the field of ophthalmic research, outside that of COAG, which have applied Bayesian statistical analyses.

Year	Brief description of topic studied
1978	Computer assisted diagnosis in ophthalmology (3)
1988	Post-operative refraction following cataract surgery (4)
1992	Visual outcomes in diabetics with vitreous haemorrhage (5)
1995	Assessment of retinal-neural function prior to laser treatment (6)
1998	Lifetime prevalence of uveal melanoma (7)
2003	Prediction of Snellen visual acuity (8)
2003	Effect of steroids on aqueous humour outflow (9)
2012	Effect of steroids on corneal ulcers (10)

Table 1.1 Chronological list of studies in the field of ophthalmic research, outside the field of COAG, which has made use of Bayesian statistical analyses.

At Aston University, Sagar investigated the accuracy of a diagnostic support system, based on Bayes, which was applied to making differential diagnoses in primary care optometry (1). Her study explored how accuracy was influenced by circularity (testing a diagnostic support system using the same data used to build it), comorbidity (the coexistence of eye disease in one individual), prevalence and presentation variation (natural variations in symptoms and signs that occur with any eye condition). She analysed 1422 of the optometric records collected in her practice. In a subset of her data, including 15 clinical tests and 10 eye conditions, her results indicated that Bayes could be applied with 100% accuracy as long as presentation variation was absent but fell to 94% when presentation variation existed. Presentation variation is where a disease has a non-text book presentation for example not all patients with cataract will report reduced vision or glare. This could confound the link between test findings and a specific diagnosis thus reducing accuracy (1). Surprisingly, circularity only artificially elevated accuracy by 0.5%. Sagar also trialled Chi-square filtering which removed weak associations between diagnostic tests and eye conditions but only increased accuracy by 0.4%. When her analyses were extended to 105 clinical tests and 35 eye conditions, accuracy dropped to 72% and this fall in accuracy was most noticeable when prevalence was low and both comorbidity and presentation variation was high. The largest contributor to the fall in accuracy was, however, considered to be that some eye conditions lacked strong diagnostic signs. An important feature of her study was that only positive clinical test findings were recorded.

These positive clinical test findings could also have contributed to the reduction in accuracy and so her findings led to the notion that higher levels of accuracy could be achieved if both positive and negative test findings were recorded in clinical tests that belonged to a well-developed SOP.

1.3 Bayes in glaucoma research

Bayes has been used in glaucoma research. In visual field research Bayes has been used extensively, linking structural and functional loss in glaucoma, developing visual field algorithms, software to detect glaucomatous field loss and assist clinicians determining progression in patients with glaucoma. Table 1.2 lists these studies in chronological order. However, none of these studies have considered the use of Bayes for clinical decision support in relation to the referral refinement of COAG.

Year	Brief description of the purpose of the study
1994	Test strategies for detecting colour vision defects in patients with glaucoma (11)
1997	Development of new SITA algorithms (12)
2000	Determination of patient compliance in glaucoma cases (13)
2003	Development of a rapid threshold algorithm for short wavelength perimetry (14)
2005	Investigation of visual field deterioration (15)
2007	Classification of glaucoma from fundus photographs (16)
2007	Prediction of night time intraocular pressure peaks (17)
2008	Classification of glaucoma from structural and functional measurements (18)
2009	Evaluation of quality of life and priorities in people with glaucoma (19)
2009	Linkage of retinal structure and visual function in glaucoma (20)
2011	Detection of progression from structural and functional measurements (21)
2012	Machine learning classifiers and optical coherence tomography in glaucoma (22)
2013	Detection of progression from Heidelberg Retina Tomograph images (23)
2013	Prediction of the prevalence of glaucoma in Asia (24)
2013	Detection of progressing glaucoma from retinal nerve fibre layer thickness (25)
2015	Detection of glaucomatous progression (26, 27)
2016	Effect of cataract surgery on intraocular pressure in glaucoma (27)

Table 1.2 Chronological list of glaucoma studies that have involved the use of Bayes' theorem.

1.4 Bayes in the present study

Bayes was applied in the present study. The assumption is that all diagnostic tests, carried out for the purpose of detecting COAG, are independent of each other (28). In reality this is not always the case. However, Bayes in previous studies has been shown to render high levels of accuracy even when clinical tests are inter-dependent (1, 29). The data within this study was highly structured with a sophisticated SOP so despite tests not being independent high levels of accuracy were anticipated.

1.4.1 Diagnostic matrices

Bayes can be applied to a single diagnostic test in relation to a single diagnosis or diagnostic outcome by using the diagnostic matrix shown in Figure 1.1 (30). The figure shows a screenshot of the CatMaker program adopted by the Oxford Centre for Evidence Based Medicine (http://www.cebm.net/catmaker-ebm-calculators/) at the time of the present study. CatMaker simplified the application of naïve Bayes by practitioners who may have limited knowledge of statistics. Sagar applied the formulae shown in Figure 1.1 for her research (1). Cat Maker has been included to demonstrate how Bayes probabilities may be calculated .The present study used these formulae, however Cat Maker was not used to analyse the data in this study (see section 1.4.3).

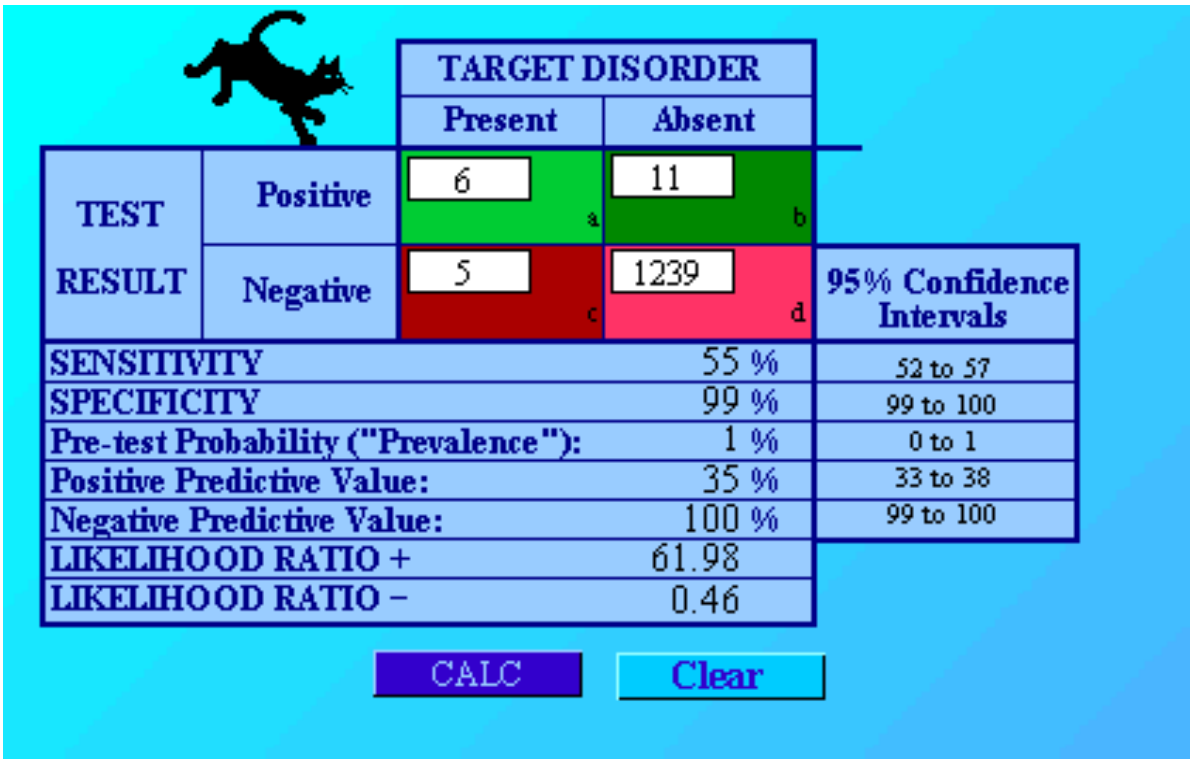


Figure 1.1. Screen shot of the CatMaker program currently adopted by the Oxford Centre of Evidence Based Medicine. Naïve Bayes' theorem is applied to a single diagnostic test in relation to a single diagnosis or diagnostic outcome. In this case, the diagnostic test is an intraocular pressure reading of >21 mm Hg and the diagnosis outcome is suspected COAG. The formulae shown were those used in the present study.

Figure 1.1 shows a 2 x 2 decision matrix made up of cells a (true positives i.e. 6 cases with raised intraocular pressure and suspected COAG), b (false positives i.e. 11 cases with raised intraocular pressure did not have suspected COAG), c (false negatives i.e. 5 cases without raised intraocular pressure but with suspected COAG) and d (true negatives i.e. 1239 cases without raised intraocular pressure or suspected COAG). Sagar's data was extracted by the author from Sagar's DOptom thesis (1). Figure 1.1 shows calculations involved in the application naïve Bayes. The first of these is the sensitivity (the probability of raised intraocular pressure if COAG is suspected) and specificity (the probability of normal intraocular pressure if COAG is not suspected). In this case the sensitivity and specificity, expressed as a percentage, is 55% and 99%, respectively.

The problem with these values is that they say more about the probability of a test result if an eye condition is present or absent than the probability of having an eye condition if there is a positive or negative test result; the latter being what is required from the application of naïve Bayes theorem. Positive and negative likelihood ratios (LR) are calculated as an intermediate step towards getting the required answer. It can be seen that likelihood ratios are derived from sensitivity and specificity and are the ratio of the probability of the diagnostic test being right over the probability of it being wrong when the test result is either positive or negative. These ratios are used to raise or lower the post-test odds of a disorder being present. It follows that if these ratios have a value of 1 then the test is of no diagnostic value at all (31-33). On the other hand, a very high positive likelihood ratio indicates a diagnostically useful test because it dramatically elevates the post-test odds of an eye condition being present when the test result is positive. A negative likelihood ratio close to zero also indicates a diagnostically useful test as it dramatically reduces the post-test odds that an eye condition is present when the test result is negative. Figure 1.1 shows positive and negative likelihood ratios of 62 and 0.46, respectively.

Note that the equation for negative likelihood ratio has been turned upside down so that it indicates the alteration to the post-test odds of suspected COAG rather than COAG not suspected. In other words, a positive test result raises the post-test odds of referral by 62 times while a negative test result reduces the post-test odds of referral by 0.46 times. The treatment of the negative likelihood ratio is very useful when multiple test outcomes are being considered (34, 35), as was the case in the present study. This is because several positive and negative likelihood ratios could be combined into a single likelihood ratio by taking their product. This would not have been possible had the negative likelihood ratios not have been turned upside down. It is worth pointing out here that Sagar (1) only used positive likelihood ratios in her study as she only recorded positive test outcomes (see section 2.1).

According to Bayes, the post-test odds of an eye condition being present or absent is calculated by multiplying the pre-test odds (which is equal to the pre-test probability / [pre-test probability – 1] and is also known as the prior odds) by the appropriate likelihood ratio (i.e. positive or negative likelihood ratio depending on the test result). The post-test probability is simply derived from the post-test odds (post-test probability = post-test odds / [1 + post-test odds]). The method of generating the post-test probability is very useful when multiple test outcomes are being considered as it is derived from the pre-test odds multiplied by the product of all likelihood ratios from all tests carried out.

Figure 1.1, however, shows an alternative way of calculating post-test probability; use of positive (PPV) and negative predictive values (NPV) (36, 37). These predictive values are equal to the post-test probability. They were not used in the present study but are included here for the sake of completion. The equation used to calculate the negative predictive value shown in Figure 1.1 was not turned upside down (as was the case for the equation used to calculate the negative likelihood ratio). Nevertheless, this is effectively achieved by subtracting the negative predictive value shown from 100. Expressed as a percentage, the predictive values show that a positive test result raises the probability of suspected COAG from 1% (the pre-test probability) to 35% while a negative test result lowers the probability of referral to 0% (100 – 100%).

The data shown in Figure 1.1 can be used to make a well-known observation about screening for rare conditions (38). The prevalence of suspected COAG was only 1%, making it relatively rare. What if referral for suspected COAG was purely based on intraocular pressure of greater than 21 mmHg which had, in this case, a sensitivity of 55% and a specificity of 99%? Figure 1 shows that 17 (cells a + b in the diagnostic matrix) referrals will have been made of which 65% (11 referrals, cell b in the diagnostic matrix) will have been false positives. This would lead to significant unnecessary burden on the Hospital Eye Service (HES) (see section 1.7.2).

1.4.2 Laplacian correction

The diagnostic matrix shown in Figure 1.1 lacked cells with zero counts. Sagar, however, noted that these frequently occurred and complicated the application of Bayes to her clinical dataset (1). The biggest problem arose if zero counts gave rise to likelihood ratios of zero. Recall from section 1.4.1 that, when using multiple tests to determine the probability of a diagnosis or diagnostic outcome, several positive and negative likelihood ratios could be combined into a single likelihood ratio by taking their product. If just one of those likelihood ratios was zero, then the combined product would also be zero regardless of whether the likelihood ratios for the other tests were very large effectively ruling out the diagnosis or diagnostic test outcome. The solution to the problem is to make a Laplacian correction (35). Usually this involves adding 1 to the counts in each cell of the diagnostic matrix which eliminates cells with zero counts together with the possibility of likelihood ratios of zero (35). The philosophical rationale for the Laplacian correction is that a likelihood ratio of zero absolutely rules out a diagnosis or diagnostic outcome when, in reality, no statistical model is perfect enough to do this. The Laplacian correction effectively remedies this at the cost of small artificial alterations to the calculated likelihood ratios. Sagar used a Laplacian correction of 0.001 (1) to ensure that this caused minimal alterations to likelihood ratios calculated for rare test outcomes or eye conditions. The same Laplacian correction was adopted in the present study.

1.4.3 Generation of Bayes learning curves

Sagar created Microsoft Excel spreadsheets to apply Bayes to multiple tests and diagnoses (1). These dramatically sped up the process of carrying out the calculations described in Figure 1.1 for the large number of diagnostic matrices relating to each of her test / diagnosis combinations (105 tests x 35 diagnoses = 3675 diagnostic matrices). The spreadsheets allowed her to calculate the accuracy with which this form of artificial intelligence could match her own diagnoses (1).

While use of spreadsheets was sufficient for her purposes they did not readily lend themselves to the generation of learning curves that show the speed with which maximum accuracy is achieved when applying Bayes. Fortunately, Aston Eyetech Limited (AEL), a spin out company of Aston University, had developed a computer program (the AEL Bayes application) that applied Sagar's computing scheme to measure accuracy (1) and could construct Bayes learning curves. The software adopted a Laplacian correction of 0.001 and was used in the present study.

The Aston Bayes application could divide any clinical dataset into 20 approximately equal cohorts. It learned from the first cohort of data.

Learning consisted of:

(a) Calculating pre-test odds for each of the diagnostic outcomes;

(b) Generating diagnostic matrices for every combination of clinical test and diagnostic outcome;

(c) Calculating the sensitivity, specificity and likelihood ratios for every diagnostic matrix.

System accuracy was then calculated by:

(a) Applying Bayes' to all remaining cases, those not used for learning, by multiplying the pre-test odds of each diagnostic outcome by the corresponding positive or negative likelihood ratio for every clinical test finding, in order to determine post-test probabilities for all diagnostic outcomes;

(b) Selecting, for each case, the diagnostic outcome with the highest post-test probability as the chosen diagnostic outcome;

(c) Comparing, for each case, the chosen diagnostic outcome with the actual one made by the practitioner and;

(d) Calculating the percentage of cases for which the chosen and actual diagnostic outcome matched.

The process was repeated on the remaining cohorts and continued until all cohorts had been used for learning.

System accuracy was then plotted as a function of the number of records used for learning. The resulting learning curve could show how many cases were needed before the graph reached an asymptote representing maximum system accuracy.

Figure 1.2 shows learning curves plotted using the AEL Bayes application. The author ran analysis on 1261 of Sagar's eye examination records (1). Sagar's original dataset included 1422 records and application of naïve Bayes theorem to this data gave rise to an accuracy of 72% (see section 1.2). Subsequent detailed analyses of the dataset by Malcolm Maciver, Edward Kundzikz, Arti Patel, Komail Ladha (undergraduates, working towards their third year dissertations on the optometry degree course), revealed that the diagnoses lacked supporting signs in 161 records. This was considered to be due to the lack of an SOP for recording clinical findings and almost certainly reduced accuracy.

Sagar was not able to generate a learning curve but Figure 1.2 shows that after removal of 161 aberrant records by the author, the AEL Bayes application rapidly learned and levelled off at an accuracy of 97% 'by record' and 98% 'by eye condition'. Accuracy 'by record' represented the percentage of times that the AEL Bayes application identified all eye conditions manifest in each record; comorbidity was common. Accuracy 'by condition' represented the percentage of times a specified eye condition was identified in every record, regardless of whether or not other comorbid eye conditions were identified.

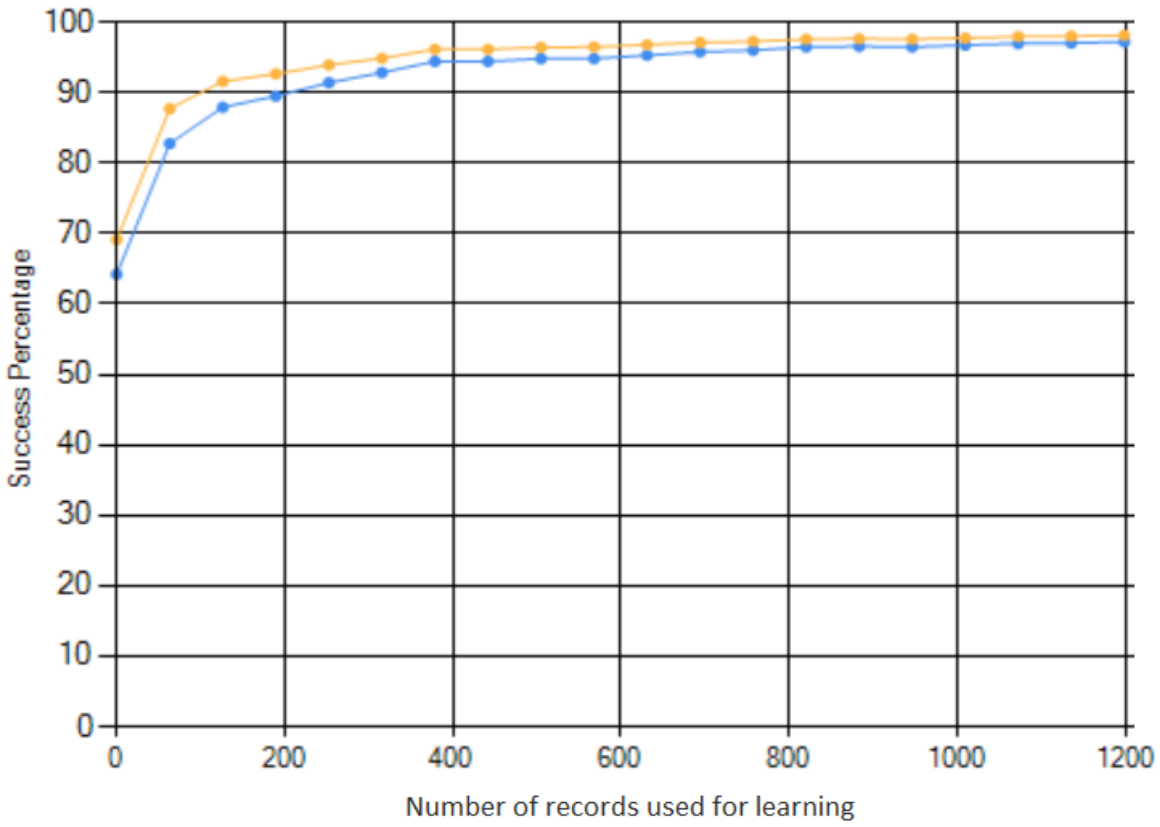


Figure1.2. Screen shot of Bayes learning curves generated by the AEL Bayes Application. The analysis was carried out on 1261 records collected by Sagar (1) which included 105 diagnostic test findings and 27 suspected eye conditions. Learning curves were analysed 'by record' (yellow line) and 'by eye condition' (blue line).

Recall that Sagar only recorded positive test results (1). A major assumption made in the author's re-analysis of her data, shown in Figure 1.2, was that the lack of a positive test result represented a negative test result. This allowed both positive and negative likelihood ratios to be used. The lack of a positive test result could actually have meant that a test had not been performed at all.

The promising levels of accuracy shown in Figure 1.2 might, therefore, be fictitious, which reinforced the need for the present study in which the speed of learning and the accuracy level achieved could be re-examined in clinical data collected using an SOP in which positive and negative test findings were recorded.

1.4.4 Use of likelihood ratios for multiple cut-off points

Figure 1.1 considered a single cut-off point of >21 mmHg for intraocular pressure as a means of detecting COAG. While receiver operator curves can be used to identify an optimum single cut-off point (39, 40), Parikh's research group recommended the use of multiple cut-off points in order to individualise clinical decisions (31). Indeed, it seemed intuitively obvious to the author that all intraocular pressure values signal some risk of POAG. Why not calculate likelihood ratios for multiple cut-off points? Application of Bayes would then involve selecting the likelihood ratio that matched an individual's intraocular pressure. Although Sagar had not used multiple cut-off points in her study, she did recommend that future research might consider this for parameters such as age, intraocular pressure and cup-to-disc ratios (1). For this reason, multiple cut-off points were adopted in the present study. The author initially considered using multiple cut-off points in the form of, for example, >21mmHg, >25mmHg, >30mmHg, and so on, for measurements of intraocular pressure. However, objections were raised against this approach by one of Aston's Bayes experts (Professor Ian Nabney, Centenary Professor of the System Analytics Research Institute, SARI). These objections were based on the fact that use of these cut-off points would further violate the assumption of Bayes, that test results were independent (see section 1.4). For example, an individual pressure reading of 32mmHg would actually fall into all three groups mentioned above. On the other hand, the expert advisor had no objection to using multiple mutually exclusive intraocular pressure categories, such as 21 to 24mmHg, 25 to 29mmHg, 30 to 35mmHg, and so on, as now a pressure of 32mmHg would only fall into one group. The author adopted this alternative form of multiple cut-off point (see chapter 3).

1.5 Glaucoma

1.5.1 Prevalence of glaucoma

Population studies had shown that the prevalence of glaucoma for all age groups ranged from 0.8% (41) to 7% (42). Table 1.3 summarises the prevalence of glaucoma in the major epidemiological studies**.**

Study	**Prevalence**
Rotterdam study (41) 1990-1993	0.8%
PVER study (43) 1999-2000	1.8%
Melbourne VIP (44) 1991-1998	1.8%
Beaver Dam study (45) 1988-1990	2.1%
Baltimore eye survey (46) 1985-1988	2.5%
Blue Mountains study (47) 1992-1994	2.8%
Kongwa Eye Project (48) 1996	3.0%
Barbados study (42) 1988-1992	7.0%

Table 1.3 Prevalence of glaucoma estimated from major epidemiological studies. The studies have been arranged in order of increasing prevalence to show the range of values found.

The global estimate of the prevalence of glaucoma for people of 40 to 80 years is 3.5% (49). Declining fertility rates and reducing mortality rates mean that there is an ever increasing number of people in this age group (50), making glaucoma a growing major public health concern.

Due to the worlds increasing population by 2020 it is expected that 79.6 million people worldwide will have glaucoma (51) and this will rise to 111.8 million by 2040 (49). Glaucoma is the second leading cause of blindness in the world (51). Nearly three quarters of these individuals will have open angle glaucoma resulting in bilateral blindness for 5.9 million people (51).

In the UK, COAG affects 2% of the population over 40 rising to 10% in those over 75 years of age (52). The number of individuals with open angle glaucoma in England is almost half a million (52). Every year there is estimated to be 11,054 new cases of POAG in the UK for people between 40-70 years of age (53), 9263 cases in England alone (53). Approximately 20% of referrals to the Hospital Eye Service (HES) in the UK are due to glaucoma (54). Individuals with glaucoma will require lifelong monitoring for signs of progression and disease control as any sight lost is permanent and cannot be restored (52). The yearly cost of monitoring these patients is estimated at £22,469,000 (52).

Glaucoma is the third most common cause responsible for severely sight impaired (SSI) registrations and second most common cause for sight impaired (SI) registrations in England and Wales (55). Figure 1.3 shows the causes of SSI registrations in the UK from April 1999 to March 2000. Figure 1.4 shows the causes of SI registrations in the UK over the same period.

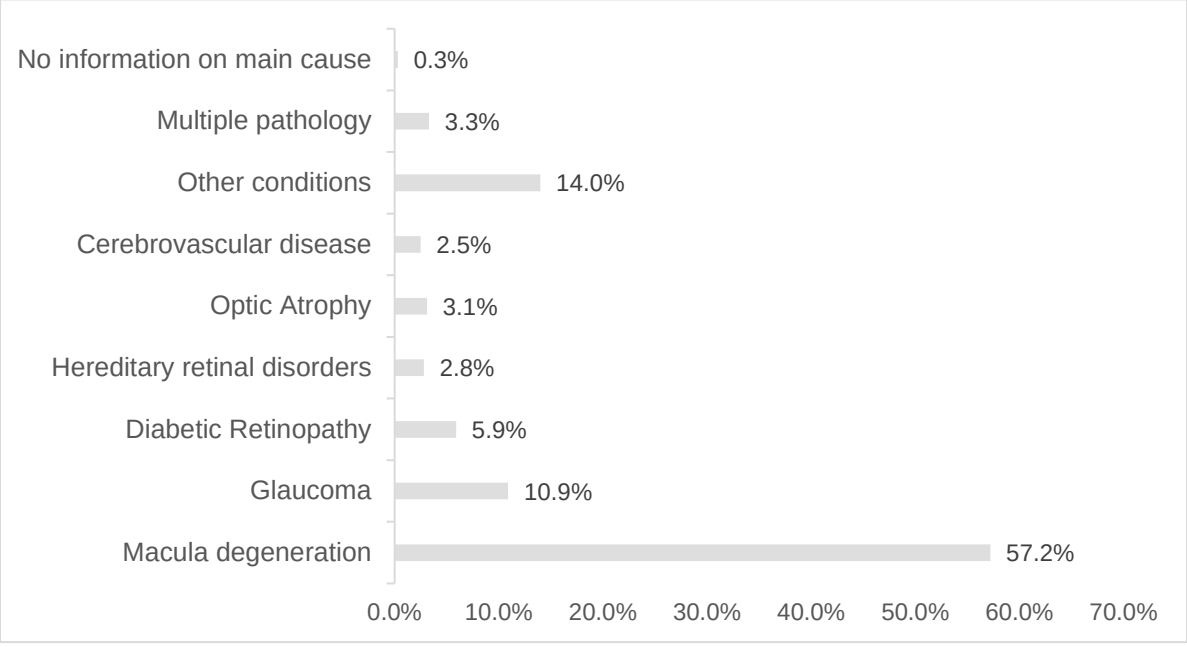


Figure 1.3 Causes of certifications for Severely Sight Impaired Registrations in England and Wales (April 1999 to March 2000). Adapted from data by Bunce and Wormald (55).

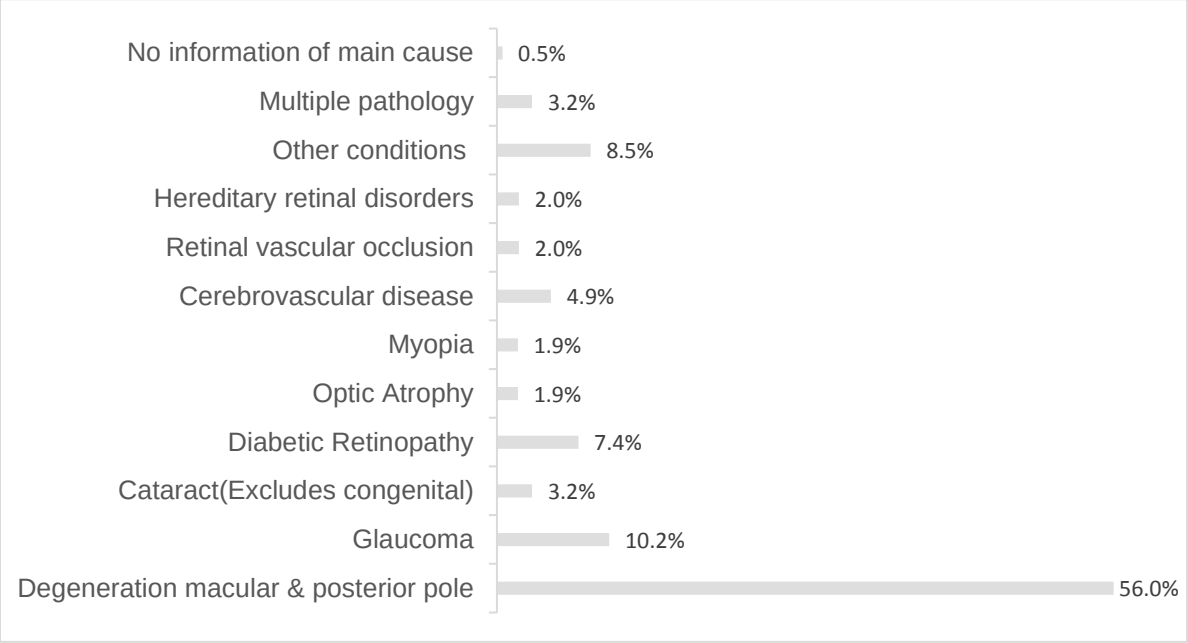


Figure 1.4 Causes of certifications for Sight Impaired Registrations in England and Wales (April 1999 to March 2000). Adapted from data by Bunce and Wormald (55).

1.5.2 Classification of glaucoma

Glaucoma is usually classified by either the aetiology or the mechanism of damage. Glaucoma may be primary when no other disease is implicated or secondary where it is due to another detectable comorbidity. Glaucoma may be further divided clinically into open and closed angle on the basis of gonioscopic examination of the drainage angles. The exact mechanism for variations in susceptibility and patterns of damage seen in glaucoma is not known although the pathogenesis is likely to be multifactorial (56).

1.5.3 Chronic Open Angle Glaucoma (COAG)

1.5.3.1 Clinical features of COAG

The present study was primarily concerned with the application of Bayes to the referral refinement of COAG. For this reason, COAG is described in more detail than other forms of glaucoma, which are included later for the sake of completion. Figure 1.5 shows a normal optic nerve head with a healthy neuroretinal rim.

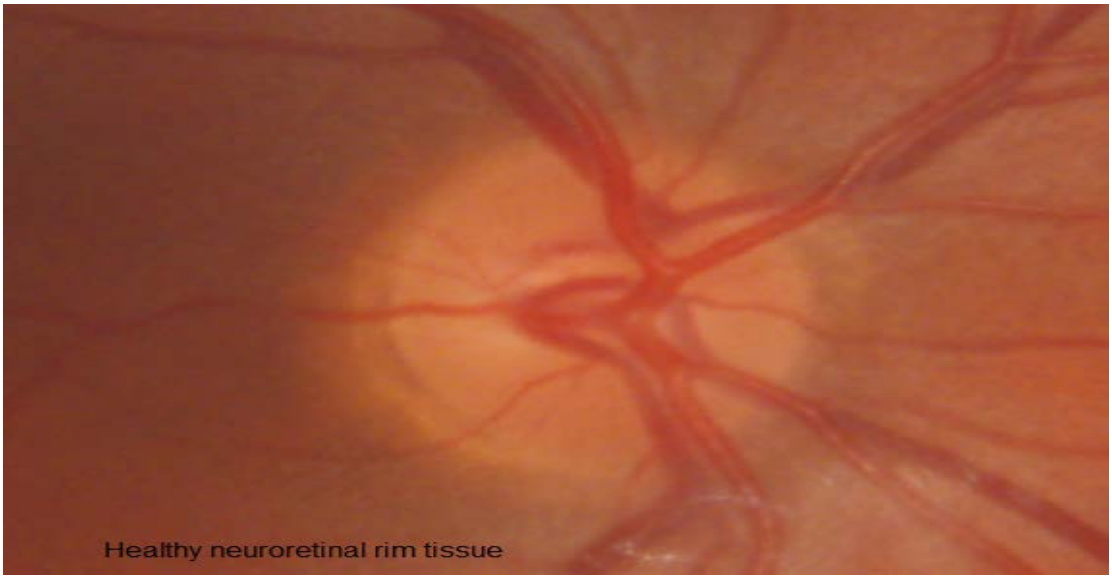


Figure 1.5 Colour photograph of an optic nerve with healthy neuroretinal rim tissue (taken by the author).

Chronic Open Angle Glaucoma is a progressive chronic optic neuropathy with characteristic changes at the optic nerve head and in the retinal nerve fibre layer which result from a loss of ganglion cells and are associated with progressive visual field loss. Figure 1.6 shows a wedge defect resulting from localised loss of the retinal nerve fibre layer in a person with COAG.

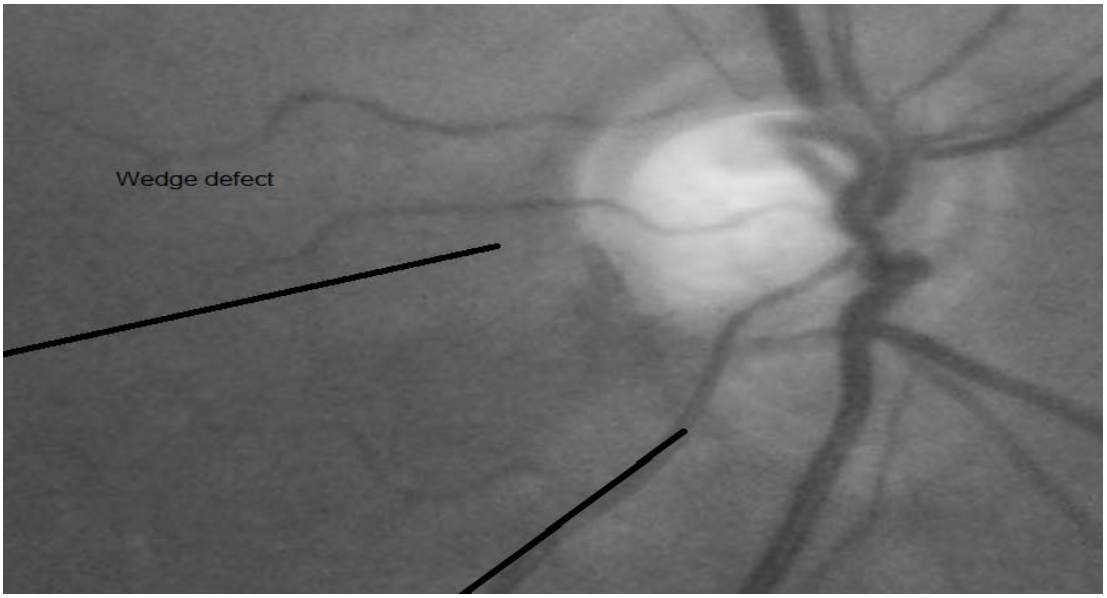


Figure 1.6 Red free photograph of a wedge defect of the Retinal Nerve Fibre layer seen in the right eye of a person with COAG (taken by the author).

Chronic Open Angle Glaucoma exhibits increased cupping of the optic nerve head which is often termed Glaucomatous Optic Neuropathy (GON). Figure 1.7 shows an optic nerve head exhibiting GON.

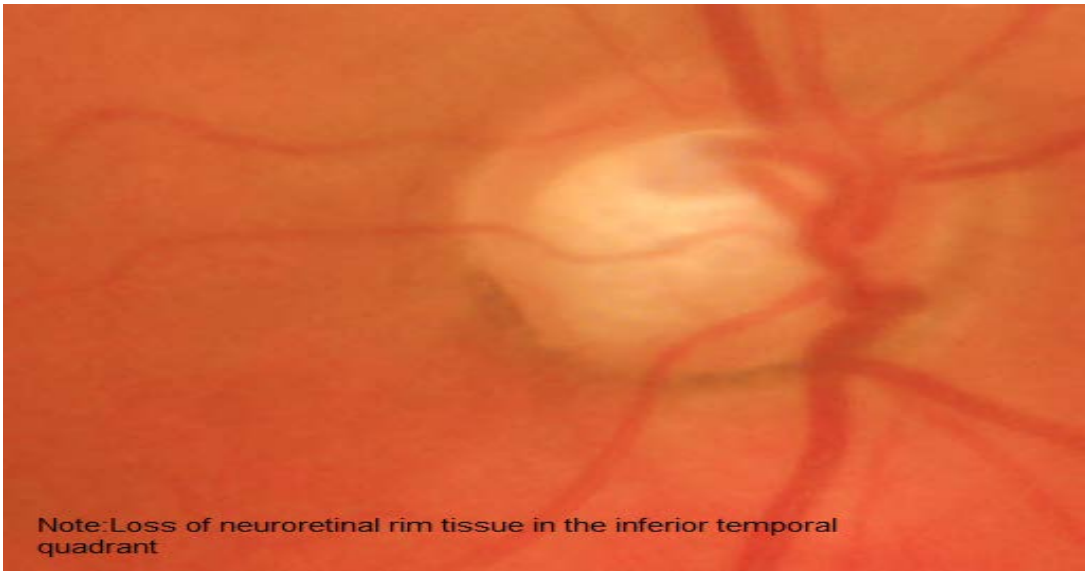


Figure 1.7 Colour photograph showing loss of neuroretinal rim tissue in the same right optic nerve head as shown in Figure 1.6 (taken by the author).

Glaucomatous damage starts with thinning of the neuroretinal rim of the optic nerve head progressing to advanced cupping which is associated with visual field defect loss. The earliest glaucomatous visual field loss is often in the paracentral area (57) and progresses to a nasal step, arcuate and advanced visual field loss. Figure 1.8 shows a superior arcuate visual field loss as a result of COAG.

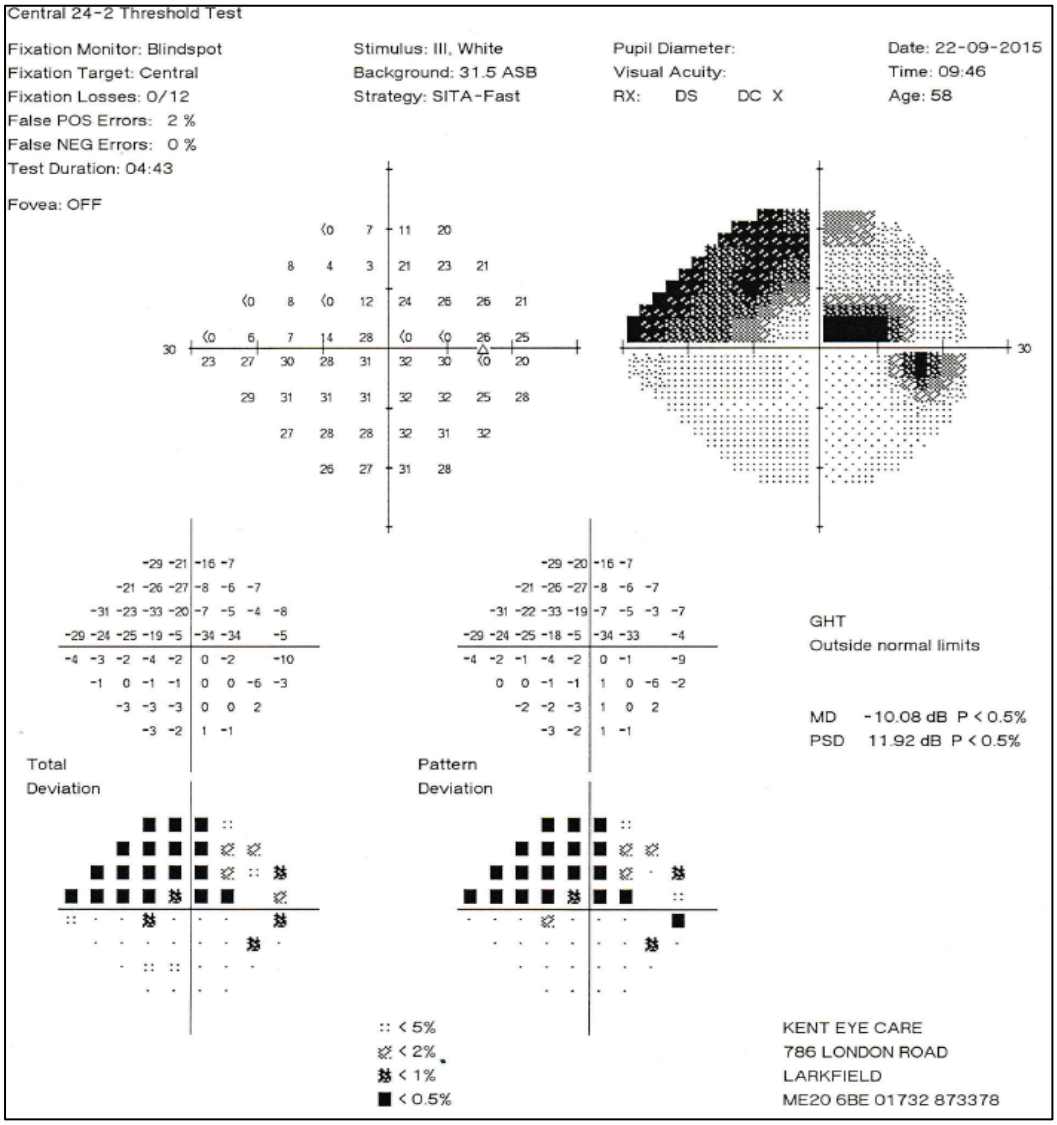


Figure 1.8. Visual field plot showing superior arcuate visual field loss in the same right eye as shown in Figures 1.6 and 1.7 (recorded by the author).

Often only a temporal island of vision remains in advanced visual field loss rendering the patient almost totally blind. Approximately 40-50% of the retinal nerve fibres need to be lost before there is an impact on visual function (58). COAG is responsible for 50% of blindness associated with glaucoma (53) and is often but not always associated with raised intraocular pressure (IOP) (41) due to resistance of aqueous outflow in the trabecular meshwork.

Primary glaucomas make up 80% of the glaucoma sufferers in the world (59). Open angle glaucoma makes up two thirds of the glaucoma population worldwide (51). In the UK it is estimated that 67% of open angle glaucoma goes undetected (53) because it remains asymptomatic until it becomes advanced.

1.5.3.2 Risk Factors for COAG

Increasing age is the single most important risk factor for COAG. Other risk factors include raised intraocular pressure (IOP), central corneal thickness and family history. These factors and their inclusion in the present study are explained further in chapter 2.

1.5.3.3 Pathophysiology of COAG

COAG is a bilateral asymmetrical condition at diagnosis and throughout its clinical course (60). It is known that retinal ganglion cell axons undergo apoptosis but the exact cause of this is not completely clear (61). Three main theories emerge from the literature; mechanical, vascular and biomechanical.

1.5.3.3.1 Mechanical Theory

The mechanical theory proposes that increased IOP is caused by obstruction and resistance of the aqueous humour outflow at the level of the trabecular meshwork. The resultant increased pressure imparts mechanical stress causing stretching of the lamina cribrosa. The ganglion cell axons pass through the lamina cribrosa so that they are damaged and lost directly because of this stress (62). Elasticity of the lamina cribrosa connective tissue reduces with age so any deformation will cause pressure on the nerve axons passing through it (63). The thickness of the lamina cribrosa increases with age which may also contribute to pressure on nerve axons (64).

1.5.3.3.2 Vascular Theory

The vascular theory relates to reduced or insufficient ocular blood flow. Reduced ocular blood flow at the optic nerve head may result from faulty autoregulation, unstable ocular perfusion, local vasospasm, and high blood viscosity; all of which lead to ischemia and hypoxia (65). Raised IOP may reduce ocular blood flow to the optic nerve by compression of the capillaries causing localised ischemic damage (66). In support of this theory, ocular blood flow tends to be reduced in glaucoma (67-69).

1.5.3.3.3 Biochemical Theory

The biochemical theory suggests that COAG results from autoimmune damage. Mechanical stress is thought to cause the release of neurotoxins such as nitric oxide from retinal glial cells. Exposure to high levels of nitric oxide may damage the axons of the retinal ganglion cells (70). Mitochondrial dysfunction is often present in glaucomatous eyes and it is thought that the resulting oxidative stress may also cause glaucoma (71). The exact mechanism by which oxidative stress causes glaucoma is not yet fully understood.

1.5.4 Normal tension glaucoma

Normal tension glaucoma (NTG) or low tension glaucoma (LTG) occurs with normal IOP; less than 21 mmHg (72). There is current disagreement in the literature on whether NTG is a separate entity or a variant of COAG (73). NTG may also be described as a condition of exclusion where the optic disc damage and subsequent field loss cannot be attributed to any other pathological condition (74). NTG was once considered quite rare but several recent studies have revealed that it may account for between 20-40% of all open angle glaucoma (75).

Patients who suffer from vascular dysregulation syndrome tend to have low blood pressure (systemic hypotension) (76). They suffer from cold extremities of the hands and feet and conditions such as Raynauds syndrome and migraines (77). These patients have a higher prevalence of NTG which seems to be related to their ocular blood flow. People with vascular dysregulation syndrome tend to have stiffer retinal blood vessels with a higher spatial irregularity so that vasodilatation is reduced, leading to reduced ocular perfusion of the optic nerve head (77, 78). The reduced perfusion to the optic nerve head is believed to be the cause of NTG.

1.5.5 Ocular hypertension

Ocular Hypertension (OHT) is defined as consistently or recurrently elevated IOP of greater than 21mmHg in the absence of optic nerve damage, visual field loss or closed aqueous drainage angles (52). In England it is estimated that one million people have OHT, 3-5 % of the people aged over 40 years (52). OHT can progress to COAG (79). The risk of conversion from OHT to COAG decreased from 9.5% to 4.4% after 5 years with ocular hypotensive medication (79).

1.6 NICE guideline CG85

The National Health Service (NHS) receives guidance from NICE that is gathered by established literature search methods that respect a hierarchy of research evidence (80-82). In 2009, NICE published guideline CG85 for the diagnosis, management and treatment of COAG and OHT (52). The sections that follow (1.6.1 to 1.6.8) cover those aspects of the NICE guidelines that inform the present study (52).

1.6.1 Goldmann applanation tonometry

Goldmann applanation tonometry (GAT) is considered by NICE to be the reference standard for measuring IOP (52). Although non-contact tonometry (NCT) was less costly, experts were of the opinion that GAT was more accurate and that its use for detecting COAG and OHT would save more money by reducing inappropriate treatment (83).

1.6.2 Central corneal thickness

Central Corneal Thickness (CCT) is a risk factor for conversion of OHT to COAG (79, 84, 85). While NICE recommended measurement of CCT, no specific method was specified (52).

1.6.3 Anterior chamber angle assessment

Despite being more costly than alternative techniques, gonioscopy was considered by NICE to be the reference standard for assessment of the anterior chamber angle (52). The clinical evidence reviewed indicated that gonioscopy revealed more detailed information than van Herick's test (86), the flashlight test (87), the scanning Peripheral Anterior Chamber Depth analyser (86) and non-contact anterior segment optical coherence tomography (88). However, NICE recommended that van Herick's test was suitable if gonioscopy was not possible due to physical or mental health concerns relating to the patient (52).

1.6.4 Visual field assessment

The SITA (Swedish interactive thresholding algorithm) 24-2 testing strategy was considered by NICE to be the reference standard for visual field assessment but no perimeter was specified (52). Literature comparing SITA 24-2 with other strategies were lacking but NICE recommended that a threshold technique, such as SITA, was used for any patient suspected of having glaucoma (52).

1.6.5 Optic nerve head assessment

Dilated stereoscopic evaluation with a slit lamp biomicroscope by a trained clinician was considered by NICE as the reference standard for optic nerve head assessment (ONHA) (52).

Slit lamp biomicroscopy was less expensive than alternative techniques which included Laser Polarimetry, optical coherence tomography and the Heidelberg Retina Tomograph (52). NICE recommended slit lamp biomicroscopy on its own as combining it with stereoscopic optic disc photography increased the cost (52).

Although NICE recognized that a stereoscopic optic disc photograph taken at the first visit was a useful means of establishing any progression of COAG or conversion of OHT to COAG at subsequent visits, the NICE guidelines were written when this technology was not commonly available in clinical practice.

1.6.6 Training for referral refinement

1.6.6.1 Requirements of NICE

The NICE guidelines stated that clinicians involved in referral refinement should:

a. Have a specialist qualification or be working under the supervision of a consultant ophthalmologist.
b. Have enough experience to perform and interpret the relevant clinical tests, such as those described in sections 1.6.1 to 1.6.5 above.

While these guidelines allowed clinicians with a specialist qualification to carry out referral refinement without the need to be supervised by a consultant ophthalmologist (as was the case for the COT optometrists that took part in the present study), the specialist qualifications were not specified.

1.6.6.2 Development of specialist qualifications for optometrists

Over the past 20 years optometrists have been increasingly involved not only in case detection of OHT and COAG but in diagnosis, monitoring and treatment of these conditions (89).

To help develop and expand the role of optometrists in this area, specialist qualifications have been developed by the College of Optometrists (COO).

In 1999, the COO glaucoma certificate A was available followed in 2004 by certificate B (90). The certificates were developed by both specialist ophthalmologists and optometrists working in the field of glaucoma.

The aim of the certificates was to allow optometrists to further develop skills, experience and understanding of glaucoma (90). Part A was designed for optometrists working in a glaucoma referral refinement scheme while Part B was designed to allow optometrists to work independently in the management of glaucoma and OHT (90). Candidates who passed part A were awarded the Diploma in Ocular Conditions (Dip Oc). Those who passed Parts A and B were awarded the Diploma in Glaucoma (Dip Glauc). Due to lack of demand, the COO now only offers Part B for those candidates still needing to complete their examinations (91).

These qualifications were replaced by three new qualifications in glaucoma: the Professional Certificate in Glaucoma (Prof Cert Glauc), the Professional Higher Certificate in Glaucoma (Higher Cert Glauc) and the new Diploma in Glaucoma (Dip Glauc). The Professional Certificate was developed for optometrists to participate in formal referral refinement and OHT/suspect COAG monitoring schemes (92). The Professional Higher Certificate prepared optometrists to participate in community or hospital-based schemes involving the diagnosis of OHT and preliminary diagnosis of COAG (92). The Diploma was designed for optometrists to participate in community or hospital-based schemes for the management of patients with established COAG (92). To obtain the Higher Certificate, 150 patients of varying degrees of clinical complexity needed to be examined in a clinical placement under an ophthalmologist mentor. For the Diploma a further 250 cases needed to be seen (90).

The higher qualifications in glaucoma were available at several UK institutions including University College London and Cardiff University. The COO provided training institutions with guidance. These providers also needed accreditation by the COO (93). The rigorous accreditation process and clinical placement ensured that optometrists awarded these higher qualifications had the necessary skills, expertise and experience in the field of glaucoma that was required by NICE.

1.6.7 Consultant ophthalmologists are responsible for the final management plan

An evaluation was carried out by NICE of several studies comparing a range of healthcare professionals in the monitoring of patients with OHT and COAG (94-98). These healthcare professionals included optometrists and ophthalmologists who were non-specialists and specialists. Agreement between optometrists and ophthalmologists varied from fair to substantial, depending upon the degree of specialisation. From this evidence, NICE concluded that the consequence of either failing to identify COAG or incorrect diagnosis may lead to irreversible blindness and visual disability. Therefore, COAG had to be diagnosed by a consultant ophthalmologist who could form the clinical management plan and reduce this risk to a minimum (52).

1.6.8 Recommended treatment plan

The treatment plan recommended by NICE for OHT and suspected COAG (Table 1.4) was based on the risk for conversion of OHT to COAG. It acknowledged the relationship between IOP and CCT (see section 1.6.2) and shows how the decision to treat or not depends on IOP, CCT and age. Table 1.4 will be referred to again in chapter two when multiple cut-off points for clinical tests is revisited.

As the study described in this thesis was concerned with referral refinement, further treatment and follow-up options recommended by NICE are not covered.

IOP	**CCT <555**	**CCT 555-590**	**CCT>590**
>21mmHg to 25mmHg	treat with PGA until 65	no treatment	no treatment
>25mmHg to 32mmHg	treat with PGA until 80	treat with BB until 60	no treatment
>32mmHg	PGA	PGA	PGA

Table 1.4 Treatment plan recommended by NICE (52). Note that the decision whether or not to treat with BB (beta-blockers) or PGA (prostaglandin analogue) is dependent on IOP, CCT and age.

1.7 Referral Refinement

Glaucoma sufferers are often unaware of their progressive loss of visual field until late in the disease process when it affects the central vision (99). Late presentation is a major risk factor for blindness attributed to glaucoma (100). Early detection of the disease is therefore vital.

Currently there is no formal screening programme for COAG or OHT. There is a lack of high quality research evidence to support formal screening (101) and it is not considered to be cost effective (53, 102), though screening patients in high risk groups, such as those of black ethnic origin, may be more beneficial (53, 100). The most recent policy review by the UK National screening committee reported in 2016 and confirmed that formal screening for glaucoma is still not currently recommended (103). A further review is due in 2018/19.

Because of the absence of a formal screening programme, most cases of COAG and OHT are found opportunistically by community optometrists when patients attend for routine eye examinations (104). Patients suspected of having OHT or COAG are usually referred into the HES for formal diagnosis, treatment and monitoring (105).

Following publication of NICE guideline CG85 (see section 1.6), advice was issued by the COO and Association of Optometrists (AOP) stating that, in the absence of funding for repeat IOP measurements using GAT, optometrists had no choice other than to refer a patient who had raised IOP (indicating OHT) using whatever tonometer they chose. Furthermore, they advised that failure to refer a case with suspected OHT may be considered unprofessional (106).

In response to this advice, referrals to the HES for suspected COAG and OHT doubled (107). A web-based questionnaire was sent to 9386 optometrists on the COO mailing list asking about referral behaviour. Extrapolation of the results to reflect all optometrists on the General Optical Council (GOC) register revealed that the advice given had resulted in 540,000 extra referrals to the HES per year (108).

Schemes started to develop in different parts of the country (38, 105, 111-114). These developed local agreed protocols and usually involved multidisciplinary teams. Some schemes were simply repeat measures where practitioners repeated IOP measurements with GAT or carried out repeated visual field tests. Some utilised specialist optometrists for enhanced case finding. Others, including COT, involved specialist optometrists in true Glaucoma Referral Refinement (GRR) where the undertaking of tests are sufficient to diagnose OHT and suspect COAG (115, 116). In addition to reducing false-positive referrals to the HES by 35-77% (Table 1.5), these also reported cost savings of up to £117 per patient (113) and reduced waiting times (38).

Scheme	***Reduction of referrals (%)***
Huntington (114)	35
Manchester (38)	40
Carmarthenshire (113)	53
Nottingham (111)	54
RCAS (105)	71
EGRM (105)	76
Stockport, Tameside, Glossop (112)	77

Table 1.5 Reduction in false-positive referrals to the HES due to various schemes operating across the UK. Studies are shown in ascending order of reduction in referrals. Key: RCAS = Refinement by the Community Team after Clinical Assessment Scheme, EGRM = Enhanced Glaucoma Repeat Measurement scheme.

One of the studies indicated that a scheme involving repeat measurements by non-specialist optometrists was more cost-effective and reduced false positives just as effectively as a scheme involving a team of specialist optometrists (105).

1.8 Community Ophthalmology Team

1.8.1 History and regional coverage

The COT was set up in 2006 in conjunction with consultant ophthalmologists from the Maidstone and Tunbridge Wells NHS Trust and the West Kent Primary Care Trust.

The COT allows optometrists the opportunity to extend their role in providing a range of ophthalmology services in their local community. It is a multidisciplinary team which uses the different skills of its members to enhance patient care. The advantages of optometrists working so closely with the ophthalmologists is the development of excellent working relationships, continuing education and a mutual understanding of each other's skill sets. The benefit to the patient is that they are seen more quickly by the appropriate clinician. Allowing the lower risk patients to be seen in the community by the COT facilitates more consultant ophthalmologist hospital clinic time for those patients deemed at greatest risk of sight loss due to glaucoma.

On the 1st April 2013, Strategic Health Authorities and Primary Care Trusts were abolished by the government in favour of the NHS commissioning board (NHSCB) and Clinical Commissioning Groups (CCGs). Following this change, the COT was commissioned in three CCGs:

(i) West Kent - 463,000 residents and 62 general practitioner (GP) practices (117);

(ii) Medway - 283,000 residents and 56 GP practices (118);

(iii) Dartford, Gravesham and Swanley - 249,000 residents and 34 GP practices (119).

The commissioning of the COT across the 3 CCGs in Kent is unique within England and Wales and the COT continues to flourish. The CCGs and the Maidstone and Tunbridge Wells NHS Trust are looking at measures to expand the COT further with more glaucoma and medical retina patients being discharged into the community for management by the COT. Recently the Clinical Council for Eye Health Commissioning has published a Community Ophthalmology Framework which is very similar to the COT and endorses its approach (120).

1.8.2 Specialist accreditation and continued training

The COT currently employs the services of 15 Optometrists and one GP. The COT is mentored by Lead Consultant Ophthalmologist, Professor Ejaz Ansari. All optometrists working in the COT receive training in hospital clinics and, after accreditation by Professor Ansari, are allowed use of the affix Optometrist with Special Interest in Ophthalmology (OPwSI).

Accredited optometrists receive annual appraisal with Professor Ansari and are required to undertake any training facilitated by the commissioners. The training is regularly repeated to ensure patient safety, continuity with the HES glaucoma service and compliance with NICE guidelines (see section 1.6). In addition, COT practitioners have direct access to advice from Consultant Ophthalmologists via telephone or NHS secure email regarding the management of any patient they are seeing on behalf of the COT.

The author and two of his colleagues, Deacon Harle (DH) and Niall O'Kane (NOK), were all involved in clinical data collection for the study described in within the thesis. All are Independent Prescribing (IP) optometrists. The author also holds the Higher Cert Glauc and DH the Dip OC.

1.8.3 Operating procedures

All non-emergency ophthalmology referrals (excluding any patient under 16, for which there is a different pathway) by GPs, pharmacists and non-specialist community optometrists working in the three CCGs are sent to the Primary Care Booking Service (PCBS) in Maidstone.

Referrals are uploaded onto a secure NHS system by the PCBS staff. Each month one of the members of the COT triages the referrals on a daily basis from this system and advises the PCBS whether the patient is suitable to be seen by a member of the COT or by their HES colleagues. Patients that can be seen safely in the COT are contacted by the PCBS and offered an appointment.

Elements of the SOP used by the COT that have direct relevance to the present study include:

i. Taking family history of glaucoma (FHG);
ii. Gonioscopy to establish that the drainage angle is open;
iii. GAT;
iv. CCT by ultrasound pachymetry;
v. Dilated stereoscopic ONHA with Volk lens to enable;
    a) Measurement of vertical optic disc size (VDS) with thin slit lamp beam directed onto the disc with the slit lamp graticule used to reduce height of beam until it corresponds with the size of the disc. The VDS can then be read directly from the graticule and a correction factor applied if required;
    b) Estimation of the vertical cup-to-disc ratio (VCDR);
vi. Visual field assessment using the Humphrey Visual Field Analyser SITA fast 24-2 strategy.

More details about the ophthalmic instrumentation used by the three COT members who participated in the study and grouping of clinical data into multiple cut-off points are provided in Chapter 2.

The management options, of direct relevance to the study described in this thesis, following the COT appointment were as follows:

i. Discharge(Dis);

ii. Follow-up in the COT for suspected COAG(Fup);

iii. Refer to the HES for COAG (Refer).

1.9 Aim of the present study

The aim of the study was to determine whether application of naïve Bayes to clinical data collected using the established SOP of the COT could accurately replicate clinical decisions relating to the referral and management of COAG made by three IP optometrists with a special interest in ophthalmology. A major feature of this study was its adoption of likelihood ratios for multiple cut-off points (see section 1.4.4). As three optometrists were involved, the opportunity also arose to study the consistency of system accuracy and speed of learning. Likelihood ratios generated for each component of the SOP were also used to determine whether there was any redundancy; that is, could the SOP be further refined?

As far as the author and his supervisory team were aware, this was the first investigation of naïve Bayes as a means of providing clinical decision support for refinement of referrals relating to suspected COAG. As such, the scope of the study was believed to constitute a substantial and original contribution to knowledge in this field.

1.10 Summary and chapter outline

Chapter 1 has outlined the background and aims of this study. Chapter 2 provides a more detailed description of the ophthalmic instrumentation used by the three COT members who participated in this study and grouping of clinical data into multiple cut-off points. A preliminary analysis of clinical data is described in Chapter 3. Application of naïve Bayes is evaluated in Chapter 4 in terms of its accuracy, speed of learning, consistency and the presence of any redundancy. Chapter 5 ends the thesis with a review of key findings, a critique and recommendations for further research.

Chapter 2: Methods

2.1 Introduction

The purpose of this study was to determine the accuracy of clinical decision support based on naïve Bayes when applied to a well-developed SOP. The SOP, used by the author and two of his colleagues for referral refinement of COAG, was introduced in chapter 1 (see section 1.8.3). Chapter 2 describes the SOP in more detail. The recommendation by Parikh's research group to use likelihood ratios for multiple cut-off points was also introduced in chapter 1 (see section 1.4.4). Chapter 2 also provides a description of how the clinical data was grouped to enable use of multiple cut-off points.

2.2 Ethics

Approval was granted from the Life and Health Sciences Research Ethics Committee at Aston University (Project 495). Ethical clearance was based on data collection being treated as a clinical audit. Treating the study as an audit allowed collection of fully anonymised data without the consent of the individuals being examined, ensuring that all referral refinements carried out by the COT during the study period (see section 1.8.2) could be included. Failure to consent would otherwise have distorted the prior odds (see section 1.4.1) upon which naïve Bayes depended.

2.3 Data Collection

Clinical data for this study represented all referrals for suspected COAG seen by the author (JG) and two of his colleagues (DH and NOK, see section 1.8.2) between 1st October 2014 and 1st October 2015.

Patients referred or followed up for any other reason were excluded as were those unable to undertake visual fields for medical reasons such as dementia.

A total of 1069 patients were examined and 63 cases were excluded for the reasons stated above which left 1006 cases; 451 from JG, 486 from DH and 69 from NOK.

Data was taken from the worst affected eye or, if there was no asymmetry in severity, from the right eye. The sections that follow detail the clinical data and its grouping for the purpose of using likelihood ratios for multiple cut-off points (see section 1.4.4).

2.3.1 Age

Increasing age is a major risk factor for the development of COAG. Table 2.1 shows the increased prevalence of COAG with age in some of the major epidemiological studies.

Study	Age Range (years)	Prevalence (%)
Beaver Dam(45)	43-54	0.9
Beaver Dam(45)	over 75	5.0
Barbados(42)	over 50	9.0
Barbados (42)	over 70	17.0
Rotterdam(41)	55-59	0.3
Rotterdam(41)	85-89	3.3
Baltimore Whites (120)	40-49	0.9
Baltimore Whites (120)	over 80	2.2

Table 2.1 Prevalence of COAG estimated from some major epidemiological studies.

Table 2.2 shows how age was grouped in these studies compared to the present study.

Study	Age Groups (years)
Beaver Dam (45)	43-54, 55-64, 65-74, 75+
Blue Mountains(47)	<60, 60-69, 70-79, 80+
Baltimore (46)	40-49, 50-59, 60-69, 70-79, 80+
LA Latino (121)	40-54,55-64,65-74,>75,>80
Barbados(42)	40-49,50-59,60-69,>70
Tanjong Pagar(122)	40-49,50-59,60-69,70-79
Present study	<40, 40-49, 50-59, 60-69, 70-79, 80+

Table 2.2 Age groups adopted in major epidemiological studies.

2.3.2 Race

Studies have shown that race influences the prevalence of COAG (Table 2.3). The prevalence of COAG ranges from 0.8% in Caucasians to 8.8% in Afro Caribbeans. The racial groups adopted in the present study were those shown in Table 2.3.

Study	Caucasian	Asian	Afro Caribbean	African	Hispanic	Prevalence (%)
Rotterdam(41)	X					0.8
Melbourne VIP(44)	X					1.7
Proyecto VER (123)					X	2.0
Blue Mountains (47)	X					2.0
Mongolia(124)		X				2.1
Liwan (125)		X				2.1
Thailand (126)		X				2.3
Southern India (127)		X				2.6
West of Ireland (128)	X					2.8
Chennai (129)		X				3.2
Tanjong Pagar (122)		X				3.2
Kongwa district (48)				X		4.2
LA Latino study (121)					X	4.7
Barbados (42)			X			6.7
Ghanaian (130)				X		7.7
St Lucia (43)			X			8.8

Table 2.3 Prevalence of COAG in different racial groups estimated from major epidemiological studies. The studies have been arranged in order of increasing prevalence to show the range of values found. The symbol X shows the majority race within each study. The racial groups shown were those adopted in the present study.

2.3.3 Sex

Opinions differ as to whether sex is a risk factor in COAG (45, 131, 132). Some studies have shown that males are more at risk (42, 44, 133), while others indicate that females have a greater risk (134). Sex was therefore included in this study to see if a link could be established.

2.3.4 Family history

Some studies have reported lifetime risk of developing COAG in first degree relatives is approximately 10x greater than those without a family history of the condition (135). Having siblings with the disease raises the risk more than having a parental family history (136). More risk is also associated with a maternal rather than a paternal family history (137). Given these findings, the present study included any recorded family history involving the mother, father and/or one or more siblings.

2.3.5 Reason for referral

Reason(s) for referral (RFR) recorded in the present study were raised IOP, suspect optic discs and suspect visual fields. The purpose of this was to follow up the recommendation of a previous study to minimise false positives by combining test data relating to IOP, optic disc and visual field assessment (138). It has been suggested that basing referrals on combined test results may provide a more cost effective means of reducing false positives than using accredited community optometrists (105).

2.3.6 Intra Ocular Pressure

Raised IOP is an important risk factor for COAG (139) and its reduction and control is the goal of treatment (99). Reducing the IOP by 20% lowers the risk of progression to COAG within 5 year period by 60% (140). Each 1mmHg rise in IOP is associated with a 10% increased risk of COAG progression (141).

The author and his two colleagues measured IOP using a Haag Streit AT900 Model T GAT and non-disposable prisms with a slit lamp biomicroscope.

The IOP values were grouped (<21mmHg, 21-25mmHg, 25-32mmHg, >32mmHg) according to the NICE guidelines (see table 1.4 in section 1.6.8).

The inter-ocular difference in IOP (IOP diff) in both eyes was grouped as <3mmHg, 3-6mmHg, >6mmHg as the risk of POAG increases by 6% and 57% for IOP diffs of 3mmHg and greater than 6mmHg, respectively (142).

Even in NTG, it has been found that an IOP difference, in 86% of cases, causes greater damage in the eye with the highest pressure (143).

2.3.7 Optic Nerve Head Assessment

Optic nerve head assessment involved the use of dilated stereoscopic slit lamp biomicroscopy (JG and NOK – Haag Streit BQ900, DH – Topcon PS30) with a Volk lens (JG and DH – 66D, NOK – Digital 1x) (see section 1.8.3).

The vertical disc size (VDS) was recorded because larger optic discs tend to have larger cup to disc ratios (144). The measurement of VDS is particularly important when assessing both the amount of cupping and neuroretinal rim tissue in a disc for possible COAG (144-146). Those of African descent have been shown to have larger optic discs than whites (147, 148) which, combined with the higher prevalence of the disease in Africans, has led to the notion that larger optic discs are a risk factor for COAG (149). However, there is no evidence that the size of an optic disc is an independent risk factor for COAG (150).

As patients seen by the author and his two colleagues were more often Caucasian, VDS was grouped (<1.4mm, 1.4-1.7mm, >1.7mm) according to a previous study carried out on a predominantly Caucasian population (151).

The vertical cup-to-disc ratio (VCDR) was expressed as a percentage. The superior and inferior poles of the neural retinal rim are preferentially damaged in COAG (152). COAG causes the cup-to-disc ratio to increase most in the vertical meridian (153). In keeping with this, the VCDR is a slightly better predictor for the development of COAG than the horizontal cup disc ratio (140).

Every 0.1mm increase in vertical cupping increases the risk of COAG by 1.3 times (140). The VCDR also increases at a faster rate in early stages of COAG compared to the horizontal cup-to-disc ratio (154).

Approximately 98% of healthy eyes have a VCDR of less than 65% (155). Cut-off values that enable VCDR to distinguish healthy eyes from those with COAG have ranged from 50% (144) to 70% (156). The VCDR groups (50%, 50-70%, >70%) adopted in the present study broadly encompassed these suggested cut-off values.

COAG is a bilateral condition which is frequently asymmetric (157). Inter-ocular differences in VCDR (VCDR diff) can occur with COAG (158, 159). The VCDR diff was expressed as a percentage in the present study. It was grouped (<20%, 20-30%, >30%) to reflect previous findings that VCDR diffs of 20% and 30% were found three and ten times more often, respectively, in COAG compared to healthy eyes (158).

2.3.8 Central Corneal Thickness

Central corneal thickness (CCT) may influence the accuracy of GAT in the diagnosis and management of COAG (160). Thin and thick corneas can result in artificially low and high IOP readings, respectively (161). CCT is also an important risk factor for the development of COAG (140) with an increased risk of 1.4 for every 40 microns reduction in CCT (162). Measurement of CCT is used in conjunction with GAT and age to decide whether treatment for COAG or OHT should begin (52) (see Table 1.4 in section 1.6.8).

The author and his two colleagues measured CCT (see section 1.8.3) using handheld ultrasound pachymetry (JG - Accutome Pachpen, DH – Pachmate, NOK – Pachmate 2). The CCT measurements were grouped (<555 microns, 555-590 microns,>590 microns), in the present study, according to the NICE guidelines (see Table 1.4 in section 1.6.8).

2.3.9 Visual Fields

Visual field (VF) assessment is important in the diagnosis and management of COAG (see section 1.5.3.1). NICE recommended the SITA 24-2 testing strategy as the reference standard for measuring the VF (see section 1.6.4). The author and his two colleagues measured the VF using the Zeiss Humphrey Visual Field Analyser (VFA: JG- model 720i, DH – model 720, NOK – model 720i) with the SITA FAST 24-2 testing strategy. This testing strategy was the method preferred by the Lead Consultant Ophthalmologist overseeing the COT (Prof Ejaz Ansari, see section 1.8.2) and was also used in the HES glaucoma clinic in the area covered by the COT. Grouping of VF defects (mild, medium, severe) followed the Hodapp-Anderson-Parrish (HAP) grading system (163). The original HAP grading system was applied to the SITA standard 30-2 testing strategy, but other studies have applied it to the SITA fast 30-2 testing strategy (164). A description of the application of the HAP grading system to the SITA fast 24-2 testing strategy now follows.

A mild defect was recorded if (i) the mean deviation was no worse than -6db and, on the pattern deviation plot, (ii) less than 25% of points were depressed below the 5% level, (iii)less than 15% of points were depressed below the 1% level and (iv) no point within the central 5 degrees had a sensitivity of less than 15db (163).

A moderate defect was recorded if (i) the mean deviation fell between -6db and -12db and, on the pattern deviation plot, (ii) less than 50% of the points were depressed below the 5% level, (iii) less than 25% of the points were depressed below the 1% level, (iv) no point within the central 5 degrees had a sensitivity of less than 0db and (v) only one hemi field contained a point with a sensitivity of less than 15db within 5 degrees of fixation (163).

A severe defect was recorded if (i) the mean deviation was worse than -12db and, on the pattern deviation plot, (ii) more than 50% of the points were depressed below the 5% level, (iii) more than 25% of points were depressed below the 1% level, (iv) any point within the central 5 degrees had a sensitivity of less than 0db and (v) both hemi fields contained points with a sensitivity of less than 15db within 5 degrees of fixation (163).

2.3.10 Management

Management decisions were grouped (Dis, Fup, Refer) as described in section 1.8.3.

2.4 Summary

This chapter has provided a detailed description of the methods used by the three COT members who participated in this study and the grouping of clinical data to allow the use of likelihood ratios with multiple cut-off points. A preliminary analysis of the data is described in the next chapter.

Chapter 3: Preliminary analyses

3.1 Introduction

A preliminary analysis of the data collected, as described in chapter 2, is provided in this chapter. The broad objectives of the chapter are to:

1. Show the distribution of the COT's clinical test findings according to the multiple cut-off points described in chapter 2;
2. Determine whether the clinical test findings were equally distributed across the three COT members (JG, DH and NOK);
3. Present a series of analyses describing the association between each clinical test and COT management outcomes (Dis, Fup or Refer);
4. To examine the influence of CCT on GAT-IOP and to show how it might impact on the agreement between the referrals made to the COT for raised IOP and COT measurements of IOP;
5. To examine the influence of VDS on VCDR;
6. To compare RFRs for suspect VF and COT measurements of HPA;
7. To examine First Discharge Rates (FDRs) and cost savings of the COT;
8. To rank COT tests in order of the strength of their statistical associations with COT management outcomes.

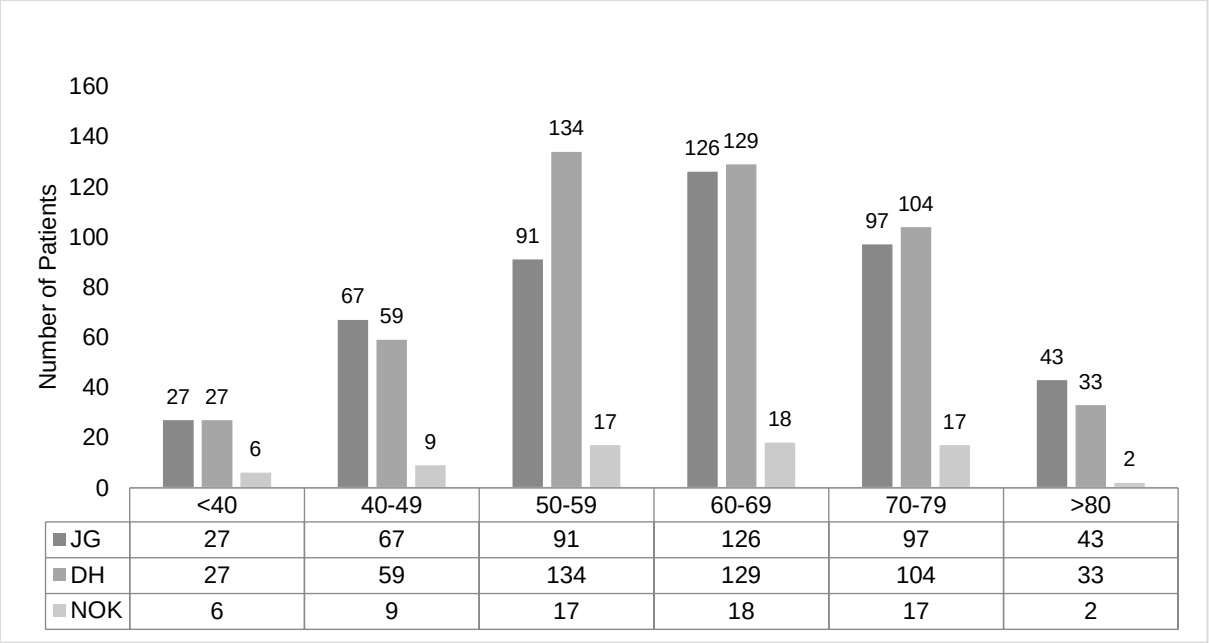
3.2 Statistical methods

Frequencies were compared using Chi-square for R x C contingency tables (165). This form of Chi-square test was also used to carry out analyses exploring the strength of associations between each clinical test and the COT management outcomes (Dis, Fup or Refer – chapter 2 section 2.3.10).

Tables of contributions, the degrees of freedom associated with these Chi-square tests varied from 1 to 10. Power analyses, conducted using GPower3.1 (166), indicated that Chi-square tests could detect medium size effects at the conventional alpha and beta levels of, respectively, 0.05 (equating to the 95% level of statistical significance) and 0.2 (equating to 80% power), for total sample sizes of between 88 and 181 people for, respectively, 1 to 10 degrees of freedom. All Chi-square tests presented in this chapter exceeded these minimum sample size requirements. Pearson's correlation coefficient (r), the coefficient of determination (r^2), analysis of variance (ANOVA) and linear regression by the method of least mean squares (167) was carried out to determine relationships between CCT versus GAT-IOP and VDS versus VCDR. Power calculations (also using GPower3.1) revealed that linear regression could detect medium size effects at the alpha and beta levels mentioned above for a sample sizes of 82. The minimum sample requirement was far exceeded in the analyses presented.

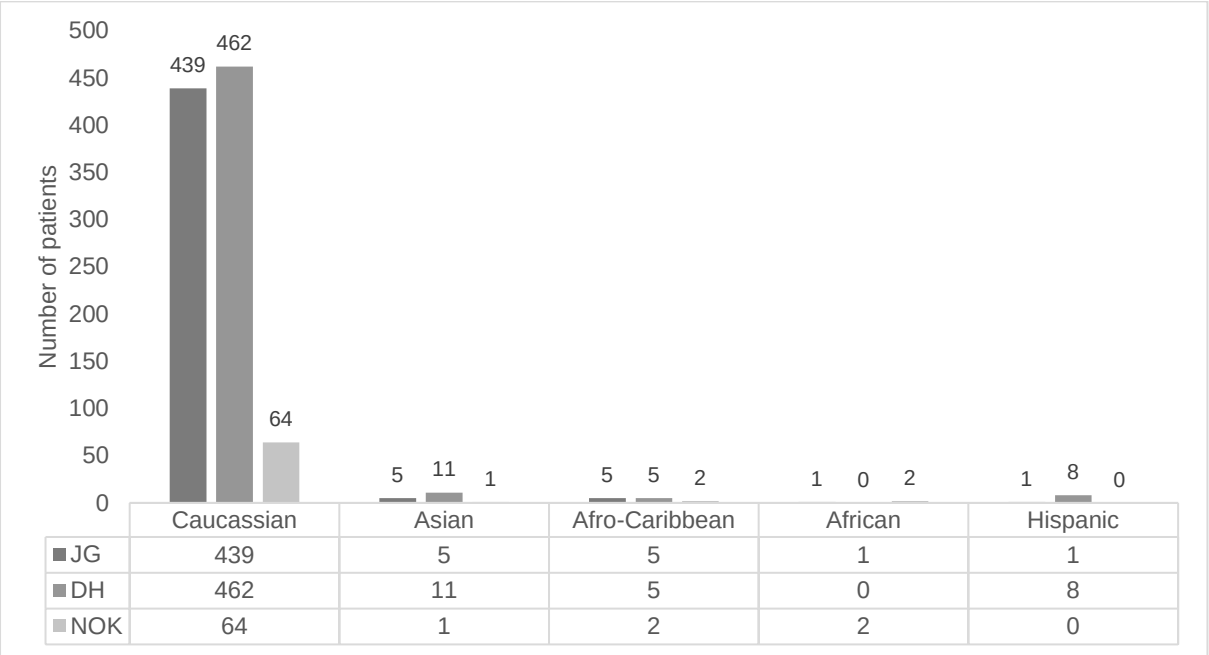
3.3 Distribution of the COT clinical test findings

The first objective was to show the distribution of the COT clinical test findings according to the multiple cut-off points described in chapter 2. Figures 3.1 to 3.13 show the distribution of clinical test findings across the samples seen by each COT optometrist (JG, DH, NOK) contributing data to the study. Table 3.1 summarises the findings in Figures 3.1 to 3.13 by showing the percentage of cases falling in each of the multiple COT test categories after pooling the data from the 3 COT optometrists.



	<40	40-49	50-59	60-69	70-79	>80
JG	27	67	91	126	97	43
DH	27	59	134	129	104	33
NOK	6	9	17	18	17	2

Figure 3.1. Distribution of age across the sample seen by each COT optometrist (JG, DH, NOK) contributing data to this study N =1006.



	Caucassian	Asian	Afro-Caribbean	African	Hispanic
JG	439	5	5	1	1
DH	462	11	5	0	8
NOK	64	1	2	2	0

Figure 3.2 Distribution of race across the sample seen by each COT optometrist (JG, DH, NOK) contributing data to this study N=1006.

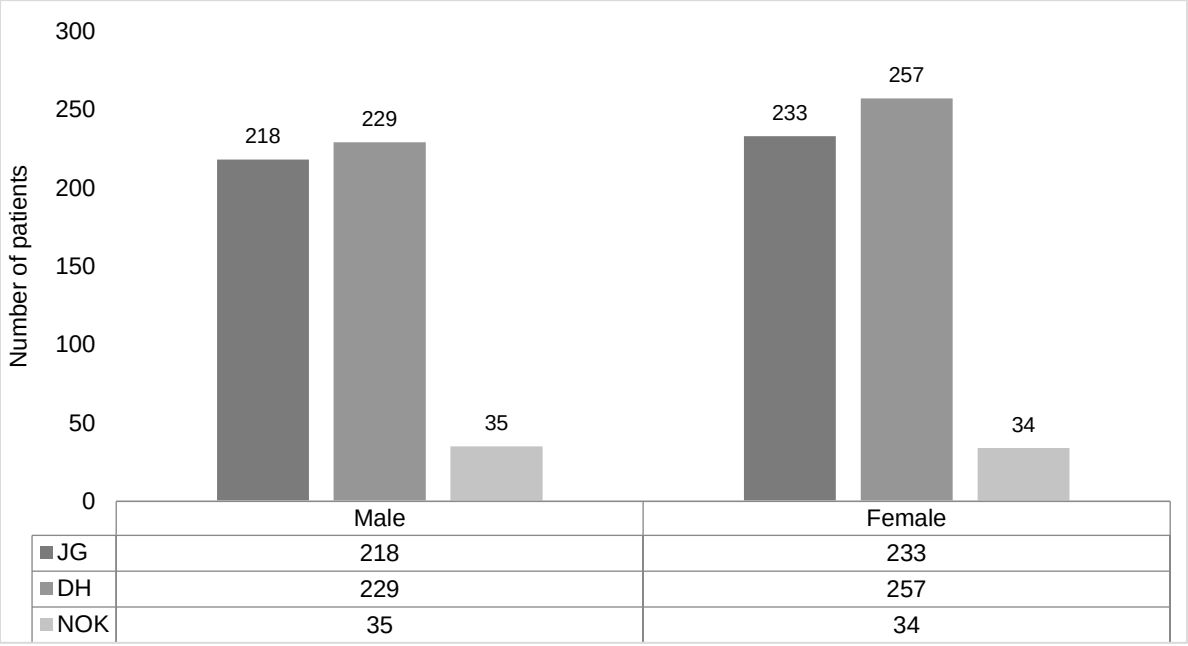


Figure 3.3 Distribution of sex across the sample seen by each COT optometrist (JG, DH, NOK) contributing data to this study N=1006.

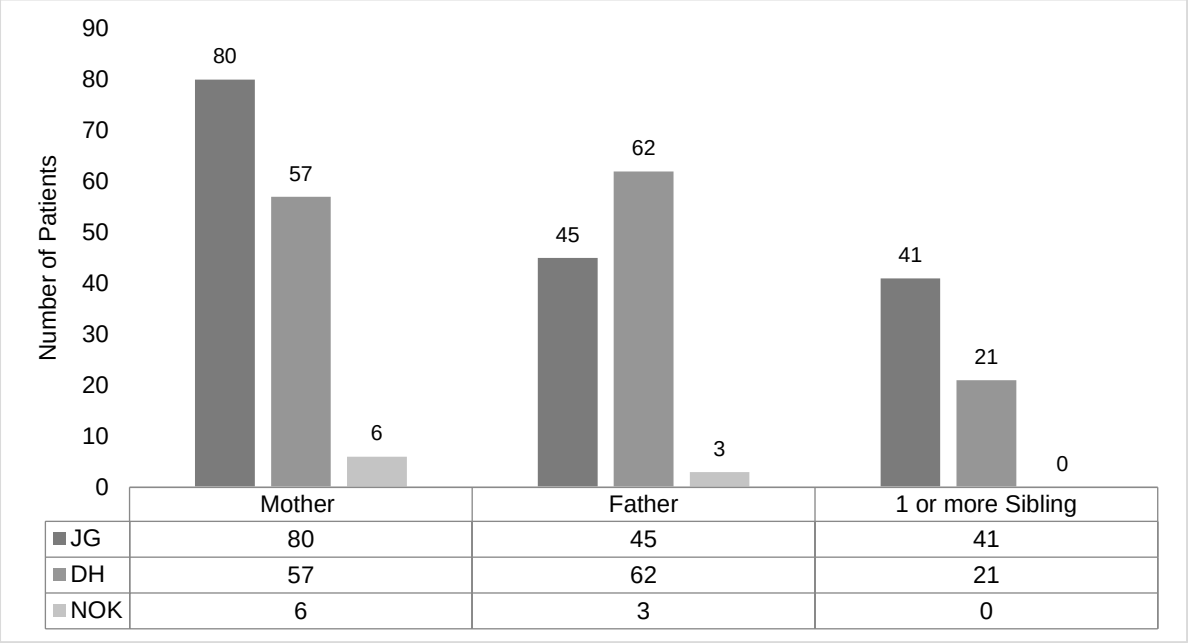
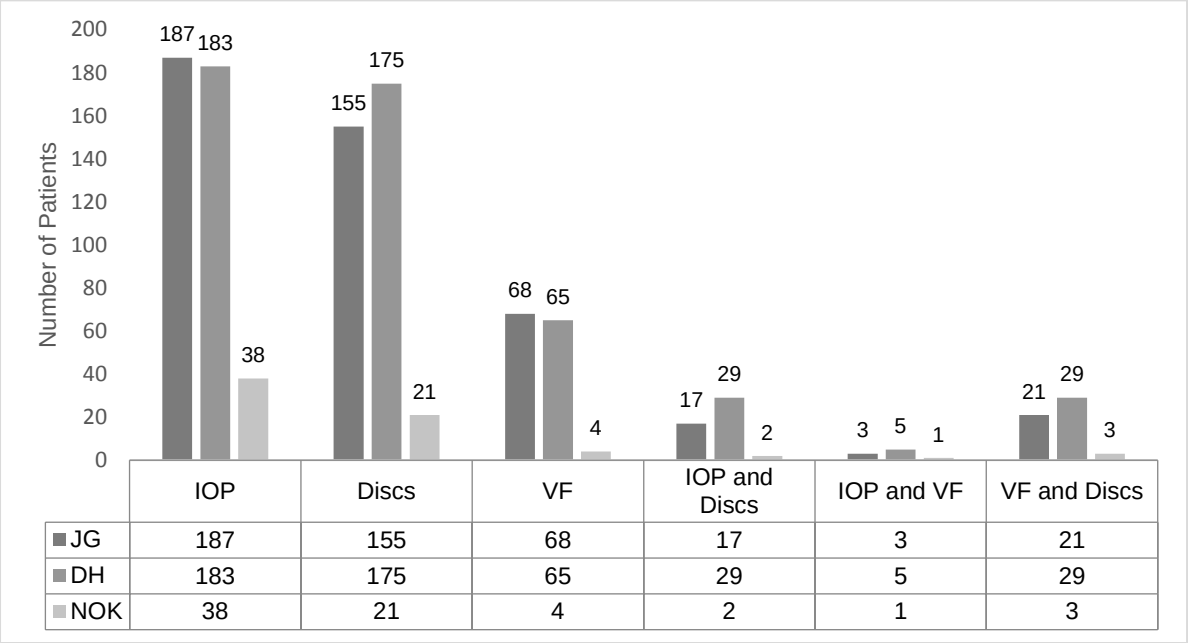
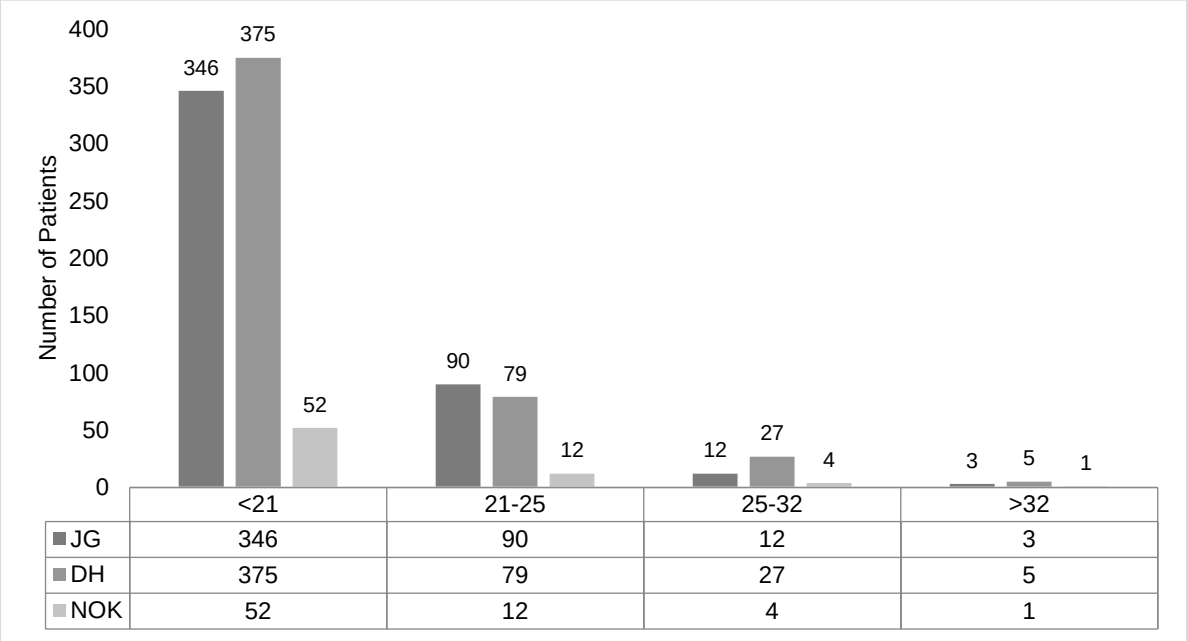


Figure 3.4 Distribution of Family history across the sample seen by each COT optometrist (JG, DH, NOK) contributing data to this study N=315.



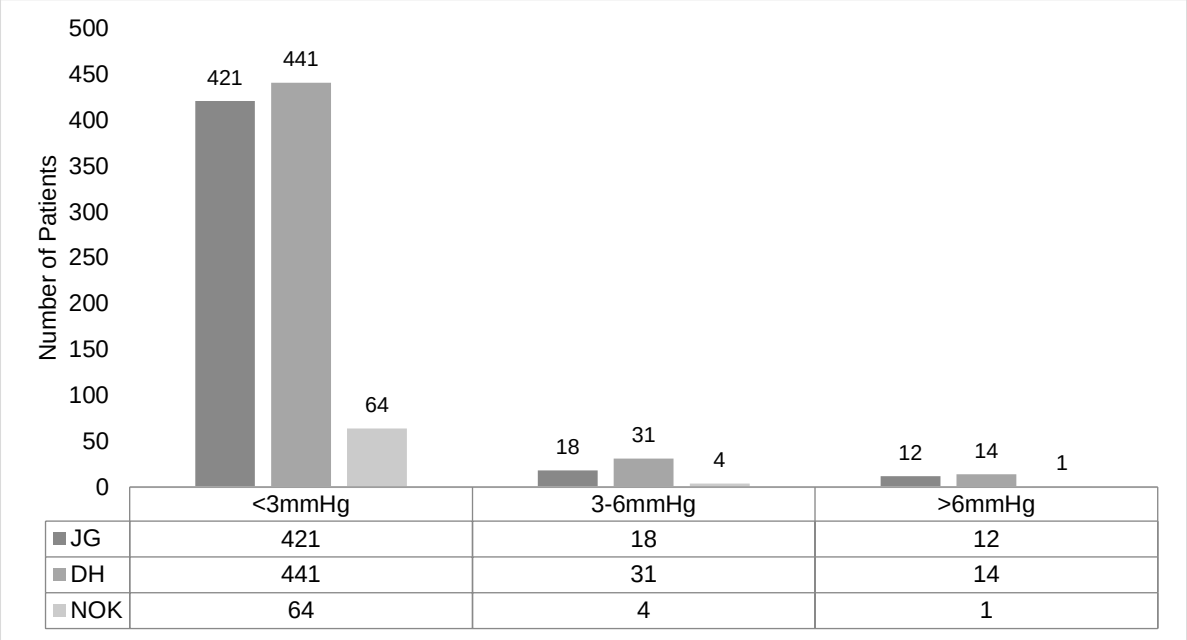
	IOP	Discs	VF	IOP and Discs	IOP and VF	VF and Discs
JG	187	155	68	17	3	21
DH	183	175	65	29	5	29
NOK	38	21	4	2	1	3

Figure 3.5 Distribution of RFR across the sample seen by each COT optometrist (JG, DH, NOK) contributing data to this study N=1006.



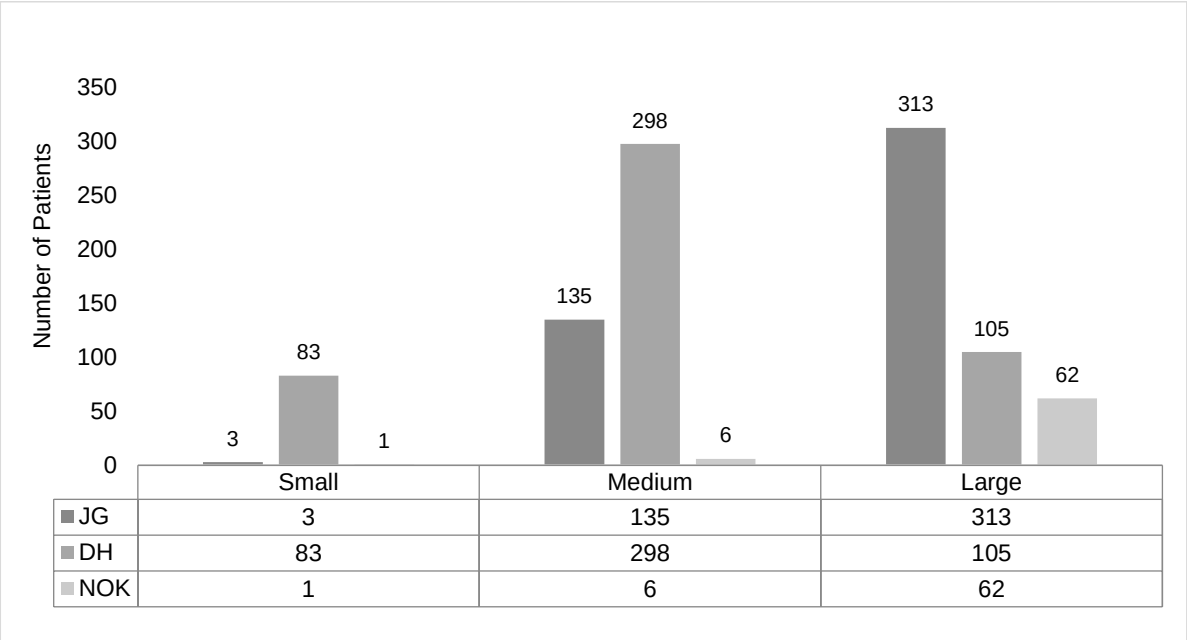
	<21	21-25	25-32	>32
JG	346	90	12	3
DH	375	79	27	5
NOK	52	12	4	1

Figure 3.6 Distribution of GAT IOP across the sample seen by each COT optometrist (JG, DH, NOK) contributing data to this study N=1006.



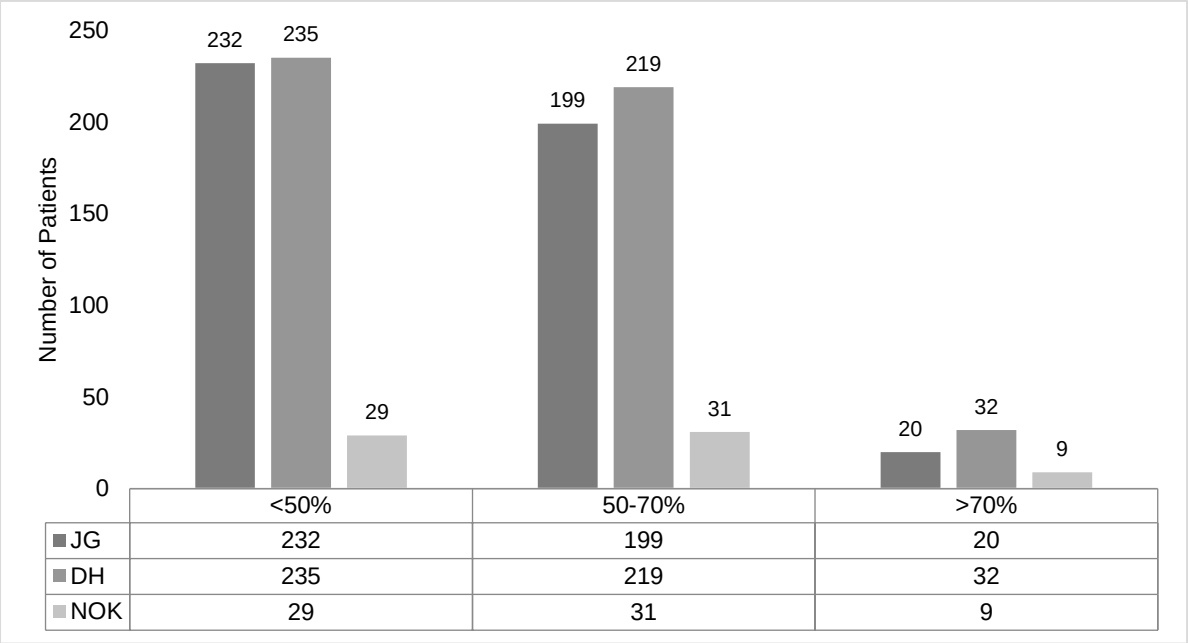
	<3mmHg	3-6mmHg	>6mmHg
JG	421	18	12
DH	441	31	14
NOK	64	4	1

Figure 3.7 Distribution of IOP diff across the sample seen by each COT optometrist (JG, DH, NOK) contributing data to this study N=1006.



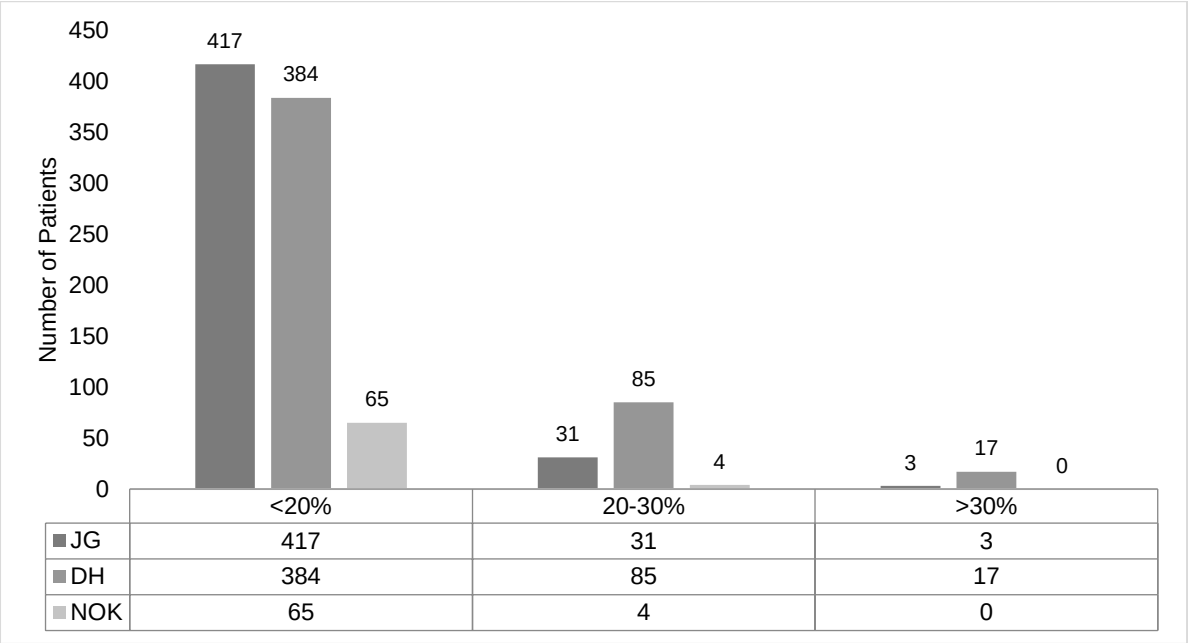
	Small	Medium	Large
JG	3	135	313
DH	83	298	105
NOK	1	6	62

Figure 3.8 Distribution of VDS across the sample seen by each COT optometrist (JG, DH, NOK) contributing data to this study N=1006.



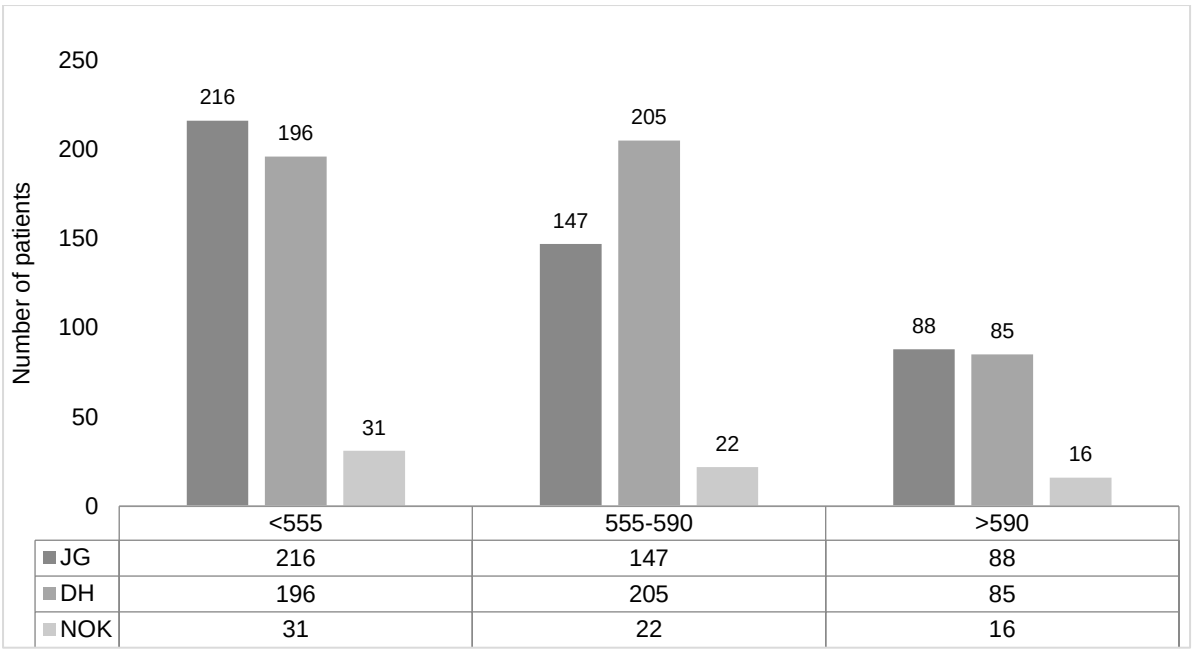
	<50%	50-70%	>70%
JG	232	199	20
DH	235	219	32
NOK	29	31	9

Figure 3.9 Distribution of VCDR across the sample seen by each COT optometrist (JG, DH, NOK) contributing data to this study N=1006.



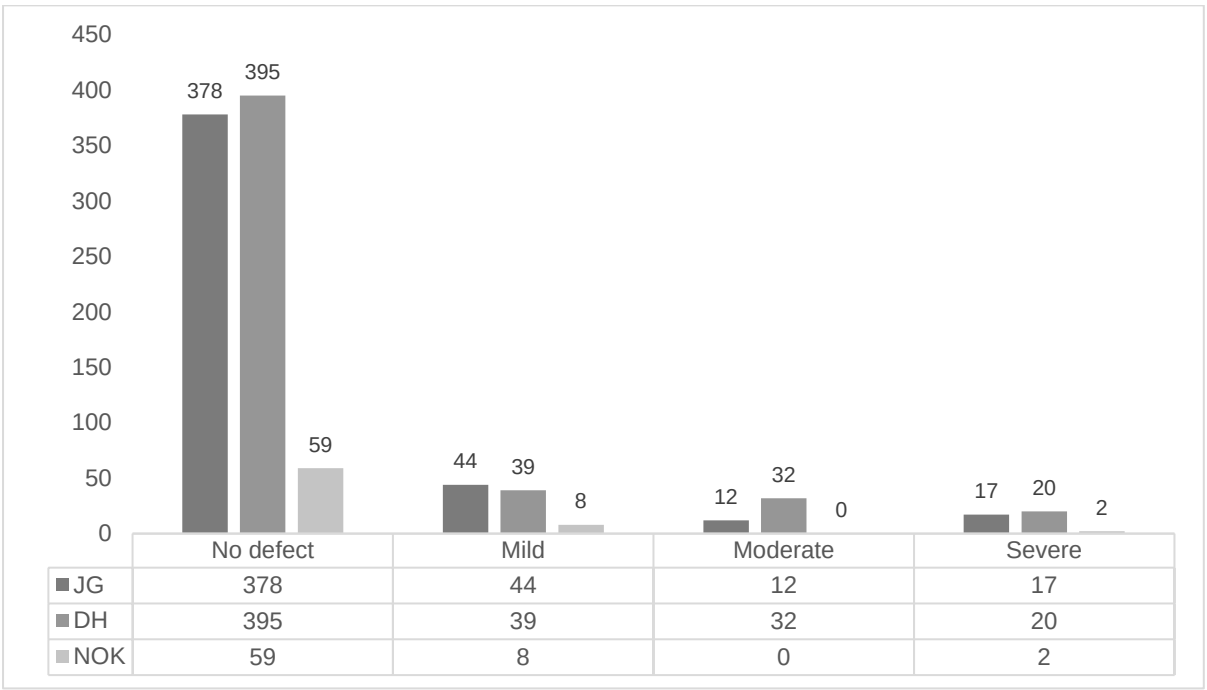
	<20%	20-30%	>30%
JG	417	31	3
DH	384	85	17
NOK	65	4	0

Figure 3.10 Distribution of VCDR diff across the sample seen by each COT optometrist (JG, DH, NOK) contributing data to this study N=1006.



	<555	555-590	>590
JG	216	147	88
DH	196	205	85
NOK	31	22	16

Figure 3.11 Distribution of CCT across the sample seen by each COT optometrist (JG, DH, NOK) contributing data to this study N=1006.



	No defect	Mild	Moderate	Severe
JG	378	44	12	17
DH	395	39	32	20
NOK	59	8	0	2

Figure 3.12 Distribution of HPA classification of VFA test across the sample seen by each COT optometrist (JG, DH, NOK) contributing data to this study N=1006.

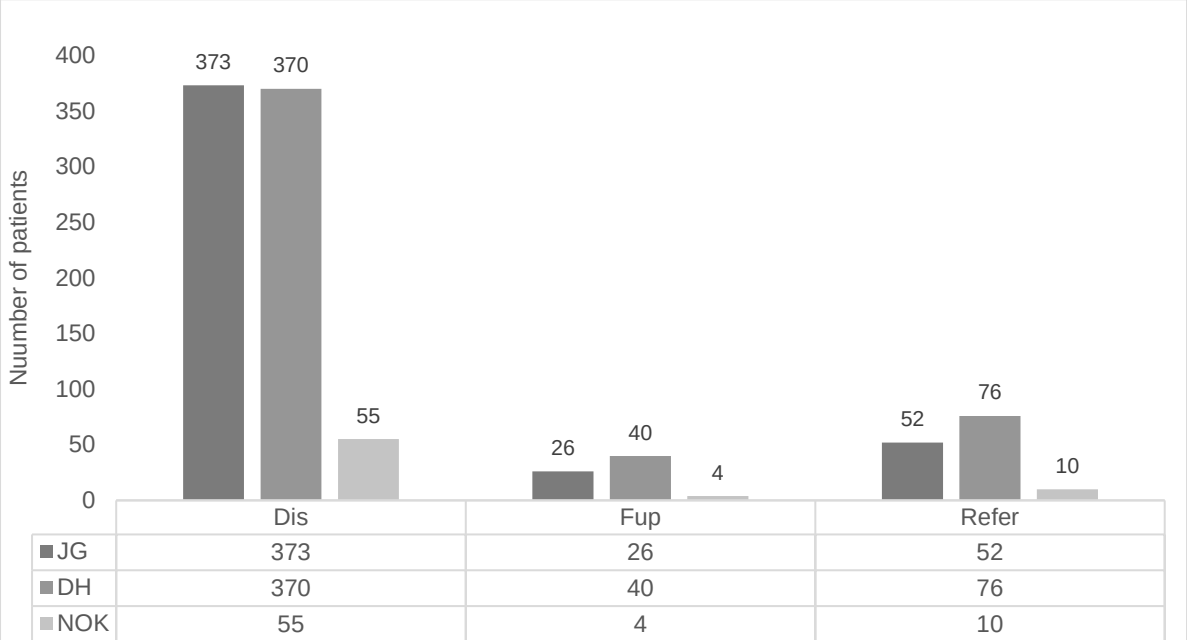


Figure 3.13 Distribution of COT management decisions across the sample seen by each COT optometrist (JG, DH, NOK) contributing data to this study N=1006

.

Clinical Test	Percentage of cases falling in each category
Age (years)	<40 (6%), 40-49 (13.4%), 50-59 (24.1%), 60-69 (27.1%), 70-79 (21.7%), 80+ (7.8%)
Race	Caucasian (95.9%), Asian (1.7%), Afro-Caribbean (1.2%), African (0.3%), Hispanic (0.9%)
Sex	Female (52.1%), Male (47.9%)
FHG	Mother (14.2%), Father (10.9%), Sibling (6.2%)
RFR	IOP (40.6%), Discs (34.9%), VF (13.6%), IOP/Discs (4.8%), IOP/VF (0.9%), VF/Discs (5.3%)
GAT IOP (mmHg)	<21 (76.8%), 21-25 (18.0%), >25-32 (4.3%), >32 (0.9%)
IOP diff (mmHg)	<3 (92.0%), 3-6 (5.3%), >6 (2.7%)
VDS (mm)	<1.4 (8.6%), 1.4-1.7 (43.6%), 1.8+ (47.7%)
VCDR	<50 (49.3%), 50-70 (44.6%), >70 (6.1%)
VCDR diff	<20 (86.0%), 20-30 (12.0%), >30 (2.0%)
CCT (nm)	<555 (44.0%), 555-590 (37.2%), >590 (18.8%)
VFA HPA	Mild (9.0%), Moderate (4.4%), Severe (2.9%)
Management	Dis (79.3%), Fup (7.0%), Refer (13.7%)

Table 3.1 Percentage of cases falling in each of the multiple categories for each COT test pooled across the 3 COT optometrists (JG, DH, NOK) contributing data to this study.

3.4 Were clinical and biographical test findings equally distributed across the 3 COT optometrists?

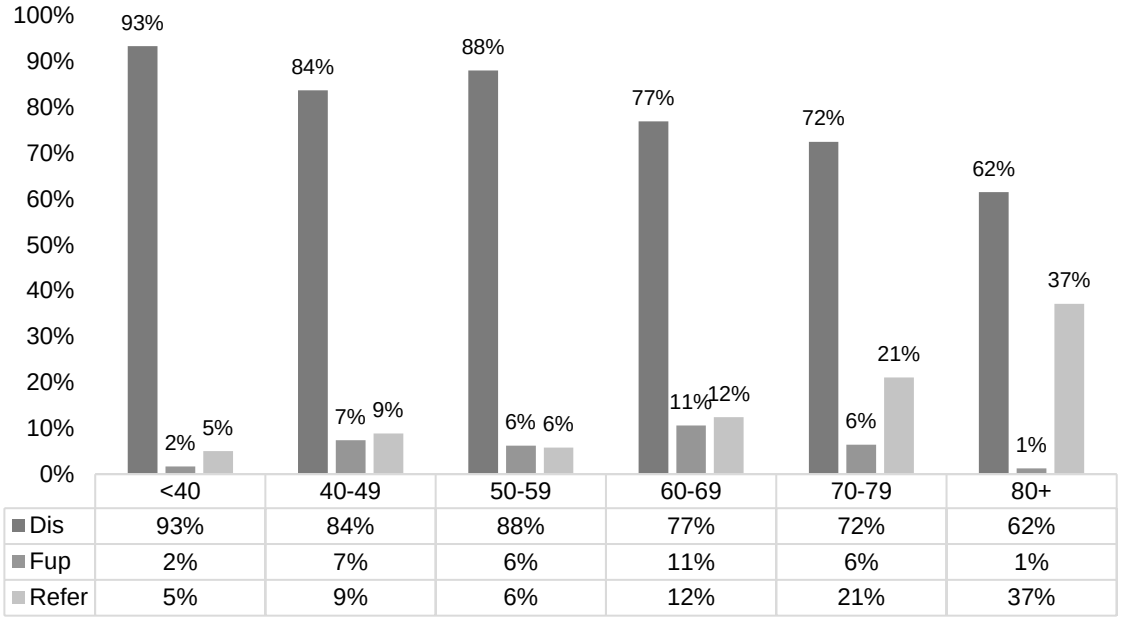
The second objective was to determine whether the clinical and biographical test findings shown in Figures 3.1 to 3.13 were equally distributed across the three COT optometrists (JG, DH and NOK). Analysis was carried out to explore any inhomogeneity between the samples seen by each COT optometrist as this could affect the outcomes of investigations into the consistency of naïve Bayes applied separately to their data (see chapter 4). Table 3.2 summarises the findings and shows that 6 out of the 13 clinical tests showed inhomogeneous distributions between the COT optometrists.

Clinical test	Chi-square, degrees of freedom (df), P-value
Age	12.65, df = 10, P = NS
Race	**27.05, df = 8, P <0.001**
Sex	0.37, df = 2, P = NS
FHG	**13.74, df = 4, P <0.01**
RFR	13.51, df = 10, P = NS
GAT IOP	7.41, df = 6, P = NS
IOP diff	2.58, df = 4, P = NS
VDS	**290.40, df = 4, P <0.0001**
VCDR	9.03, df = 4, P = NS
VCDR diff	**40.83, df = 4, P <0.0001**
CCT	**10.67, df = 4, P <0.05**
HPA class	**12.26, df = 4, P <0.05**
Management	6.37, df = 4, P = NS

Table 3.2 Summary of Chi-square tests carried out to determine the homogeneity of the samples seen by the 3 COT optometrists (JG, DH, NOK) contributing data to this study. Instances of statistically significant inhomogeneity are shown in bold.

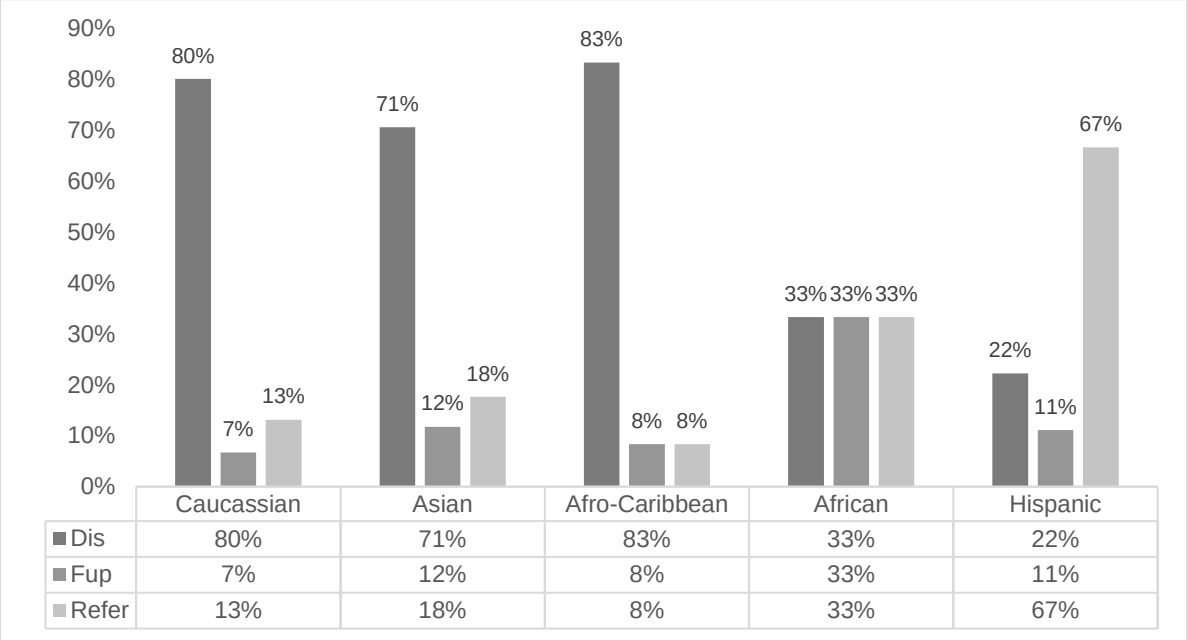
3.5 Associations between clinical tests and COT management decisions

The third objective was to present a series of analyses describing the associations between each clinical test and the COT management outcomes. Figures 3.14 to 3.25 show the percentage discharge (Dis), follow-up (Fup) and (Refer) for each of the multiple COT test categories after pooling the data from the 3 COT optometrists. Table 3.3 summarises the findings of these statistical tests.



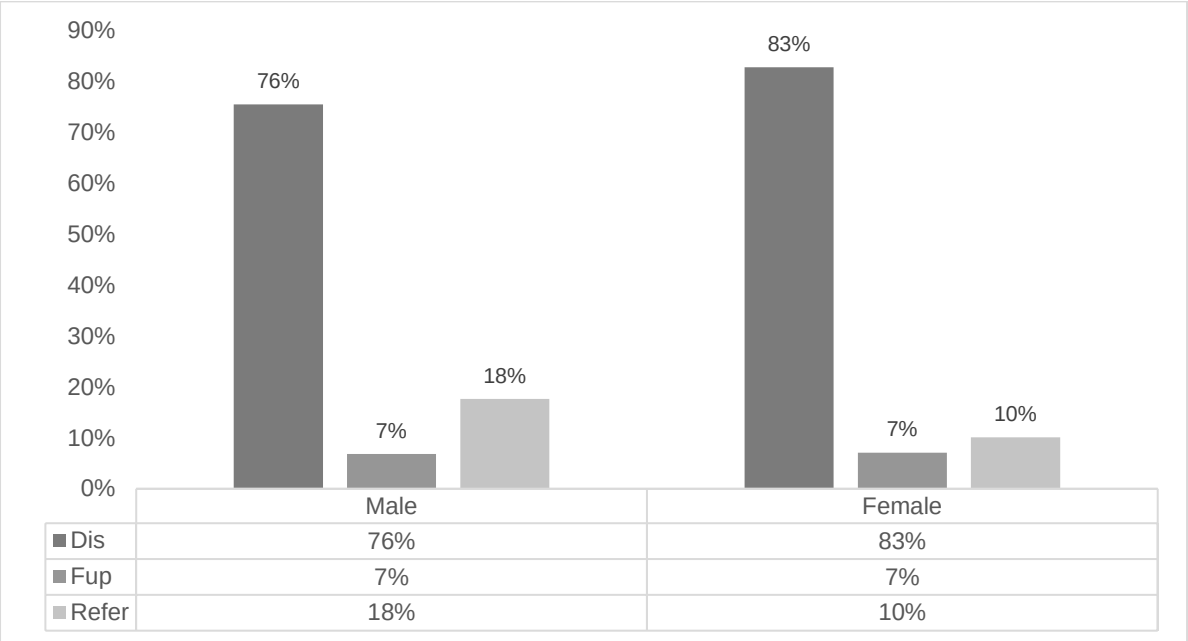
	<40	40-49	50-59	60-69	70-79	80+
Dis	93%	84%	88%	77%	72%	62%
Fup	2%	7%	6%	11%	6%	1%
Refer	5%	9%	6%	12%	21%	37%

Figure 3.14 Percentage Dis, Fup and Refer for each age group, pooled across the three COT optometrists (JG, DH, NOK) contributing data to this study N=1006.



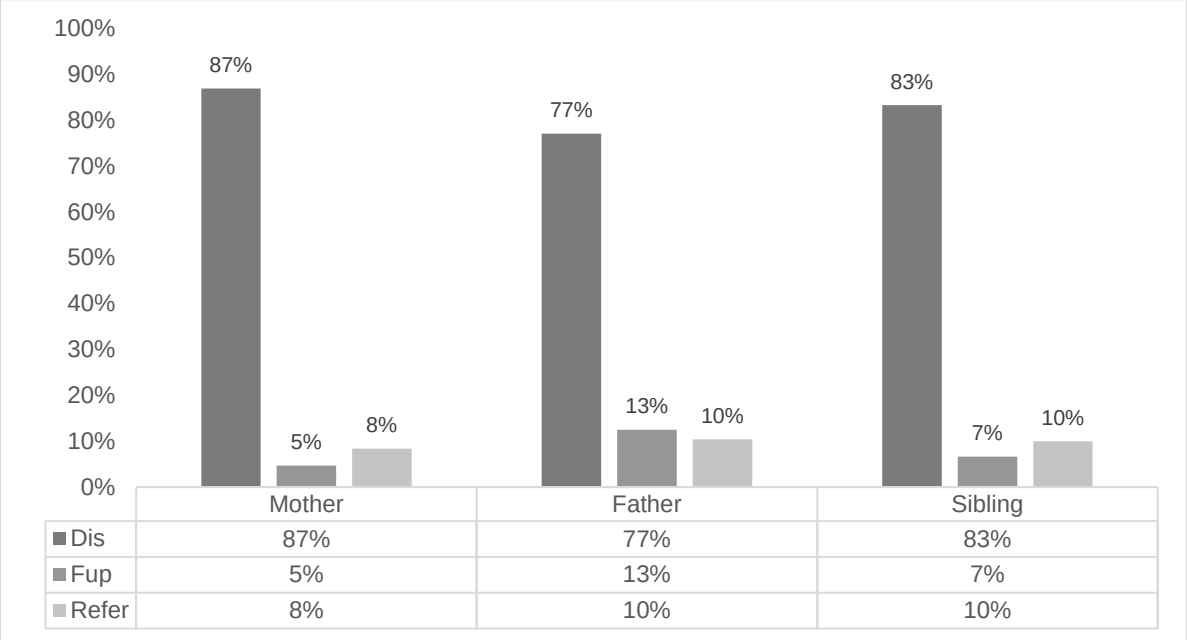
	Caucassian	Asian	Afro-Caribbean	African	Hispanic
Dis	80%	71%	83%	33%	22%
Fup	7%	12%	8%	33%	11%
Refer	13%	18%	8%	33%	67%

Figure 3.15 Percentage Dis, Fup and Refer for each racial group, pooled across the three COT optometrists (JG, DH, NOK) contributing data to this study N=1006.



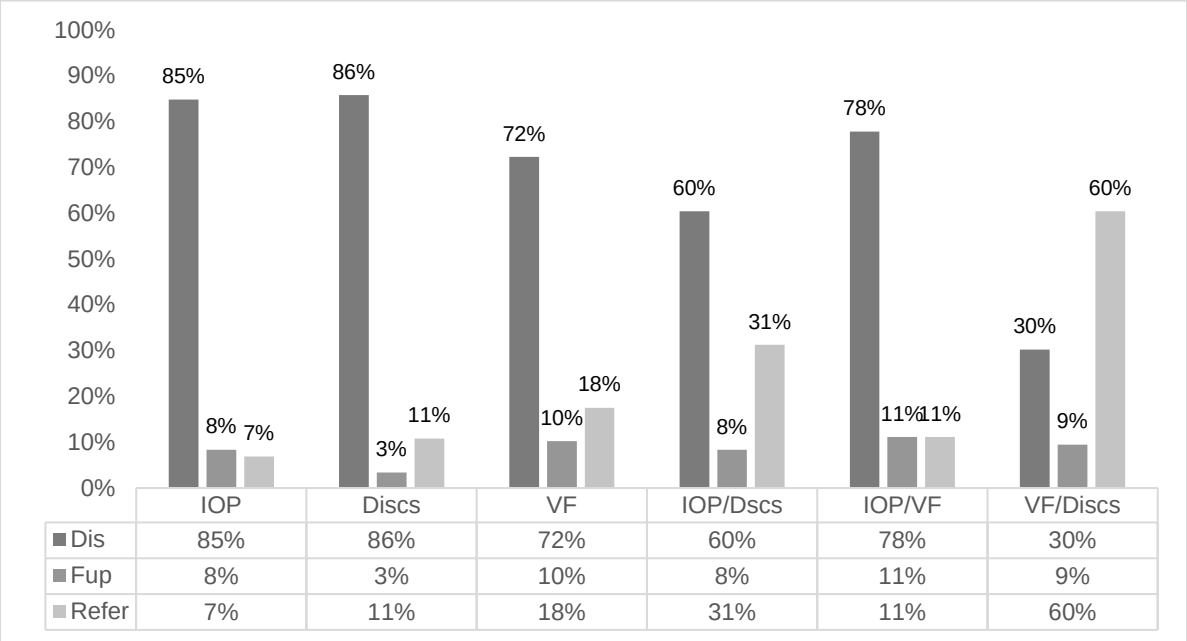
	Male	Female
Dis	76%	83%
Fup	7%	7%
Refer	18%	10%

Figure 3.16 Percentage Dis, Fup and Refer for males and females, pooled across the three COT optometrists (JG, DH, NOK) contributing data to this study N=1006.



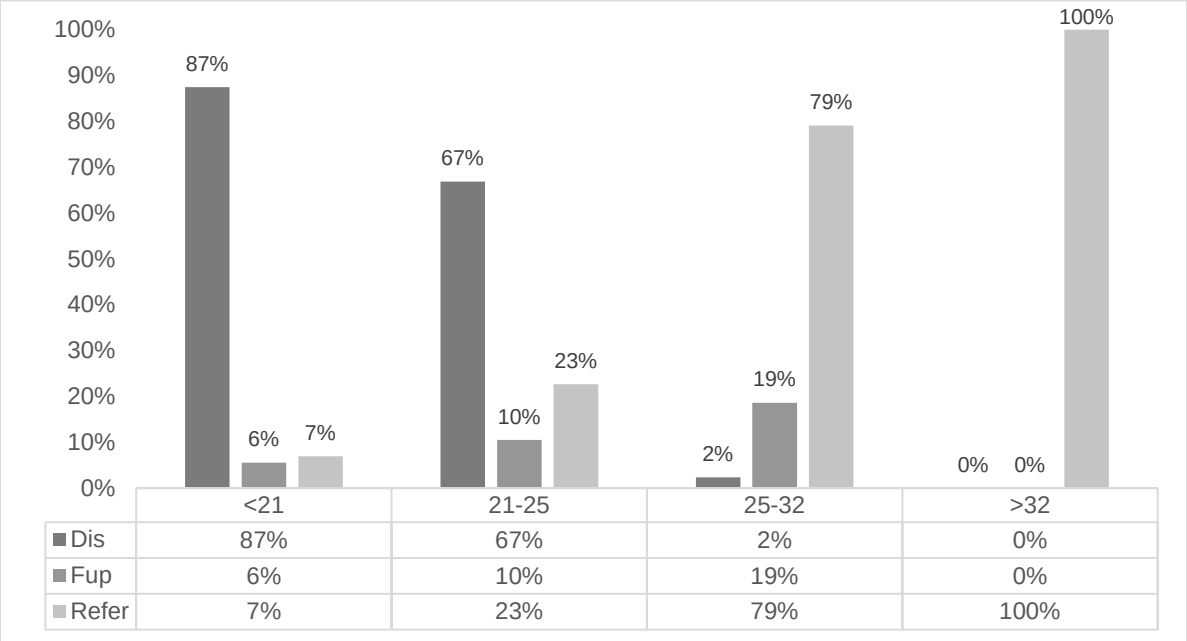
	Mother	Father	Sibling
Dis	87%	77%	83%
Fup	5%	13%	7%
Refer	8%	10%	10%

Figure 3.17 Percentage Dis, Fup and Refer for each family history category, pooled across the three COT optometrists (JG, DH, NOK) contributing data to this study N=1006.



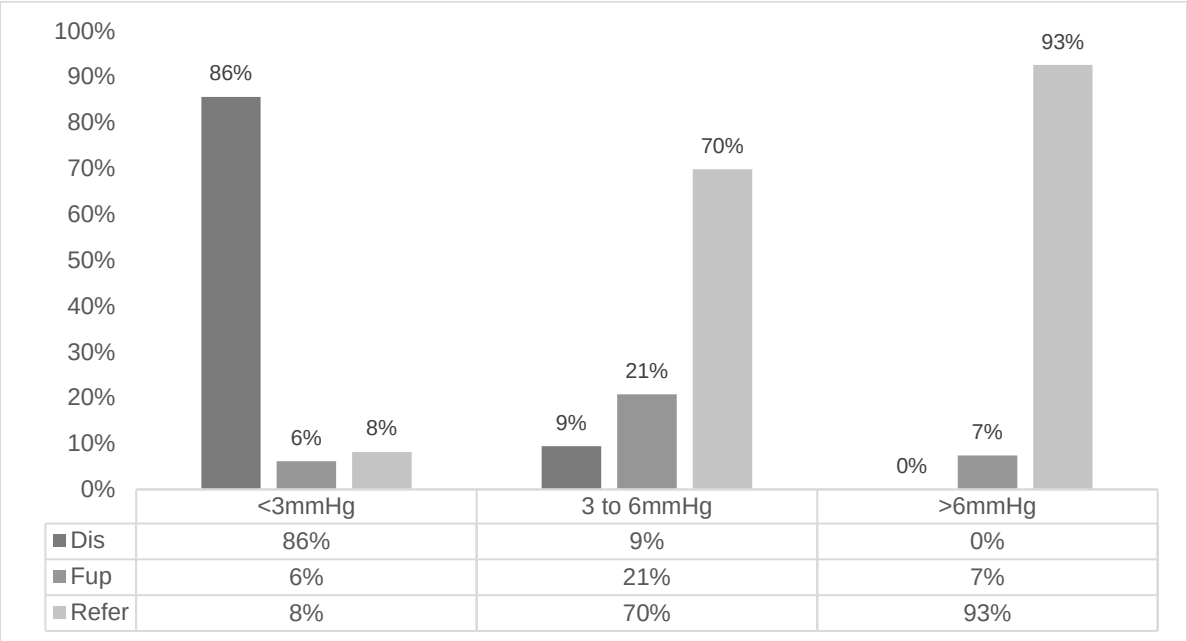
	IOP	Discs	VF	IOP/Dscs	IOP/VF	VF/Discs
Dis	85%	86%	72%	60%	78%	30%
Fup	8%	3%	10%	8%	11%	9%
Refer	7%	11%	18%	31%	11%	60%

Figure 3.18 Percentage Dis, Fup and Refer for each RFR category, pooled across the three COT optometrists (JG, DH, NOK) contributing data to this study N=1006.



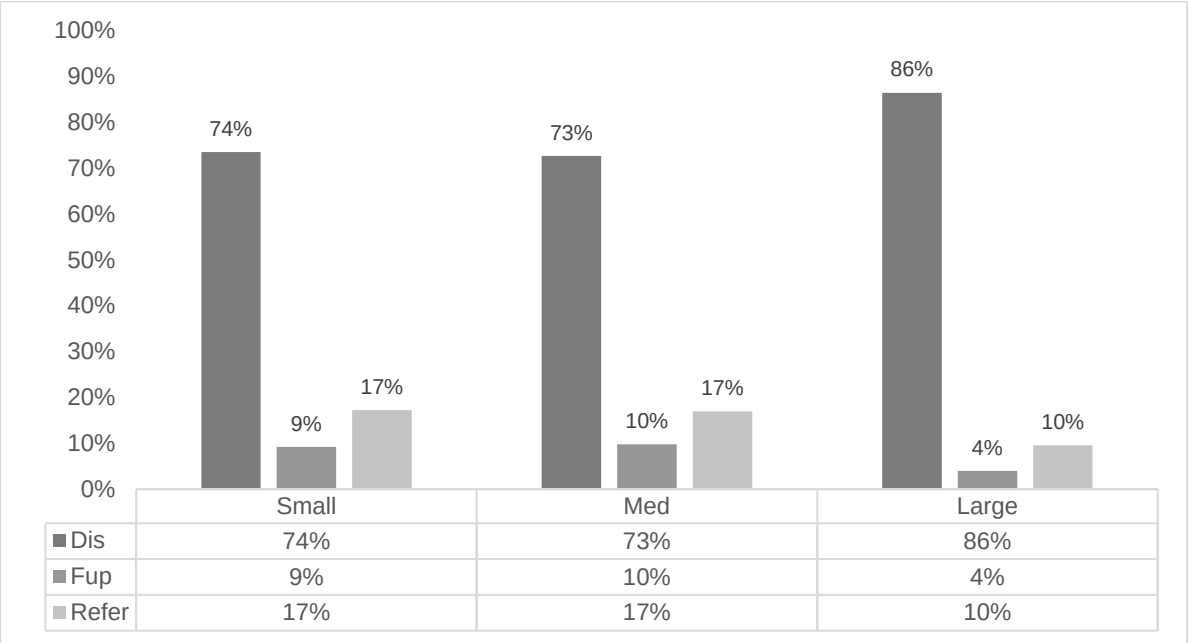
	<21	21-25	25-32	>32
Dis	87%	67%	2%	0%
Fup	6%	10%	19%	0%
Refer	7%	23%	79%	100%

Figure 3.19 Percentage Dis, Fup and Refer for each GAT IOP category, pooled across the three COT optometrists (JG, DH, NOK) contributing data to this study N=1006.



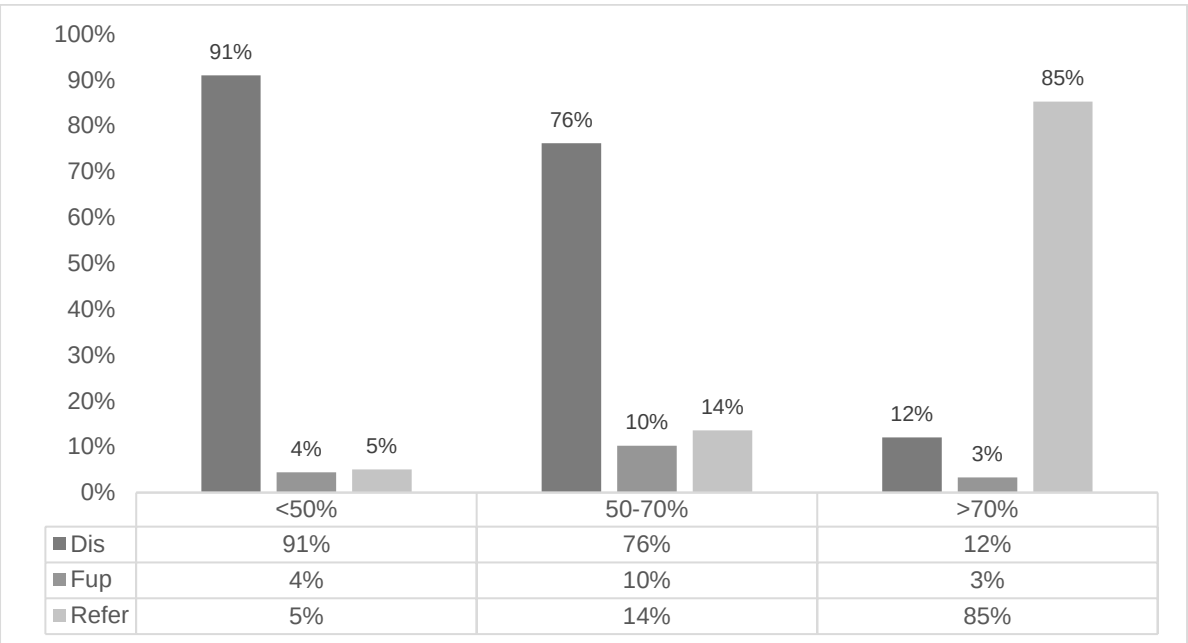
	<3mmHg	3 to 6mmHg	>6mmHg
Dis	86%	9%	0%
Fup	6%	21%	7%
Refer	8%	70%	93%

Figure 3.20 Percentage Dis, Fup and Refer for each IOP diff category, pooled across the three COT optometrists (JG, DH, NOK) contributing data to this study N=1006.



	Small	Med	Large
Dis	74%	73%	86%
Fup	9%	10%	4%
Refer	17%	17%	10%

Figure 3.21 Percentage Dis, Fup and Refer for each VDS category, pooled across the three COT optometrists (JG, DH, NOK) contributing data to this study N=1006.



	<50%	50-70%	>70%
Dis	91%	76%	12%
Fup	4%	10%	3%
Refer	5%	14%	85%

Figure 3.22 Percentage Dis, Fup and Refer for each VCDR category, pooled across the three COT optometrists (JG, DH, NOK) contributing data to this study N=1006.

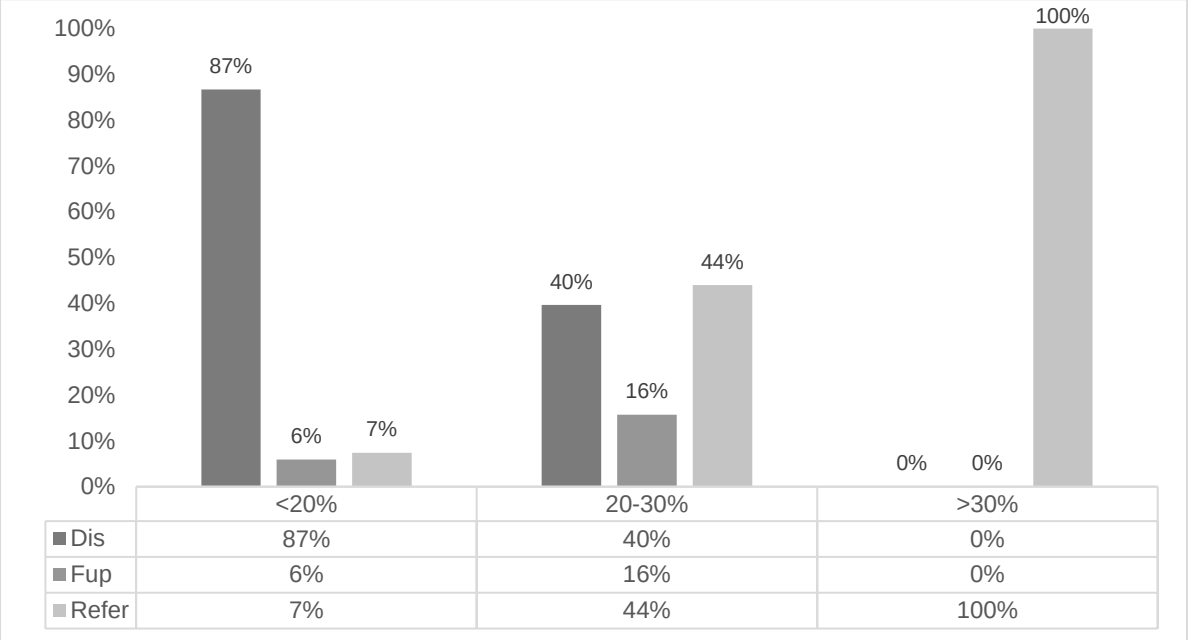


Figure 3.23 Percentage Dis, Fup and Refer for each VCDR diff category, pooled across the three COT optometrists (JG, DH, NOK) contributing data to this study N=1006.

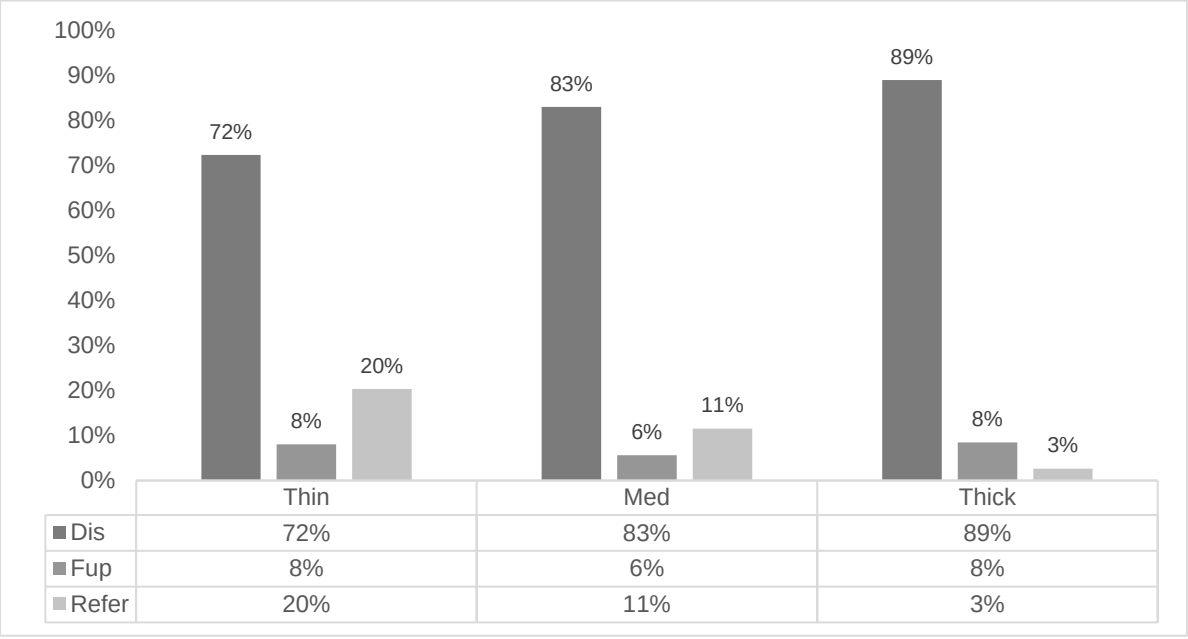


Figure 3.24. Percentage Dis, Fup and Refer for each CCT category, pooled across the three COT optometrists (JG, DH, NOK) contributing data to this study N=1006.

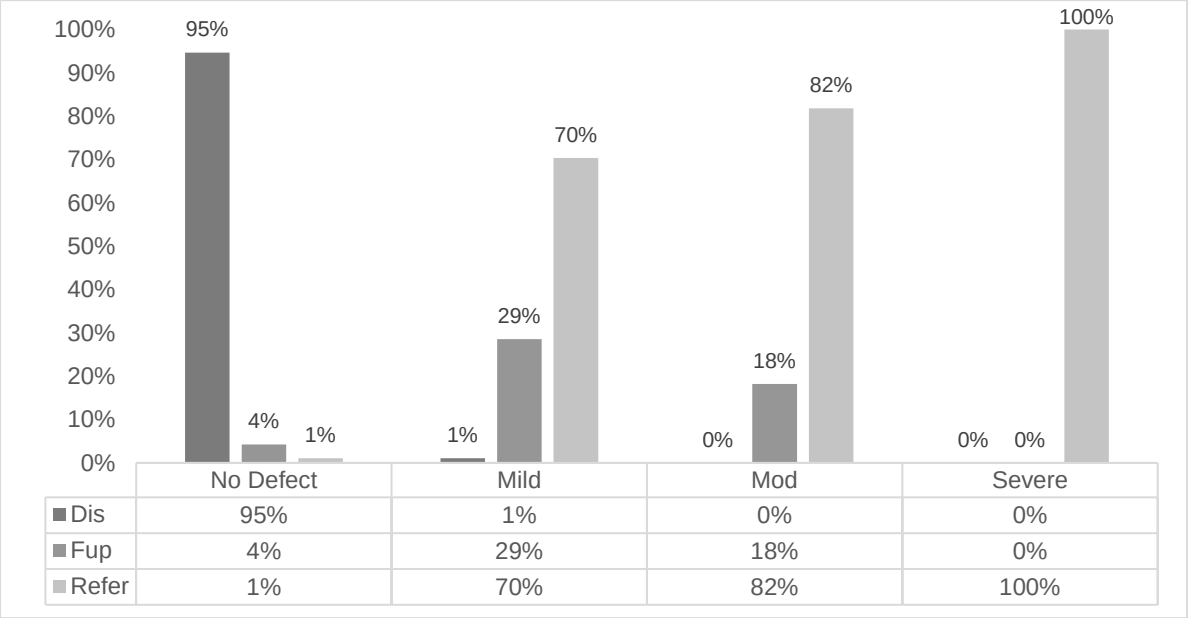


Figure 3.25 Percentage Dis, Fup and Refer for each HPA classification, pooled across the three COT optometrists (JG, DH, NOK) contributing data to this study N=1006.

Clinical tests	Influence on referral rates (Chi-square, degrees of freedom (df), P-value)	Agrees with other studies	Disagrees other studies
Age	Referral rates increased with age (Chi-square = 77.34, df = 10, P = <0.0001)	(168)	
Race	Hispanics and Africans were referred most often (Chi-square = 28.55, df = 8, P = <0.001)	(41, 47, 48, 121, 130)	
Sex	Males were referred most often (Chi-square = 12.06, df = 2, P = <0.01)	(42, 44, 133)	(134)
Family history	Family history of COAG did not influence referrals (Chi-square = 4.72, df = 4, P = NS).	(140)	(135, 136)
RFR	Multiple RFRs were referred most often (Chi-square = 97.92, df = 2, P = <0.0001)	(110, 169)	
GAT IOP	Referral rates increased with GAT IOP (Chi-square = 282.48, df = 6, P = <0.0001)	(79, 139, 140, 170-172)	
IOP Diff	Referral rates increased with IOP diff (Chi-square = 338.59, df = 4, P = <0.0001)	(142)	
VDS	Referral rates increased as VDS reduced (Chi-square = 29.25, df = 4, P = <0.0001)		(150)
VCDR	Referral rates increased with VCDR (Chi-square = 311.24, df= 4, P = <0.0001)	(140)	
VCDR Diff	Referral rates increased with VCDR Diff (Chi-square = 279.88, df = 4, P = <0.0001)	(158)	
CCT	Referral rates increased as CCT decreased (Chi-square = 39.62, df = 4, P = <0.0001)	(140, 162)	
HPA Class	Referral rates increased with HPA severity (Chi-square = 849.79, df = 6, P = <0.0001)	(163)	

Table 3.3 Associations between each clinical test and COT management outcomes showing agreement or disagreement with previous studies.

Table 3.3 shows that all but one of the findings of the present study agreed with at least one previous study. The association between VDS and COT management decisions was the only finding that disagreed with previous research. Figure 3.21 shows that referral rates increase as VDS is reduced. These results disagree with a previous study (149) that found no evidence for optic disc size being an independent risk factor for COAG (150). It must be remembered that only univariate analyses are shown in this chapter. Analyses of this type are prone to showing erroneous associations caused by hidden covariations between variables. There is, however, another possible explanation for the association found in the present study. Table 3.2 shows that there was a highly statistically significant degree of inhomogeneity between the VDS distributions (see Figure 3.8) found in the samples seen by each COT optometrist. This finding was surprising as all three used the same equipment for recording VDS (see section 2.3.7).

Despite agreement with previous studies, the association between racial groups and COT management decisions must be interpreted cautiously as the study sample was predominantly Caucasian.

The association between RFR and COT management decisions was revisited. Figure 3.26 shows that referrals into the COT with multiple RFR had lower percentage discharge and higher percentage referral rates than those with a single RFR. The association was statistically significant (Chi-square = 97.92, degrees of freedom = 2, P = <0.0001) and suggests that false positive referrals would reduce if optometrists referred patients with more than one suspicious test finding.

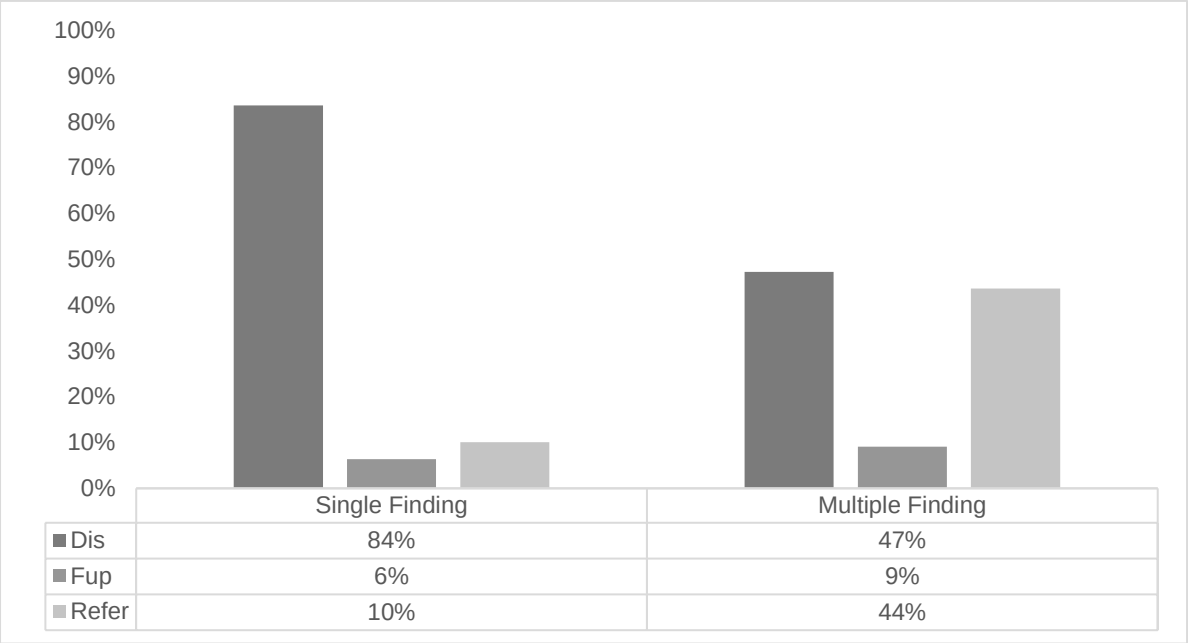


Figure 3.26 Percentage Dis, Fup and Refer for single and multiple RFRs, pooled across the three COT optometrists (JG, DH, NOK) contributing data to this study N=1006.

3.6 The influence of CCT on GAT IOP

The fourth objective was to examine the influence of CCT on GAT IOP in order to show how it might impact on the agreement between (a) the referrals made to the COT for raised IOP and (b) COT measurements of IOP. Higher IOP readings are associated with increased CCT (173-175). Figure 3.27 shows the statistically significant linear relationship between GAT IOP and CCT found in the present study ($r = 0.027$, $p < 0.001$). For every 50 microns increase in CCT there was a 0.96 mmHg increase in IOP. The rate of increase found is in keeping with previous studies which reported between 0.90 to 1.00 mmHg increase IOP for every 50 microns of CCT increase (174).

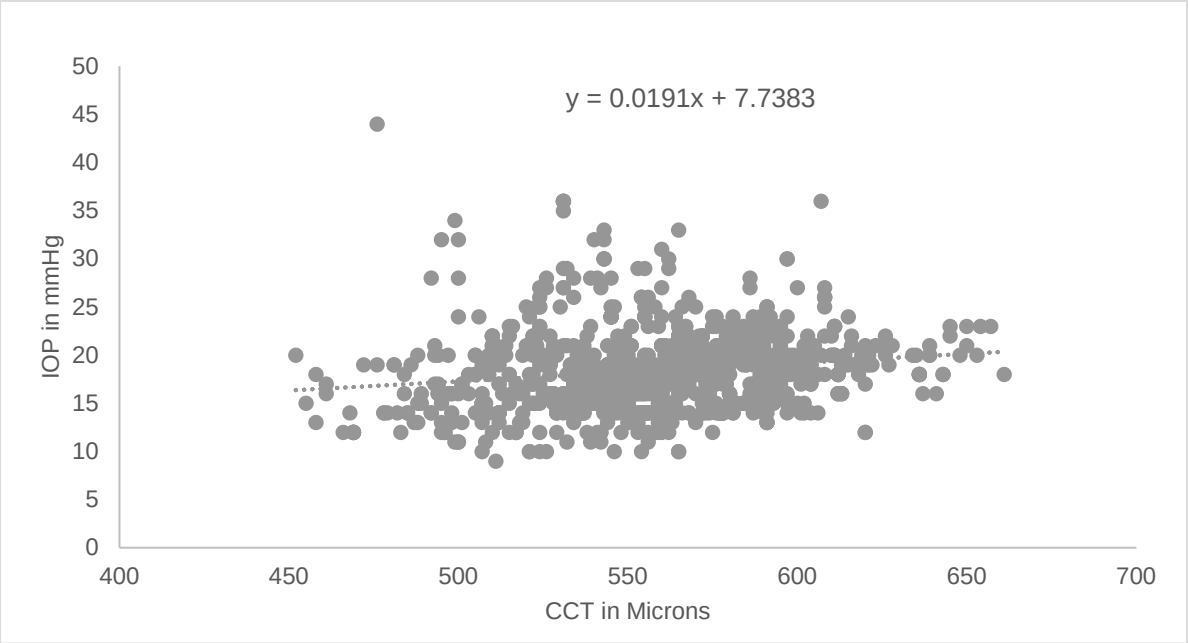


Figure 3.27 Relationship between GAT IOP and CCT measured in the present study N=1006.

The question was, could variations in CCT account for the occurrence of cases referred into the COT with raised IOP that were subsequently found to have normal IOP? Figure 3.28 shows that 405 cases were referred in to the COT with raised IOP. Of these, only 159 (39%) had raised GAT IOPs (>21mmHg).

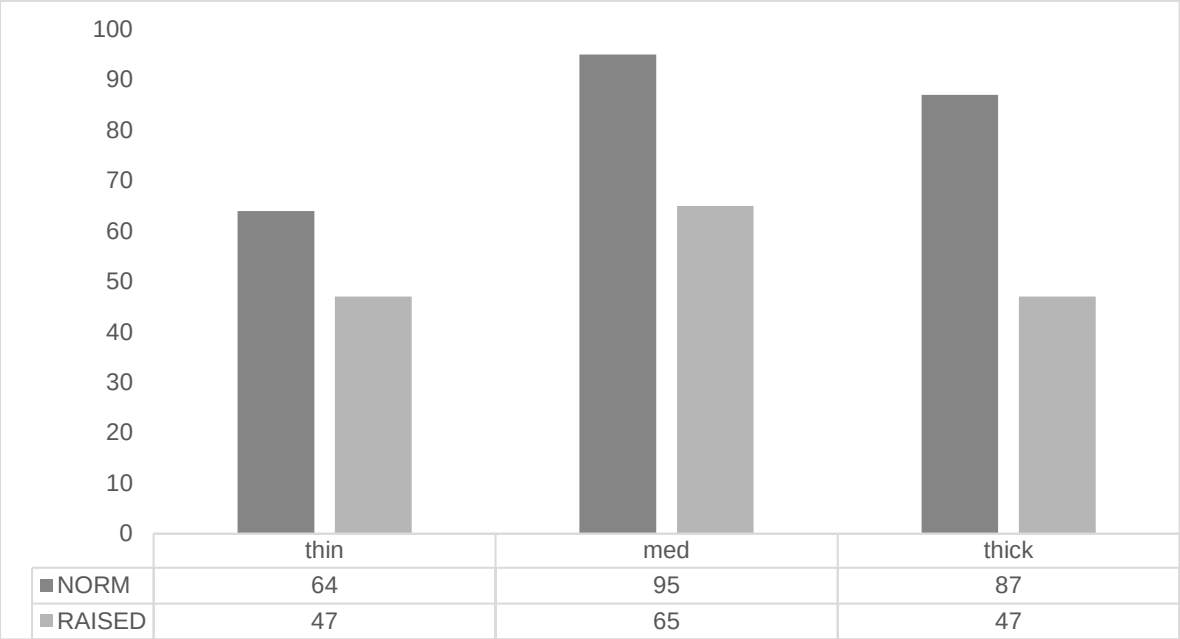


Figure 3.28 Frequencies of cases that had been referred in to the COT with raised IOP but were subsequently found to have normal (norm) or raised GAT IOPs (>21mmHg) after dividing the sample into those with thin, medium (med) or thick CCTs N=405.

Several reasons may account for this finding. Firstly, Non-Contact Tonometry (NCT) tends to read higher than GAT (176, 177). Differences of at least 2 mmHg have been reported when comparing various tonometers to GAT (178). NCT is also more affected by CCT than is GAT (179) and gives a greater overestimation of IOP with thicker CCT and steeper curvatures than occurs with GAT (180). The majority of community optometrists use NCT as the primary device for recording IOP (181, 182). Indeed, a study on the accuracy of referrals to a glaucoma clinic from community optometrists revealed that only 2% were based on contact tonometry (109). All COT tests were carried out at a single COT visit, importantly this leads to limitation with respect to regression to the mean (RTM). RTM may bias an investigation when it is based upon an initial value without a control group (183). For example in this study IOP was measured at one given time. GAT IOP is known to have diurnal variations (184).

An attempt was made to determine whether variations in CCT could explain the observed differences between (a) the referrals made to the COT for raised IOP and (b) COT measurements of IOP. Chi-square was used to compare frequencies of cases that had been referred in to the COT with raised IOP but were subsequently found to have normal (norm) or raised GAT IOPs (>21mmHg) after dividing the sample into those with thin, medium (med) or thick CCTs. The expectation was that, if CCT variations were influential, a greater proportion of cases with thicker corneas would be found to have normal GAT IOP. Figure 3.28 shows some evidence for this as the proportion of cases with normal GAT IOP increased from (64 / [64 + 47] =) 0.577 in those with thin CCTs, through (95 / [95 + 65] =) 0.594 in those with medium CCTs to (87 / [87 + 47] =) 0.649 in those with thick CCTs. However, the trend was not statistically significant (Chi-square = 1.55, degrees of freedom = 2, P = NS).

3.7 The influence of VDS on VCDR

The fifth objective was to examine the influence of VDS on VCDR. Previous research has found that VCDR increases with VDS (144, 151, 185). Figure 3.29 shows the statistically significant linear relationship between VCDR and VDS found in the present study ($r = 0.335$, $p < 0.0001$). For every 1 mm increase in VDS there was a 24% increase in the VCDR. The rate of increase found was similar to that reported in a previous study (27%) (186).

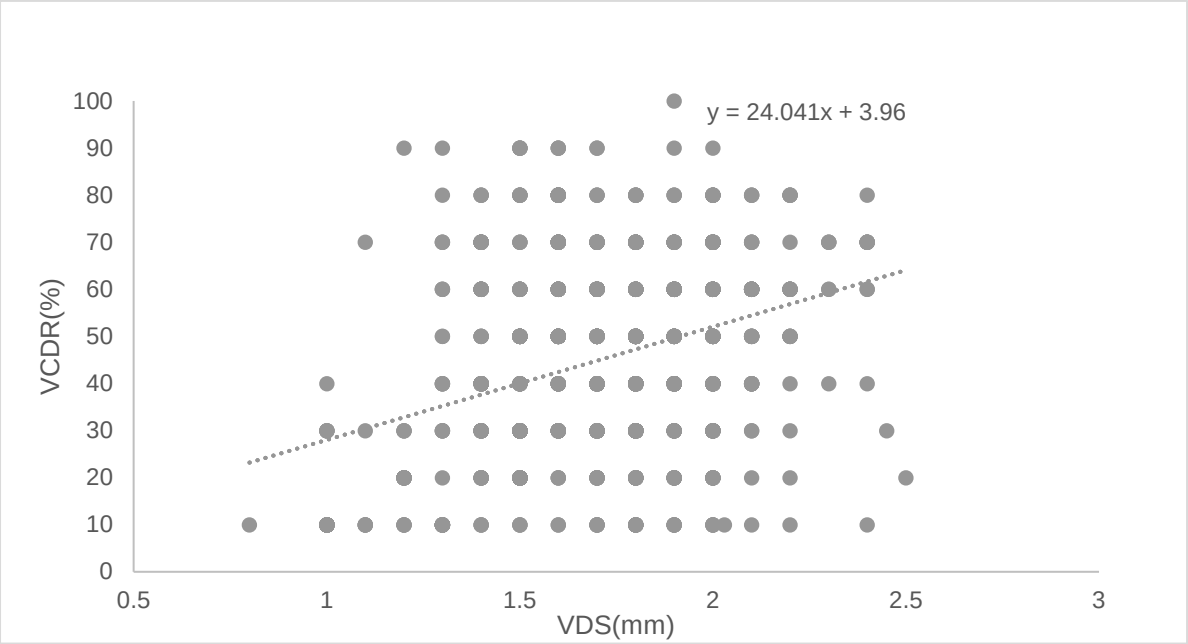


Figure 3.29 Relationship between VCDR and VDS N=1006.

3.8 Comparison between RFRs for suspect VF and COT measurements of HPA

The sixth objective was to compare reasons for referrals in to the COT (RFRs) for suspect VF and subsequent COT measurements of HPA. Some interesting observations arose. On the one hand, 16% of the cases referred in to the COT with suspect VF subsequently had normal VFs according to the HPA classification system. On the other hand, some (9%) of cases had HPA defects having not been referred in to the COT with suspect VF.

One reason for these discrepancies could be learning affects (187). The COT effectively repeats the VF test for at least a second time. Previous literature has recommended repeat VF testing for borderline cases (187). Simply repeating a VF test can improve patient understanding leading to greater reliability (188).

Another reason for the observed differences was likely to be that a variety of VF tests were used by optometrists who made referrals in to the COT (the VF test used was not recorded as part of this study). Variations in VF test design are known to influence their ability to detect defects (189). It might also be that no VF test had not been performed.

3.9 FDRs and cost savings of the COT

The seventh objective was to examine FDRs and cost saving bought about by the activities of the COT. As explained in chapter 1 (section 1.8.1), the COT was established to allow a multi-disciplinary approach to benefit patient care. Due to increasing population demands and longevity, cost benefits are also important to the NHS and CCGs.

A key finding of the present study relates to its high FDRs; where the patient is discharged following a COT appointment as no disease is present. The FDR in the present study (79%, see Figure 3.13 and Table 3.1) compares favourably with FDRs reported for other schemes (see Table 1.5) operated by specialist optometrists (38, 105, 113, 114) and non-specialist optometrists (111). Discharge rates may well have been higher than many schemes which use repeat measures or enhanced case finding as the COT truly does refine the referrals in accordance with NICE recommendation(115) with more extensive testing involving gonioscopy. The extra evidence gathered by the COT would lead to some patients with the lower risk of conversion to COAG not being retained within the COT system compared to other schemes. Interestingly, schemes which simply repeat measurements have a similar FDRs (105, 112). So, although the COT can save considerable costs, it might be that the same or even greater savings can be made from simpler GRR schemes. RTM may explain the tendency for high FDRs in simpler schemes.

The COT practitioner receives £65.65 per patient from the CCG for the first GRR visit. The cost of equipment and other sundries such as ocular drugs are borne by the COT practitioner. In comparison, the CCG currently pays the HES a consultant-led first attendance multi professional HES tariff of £125 x 1.1095 MFF; where MFF is the Market Forces Factor and amounts to £138.69 (190). The MFF is an estimate of unavoidable cost differences between healthcare providers dependent upon where they are located. The £65.65 received by COT practitioners includes the MFF. These figures allow a direct comparison of the costs of referral to the COT versus the HES (rounded down to the nearest pound). The cost for the first visit of 1006 cases seen in the COT was (1006 x 65.65 =) £66,043 added to this were the costs of COT referrals to the HES of 138 cases for confirmation of POAG. At £138.69 per case, which amounted to (138 x 138.69 =) £19,139. So the total cost amounted to (66043 + 19139 =) £85,182. In comparison, the cost of 1006 cases seen by the HES would have been (1006 x 138.69 =) £139,522. The total cost saving was, therefore, (139,522 – 85,182 =) £54,340 which was (54,340 ÷ 1006 =) £54 per case. Administration is fairly cost neutral as the same PCBS staff who are employed by the CCG are responsible for bookings on behalf of both the COT and HES.

Previous studies (38, 105, 113) have reported cost savings varying from £9 (106) to £117 (113) per case. Previously reported cost analyses have accounted for the cost of follow up appointments (105) or have made assumptions about the number of patients being seen and the minimum percentage that will not need to have been referred to the HES (38). A limitation of the present study was that neither of these factors were built into the cost analysis for the COT.

3.10 Ranking of COT tests

The final objective was to rank COT tests in order of the strength of their statistical associations with COT management outcomes. The simplest method of doing this (Method A) was to rank tests in order of the Chi-square values shown in Table 3.3. However, the degrees of freedom (df) of a Chi-square test is also important when establishing whether an association is statistically significant. Therefore, after seeking statistical advice (191), a second method of ranking tests (method B) was used in which Chi-square was divided by the degrees of freedom; Chi-square per degree of freedom. The rationale behind the second method follows. With 1 degree of freedom, a Chi-square value of >3.841 indicates a statistically significant association at the 95% level. With 10 degrees of freedom, a Chi-square value of >18.31 indicates the same. So, considering Chi-square alone (method A) could lead to overestimated strength of an association and corrupt test rankings. Dividing 18.31 by 10 gives a Chi-square per degree of freedom of 1.831 which is closer to 3.841. Though method B is not perfect, greater confidence in the outcomes can, at least, be inferred if its rankings match those of method A.

Table 3.4 shows all COT tests ranked according to methods A and B. Both methods gave a fairly consistent indication of the ranking positions of each test; with differences of no more than one ranked position. The test ranks shown in Table 3.4 are compared to those based on likelihood ratios in chapter 4 (see Table 4.7).

Method A Test (Chi-square)	Method B Test (Chi-square per degree of freedom)
HPA class (850)	HPA class (142)
IOP diff (339)	IOP diff (85)
VCDR (311)	VCDR (78)
GAT IOP (282)	**VCDR diff (70)**
VCDR diff (280)	**GAT IOP (47)**
RFR (98)	RFR (15)
Age (77)	**CCT (10)**
CCT (40)	**Age (8)**
VDS (29)	VDS (7)
Race (29)	**Sex (6)**
Sex (12)	**Race (4)**
Family History (5)	Family History (1)

Table 3.4 COT tests ranked in order of the strength of their statistical associations with COT management outcomes. Rankings are compared for two methods: Method A based on Chi-square and method B based on Chi-square per degree of freedom. Differences in ranking between the two methods are highlighted in bold.

3.11 Summary

Key findings arising from the 8 objectives outlined in section 3.1 of this chapter are listed below:

- Figures 3.1 to 3.13 and Table 3.1 show the distribution of the COT's clinical test findings, according to the multiple cut-off points introduced in chapter 2, for each of the optometrists that contributed data to this study.
- Findings summarised in Table 3.2 indicated that 6 out of the 13 COT clinical tests showed inhomogeneous distributions between the COT optometrists. The inhomogeneity between observers, particularly with VDS is an important limitation of the study. This could also affect the outcomes of investigations into the consistency of naïve Bayes applied separately to their data in chapter 4.

- Figures 3.14 to 3.25 and Table 3.3 show the relationship between each clinical test and the COT management outcomes. All but one of the findings (that relating to VDS) corroborated previous research. Study of the influence of race was limited as the sample was predominantly Caucasian. The ethnicity of the sample does not allow for judgements to be made on any other ethnic group than Caucasian.
- Referrals to the COT were more likely to indicate COAG if based on multiple RFRs.
- GAT IOP increased with CCT (Figure 3.27), as had been reported in previous studies. However, variations in CCT did not explain observed differences in referrals made to the COT for raised IOP and COT measurements of IOP (Figure 3.28).
- VCDR increased with VDS (Figure 3.29), as had been reported in the previous literature.
- Observed differences between referrals made to the COT for suspect VF and COT measurements of HPA were considered to arise from the use of a variety of different VF tests by community optometrists.
- FDRs compared favourably to those reported for other GRR schemes. Estimated cost savings of £54 per case arose from the activities of the COT.
- COT tests were ranked in order of the strength of their statistical associations with COT management outcomes (Table 3.4) for later comparison with ranks based on likelihood ratios in chapter 4 (see Table 4.7).

Naïve Bayes is applied in the next chapter in order to determine whether it can provide useful diagnostic support in the referral refinement of COAG.

Chapter 4: Can naïve Bayesian artificial intelligence predict COT decisions?

4.1 Introduction

As described in chapter 1, the primary aim of this study was to determine whether naïve Bayesian artificial intelligence (Bayes) could be used to predict the decisions of COT optometrists (JG, DH and NOK). This chapter provides an account of how the primary aim was achieved. Bayes was applied as described in section 1.4. The clinical data that Bayes was applied to is that summarised in Table 3.1 (section 3.3). The material presented in this chapter represents completion of the following objectives:

- To establish the best means of determining the accuracy of Bayes from current literature on machine learning (section 4.2);
- To determine the extent to which COT clinical test findings were truly independent of each other; the key assumption of naïve Bayes (section 4.3);
- To plot learning curves in order to determine how much data is required to achieve maximum learning when using Bayes (section 4.4);
- To investigate the transferability of likelihood ratios generated during the application of Bayes (sections 4.5 and 4.7);
- To determine the accuracy of Bayes and, in the event that accuracy was insufficient, to explore any improvement of accuracy using cost-sensitive learning (section 4.6);
- To determine, using likelihood ratios, whether redundancy existed in the COT clinical data (section 4.8);
- To rank COT clinical tests (for comparison with those shown in section 3.10) based upon their likelihood ratios (section 4.9).

4.2 Evaluating accuracy

There are numerous methods of evaluating the accuracy of learning schemes. These are all based on confusion matrices that show the frequency of correct and incorrect decisions made by the learning scheme (see Table 4.4 later).

Baseline performance of any learning scheme is based on ZeroR; the prior probability of the commonest clinical decision. Evaluation of the learning scheme itself involves learning from a training dataset and testing what has been learned on a testing dataset. Using the same dataset for training and testing leads to optimistic predictions of accuracy but is considered to provide an estimate of the best possible performance of a learning scheme (192). Ideally, the testing dataset should not be involved in training but an acceptable alternative is to perform randomised stratified tenfold cross-validation. Here, the dataset is divided into 10 folds. The cases included in each fold are selected randomly and stratification ensures that each clinical decision (Dis, Fup and Refer) is equally represented in each fold. Each fold in turn is then assigned as the testing dataset with all other folds used for training. Treating the data in this way is considered to offer the most realistic prediction of the accuracy of the learn scheme (192).

Measures of accuracy are also numerous (193). The simplest of these is Rand accuracy and is the probability that the learning scheme makes a correct decision (193). Cohen's Kappa is ubiquitous and provides an estimate of the agreement between the learning scheme and the COT optometrists after removing the proportion of agreement considered to be due to chance.

Accuracy of a learning scheme can also be expressed in the form of sensitivity (AKA recall, true positive rate and hit rate) and specificity (AKA inverse recall and true negative rate), both terms being well known to clinicians (193). These can be combined into a single term called informedness which is an estimate of the ability of a learning scheme to correctly test positive or negative above chance (193).

Informedness cannot, however, be used to estimate how well the learning scheme predicts an outcome. To do this, a term called markedness combines predictive values for positive (PPV, AKA precision) and negative (NPV, AKA inverse precision) findings and indicates how well the learning scheme makes these predictions above chance (193). Other frequently used estimates of learning scheme performance are the area under the receiver operating curves (AUC) and the Matthew's correlation coefficient (MCC). The F-measure, which combines sensitivity and PPV, has been considered biased to predictions of positive outcomes; while the machine learning community seems to be solely interested in detection of positive outcomes, the healthcare community is just as interested in negative outcomes (193).

Learning curves (see section 1.4.3) provide a means of determining how much data is required before the learning scheme achieves maximum accuracy (193).

Incorrect decisions made by learning schemes carry costs. These can be overcome by adopting cost sensitive learning (190). Cost learning takes misclassification of costs into account ,treating each of these misclassifications in a different manner in an effort to reduce the total cost (194). A person discharged when they should have been followed up or referred is at risk of avoidable blindness. On the other hand, a person referred when they should have been followed up or discharged suffers avoidable distress and the receiving HES glaucoma unit suffers avoidable burden. These costs are difficult to estimate exactly but adding differential costs to the learning scheme may reduce the costs of incorrect decisions (192).

4.3 Correlations between clinical tests

Table 4.1 shows that many of the clinical tests were inter-correlated. The data, therefore, did not obey the primary assumption of naïve Bayes'; that all clinical tests are independent of each other. Despite tests not being independent high levels of accuracy were anticipated as described (section 1.4).

Kendall's rank correlation was used for all tests as the data variables were discrete (bucketed) in nature. Spearman's correction was not used as this would be affected by how far the ranks were apart. A Bonferoni correction was considered which adjusts probability due to the risk of increased type 1 errors when making multiple statistical tests (195). The use of this correction has been considered to be contentious (195) as If applied to this data it was felt the number of false negatives would increase and reduce power. Furthermore this correction can often be regarded as being too conservative (195) and that if it were applied to this data then it could look like it was less inter related than initially thought so was not used in this study. Had this correction been used in this study the Bonferoni correction alpha adjusted level would be 0.05/77= p=0.00065.

	age	FH	RFR	IOP	IOP diff	VDS	VCDR	VCDR diff	CCT	HPA	Management
race	-0.11				0.09		0.10		-0.07	0.07	0.09
sex	-0.07	0.08			-0.07		-0.13			-0.12	-0.09
age			0.10	0.09	0.15		0.17	0.10	-0.15	0.20	0.18
FH				-0.04		0.11		-0.07		-0.10	-0.08
RFR				0.12	0.12		0.26	0.16		0.28	0.27
IOP					0.36	-0.14	0.09	0.16		0.32	0.39
IOP diff						-0.07	0.17	0.33	-0.10	0.47	0.54
VDS							0.16	-0.08		-0.12	-0.15
VCDR								0.20	-0.19	0.31	0.35
VCDR diff									-0.12	0.42	0.46
CCT										-0.16	-0.16
HPA											0.85

Table 4.1 Matrix showing statistically significant (P < 0.05) inter-correlations (Kendall's tau for 2-tailed tests) between the clinical tests included in this study. Variables included race (Caucasian, Asian, Afro-Caribbean, African, Hispanic), sex (male, female), age (<40, 40-49, 50-59, 60-69, 70-79, 80+), family history of COAG (FH, a count from 1 to 3 based on how many types of relative - mother, father and sibling – had glaucoma), reason for referral (RFR, a count from 1 to 3 based on how many suspicious signs - IOP, optic discs and/or VF – were present in the referral), GAT intra-ocular pressure (IOP, mmHg: <21, 21-25, >25-32, >32), GAT inter-ocular difference in intraocular pressure (IOP diff, mmHg: <3, 3-6, >6), vertical disc size (VDS, mm: <1.4, 1.4-1.7, 1.8+), vertical cup disc ratio (VCDR, %: <50, 50-70, >70), inter-ocular difference in VCDR (VCDR diff, %: <20, 20-30, >30), Central corneal thickness (CCT, <555, 555-590, >590) and visual field loss (HPA classification: mild, moderate, severe).

4.4 Learning curves

As one of the COT optometrists (NOK) saw 69 cases, learning curves were constructed, using the AEL Bayes' application (see section 1.4.3), for the first 69 case of all three COT optometrists. Learning was initially carried out on the first case followed by testing on all 67 cases. Learning was repeated a further 22 times, adding three more cases to the learning set each time prior to testing on all 69 cases. Figure 4.1 shows that Bayes learned rapidly and achieved maximum Rand accuracy (between 94 and 99%) even before including all 69 cases in the learning process. This suggested that lack of adherence to the naïve Bayes' assumption of independence (see section 1.4) had little effect on accuracy.

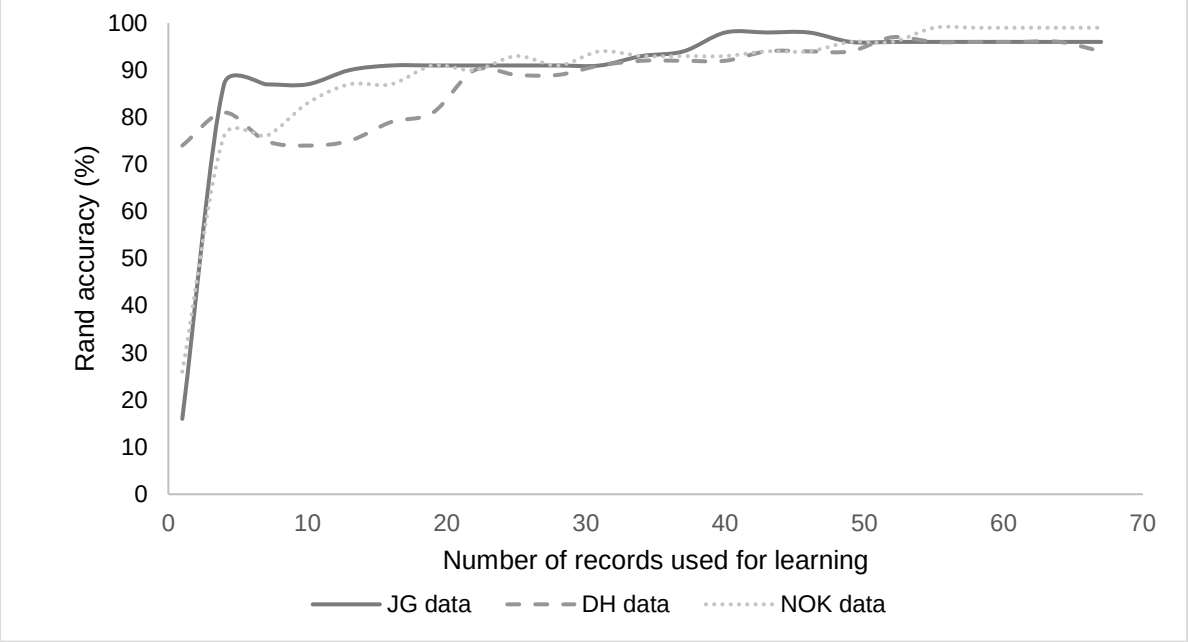


Figure 4.1 Learning curves constructed using the AEL Bayes' application for the first 69 cases seen by each COT optometrist (JG, DH and NOK).

4.5 Initial investigation of transferability of LRs

The three sets of likelihood ratios generated by the AEL Bayes' application after learning from all 69 cases seen by each COT optometrist should, in theory, have been transferable. That is, the high Rand accuracy shown in Figure 4.1 should have been maintained when the likelihood ratios from one optometrist were tested on the clinical data of another, which was investigated and the results shown in Table 4.2.

Training set	Testing set	Weighted average Rand accuracy	Kappa
JG	JG	0.97	0.83
	DH	0.81	0.37
	NOK	0.90	0.60
DH	DH	0.97	0.87
	JG	0.81	0.18
	NOK	0.85	0.40
NOK	NOK	0.98	0.91
	JG	0.94	0.63
	DH	0.84	0.41

Table 4.2. Investigation of the transferability of likelihood ratios generated from the 69 cases seen by each COT optometrist (the training set) and tested on their own clinical data and that of their colleagues (the testing set).

The weighted average Rand accuracy was calculated from the three separate estimates of Rand accuracy for each COT management outcome (Dis, Fup, Refer). Kappa is also shown in the table. It can be seen that Rand accuracy and kappa reduced when one optometrist's likelihood ratios were applied to another's clinical data. These findings was disappointing as it indicated that likelihood ratios were not transferable. The decision was made to re-calculate likelihood ratios on a dataset comprising 1006 cases pooled from all three COT optometrists. Maybe these would be more transferable.

4.6 Baseline, maximum and realistic predicted accuracy

The purpose of this section was to determine baseline accuracy, maximum accuracy and the most realistic estimate of accuracy for 1006 cases, pooled from all three COT optometrists.

The machine learning community suggest that ZeroR is the baseline accuracy against which the performance of a learning scheme, such as Bayes, should be judged (190). ZeroR is equal to the prevalence of the most commonly occurring outcome. In this case, it was discharge which occurred for 798 of the 1006 cases; a ZeroR of 0.79.

Although training and testing on the same dataset is known to give over optimistic estimates of accuracy, the machine learning community suggest that this process does at least provide the maximum possible accuracy of a learning scheme (190).

Randomised stratified tenfold cross-validation is an established method of estimating the most realistic accuracy of a learning scheme.

Table 4.3 shows that surprisingly little difference arose between maximum and most realistic accuracy measured in terms of Rand accuracy, Kappa, informedness, markedness, AUC and MCC. The table also clearly shows that Fup was predicted relatively poorly compared to the other COT outcomes (Dis and Refer). Rand accuracy appeared to be relatively unresponsive to variations in performance. The other measures (informedness, markedness, AUC and MCC) varied a lot more and in different ways. For example, Rand accuracy indicated that performance was about equal for the outcomes Dis and Refer. However, informedness and AUC were highest for the outcome Refer while markedness and MCC was highest for the outcome Dis. Variations in informedness reflected those in AUC; likewise, for variations in markedness and MCC. Therefore, performance could be judged based on informedness and markedness alone.

ZeroR 0.79	Rand accuracy	Kappa	informedness	markedness	AUC	MCC
Maximum accuracy (trained and tested on same dataset)						
Overall	0.96	0.79	0.80	0.88	0.90	0.83
Discharge	0.96	-	0.83	0.91	0.91	0.87
FUP	0.94	-	0.31	0.68	0.65	0.46
Refer	0.96	-	0.88	0.81	0.94	0.84
Most realistic accuracy (randomised stratified tenfold cross-validation)						
Overall	0.95 (0.02)	0.76 (0.05)	0.78 (0.07)	0.85 (0.06)	0.89 (0.03)	0.82 (0.06)
Discharge	0.96 (0.02)	-	0.82 (0.08)	0.90 (0.05)	0.91 (0.04)	0.86 (0.06)
FUP	0.93 (0.01)	-	0.22 (0.19)	0.38 (0.37)	0.61 (0.09)	0.31 (0.23)
Refer	0.95 (0.02)	-	0.85 (0.08)	0.79 (0.08)	0.93 (0.04)	0.82 (0.07)

Table 4.3 Maximum and most realistic accuracy of naïve Bayes' against the ZeroR baseline for 1006 cases seen by the COT, pooled from the cases seen by JG, DH and NOK. Rand accuracy, informedness, markedness, area under the receiver operating curve (AUC) and Matthew's correlation coefficient (MCC) are shown for all three COT outcomes (discharge, FUP and Refer) along with an overall weighted average. Kappa only applies to overall performance. Standard deviations arising from tenfold cross-validation are shown in brackets.

This is interesting, given that Bayes' theorem illustrates the 'base rate fallacy' (196) in which it is assumed that forward probability (measured in terms of sensitivity and specificity) is equal to backward probability (measured in terms of positive and negative predictive value). In other words, 'base rate fallacy' refers to misinterpretation of clinical data.

In the context of, COAG, this refers to the mistaken belief that a positive result, for a COAG test with 90% sensitivity and specificity, means that a patient has a 90% probability of having COAG. At least one study has shown that 50% of medical practitioners make this mistaken interpretation (197).

In this context, informedness relates to forward probability (how probable is it that a learning scheme will predict an event given that the event is present) while markedness relates to backward probability (how probable is it that an event is present if a learning scheme has predicted that it is present). Of the two, markedness is most important if the clinician wants to be certain that a prediction of referral is a good one and, in that sense, Bayes appears to work best for identifying individuals to be discharged. Perhaps it is not surprising, given that far more cases were discharged than referred. The poor performance for cases requiring Fup may also be explained by these being rarest of all of the COT outcomes. Fups are the least common outcome which may be an artefact of the COT system because as described (section 1.7) the COT is a true GRR service with highly trained practitioners who can diagnose both OHT and suspect COAG, allowing patients of low risk of developing COAG to be discharged back to their community optometrists. Equally local community optometrists know that they can re-refer patients back into the COT system if concerned at any subsequent examination.

Arguably, a confusion matrix provides an even clearer indication of performance and one is shown in Table 4.4.

		Actual			
		Dis	Fup	Refer	Totals
Predicted	Dis	**789**	**28**	**6**	*823*
	Fup	**2**	**22**	**6**	*30*
	Refer	**7**	**20**	**126**	*153*
	Totals	*798*	*70*	*138*	*1006*

Table 4.4 Confusion matrix for the predictions of naïve Bayes' in which training and testing was carried out on all 1006 cases (giving rise to maximum accuracy) seen by the COT, pooled from the cases seen by JG, DH and NOK. Actual COT outcomes (discharge, Fup and Refer) are shown against those predicted.

Only one matrix is shown (corresponding to maximum accuracy in Table 3.3) because this represented the best possible performance of Bayes. Table 4.4 shows that 27 (false referral rate = 2.7%) cases were unnecessarily referred to the HES and 34 (false discharge rate = 3.4%) cases were discharged when they should have been followed up or referred. Only 22 of the 70 Fups were correctly classified. Randomised stratified tenfold cross-validation, a better indicator of the performance of naïve Bayes', gave rise to slightly higher false referral and discharge rates (and standard deviations) of 3.1(1.5)% and 3.4 (1.6)%, respectively. While false referrals would have led to avoidable burden on HES resources, false discharges are arguably more serious as these had the potential to lead to avoidable blindness.

Could it be that making Bayes cost sensitive would remove the false discharges and referrals? This was investigated by altering the weights of the naïve Bayes' probability estimates for each COT outcome. This was achieved by increasing the post-test probabilities by an ever increasing number until all the false discharges disappeared. For example, one case that naïve Bayes' incorrectly classed as requiring discharge but should have been classed as Refer had probability estimates of 0.54 (Dis), 0.08 (Fup) and <0.01 (Refer). Increasing the weight of the probability for Refer would have removed the error but will have increased false referrals to the HES and was, therefore, not considered to be an option. More acceptable options were to reduce the weight of the probability for discharge or to increase the weight of that for Fup.

Figure 4.2 shows that all 34 false discharges could be removed when the weight of the probability estimate for discharge was reduced by multiplying it by 0.005 (which appears as a value of 1 – 0.005 = 0.995 on the horizontal axis of Figure 3.3). However, this dramatically increased the number of false Fups (from 8 to 530) and refers (from 27 to 100) and resulted in a reduction in percentage Kappa (from 79 to 19%). While this removed the risk of avoidable blindness, its costs in terms of a 60-fold increase in Fups and a 4-fold increase in referrals to the HES rendered the solution unworkable.

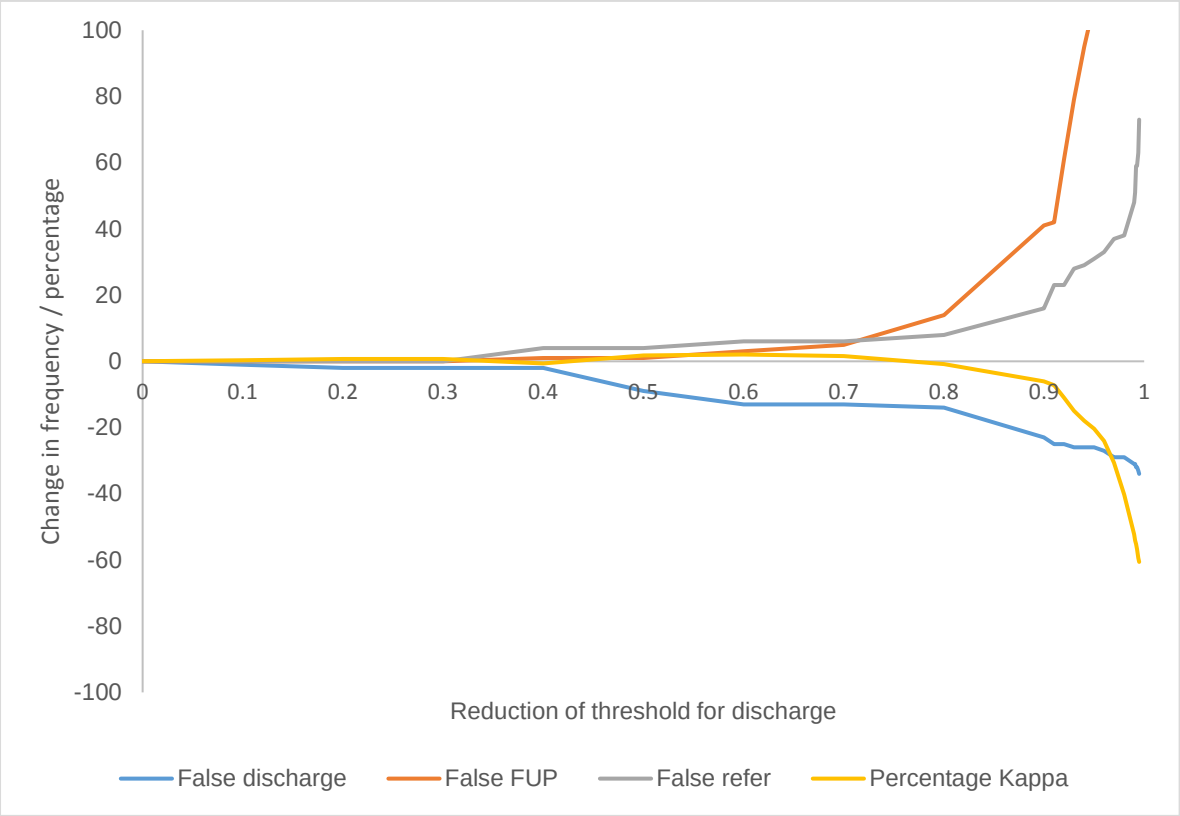


Figure 4.2 Effect of reducing the weight of the naïve Bayes' probability estimate for discharge upon the number of false discharges, false Fups, false refers and percentage Kappa. Reduction of weight was brought about by multiplying the probability estimate by between 1 and 0.005. In order to depict progressively reducing weight from left to right, the value on the horizontal axis is equal to 1 minus the multiplier. The vertical axis has been truncated at 100 to make key elements of the graph clearer (obscuring the full increase in false Fups).

Figure 4.3 shows that all 34 false discharges and 27 false refers could be removed when the weight of the probability estimate for Fup was increased by multiplying it by 180. This, again, dramatically increased the false Fups (from 8 to 649) and reduced percentage Kappa (from 79 to 15%). Although this solution was safer than the previous one, the 80-fold increase in Fups rendered it just as unworkable.

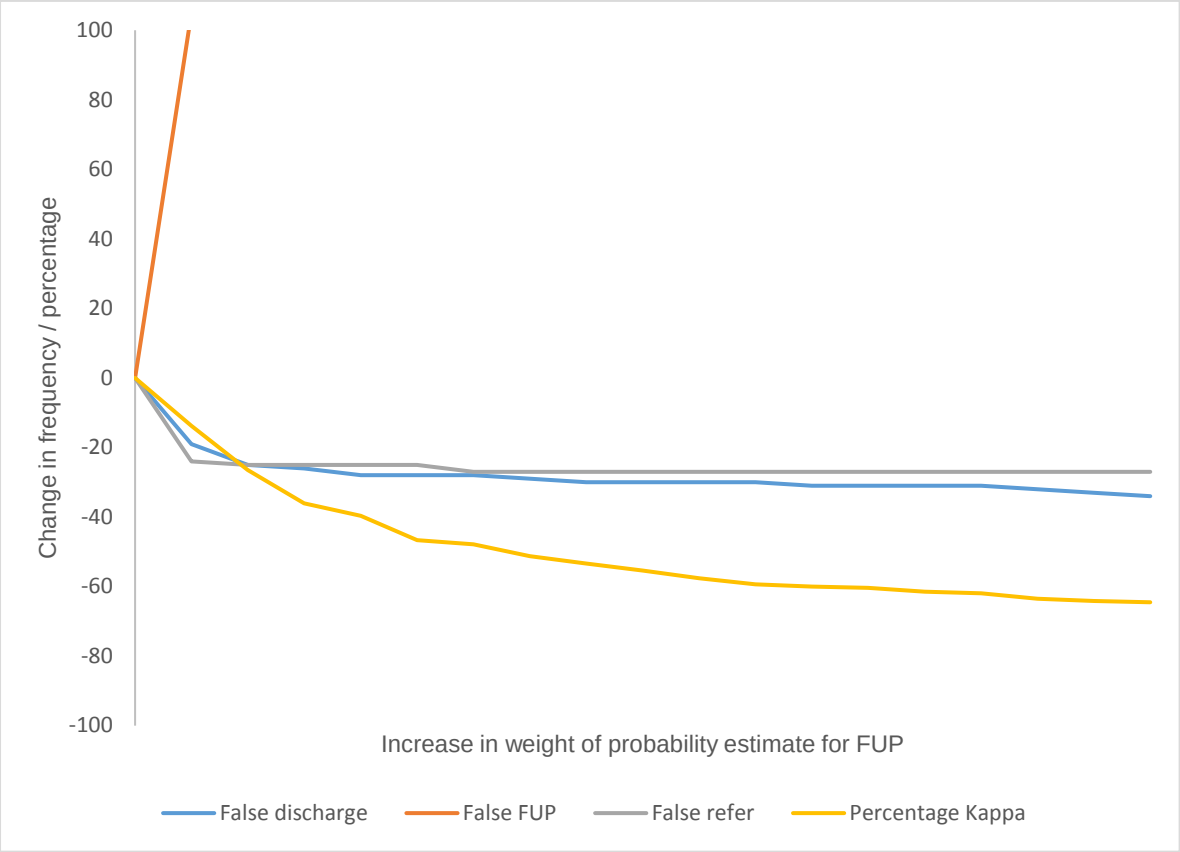


Figure 4.3 Effect of increasing the weight of the naïve Bayes' probability estimate for Fup upon the number of false discharges, false Fups, false refers and percentage Kappa. Increasing of weight was brought about by multiplying the probability estimate by between 1 and 180. The vertical axis has been truncated at 100 to make key elements of the graph clearer (obscuring the full increase in false Fups).

At this point in the study it became apparent that Bayes could not be safely applied in a referral refinement centre. The remaining sections of this study were, therefore, included for academic interest alone.

4.7 Further investigation of transferability of LRs

Transferability was revisited by applying the LRs generated from the 1006 pooled cases to the subset of 69 cases of each COT optometrist (see section 4.4). Here, the 10 estimates of each LR arising from randomised stratified tenfold cross-validation (in section 4.6) were averaged. Table 4.5 shows that LRs generated from the pooled data were more transferable as they resulted in higher weighted average Rand accuracy and Kappa values than were found in Table 4.2 when likelihood ratios generated from the 69 cases seen by each COT optometrist were tested on their colleagues' cases. However, unacceptably high false discharge (1.4 to 7.2%) and referral (1.4 to 4.3%) rates still remained. Ultimately the lack of transferability between COT optometrists is most likely due to inhomogeneity between them.

Training set	Testing set	Weighted average Rand accuracy	Kappa
pooled	JG	0.98 (0.81 - 0.94)	0.82 (0.18 - 0.63)
	DH	0.91 (0.81 - 0.84)	0.65 (0.37 - 0.41)
	NOK	0.95 (0.85 - 0.90)	0.76 (0.40 - 0.60)

Table 4.5. Re-investigation of the transferability of likelihood ratios generated from randomised stratified tenfold cross-validation on 1006 pooled cases (the training set) and tested on the 69 cases, shown in Table 4.2, seen by each COT optometrist (the testing set). Resulting weighted average Rand accuracy and Kappa values are shown next to those that arose in Table 4.2 for ease of comparison.

4.8 Redundancy

Likelihood ratios were the means by which new evidence from each COT test raised or lowered the degree of belief that any one of the three COT outcomes (Dis, Fup or Refer) was most appropriate for a given case.

The size of each LR therefore indicated the relatively usefulness of each COT test. Those tests with LRs close to 1 (indicating that they were of no clinical value) could possibly be removed as they were, arguably, redundant. Redundancy was explored on the 1006 pooled COT cases by using positive LRs from randomised stratified tenfold cross-validation to rank clinical tests in descending order. Naïve Bayes' was repeated after removing clinical tests, one at a time, starting with those ranked lowest. Kappa, false discharge rate and false referral rate were calculated for each run in order to determine how many clinical tests could be removed before performance reduced. As three sets of positive LRs arose, relating to the three possible COT outcomes (Dis, Fup and Refer), the decision was made to rank clinical tests on the basis of LRs for the Refer outcome; as correct identification of cases requiring referral was considered most important. Within this thesis we are concerned with the detection of COAG and redundancy explored which tests were most important and which could be excluded for this purpose, thus redundancy for the Dis and Fup outcomes were not ranked. Systematically removing tests, those with lowest LR first, without a drop in accuracy means that those tests added nothing to diagnosis i.e. were redundant in our evidence for diagnosis of COAG.

Figure 4.4 shows that percentage Kappa stayed within 76 to 82% until 33 of the 41 COT tests had been excluded, at which point it fell dramatically.

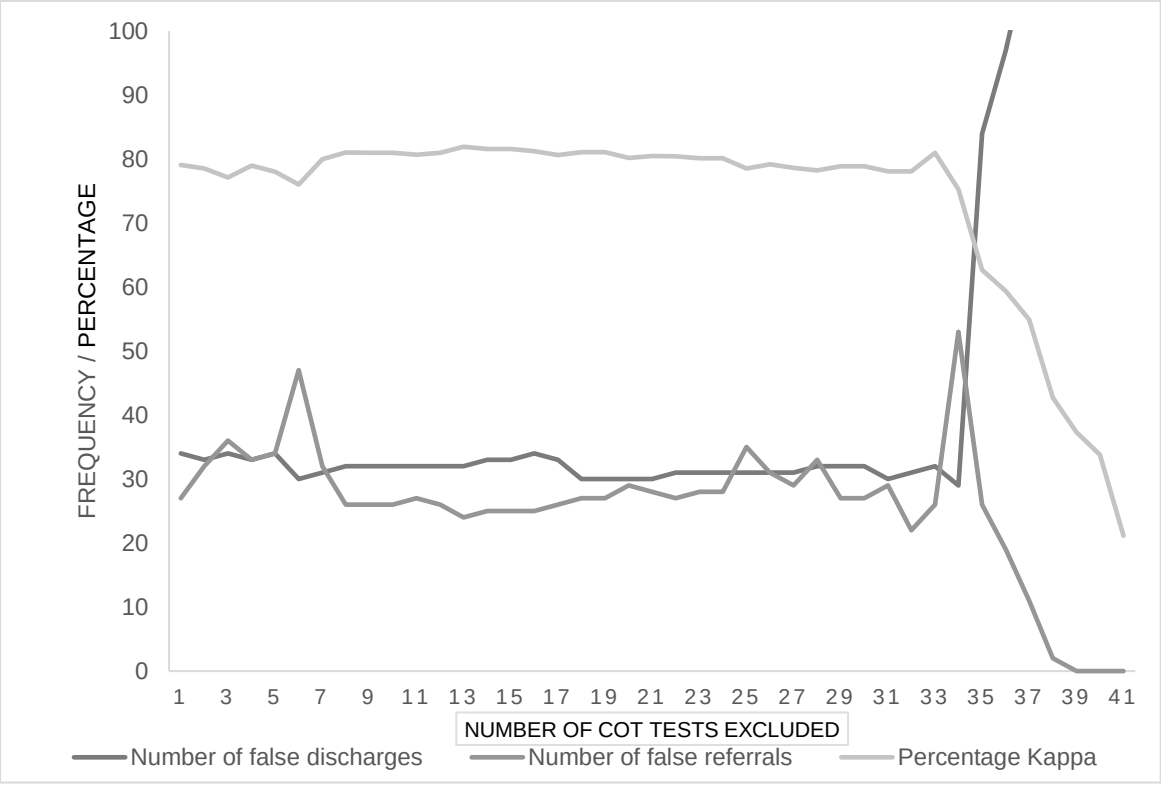


Figure 4.4 Effect of excluding COT tests one by one starting with those ranked lowest in terms of their positive LR for the outcome Refer. Naïve Bayes was repeated after each exclusion. The number of false discharges, false referrals and percentage Kappa are shown after each exclusion.

The number of false discharges, stayed at between 29 and 34 cases until 34 COT tests had been excluded, after which point it dramatically rose. False referrals showed more variation, of between 22 and 53 cases, until, after 34 COT test removals, their number fell dramatically to zero. Performance prior to removal of any COT tests (79% percentage Kappa, 32 false discharges and 27 false referrals) was closely matched (78% percentage Kappa, 30 false discharges and 29 false referrals) after 30 removals, leaving just 11 COT tests. Table 4.6 shows these tests, ranked in descending order, with their respective positive and negative LRs for all outcomes (shown also with their respective standard deviations).

Table 4.6 shows positive and negative LRs for the 11 COT tests remaining after removal of 30 redundant tests. Standard deviations (SD), expressed as a percentage of the mean, varied from <1 to 52% (mean = 12%) for positive LRs and from <1 to 21% (mean = 3%). So LRs could vary substantially across the 10 folds of the cross-validation procedure. The variation may have arisen due to differences in the numbers of cases exhibiting various test results in each fold; each fold being stratified for COT outcomes (Dis, Fup and Refer) rather than test outcomes.

COT test		Positive LR			Negative LR		
		Refer	Fup	Dis	Refer	Fup	Dis
HPA class: severe	*mean*	164175	0.00	0.00	0.79	1.03	1.16
	SD	(9235)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)
VCDR diff: >30	*mean*	113214	0.00	0.00	0.86	1.02	1.11
	SD	(6475)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)
GAT IOP: >32	*mean*	50954	0.00	0.00	0.93	1.01	1.05
	SD	(3591)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
IOP diff: >6	*mean*	84.88	1.07	0.00	0.82	1.00	1.15
	SD	(30.15)	(0.26)	(0.00)	(0.01)	(0.01)	(0.01)
VCDR: >70	*mean*	36.60	0.46	0.03	0.63	1.04	1.34
	SD	(3.40)	(0.11)	(0.00)	(0.01)	(0.01)	(0.01)
HPA class: moderate	*mean*	28.78	2.98	0.00	0.75	0.92	1.27
	SD	(4.52)	(0.45)	(0.00)	(0.02)	(0.02)	(0.02)
GAT IOP: >25-32	*mean*	24.10	3.06	0.01	0.76	0.92	1.25
	SD	(3.20)	(0.38)	(0.00)	(0.01)	(0.01)	(0.01)
HPA class: mild	*mean*	14.94	5.35	0.00	0.55	0.68	1.76
	SD	(0.90)	(0.37)	(0.00)	(0.02)	(0.03)	(0.05)
IOP diff: 3 to 6	*mean*	14.62	3.50	0.03	0.75	0.88	1.29
	SD	(1.38)	(0.19)	(0.00)	(0.01)	(0.01)	(0.02)
Race: Hispanic	*mean*	13.21	1.73	0.08	0.96	0.99	1.03
	SD	(4.43)	(0.69)	(0.03)	(0.01)	(0.01)	(0.01)
VCDR diff: 20-30	*mean*	5.08	2.49	0.17	0.66	0.82	1.45
	SD	(0.33)	(0.18)	(0.01)	(0.02)	(0.02)	(0.02)

Table 4.6 Positive and negative likelihood ratios (LR) for the 11 COT tests after removal of 30 redundant tests. The mean and standard deviation (SD) values shown were derived from the 10 estimates of each likelihood ratio from randomised stratified tenfold cross-validation on the 1006 pooled COT cases. COT tests are ranked on the basis of positive likelihood ratios for the Refer outcome. Key (alphabetical order): GAT = Goldmann Applanation Tonometry, HPA = Hodapp-Parrish-Anderson classification, IOP = Intra-Ocular Pressure (mmHg), IOP diff = Intra-Ocular Pressure difference between eyes (mmHg), VCDR = Vertical Cup Disc Ratio (%), VCDR diff = Vertical Cup Disc Ratio (%) difference between eyes.

4.9 Ranking of COT tests

Table 4.7 shows all COT tests listed in order of their highest ranking test outcome. Those included in the top 11 appear in bold. Table 4.7 also shows ranked positions and the frequencies for each tests result, allowing the reader can see the size of the samples from which LRs, used to establish ranks, were based. The key COT test methods were VFA, VCDR, GAT and CCT. Of these, only CCT could be considered redundant. This is surprising, given its influence on GAT. All HPA classifications were included in the top 11 tests. It was surprising was that VDS did not make it into the top 11, given its influence on the interpretation of VCDR. The inhomogeneity between the 3 optometrists for VDS may account for this. The GAT test method included observations relating to IOP and IOP diff. Both of these made it into to the top 11 but the finding that IOPs of lower than 25mmHg were irrelevant was surprising; maybe the referral cut-off for IOP of 21mmHg is too low. That race (Hispanic) made it into the top 11 was a surprise given that vast majority of cases were Caucasian. Perhaps the low frequency of Hispanic cases was responsible for this. Likewise, the ranking of Afro-Caribbeans below Caucasian seems to be an aberration, perhaps also due to the low frequency of the former. It appears that age, RFR, sex and FH could all be considered as redundant to COT decisions, though age (80+ years) narrowly missed redundancy, being ranked 12th. It is interesting that RFR classes are ranked in roughly the same order as the equivalent COT test; being, in descending order, suspected visual fields, optic discs and IOPs. The real surprise is that FH falls lowest in the list.

The rankings shown on Table 4.7 broadly agree with those based on Chi-square (Table 3.4) but differences of up to five ranking positions arose.

Rankings based on likelihood ratios (Table 4.7) showed slightly better agreement with those based on Chi square (Table 3.4, method A; mean and range of differences in ranked position = 1.42, 0 to 4) than those based on Chi-square per degree of freedom (Table 3.4, method B; mean and range of differences in ranked position = 1.58, 0 to 5).

COT test	Test outcome rank, frequency					
HPA class	**Severe 1, 29**	**Moderate 6, 44**	**Mild 8, 91**			
VCDR diff	**>30 2, 20**	**20-30 11, 121**	<20 36, 865			
GAT IOP	**>32 3, 9**	**>25-32 7, 43**	21-25 15, 181	<21 39,773		
IOP diff	**>6 4, 27**	**3-6 9, 53**	<3 35, 926			
VCDR	**>70 5, 71**	50-70 23, 449	<50 39, 496			
Race	**Hispanic 10, 9**	African 13, 3	Asian 19, 17	Caucasian 24, 965	Afro-Caribbean 33, 12	
Age	80+ 12, 78	70-79 16, 218	60-69 25, 275	40-49 32, 135	50-59 38, 242	<40 40, 60
RFR	Suspect visual fields 14, 199		Suspect optic discs 18, 452		Suspect IOPs 30, 462	
CCT	<555 17, 443	555-590 37, 374	>590 41, 189			
Sex	Male 20, 482	Female 27, 524				
VDS	1.4-1.7 21, 439	<1.4 22, 87	1.8+ 29, 480			
FH	Sibling 28, 62	Father 31, 110	Mother 34, 143			

Table 4.7 COT tests listed in order of their highest ranking test outcome. Ranks shown are based on positive likelihood ratios for the Refer outcome. The frequency of each test outcome is also shown. Test outcomes included in the top 11 (Table 4.6) are shown in bold. Key (alphabetical order): CCT = Central Corneal Thickness (microns), GAT = Goldmann Applanation Tonometry, HPA = Hodapp-Parrish-Anderson classification, IOP = Intra-Ocular Pressure (mmHg), IOP diff = Intra-Ocular Pressure difference between eyes (mmHg), RFR = Reason for Visit, VCDR = Vertical Cup Disc Ratio (%), VCDR diff = Vertical Cup Disc Ratio (%) difference between eyes, VDS = Vertical Disc Size (mm).

All three ranking methods agreed in placing HPA class at the top of the list and family history at the bottom. The low ranking of RFR (ranked 6th to 8th), CCT (ranked 7th to 9th), sex (ranked 10th to 11th) and VDS (ranked 9th to 11th) was also consistently found. The biggest discrepancy related to race in which rankings based on Chi-square (10th to 11th) seemed more representative than that based on likelihood ratios (6th); given that races other than Caucasians were poorly represented in the populations seen by the COT optometrists.

4.10 Summary

The findings presented in this chapter indicate that:

COT tests did not adhere to the naïve Bayes' assumption of independence. Section 4.3 showed that the 12 COT tests included in this study showed statistically significant inter-correlations. As such the basic assumption of independence upon which naïve Bayes' assumes was violated.

Naïve Bayes learned rapidly. Learning curves shown in section 4.4 showed that, despite violations of the assumption of independence, naïve Bayes learned rapidly and achieved maximum Rand accuracy, consistently requiring fewer than 69 cases to do so.

Likelihood ratios generated from just 69 cases had limited transferability Section 4.5 showed that LRs generated from the first 69 cases of each COT were not as transferable as had been hoped for. They caused a drop in Rand accuracy from 0.97-0.98 to 0.81-0.90 and Kappa from 0.83-0.91 to 0.18 to 0.63.

Bayes cannot safely predict the decisions of COT optometrists. Findings presented in section 4.6 indicated that, based on the 1006 cases pooled from all three COT optometrists, the maximum Rand accuracy of Bayes was 0.96 (Kappa 0.79) but a more realistic estimate, from randomised stratified tenfold cross-validation, amounted to a Rand accuracy of 0.95 (Kappa 0.75).

Informedness reflected variation in the area under the ROC curve while markedness (which of greater importance to making predictions) reflected variations in Matthew's correlation coefficient. These alternative measures of performance showed greater variations than seen for Rand accuracy. These and a confusion matrix indicated that false discharge and referral rates were, despite being less than 5%, unacceptably high. Attempts to remedy this by making Bayes cost sensitive failed as they caused too many unnecessary follow ups and would have rendered the COT not cost effective.

Likelihood ratios derived from the pooled COT cases were more transferable. Transferability was revisited in section 4.7. Findings indicated that LRs generated from the 1006 pooled cases, and tested on the first 69 cases of each COT optometrist, were more transferable than had been found in section 4.5. Nevertheless, false discharges and referrals were still unacceptably high.

Likelihood ratios are a useful means of investigating redundancy. Work presented in section 4.8 showed how LRs could be used to identify redundancy in the COT test outcomes. Thirty COT test findings could be removed with little change to Rand accuracy or Kappa. Visual field analysis, optic nerve head assessment and GAT were, as might be expected, most useful. Apart from race, which might have been an aberration, none of the other COT tests needed to have been included.

That this included central corneal thickness and vertical optic disc size came as a surprise, as did finding family history of POAG at the bottom of the list.

Broad agreement was found when ranking COT tests based on likelihood ratios (Table 4.7) or Chi-square (Table 3.4). Chi-square ranking methods did, however, appear to better reflect the influence of race that was poorly represented in the study sample.

Chapter 5 provides a summary of the key findings of this thesis, its limitations and recommendations for future work.

Chapter 5: Study findings and limitations with recommendations for further work

5.1 Introduction

A summary of the key findings and limitations of the present study is provided in this chapter together with recommendations for future research.

5.2 Key findings

This study was designed to determine whether artificial intelligence based on naïve Bayes could replicate the clinical decisions made by experienced optometrists using a structured SOP. Previous research carried out Aston (1) had shown that use of unstructured clinical data limited the accuracy of naïve Bayes.

The structured clinical data presented in this thesis was that collected by three highly trained optometrists working for the West Kent CCG COT. An opportunity arose to determine how effective these optometrists were in the referral refinement of COAG.

Naïve Bayes had been used before in ophthalmic research (section 1.2) and, more specifically, in glaucoma research (section1.3) but had not, as far as the author and his supervisory team were aware, been applied to the referral refinement of COAG (section 1.4). This was one element of the study that is claimed to have made a substantial and original contribution to the literature.

The application of naive Bayes in this study was based on calculations described by the Oxford Centre for Evidence Based Medicine (section 1.4.1) but adds Laplacian correction (section 1.4.2), the evaluation of learning curves using the AEL Bayes application (section 1.4.3) and multiple cut-off points (section 1.4.4).

Laplacian correction has been applied many times before in the literature (1) but the study of learning curves and the use of multiple cut-off points has not been described before, especially in relation to clinical decision support in the referral refinement of COAG.

Reviews of the glaucoma research deemed relevant to the present study (section 1.5), NICE guidelines for the diagnosis, management and treatment of COAG (section 1.6) and existing glaucoma referral refinement schemes (section 1.7) were influential in the development of the structured SOP used by West Kent CCG COT practitioners (section 1.8).

The study described in this thesis was treated as a clinical audit for the purposes of ethical data collection without the consent of the people referred into the COT (section 2.2). Data analysed in this study were collected from 1006 people seen by the three COT optometrists over one year (section 2.3). The rationale behind the structured SOP and multiple cut-off points adopted were described in detail (sections 2.3.1 to 2.3.10).

Preliminary analyses of the data were carried out in chapter 3. These analyses explored the distribution of the COT's clinical test findings, according to the multiple cut-off points introduced in chapter 2, for the three optometrists contributing data to this study (sections 3.3 and 3.4). Six out of the 13 COT clinical tests showed inhomogeneous distributions between the COT optometrists that could, potentially, have reduced the consistency of naïve Bayesian learning (Table 3.2).

Associations between each clinical test and the COT management outcomes were also explored (section 3.5).

All but one of the findings (that relating to VDS) corroborated previous research (Table 3.3). Here, referral rates increased as VDS reduced, which disagreed with one study (149) that had found no evidence for optic disc size being a risk factor for COAG. The discrepancy may have resulted from the use of univariate analyses in the present study that could be prone to confounding. Inhomogeneity in VDS distributions found by each COT optometrist also suggests that differences in the criteria used by each optometrists to judge VDS may have contributed to this discrepancy. Findings summarised in Table 3.3 also indicated that study of the influence of race on COT decisions was limited as the sample was predominantly Caucasian. Referrals to the COT were more likely to indicate COAG if based on multiple RFRs.

A cluster of investigations were carried out to explore relationships between various COT tests (sections 3.6 to 3.8). These showed that (a) GAT IOP increased with CCT, as had been reported in previous studies, but not explain observed differences in referrals made to the COT for raised IOP and COT measurements of IOP (section 3.6), (b) VCDR increased with VDS, as had been reported in the previous literature (section 3.7) and (c) observed differences between referrals made to the COT for suspect VF and COT measurements of HPA were considered to arise from the varied VF testing methods used by community optometrists (section 3.8).

Investigation of the effectiveness of COAG referral refinement by the COT optometrists contributing data to the present study (section 3.9) showed that FDRs compared favourably to those reported for other GRR schemes, with estimated cost savings of £54 per case.

Naïve Bayes was applied to the COT data in chapter 4. The COT tests showed statistically significant inter-correlations (section 4.3) and, therefore, violated the basic assumption of independence upon which naïve Bayes' rests.

Nevertheless, naïve Bayes learned rapidly and consistently achieved maximum Rand accuracy of at least 97% (Kappa 0.83) when trained on as few 69 cases (section 4.4). However, likelihood ratios generated from the learning process had limited transferability (section 4.5) and caused a drop in Rand accuracy to 81% (Kappa 0.37), which could, in part, have been due to the inhomogeneous distributions in COT test results described earlier (as detailed in Table 3.2). Transferability improved when likelihood ratios were based on learning from all 1006 cases (section 4.7).

Findings presented in section 4.6 indicated that naïve Bayes could not safely predict the decisions of COT optometrists. Based on all 1006 cases, the maximum Rand accuracy was 0.96 (Kappa 0.79) but a more realistic estimate, from randomised stratified tenfold cross-validation, amounted to a Rand accuracy of 0.95 (Kappa 0.75). Though accuracy was reasonably high, false discharge and referral rates were, despite being less than 5%, unacceptable. Attempts to remedy this by making naïve Bayes' cost sensitive failed as they caused too many unnecessary follow ups and would have rendered the COT not cost effective.

Interestingly, findings shown in section 4.6 also revealed that alternative measures of accuracy (informedness, the area under the ROC curve, markedness and Matthew's correlation coefficient) showed greater variations than seen for Rand accuracy and may be worth further exploration in the future.

Findings presented in section 4.8 showed how likelihood ratios could be used to identify redundancy in the COT test outcomes. Thirty COT test findings could be removed with little change to Rand accuracy or Kappa. Visual field analysis, optic nerve head assessment and Goldmann Applanation tonometry were, as might be expected, most useful.

Apart from race, which might have been an aberration, none of the other COT tests needed to have been included. That this included central corneal thickness and vertical optic disc size came as a surprise, as did finding family history of COAG at the bottom of the list.

Broad agreement was found when ranking COT tests based on likelihood ratios (section 3.10) or Chi-square (section 4.9). Chi-square ranking methods did, however, appear to better reflect the influence of race, which was poorly represented in the study sample.

5.3 Study limitations

The reader should be aware of important limitations to this study.

The Standards for Reporting of Diagnostic Accuracy (STARD) guidelines were developed to improve the reporting quality of the diagnostic accuracy of studies (198). This study was not designed with reference to these guidelines, while the use of 2x 2 tables, sensitivity and specificity do conform to STARD not all the points are adhered to which is a limitation of this study.

All three COT optometrists were Independent Prescribers with a Special Interest in Ophthalmology. They were regularly re-accredited. The optometrists confirmation of their decisions by a consultant ophthalmologist would have been desirable but was not feasible and would have made the COT unworkable.

The conclusions of this study are entirely based on naïve Bayes. Although Witten et al. (192) have indicated that a Bayes learning scheme may perform just as well as more sophisticated machine leaning methods, no attempt was been made in this study to confirm this.

The ethnicity of the sample does not allow for judgements to be made on any other ethnic group than Caucasian.

The lack of agreement between observers, particularly in VDS is an important limitation of the study.

Outcomes of the COT tests were grouped, according to the literature and NICE guidelines. This approach to the study may have limited its findings. More sophisticate machine learning methods are available to determine which groupings lead to best performance but were not explored in this study.

The level of myopia, which increases the risk of COAG (199, 200), was not included in this study as the refractive correction was not always known from the community optometrist referral. Equally, many of the more elderly subjects were pseudophakic, following cataract surgery, so that the extent of myopia that they had prior to surgery, which may have raised the risk of COAG, was also unknown.

All COT tests were carried out at a single COT visit, importantly this leads to limitation with respect to RTM which may bias an investigation when it is based upon an initial value without a control group (183). For example in this study IOP was measured at one given time. GAT IOP is known to have diurnal variations (184), with 65% of peak pressures occurring before noon (201). Therefore the analysis would have neglected to account for diurnal variations, inter-observer and intra-observer variability. This may mean that subjects classified as suspect COAG during the single visit may, at another time, be classified as normal and discharged or may have been followed up; either way, the COT management outcome will have varied.

It is possible that the SOP adopted by the COT added an issue of circularity to the findings. If COT decisions were based on strict rules, then it is possible that naïve Bayes would merely show up relationships that reflected these. No attempt was made to explore this possibility as a mechanisms for doing so could not be devised.

The exploration of redundancy in section 4.8 was based on COT tests ranked purely on the basis of positive likelihood ratios for the referral outcome.

Ranking could have been based on negative likelihood ratios or some ranking index based on a combination of both. Ranking could also have been based on the other COT outcomes (discharge and follow-up) or, again, some ranking index based on all three. Additional investigation of these alternative ranking indices would have been ideal but, given that application of naïve Bayes was considered unsafe for use by the COT (see section 4.6), restricting analyses to positive likelihood ratios for the referral outcome was deemed sufficient.

This study has shown that 79% of the referrals from community optometrists were discharged, thereby representing the high false positive rate expected when screening for a rare eye condition such as COAG. An analysis of which community optometrist referrals were least prone to false positive errors would have been valuable but the precise details of these referrals were not recorded in this study. For example, was a referral for suspected IOP based on a pressures of >21mmHg or higher? Or was referral based on intraocular differences in IOP? The same uncertainties apply to referrals for suspect optic discs and visual fields.

A gross examination of referrals revealed that those based on suspect optic discs had the greatest number of false positives (86%, 301 out of 351 referrals) while those based on suspect optic discs and visual fields had the lowest (30%, 16 out of 53 referrals). Such a study could have provided very useful indicators for how to improve referrals by community optometrists.

5.4 Recommendations for further research

The findings presented in this thesis indicate that naïve Bayes cannot safely predict the decisions of COT optometrists for referral refinement of COAG.

Naïve Bayes was only used in this study and more sophisticated Bayes and machine learning classifiers were beyond the scope of this thesis. Future research could investigate the use of more sophisticated Bayes and machine learning classifiers to determine if false discharge and referral rates may be reduced.

Future plans include extending the application of naïve Bayes to other less serious ophthalmic conditions such as in the differential diagnosis of red eye such as conjunctivitis seen by the West Kent CCG COT.

5.5 Summary

For now at least the findings of this study indicate that a naïve Bayesian classifier is not a substitute for an experienced specialist clinician when evaluating a patient for the possible presence of a sight threating condition such as COAG. This form of Bayesian classifier however may lend its self to other applications such as staff training, helping clinicians triage referrals and developing accreditation models in different levels of service.

References

1. Sagar R. Application of Naïve Bayesian sequential analysis to primary care optometry: Aston University; 2014.
2. McGrayne SB. The theory that would not die: how Bayes' rule cracked the enigma code, hunted down Russian submarines, & emerged triumphant from two centuries of controversy: Yale University Press; 2011.
3. Fritz KJ, Polascik MA, Potts AM. Computer assisted diagnosis for ophthalmology. Computers in biology and medicine. 1978;8(3):223-8.
4. Odom JV, Chao G-M, Weinstein GW. Preoperative prediction of postoperative visual acuity in patients with cataracts: a quantitative review. Documenta ophthalmologica. 1988;70(1):5-17.
5. Vadrevu V, Cavender S, Odom JV. Use of 10-Hz flash visual evoked potentials in prediction of final visual acuity in diabetic eyes with vitreous hemorrhage. Documenta Ophthalmologica. 1992;79(4):371-82.
6. McGraw PV, Brosnahan D, Winn B, Whitaker D. Assessment of retinal-neural function before neodymium: YAG laser capsulotomy. Investigative ophthalmology & visual science. 1995;36(6):1155-62.
7. Singh AD, De Potter P, Fijal BA, Shields CL, Shields JA, Elston RC. Lifetime prevalence of uveal melanoma in white patients with oculo(dermal) melanocytosis. Ophthalmology. 1998;105(1):195-8.
8. Nestares O, Navarro R, Antona B. Bayesian model of Snellen visual acuity. Journal of the Optical Society of America a-Optics Image Science and Vision. 2003;20(7):1371-81.
9. Leung YF, Tam POS, Lee WS, Lam DSC, Yam HF, Fan BJ, et al. The dual role of dexamethasone on anti-inflammation and outflow resistance demonstrated in cultured human trabecular meshwork cells. Molecular Vision. 2003;9(55):425-39.
10. See CW, Srinivasan M, Saravanan S, Oldenburg CE, Esterberg EJ, Ray KJ, et al. Prior Elicitation and Bayesian Analysis of the Steroids for Corneal Ulcers Trial. Ophthalmic Epidemiology. 2012;19(6):407-13.
11. Heron G, Erskine NA, Farquharson E, Moore AT, White H. Colour vision screening in glaucoma: the Tritan Album and other simple tests. Ophthalmic and Physiological Optics. 1994;14(3):233-8.
12. Bengtsson B, Olsson J, Heijl A, Rootzén H. A new generation of algorithms for computerized t hreshold perimetry, SITA. Acta Ophthalmologica Scandinavica. 1997;75(4):368-75.
13. Korrapati R, Mukherjee S, Chalam KV. A Bayesian Framework to Determine Patient Compliance in Glaucoma Cases. Proceedings of the AMIA Symposium. 2000:1050.
14. Bengtsson B. A new rapid threshold algorithm for short-wavelength automated perimetry. Invest Ophthalmol Vis Sci. 2003;44(3):1388-94.
15. Tucker A, Vinciotti V, Liu X, Garway-Heath D. A spatio-temporal Bayesian network classifier for understanding visual field deterioration. Artificial Intelligence in Medicine. 2005;34(2):163-77.
16. Bock R, Meier J, Michelson G, Nyúl LG, Hornegger J. Classifying glaucoma with image-based features from fundus photographs. Pattern Recognition: Springer; 2007;355-64.
17. Nordmann J-P, Berdeaux G. Use of a bayesian network to predict the nighttime intraocular pressure peak from daytime measurements. Clinical therapeutics. 2007;29(8):1751-60.
18. Bowd C, Hao J, Tavares IM, Medeiros FA, Zangwill LM, Lee T-W, et al. Bayesian Machine Learning Classifiers for Combining Structural and Functional Measurements to Classify Healthy and Glaucomatous Eyes. Investigative Ophthalmology & Visual Science. 2008;49(3):945-53.
19. Aspinall P. Evaluation of quality of life, and priorities in people with glaucoma. Acta Ophthalmologica. 2009;87(s244).
20. Zhu H, Crabb DP, Garway-Heath DF, editors. A Bayesian Radial Basis Function Model to Link Retinal Structure and Visual Function in Glaucoma. Bioinformatics and Biomedical Engineering , 2009 ICBBE 2009 3rd International Conference on; 2009 11-13 June 2009.

21. Medeiros FA, Leite MT, Zangwill LM, Weinreb RN. Combining Structural and Functional Measurements to Improve Detection of Glaucoma Progression using Bayesian Hierarchical Models. Investigative Ophthalmology & Visual Science. 2011;52(8):5794-5803.
22. Vidotti VG, Costa VP, Silva FR, Resende GM, Cremasco F, Dias M, et al. Sensitivity and specificity of machine learning classifiers and spectral domain OCT for the diagnosis of glaucoma. Eur J Ophthalmol. 2012:170-4.
23. Belghith A, Christopher B, Balasubramanian M, Weinreb RN, Zangwill LM. A Bayesian framework for glaucoma progression detection using Heidelberg retina tomograph images. Int J Adv Comput Sci Appl. 2013;4(9):223-9.
24. Li X, Chan E, Liao J, Wong T, Aung T, Cheng C-Y. Number of People with Glaucoma in Asia in 2020 and 2040: A Hierarchical Bayesian Meta-Analysis. Investigative Ophthalmology & Visual Science. 2013;54(15):2656.
25. Goldbaum M, Yousefi S, Belghith A, Zangwill L, Medeiros F, Weinreb R, et al. A Tree Classification Method for Identifying Normal Eyes, Non-Progressing Glaucoma Eyes, and Progressing Glaucoma Eyes from Spectral Domain OCT RNFL Thickness Measurements. Investigative Ophthalmology & Visual Science. 2013;54(15):4837.
26. Belghith A, Bowd C, Medeiros FA, Balasubramanian M, Weinreb RN, Zangwill LM. Learning from healthy and stable eyes: A new approach for detection of glaucomatous progression. Artificial intelligence in medicine. 2015;64(2):105-115.
27. Thomas R, Walland M, Thomas A, Mengersen K. Lowering of Intraocular Pressure After Phacoemulsification in Primary Open-Angle and Angle-Closure Glaucoma: A Bayesian Analysis. The Asia-Pacific Journal of Ophthalmology. 2016;5(1):79-84.
28. Hand DJ, Yu K. Idiot's Bayes: Not So Stupid after All? International Statistical Review / Revue Internationale de Statistique. 2001;69(3):385-98.
29. Zhang H. Exploring conditions for the optimality of naive Bayes. International Journal of Pattern Recognition and Artificial Intelligence. 2005;19(02):183-98.
30. Aspinall P, Hill A. Clinical infererences and decisions—I. Diagnosis and Bayes Theorem. Ophthalmic and Physiological Optics. 1983;3(3):295-304.
31. Parikh R, Parikh S, Arun E, Thomas R. Likelihood ratios: clinical application in day-to-day practice. Indian J Ophthalmol. 2009;57(3):217-21.
32. McGee S. Simplifying likelihood ratios. Journal of general internal medicine. 2002;17(8):647-50.
33. Sonis J. How to use and interpret interval likelihood ratios. Family Medicine-Kansas City-. 1999;31:432-7.
34. Cairoli R, Dalang RC. Bibliography. Sequential Stochastic Optimization: John Wiley & Sons, Inc.; 1996. p. 314-9.
35. Han J, Kamber M, Pei J. Data mining: concepts and techniques: Elsevier; 2011.
36. Parikh R, Mathai A, Parikh S, Sekhar GC, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian journal of ophthalmology. 2008;56(1):45.
37. Altman DG, Bland JM. Statistics Notes: Diagnostic tests 2: predictive values. Bmj. 1994;309(6947):102.
38. Henson DB, Spencer AF, Harper R, Cadman EJ. Community refinement of glaucoma referrals. Eye (Lond). 2003;17(1):21-6.
39. Aspinall P, Hill A. Clinical inferences and decisions--II. Decision trees, receiver operator curves and subjective probability. Ophthalmic & physiological optics: the journal of the British College of Ophthalmic Opticians (Optometrists). 1984;4(1):31-38.
40. Gilchrist J. QROC curves and kappa functions: new methods for evaluating the quality of clinical decisions. Ophthalmic and Physiological Optics. 1992;12(3):350-60.
41. Dielemans I, Vingerling JR, Wolfs RCW, Hofman A, Grobbee DE, de Jong PTVM. The Prevalence of Primary Open-angle Glaucoma in a Population-based Study in The Netherlands. Ophthalmology. 1994;101(11):1851-5.

42. Leske MC, Connell A, Schachat AP, Hyman L. The Barbados Eye Study: prevalence of open angle glaucoma. Archives of ophthalmology. 1994;112(6):821-9.
43. Mason RP, Kosoko O, Wilson MR, Martone JF, Cowan CL, Gear JC, et al. National Survey of the Prevalence and Risk Factors of Glaucoma in St. Lucia, West Indies. Ophthalmology. 1989;96(9):1363-8.
44. Wensor MD, McCarty CA, Stanislavsky YL, Livingston PM, Taylor HR. The prevalence of glaucoma in the Melbourne Visual Impairment Project. Ophthalmology. 1998;105(4):733-9.
45. Klein BE, Klein R, Sponsel WE, Franke T, Cantor LB, Martone J, et al. Prevalence of glaucoma: the Beaver Dam eye study. Ophthalmology. 1992;99(10):1499-1504.
46. Tielsch JM, Sommer A, Katz J, Royall RM, Quigley HA, Javitt J. Racial variations in the prevalence of primary open-angle glaucoma: the Baltimore Eye Survey. Jama. 1991;266(3):369-74.
47. Mitchell P, Smith W, Attebo K, Healey PR. Prevalence of open-angle glaucoma in Australia - The Blue Mountains eye study. Ophthalmology. 1996;103(10):1661-9.
48. Buhrmann RR, Quigley HA, Barron Y, West SK, Oliva MS, Mmbaga BBO. Prevalence of glaucoma in a rural east African population. Investigative Ophthalmology & Visual Science. 2000;41(1):40-8.
49. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology. 2014;121(11):2081-90.
50. West S, Sommer A. Prevention of blindness and priorities for the future. World Health Organization Bulletin of the World Health Organization. 2001;79(3):244-8.
51. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. Br J Ophthalmol. 2006;90(3):262-7.
52. Glaucoma N. Clinical Guidelines CG85. National Institute for Health and Clinical Excellence (www nice org uk), London, UK. 2009.
53. Burr JM, Mowatt G, Hernández R, Siddiqui M, Cook J, Lourenco T, et al. The clinical effectiveness and cost-effectiveness of screening for open angle glaucoma: a systematic review and economic evaluation. Health technology assessment (Winchester, England). 2007;11(41):1-190.
54. Ratnarajan G, Kean J, French K, Parker M, Bourne R. The false negative rate and the role for virtual review in a nationally evaluated glaucoma referral refinement scheme. Ophthalmic and Physiological Optics. 2015;35(5):577-81.
55. Bunce C, Wormald R. Leading causes of certification for blindness and partial sight in England & Wales. BMC Public Health. 2006;6:58.
56. Fechtner RD, Weinreb RN. Mechanisms of optic nerve damage in primary open angle glaucoma. Survey of ophthalmology. 1994;39(1):23-42.
57. Drance SM. The early field defects in glaucoma. Investigative Ophthalmology & Visual Science. 1969;8(1):84-91.
58. Andreou PA, Wickremasinghe SS, Asaria RH, Tay E, Franks WA. A comparison of HRT II and GDx imaging for glaucoma detection in a primary care eye clinic setting. Eye. 2007;21(8):1050-5.
59. Foster PJ, Buhrmann R, Quigley HA, Johnson GJ. The definition and classification of glaucoma in prevalence surveys. British Journal of Ophthalmology. 2002;86(2):238-42.
60. Sullivan-Mee M, Ruegg CC, Pensyl D, Halverson K, Qualls C. Diagnostic Precision of Retinal Nerve Fiber Layer and Macular Thickness Asymmetry Parameters for Identifying Early Primary Open-Angle Glaucoma. American Journal of Ophthalmology.156(3):567-77.e1.
61. Munemasa Y, Kitaoka Y. Molecular mechanisms of retinal ganglion cell degeneration in glaucoma and future prospects for cell body and axonal protection. Frontiers in cellular neuroscience. 2012;6:60.
62. Morgan JE, Jeffery G, Foss AJE. Axon deviation in the human lamina cribrosa. British Journal of Ophthalmology. 1998;82(6):680-3.
63. Albon J, Purslow PP, Karwatowski WSS, Easty DL. Age related compliance of the lamina cribrosa in human eyes. British Journal of Ophthalmology. 2000;84(3):318-23.

64. Kotecha A, Izadi S, Jeffery G. Age-related charges in the thickness of the human lamina cribrosa. British Journal of Ophthalmology. 2006;90(12):1531-4.
65. Flammer J, Mozaffarieh M. What is the present pathogenetic concept of glaucomatous optic neuropathy? Survey of Ophthalmology. 2007;52:S162-S173.
66. Findl O, Strenn K, Wolzt M, Menapace R, Vass C, Eichler HG, et al. Effects of changes in intraocular pressure on human ocular haemodynamics. Current Eye Research. 1997;16(10):1024-9.
67. Hosking SL, Harris A, Chung HS, Jonescu-Cuypers CP, Kagemann L, Hilton EJR, et al. Ocular haemodynamic responses to induced hypercapnia and hyperoxia in glaucoma. British Journal of Ophthalmology. 2004;88(3):406-11.
68. Feke GT, Pasquale LR. Retinal blood flow response to posture change in glaucoma patients compared with healthy subjects. Ophthalmology. 2008;115(2):246-52.
69. Riva CE, Salgarello T, Logean E, Colotto A, Galan EM, Falsini B. Flicker-evoked response measured at the optic disc rim is reduced in ocular hypertension and early glaucoma. Investigative Ophthalmology & Visual Science. 2004;45(10):3662-8.
70. Neufeld AH, Hernandez MR, Gonzalez M. Nitric oxide synthase in the human glaucomatous optic nerve head. Archives of Ophthalmology. 1997;115(4):497-503.
71. Chrysostomou V, Rezania F, Trounce IA, Crowston JG. Oxidative stress and mitochondrial dysfunction in glaucoma. Current Opinion in Pharmacology. 2013;13(1):12-15.
72. Anderson DR, Normal Tension Glaucoma S. Collaborative normal tension glaucoma study. Current opinion in ophthalmology. 2003;14(2):86-90.
73. Wilson MR, Creighton MS. Normal tension glaucoma. Acta Ophthalmologica Scandinavica. 2002;80:9-11.
74. Sowka J. New thoughts on normal tension glaucoma. Optometry. 2005;76(10):600-8.
75. Tielsch JM, Katz J, Singh K, Quigley HA, Gottsch JD, Javitt J, et al. A population-based evaluation of glaucoma screening: the Baltimore Eye Survey. American Journal of Epidemiology. 1991;134(10):1102-10.
76. Drance SM, Crichton A, Mills RP. Comparison of the effect of latanoprost 0.005% and timolol 0.5% on the calculated ocular perfusion pressure in patients with normal-tension glaucoma. American Journal of Ophthalmology. 1998;125(5):585-92.
77. Guthauser U, Flammer J, Mahler F. The relationship between digital and ocular vasospasm. Graefe's archive for clinical and experimental ophthalmology. 1988;226(3):224-6.
78. Kochkorov A, Gugleta K, Zawinka C, Katamay R, Flammer J, Orgul S. Short-term retinal vessel diameter variability in relation to the history of cold extremities. Investigative Ophthalmology & Visual Science. 2006;47(9):4026-33.
79. Kass MA, Heuer DK, Higginbotham EJ, Johnson CA, Keltner JL, Miller JP, et al. The Ocular Hypertension Treatment Study - A randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. Archives of Ophthalmology. 2002;120(6):701-13.
80. Philips B, Ball C, Sackett D, Badenoch D, Straus S, Haynes B, et al. Oxford Centre for evidence-based medicine levels of evidence (may 2001). Oxford Centre for Evidence-based Medicine http://www. cebm. net/levels_of_evidence. asp; 2001.
81. Dissemination C. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare. York: University of York NHS Centre for Reviews & Dissemination. 2009.
82. Steinberg E, Greenfield S, Mancher M, Wolman DM, Graham R. Clinical practice guidelines we can trust: National Academies Press; 2011.
83. Atkinson P, Wishart P, James J, Vernon S, Reid F. Deterioration in the accuracy of the pulsair non-contact tonometer with use: need for regular calibration. Eye (London, England). 1991;6:530-4.
84. Gordon MO, Ocular Hypertension Treatment S, European Prevention Study G. Validated prediction model for the development of primary open-angle glaucoma in individuals with ocular hypertension. Ophthalmology. 2007;114(1):10-19.

85. Miglior S, Zeyen T, Pfeiffer N, Cunha-Vaz J, Torri V, Adamsons I, et al. Results of the European Glaucoma Prevention Study. Ophthalmology. 2005;112(3):366-75.
86. Baskaran M, Oen FTS, Chan Y-H, Hoh S-T, Ho C-L, Kashiwagi K, et al. Comparison of the scanning peripheral anterior chamber depth analyzer and the modified van Herick grading system in the assessment of angle closure. Ophthalmology. 2007;114(3):501-6.
87. Thomas R, George T, Braganza A, Muliyil J. The flashlight test and van Herick's test are poor predictors for occludable angles. Australian and New Zealand Journal of Ophthalmology. 1996;24(3):251-6.
88. Nolan WP, See JL, Chew PTK, Friedman DS, Smith SD, Radhakrishnan S, et al. Detection of primary angle closure using anterior segment optical coherence tomography in Asian eyes. Ophthalmology. 2007;114(1):33-9.
89. Spry PG. Glaucoma co-management at Bristol Eye Hospital: translating clinical research into practice. Optometry in Practice. 2008;9:13:103-16
90. Harper RL, J.Vernon,S.Spry,P. Postgraduate specialist glaucoma training and accreditation in optometry. Optometry in Practice. 2013;14(4):125-36
91. Optometrists Co. Previous Higher Qualifications-Glaucoma 2015 [Available from: http://www.college-optometrists.org/en/CPD/hq/college-higher-qualifications/glaucoma/index.cfm.
92. Optometrists Co. Learning Outcomes 2015 [Available from: http://www.college-optometrists.org/en/CPD/hq/information-for-providers/learning-outcomes.cfm.
93. Optometrists Co. Accreditation of higher qualifications guidance for course providers2015 05/09/2015. Available from: http://www.college-optometrists.org/en/utilities/document-summary.cfm?docid=F486513B-4083-4F2D-AB939F516AB0D3B3.
94. Banes MJ, Culham LE, Crowston JG, Bunce C, Khaw PT. An optometrist's role of co-management in a hospital glaucoma clinic. Ophthalmic and Physiological Optics. 2000;20(5):351-9.
95. Gray SF, Spry PGD, Brookes ST, Peters TJ, Spencer IC, Baker IA, et al. The Bristol shared care glaucoma study: outcome at follow up at 2 years. British Journal of Ophthalmology. 2000;84(5):456-63.
96. Harper R, Radi N, Reeves BC, Fenerty C, Spencer AF, Batterbury M. Agreement between ophthalmologists and optometrists in optic disc assessment: training implications for glaucoma co-management. Graefes Archive for Clinical and Experimental Ophthalmology. 2001;239(5):342-50.
97. Harper R, Reeves B, Smith G. Observer variability in optic disc assessment: implications for glaucoma shared care. Ophthalmic and Physiological Optics. 2000;20(4):265-73.
98. Spry PGD, Spencer IC, Sparrow JM, Peters TJ, Brookes ST, Gray S, et al. The Bristol Shared Care Glaucoma Study: reliability of community optometric and hospital eye service test measures. British Journal of Ophthalmology. 1999;83(6):707-12.
99. Lawrenson J. Glaucoma: the challenge of early case detection. Ophthalmic and Physiological Optics. 2013;33(1):3-6.
100. Chen PP. Risk and risk factors for blindness from glaucoma. Current opinion in ophthalmology. 2004;15(2):107-11.
101. Hatt S, Wormald R, Burr J. Screening for prevention of optic nerve damage due to chronic open angle glaucoma. Cochrane Database of Systematic Reviews. 2006(4).
102. Rouland JF, Berdeaux G, Lafuma A. The economic burden of glaucoma and ocular hypertension - Implications for patient management: A review. Drugs & Aging. 2005;22(4):315-21.
103. NSC U. The UK NSC recommendation on Glaucoma screening in adults 2016 [Available from: http://legacy.screening.nhs.uk/glaucoma.
104. Bowling B, Chen SD, Salmon JF. Outcomes of referrals by community optometrists to a hospital glaucoma service. Br J Ophthalmol. 2005;89(9):1102-4.
105. Parkins DJ, Edgar DF. Comparison of the effectiveness of two enhanced glaucoma referral schemes. Ophthalmic Physiol Opt. 2011;31(4):343-52.
106. Optometrists Ao. NICE Glaucoma Guidelines 2010 [Available from: http://www.aop.org.uk/practitioner-advice/enhanced-services/glaucoma-nice-guidelines/.

107. Shah S, Murdoch IE. NICE - impact on glaucoma case detection. Ophthalmic Physiol Opt. 2011;31(4):339-42.
108. Edgar DF, Romanay,T.Lawrenson,J,Myint. Referral Behaviour Among Optometrists: Increase in the Number of Referrals from Optometrists Following the Publication of the April 2009 NICE Guidelines for the Diagnosis and Management of COAG and OHT in England and Wales and its Implications. Optometry in Practice. 2010;11:6:33-6.
109. Bell RWD, O'Brien C. The diagnostic outcome of new glaucoma referrals. Ophthalmic and Physiological Optics. 1997;17(1):3-6.
110. Theodossiades J, Murdoch I. Positive predictive value of optometrist-initiated referrals for glaucoma. Ophthalmic and Physiological Optics. 1999;19(1):62-7.
111. Ratnarajan G, Newsom W, Vernon SA, Fenerty C, Henson D, Spencer F, et al. The effectiveness of schemes that refine referrals between primary and secondary care--the UK experience with glaucoma referrals: the Health Innovation & Education Cluster (HIEC) Glaucoma Pathways Project. BMJ Open. 2013;3(7):1-9
112. Warburton T. Repeating pressures—an electronic reporting system. Optom Today. 2010;50:21-4.
113. Devarajan N, Williams GS, Hopes M, O'Sullivan D, Jones D. The Carmarthenshire Glaucoma Referral Refinement Scheme, a safe and efficient screening service. Eye (Lond). 2011;25(1):43-9.
114. Bourne RRA, French KA, Chang L, Borman AD, Hingorani M, Newsom WD. Can a community optometrist-based referral refinement scheme reduce false-positive glaucoma hospital referrals without compromising quality of care[quest] The community and hospital allied network glaucoma evaluation scheme (CHANGES). Eye. 2010;24(5):881-7.
115. NICE. Glaucoma in Adults, Quality Standard (QS7) 2011 [Available from: https://www.nice.org.uk/guidance/qs7/chapter/quality-statement-1-referral.
116. Ophthalmologists RCo. Glaucoma and Ocular Hypertension 2016 [Available from: https://www.rcophth.ac.uk/standards-publications-research/commissioning-in-ophthalmology/glaucoma-and-ocular-hypertension/.
117. West Kent C. [Available from: http://www.westkentccg.nhs.uk/about-us/.
118. Medway C. [Available from: http://www.medwayccg.nhs.uk/about-us.
119. Dartford GaS, CCG. [Available from: http://www.dartfordgraveshamswanleyccg.nhs.uk/about-us/our-plans-reports-and-strategies/.
120. Ophthalmologists RCo. 2015 [Available from: https://www.rcophth.ac.uk/wp-content/uploads/2015/07/Community-Ophthalmology-Framework.pdf.
121. Varma R, Ying-Lai M, Francis BA, Nguyen BB-T, Deneen J, Wilson MR, et al. Prevalence of open-angle glaucoma and ocular hypertension in Latinos: the Los Angeles Latino Eye Study. Ophthalmology. 2004;111(8):1439-48.
122. Foster PJ, Oen FT, Machin D, Ng T-P, Devereux JG, Johnson GJ, et al. The prevalence of glaucoma in Chinese residents of Singapore: a cross-sectional population survey of the Tanjong Pagar district. Archives of Ophthalmology. 2000;118(8):1105-11.
123. Quigley HA, West SK, Rodriguez J, Munoz B, Klein R, Snyder R. The prevalence of glaucoma in a population-based study of Hispanic subjects: Proyecto VER. Archives of Ophthalmology. 2001;119(12):1819-26.
124. Foster PJ, Baasanhu J, Alsbirk PH, Munkhbayar D, Uranchimeg D, Johnson GJ. Glaucoma in Mongolia: a population-based survey in Hövsgöl Province, northern Mongolia. Archives of ophthalmology. 1996;114(10):1235-41.
125. He M, Foster PJ, Ge J, Huang W, Zheng Y, Friedman DS, et al. Prevalence and clinical characteristics of glaucoma in adult Chinese: a population-based study in Liwan District, Guangzhou. Investigative ophthalmology & visual science. 2006;47(7):2782-8.
126. Bourne R, Sukudom P, Foster P, Tantisevi V, Jitapunkul S, Lee P, et al. Prevalence of glaucoma in Thailand: a population based survey in Rom Klao District, Bangkok. British journal of ophthalmology. 2003;87(9):1069-74.

127. Ramakrishnan R, Nirmalan PK, Krishnadas R, Thulasiraj R, Tielsch JM, Katz J, et al. Glaucoma in a rural population of southern India: the Aravind comprehensive eye survey. Ophthalmology. 2003;110(8):1484-90.
128. Coffey M, Reidy A, Wormald R, Xian WX, Wright L, Courtney P. Prevalence of glaucoma in the west of Ireland. British Journal of Ophthalmology. 1993;77(1):17-21.
129. Vijaya L, George R, Baskaran M, Arvind H, Raju P, Ramesh SV, et al. Prevalence of primary open-angle glaucoma in an urban south Indian population and comparison with a rural population: the Chennai Glaucoma Study. Ophthalmology. 2008;115(4):648-54. e1.
130. Ntim-Amponsah C, Amoaku W, Ofosu-Amaah S, Ewusi R, Idirisuriya-Khair R, Nyatepe-Coo E, et al. Prevalence of glaucoma in an African population. Eye. 2004;18(5):491-7.
131. Quigley HA, Enger C, Katz J, Sommer A, Scott R, Gilbert D. Risk factors for the development of glaucomatous visual field loss in ocular hypertension. Archives of Ophthalmology. 1994;112(5):644-9.
132. Tielsch JM. The epidemiology and control of open angle glaucoma: a population-based perspective. Annual review of public health. 1996;17(1):121-36.
133. Leibowitz HM, Krueger D, Maunder L, Milton R, Kini M, Kahn H, et al. The Framingham Eye Study monograph: an ophthalmological and epidemiological study of cataract, glaucoma, diabetic retinopathy, macular degeneration, and visual acuity in a general population of 2631 adults, 1973-1975. Survey of ophthalmology. 1979;24(Suppl):335-610.
134. Bengtsson B. The prevalence of glaucoma. British Journal of Ophthalmology. 1981;65(1):46-9.
135. Wolfs RC, Klaver CC, Ramrattan RS, van Duijn CM, Hofman A, de Jong PT. Genetic risk of primary open-angle glaucoma: population-based familial aggregation study. Archives of ophthalmology. 1998;116(12):1640-5.
136. Tielsch JM, Katz J, Sommer A, Quigley HA, Javitt JC. Family history and risk of primary open angle glaucoma: the Baltimore Eye Survey. Archives of ophthalmology. 1994;112(1):69-73.
137. Shin DH, Becker B, Kolker AE. Family history in primary open-angle glaucoma. Archives of ophthalmology (Chicago, Ill : 1960). 1977;95(4):598-600.
138. Salmon NJ, Terry H, Farmery AD, Salmon JF. An analysis of patients discharged from a hospital-based glaucoma case-finding clinic over a 3-year period. Ophthalmic and Physiological Optics. 2007;27(4):399-403.
139. Harper RA, Reeves BC. Glaucoma screening: the importance of combining test data. Optometry & Vision Science. 1999;76(8):537-43.
140. Gordon MO, Beiser JA, Brandt JD, Heuer DK, Higginbotham EJ, Johnson CA, et al. The Ocular Hypertension Treatment Study: baseline factors that predict the onset of primary open-angle glaucoma. Archives of ophthalmology. 2002;120(6):714-20.
141. Heijl A, Leske MC, Bengtsson B, Hyman L, Bengtsson B, Hussein M, et al. Reduction of intraocular pressure and glaucoma progression - Results from the early manifest glaucoma trial. Archives of Ophthalmology. 2002;120(10):1268-79.
142. Williams AL, Gatla S, Leiby BE, Fahmy I, Biswas A, de Barros DM, et al. The value of intraocular pressure asymmetry in diagnosing glaucoma. J Glaucoma. 2013;22(3):215-8.
143. Cartwright MJ, Anderson DR. Correlation of asymmetric damage with asymmetric intraocular pressure in normal-tension glaucoma (low-tension glaucoma). Archives of ophthalmology. 1988;106(7):898-900.
144. Garway-Heath DF, Ruben ST, Viswanathan A, Hitchings RA. Vertical cup/disc ratio in relation to optic disc size: its value in the assessment of the glaucoma suspect. British journal of ophthalmology. 1998;82(10):1118-24.
145. Heijl A, Mölder H. Optic disc diameter influences the ability to detect glaucomatous disc damage. Acta ophthalmologica. 1993;71(1):122-9.
146. Jonas JB, Bergua A, Schmitz–Valckenberg P, Papastathopoulos KI, Budde WM. Ranking of optic disc variables for detection of glaucomatous optic nerve damage. Investigative ophthalmology & visual science. 2000;41(7):1764-73.

147. Varma R, Tielsch JM, Quigley HA, Hilton SC, Katz J, Spaeth GL, et al. Race-, age-, gender-, and refractive error—related differences in the normal optic disc. Archives of Ophthalmology. 1994;112(8):1068-76.
148. Sample PA, Girkin CA, Zangwill LM, Jain S, Racette L, Becerra LM, et al. The african descent and glaucoma evaluation study (ADAGES): Design and baseline data. Archives of ophthalmology. 2009;127(9):1136-45.
149. Burk RO, Rohrschneidern K, Noack H, Völcker HE. Are large optic nerve heads susceptible to glaucomatous damage at normal intraocular pressure? Graefe's archive for clinical and experimental ophthalmology. 1992;230(6):552-60.
150. Hoffmann EM, Zangwill LM, Crowston JG, Weinreb RN. Optic disk size and glaucoma. Survey of ophthalmology. 2007;52(1):32-49.
151. Crowston JG, Hopley CR, Healey PR, Lee A, Mitchell P. The effect of optic disc diameter on vertical cup to disc ratio percentiles in a population based cohort: the Blue Mountains Eye Study. British Journal of Ophthalmology. 2004;88(6):766-70.
152. Jonas JB, Budde WM, Panda-Jonas S. Ophthalmoscopic evaluation of the optic nerve head. Survey of ophthalmology. 1999;43(4):293-320.
153. Weisman R, Asseff C, Phelps C, Podos S, Becker B. Vertical elongation of the optic cup in glaucoma. Transactions-American Academy of Ophthalmology and Otolaryngology American Academy of Ophthalmology and Otolaryngology. 1973;77(2):OP157.
154. Tsai JC, Forbes M. Medical management of glaucoma: Professional Communications; 2009.
155. Snydacker D. The normal optic disc: Ophthalmoscopic and photographic studies. American journal of ophthalmology. 1964;58(6):958-64.
156. Gloster J. Quantitative relationship between cupping of the optic disc and visual field loss in chronic simple glaucoma. British Journal of Ophthalmology. 1978;62(10):665-9.
157. Reis AS, Toren A, Nicolela MT. Clinical Optic Disc Evaluation in Glaucoma. European Ophthalmic Review. 2012;6(2):92-7.
158. Ong LS, Mitchell P, Healey PR, Cumming RG. Asymmetry in optic disc parameters: The blue mountains eye study. Investigative Ophthalmology & Visual Science. 1999;40(5):849-57.
159. Sharma B, Chaturvedi R. Disc-cup asymmetry in normal and chronic simple glaucoma. Indian journal of ophthalmology. 1982;30(3):133.
160. Brandt JD, Beiser JA, Kass MA, Gordon MO, Group OHTS. Central corneal thickness in the ocular hypertension treatment study (OHTS). Ophthalmology. 2001;108(10):1779-88.
161. Doughty MJ, Zaman ML. Human corneal thickness and its impact on intraocular pressure measures: a review and meta-analysis approach. Survey of ophthalmology. 2000;44(5):367-408.
162. Leske MC, Wu S-Y, Hennis A, Honkanen R, Nemesure B, Group BS. Risk factors for incident open-angle glaucoma: the Barbados Eye Studies. Ophthalmology. 2008;115(1):85-93.
163. Anderson D. Automated static perimetry. St Louis: Mosby–Year Book. Inc; 1992.
164. Budenz DL, Rhee P, Feuer WJ, McSoley J, Johnson CA, Anderson DR. Comparison of glaucomatous visual field defects using standard full threshold and Swedish interactive threshold algorithms. Archives of ophthalmology. 2002;120(9):1136-41.
165. Armstrong RA, Eperjesi F. Data analysis methods in optometry Part 3: the analysis of frequencies and proportions. Optometry today. 2002;2001(May):34-7.
166. Prajapati B, Dunne M, Armstrong R. Sample size estimation and statistical power analyses. Optometry Today. 2010;16(07):10-18.
167. Armstrong RA, Davies LN, Dunne M, Gilmartin B. Statistical guidelines for clinical studies of human vision. Ophthalmic and physiological optics. 2011;31(2):123-36.
168. 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. Investigative ophthalmology & visual science. 2006;47(10):4254-61.
169. Lockwood AJ, Kirwan JF, Ashleigh Z. Optometrists referrals for glaucoma assessment: a prospective survey of clinical data and outcomes. Eye (Lond). 2010;24(9):1515-19.

170. Suzuki Y, Iwase A, Araie M, Yamamoto T, Abe H, Shirato S, et al. Risk factors for open-angle glaucoma in a Japanese population: the Tajimi Study. Ophthalmology. 2006;113(9):1613-7.
171. Weinreb RN, Khaw PT. Primary open-angle glaucoma. The Lancet. 2004;363(9422):1711-20.
172. Sommer A, Tielsch JM, Katz J, Quigley HA, Gottsch JD, Javitt J, et al. Relationship between intraocular pressure and primary open angle glaucoma among white and black Americans: the Baltimore Eye Survey. Archives of ophthalmology. 1991;109(8):1090-5.
173. Wolfs R, Klaver C, Vingerling J, Grobbee D, Hofman A, Jong P. Distribution of central corneal thickness and its association with intraocular pressure: The Rotterdam Study. American journal of ophthalmology. 1997;123:767-772.
174. Foster PJ, Baasanhu J, Alsbirk PH, Munkhbayar D, Uranchimeg D, Johnson GJ. Central corneal thickness and intraocular pressure in a Mongolian population. Ophthalmology. 1998;105(6):969-73.
175. Emara B, Probst LE, Tingey DP, Kennedy DW, Willms LJ, Machat J. Correlation of intraocular pressure and central corneal thickness in normal myopic eyes and after laser in situ keratomileusis. Journal of Cataract & Refractive Surgery. 1998;24(10):1320-5.
176. Ogbuehi KC, Mucke S, Osuagwu UL. Influence of central corneal thickness on measured intraocular pressure differentials: Nidek RKT-7700, Topcon CT-80 NCTs and Goldmann Tonometer. Ophthalmic and Physiological Optics. 2012;32(6):547-55.
177. Hsu S, Sheu M, Hsu A, Wu K, Yeh J, Tien J, et al. Comparisons of intraocular pressure measurements: Goldmann applanation tonometry, noncontact tonometry, Tono-Pen tonometry, and dynamic contour tonometry. Eye. 2009;23(7):1582-8.
178. Cook JA, Botello AP, Elders A, Ali AF, Azuara-Blanco A, Fraser C, et al. Systematic review of the agreement of tonometers with Goldmann applanation tonometry. Ophthalmology. 2012;119(8):1552-7.
179. Tonnu P, Ho T, Newson T, El Sheikh A, Sharma K, White E, et al. The influence of central corneal thickness and age on intraocular pressure measured by pneumotonometry, non-contact tonometry, the Tono-Pen XL, and Goldmann applanation tonometry. British Journal of Ophthalmology. 2005;89(7):851-4.
180. Matsumoto T, Makino H, Uozato H, Saishin M, Miyamoto S. The influence of corneal thickness and curvature on the difference between intraocular pressure measurements obtained with a non-contact tonometer and those with a Goldmann applanation tonometer. Japanese journal of ophthalmology. 2000;44(6):691.
181. Strong N. How optometrists screen for glaucoma: a survey. Ophthalmic and physiological optics. 1992;12(1):3-7.
182. Khan S, Clarke J, Kotecha A. Comparison of optometrist glaucoma referrals against published guidelines. Ophthalmic Physiol Opt. 2012;32(6):472-7.
183. Bland JM, Altman DG. Regression towards the mean. BMJ: British Medical Journal. 1 994;308(6942):1499.
184. Bengtsson B, Heijl A. Diurnal IOP fluctuation: not an independent risk factor for glaucomatous visual field loss in high-risk ocular hypertension. Graefe's Archive for Clinical and Experimental Ophthalmology. 2005;243(6):513-8.
185. Britton RJ, Drance SM, Schulzer M, Douglas GR, Mawson DK. The area of the neuroretinal rim of the optic nerve in normal eyes. American journal of ophthalmology. 1987;103(4):497-504.
186. Healey P, Mitchell P, Smith W, Wang J. Relationship between cup-disc ratio and optic disc diameter: The Blue Mountains Eye Study. Australian and New Zealand journal of ophthalmology. 1997;25(4):99-101.
187. Marra G, Flammer J. The learning and fatigue effect in automated perimetry. Graefe's archive for clinical and experimental ophthalmology. 1991;229(6):501-4.
188. Fasih U, Shaikh A, Shaikh N, Fehmi M, Jafri AR, Rahman A. Evaluation of reliability of visual field examination in glaucoma patients. Pak J Ophthalmol. 2009;25(3):145-151.
189. Wild J. Techniques and developments in automated perimetry: a review. Ophthalmic and Physiological Optics. 1988;8(3):295-308.

190. England MaN. NHS National Tariff Payment System 2016/17 2016 [Available from: https://www.gov.uk/government/publications/nhs-national-tariff-payment-system-201617.
191. Armstrong R. In: Gurney J, editor. Statistical advice relating to Chi-Square statistical tests ed. Aston University.2016.
192. Witten IH, Frank E. Data Mining: Practical machine learning tools and techniques: Morgan Kaufmann; 2005.
193. Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. 2011.
194. Ling CX, Sheng VS. Cost-sensitive learning. Encyclopedia of Machine Learning: Springer; 2011. p. 231-5.
195. Armstrong RA. When to use the Bonferroni correction. Ophthalmic and Physiological Optics. 2014;34(5):502-8.
196. Bar-Hillel M. The base-rate fallacy in probability judgments. Acta Psychologica. 1980;44(3):211-33.
197. Bramwell R, West H, Salmon P. Health professionals' and service users' interpretation of screening test results experimental study. BMJ. 2006;333(7562):284.
198. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Annals of internal medicine. 2003;138(1):W1-12.
199. Chihara E, Liu X, Dong J, Takashima Y, Akimoto M, Hangai M, et al. Severe myopia as a risk factor for progressive visual field loss in primary open-angle glaucoma. Ophthalmologica. 1997;211(2):66-71.
200. Marcus MW, de Vries MM, Montolio FGJ, Jansonius NM. Myopia as a risk factor for open-angle glaucoma: a systematic review and meta-analysis. Ophthalmology. 2011;118(10):1989-94. e2.
201. David R, Zangwill L, Briscoe D, Dagan M, Yagev R, Yassur Y. Diurnal intraocular pressure variations: an analysis of 690 diurnal curves. British journal of ophthalmology. 1992;76(5):280-3.