APPLICATION OF NAÏVE BAYESIAN SEQUENTIAL ANALYSIS TO PRIMARY CARE OPTOMETRY

RAJESHWARI VIRENDRA SAGAR

Doctor of Optometry

ASTON UNIVERSITY

June 2014

"Essentially, the theory of probability is nothing but good common sense reduced to mathematics. It provides an exact appreciation of what sound minds feel with a kind of instinct, frequently without being able to account for it."

Pierre Simon Laplace (1749-1827) (1)

Bayes' ……."crack cocaine of statistics…. Seductive, addictive and ultimately destructive"

An unknown Google representative who recruited Bayesians for Google (2)

Aston University

Application of Naïve Bayesian sequential analysis to primary care optometry

Rajeshwari Virendra Sagar

Doctor of Optometry, June 2014

Abstract

The objective of this study was to investigate the effects of circularity, comorbidity, prevalence and presentation variation on the accuracy of differential diagnoses made in optometric primary care using a modified form of naïve Bayesian sequential analysis. No such investigation has ever been reported before.

Data were collected for 1422 cases seen over one year. Positive test outcomes were recorded for case history (ethnicity, age, symptoms and ocular and medical history) and clinical signs in relation to each diagnosis. For this reason only positive likelihood ratios were used for this modified form of Bayesian analysis that was carried out with Laplacian correction and Chi-square filtration. Accuracy was expressed as the percentage of cases for which the diagnoses made by the clinician appeared at the top of a list generated by Bayesian analysis.

Preliminary analyses were carried out on 10 diagnoses and 15 test outcomes. Accuracy of 100% was achieved in the absence of presentation variation but dropped by 6% when variation existed. Circularity artificially elevated accuracy by 0.5%. Surprisingly, removal of Chi-square filtering increased accuracy by 0.4%. Decision tree analysis showed that accuracy was influenced primarily by prevalence followed by presentation variation and comorbidity.

Analysis of 35 diagnoses and 105 test outcomes followed. This explored the use of positive likelihood ratios, derived from the case history, to recommend signs to look for. Accuracy of 72% was achieved when all clinical signs were entered. The drop in accuracy, compared to the preliminary analysis, was attributed to the fact that some diagnoses lacked strong diagnostic signs; the accuracy increased by 1% when only recommended signs were entered. Chi-square filtering improved recommended test selection. Decision tree analysis showed that accuracy again influenced primarily by prevalence, followed by comorbidity and presentation variation.

Future work will explore the use of likelihood ratios based on positive and negative test findings prior to considering naïve Bayesian analysis as a form of artificial intelligence in optometric practice.

(key words: Bayesian; clinical decision making; likelihood ratios; differential diagnosis; epidemiology; optometry; primary care;)

Dedication

To my father, who started with me on my journey in optometry 40 years ago with a dream,

And

To my husband, who has supported my journey for last 30 years

Thank you

Acknowledgements

Firstly I would like to note my sincere thanks to Dr Mark Dunne, my supervisor, for his knowledge, enthusiasm, guidance and immeasurable support and encouragement through emails and Skype. His expertise with Excel spread sheets has allowed the unprecedented exploration of this subject and made light work of some very complex calculations. Time, distance and information technology have been overcome to produce this thesis and would not have been possible without his unreserved dedication and totally unruffled patience in exploring an alien subject albeit with immense potential for optometry.

Many thanks also to Prof Ian Nabney and Dr Alexis Boukouvalas for their guidance on naïve Bayesian analysis.

Thanks to Dr Richard Armstrong for his guidance in many statistical matters.

Thanks to Dr Anna Sanyiwa and Eden Mashayo for their support in matters relating to Tanzania.

To my staff at Eyeline Optometrists, thank you for the wonderful support and encouragement.

Thanks to my son Brijesh, for his support in IT related matters and most of all for automating data input, making 24 hours of work complete in 15 minutes! My daughter and fellow optometrist Bansi, and son Prashant for celebrating with me each hurdle overcome during the Doctorate programme; your support has helped me to reach the end!

Contents

Abstract....3
Dedication....4
Acknowledgements....5
List of figures....9
List of tables....10
Glossary....12
1 Introduction....15
1.1 Global Action Plan 2014-2019....15
1.2 The need for effective differential diagnosis....17
1.3 Scope of the present study....17
2 Bayes Theorem....19
2.1 What is Bayes' theorem?....19
2.2 Early History of Bayes' Theorem....19
2.3 The Early 20th Century....20
2.4 Bayes' applications during World War II....21
2.5 The Cold War Years....22
2.6 Bayes' theorem comes of age....24
2.7 The future of Bayes' theorem ….....24
2.8 Previous applications of Bayes' theorem in ophthalmic research....25
2.9 Summary....25
3 Naïve Bayesian Analysis....27
3.1 Decision Matrices....27
3.2 Multiple levels of test outcomes....30
3 Worked examples of naïve Bayesian analysis....31
3.4 Factors affecting Bayesian Analysis....32
3.4.1 The effect of Prevalence....32
3.4.2 The effect of Comorbidity....35
3.4.3 The Effect of Presentation Variation....38
3.5 Circularity....38
3.6 Use of the Laplacian Correction....38
3.7 Naïve Bayesian sequential analysis....42
3.8 Summary....47

4 Methodology....49
4.1 Introduction....49
4.2 Ethical Approval....49
4.3 Data Collection....50
4.4 Methods....51
4.4.1 Splitting the database....51
4.4.2 Comorbidity....52
4.4.3 Prevalence....52
4.4.4 Presentation variation....52
4.4.5 Decision matrices with Laplacian correction....53
4.4.6 Chi-Square filtering of likelihood ratios....53
4.4.7 Naïve Bayesian sequential analysis....54
4.4.8 Accuracy....56
4.4.9 Circularity....56
4.4.10 Use of Microsoft® Excel® spreadsheets....56
4.4.11 Decision tree analysis (DTA)....65
4.5 Summary....67
5 Findings of the Preliminary Analysis....75
5.1 Selection of tests and diagnoses....68
5.2 Results....70
5.2.1 Distribution and categorization of prevalence, comorbidity and presentation variation....70
5.2.2 Overall accuracy of naïve Bayesian sequential analysis....71
5.2.3 Decision Trees Analysis....72
5.3 Summary....75
6 Findings of the Main Analysis....83
6.1 Selection of tests and diagnoses....76
6.2 The use of positive likelihood ratios to select diagnostic tests....76
6.3 Results....78
6.3.1 Overall accuracy of naïve Bayesian sequential analysis on all test items....78
6.3.2 Overall accuracy of naïve Bayesian sequential analysis on recommended test items....82
6.4 Decision tree analysis....91
6.5 Summary....86

7 Discussion…… 96
Appendix I Aston Ethical Clearance 111
Appendix II Tanzanian Ethical Clearance 113
Appendix III Ophthalmic Procedures 118
Appendix IV Comparison of LR+ between the preliminary study and the main study 120
Appendix V Comparison of recommended tests with and without Chi-square 122

List of figures

Figure 1 From Vision 2020- The Right to sight to the Global Action Plan 2014-2019 15
Figure 2 A summary of the Global action Plan 2014 -2019: Aims and Objectives 16
Figure 3 shows more than one test is used to give a final post-test probability. Here IOP, CDR measurement and presence or absence of ISNT rule determines the final probability of a positive diagnosis of POAG. 44
Figure 4 Staff of Eyeline Optometrists, Dar es Salaam 50
Figure 5 Flow chart showing the process of naïve Bayesian sequential analysis using only LR+. In the main study, the analysis was extended to include tests recommended by the use of LR+. 55
Figure 6 Dashboard of Excel spreadsheet showing how naive Bayesian sequential analysis achieves clinical accuracy by correctly identifying the same diagnoses as the clinician. Light grey bars represent the diagnoses made by the clinician and the blue bars represent the diagnosis made by naïve Bayesian sequential analysis. ... 65
Figure 7 Prevalence in Datasets A & B 70
Figure 8 Decision tree analysis for analysis AA (circular analysis with presentation variation) 72
Figure 9 Decision tree analysis for analysis AB (non- circular analysis with presentation variation) 73
Figure 10 Due to a lack of a definitive sign, the diagnosis of hypertensive retinopathy is missed completely. Similarly, the diagnosis of nuclear cataract is missed as the signs associated with this diagnosis (hazy view of fundus, small retinal haemorrhages) are equivocal at best 79
Figure 11 Naïve Bayesian sequential analysis misses a diagnosis as there is no recorded sign. 81
Figure 12 Incomplete recording of signs achieves partially accurate diagnoses, but also achieves high probability for a spurious diagnosis. In this case, the two diagnoses recorded by the clinician were "dry eye" and "allergic conjunctivitis". However, the only sign that was recorded was "conjunctival small papillae". The diagnosis for the dry eye has been made on the history and symptoms. A high diagnostic probability for "uncorrected ametropia" is also seen, which was not recorded by the clinician; thus in this case although the original diagnoses were confirmed by naïve Bayesian sequential analysis, accuracy is lost by the introduction of a spurious diagnosis. ... 82
Figure 13 Decision tree analysis for naive Bayesian sequential analysis using all 105 test items comprising of non-circular data with presentation variation 84
Figure 14 Decision tree analysis based on only recommended tests following history and symptoms comprising of non-circular data with presentation variation 85
Figure 15 ROC curves (after Aspinal and Hill) 92

List of tables

Table 1 Decision Matrix (contingency table) (109), (107) 28
Table 2 Worked example of naïve Bayesian analysis carried out on a test (reduced vision) in relation to a specific diagnosis (cataract). 31
Table 3 Worked example of naïve Bayesian analysis carried out on a test (intraocular pressure > 24 mmHg) in relation to a specific diagnosis (glaucoma) in the primary care setting in which the prevalence of glaucoma is taken to be 2.5%. 33
Table 4 Worked example of naïve Bayesian analysis carried out on a test (intraocular pressure > 24 mmHg) in relation to a specific diagnosis (glaucoma) in a secondary care setting in which the prevalence of glaucoma is taken to be 40%. 34
Table 5 Worked example of naïve Bayesian analysis carried out on a test (reported reduced vision) in relation to a specific diagnosis (uncorrected ametropia). Chi-square with Yates' correction has been applied to the diagnostic matrix and reveals a strong association between the test and the diagnosis being considered. 36
Table 6 Worked example of naïve Bayesian analysis carried out on a test (reported reduced vision) in relation to a specific diagnosis (pinguecula). Chi-square with Yates' correction has been applied to the diagnostic matrix and reveals a weak association between the test and the diagnosis being considered. 37
Table 7 Worked example of the calculation of prevalence, likelihood ratios and post-test probability before the application of the Laplacian correction. Here, if a negative test occurs the negative likelihood ratio would be applied and this effectively reduces the post-test probability to zero. In a sequential analysis, this would rule out any further diagnosis, even if the likelihood ratio applied later was of a very high magnitude. 40
Table 8 Worked example of the calculation of prevalence, likelihood ratios and post-test probability after the application of the Laplacian correction. Note how addition of a correction of 0.001 to each cell of a diagnostic matrix makes little difference to calculations apart from removing the possibility of post-test probabilities of zero. 41
Table 9 Worked example of the calculation of prevalence, likelihood ratios and post-test probability after the application of the Laplace smoothing add 1 correction. Note how addition of a correction of 1 to each cell of a diagnostic matrix makes a greater difference to the prevalence, and the post-test probability of the diagnosis being present following a negative test result compared to the example in Table 3-8. 42
Table 10 Three tests used to inform decision making about POAG with their sensitivity, specificity and associated likelihood ratios 43
Table 11 Pre-test prevalence of primary open angle glaucoma is set at 2%. Note how the post-test probability of a positive diagnosis of primary open angle glaucoma increases with each successive positive test result. 45
Table 12 Pre-test probabilities for the four conditions are shown on the left. The probabilities are changed with the application of a positive test result for itchy eye. Two conditions, ocular allergy and dry eye have significantly raised post-test probabilities. 46
Table 13 Carrying out a second test, the presence of palpebral papillae, the post-test probabilities change again and now a diagnosis of allergy becomes significantly more likely than a diagnosis of dry eye. Note how, with a positive test result for palpebral papillae, the post-test probability for a viral eye condition actually is reduced. 46
Table 14 In this example, there are two presenting symptoms, itchy eye that has a high association with ocular allergy and burning which is highly associated with dry eye. To distinguish between the two conditions, two signs are looked for and a positive result for fluorescein staining and a negative result for palpebral papillae make dry eye the more likely of the two conditions. Note, that prior to determining the presence of palpebral papillae, but after noting the presence of fluorescein staining, the probability for dry eye is 99% and ocular allergy is 80%; after determining the absence of palpebral papillae the probability of dry eye remains unchanged, but the probability of ocular allergy is reduced to 29%. This demonstrates the importance of recording negative findings during the eye examination. 47

Table 15 showing an example of how Microsoft spread sheets were used to carry out the Bayesian analysis, using the relational ability between tables. Note that the example uses a 4x7decision matrix, compared to the preliminary study which used a 10x15 matrix and the main study which used a 35x105 matrix. The table for calculating negative likelihood ratios (although not used in the study is included for completeness) *57*
Table 16 The 10 diagnoses selected for analysis together with their associated definitive signs and associated reported symptoms *69*
Table 17 Categories of prevalence, comorbidity and presentation variation used for decision tree analyses. *71*
Table 18 The overall percentage accuracy of naïve Bayesian sequential analysis achieved with and without circularity and presentation variation. *71*
Table 19 The Variation of positive likelihood ratios for uncorrected presbyopia with age. *90*
Table 20 The variation of positive likelihood ratios with age for cataract, primary open angle glaucoma, diabetic retinopathy and age-related macular degeneration. *91*
Table 21 Weighted kappa was worked out for two test/diagnosis combinations (reduced vision/uncorrected ametropia and reduced vision/pinguecula) for different values of "r". *93*

Glossary

Definitions of terms as used within the context of this study

algorithm	a sequence of steps in order to solve a problem
ban	Probability expressed in logarithm to the base 10 to ease calculations (introduced by Alan Turing).
Bayesian network	a graphical model representing the probabilities and their relationships
Chi-square filtering	Chi-square test is used to test the strength between a test item and the condition being tested. Filtering is carried out by using only strong associations (determined by by a Chi-square value that ensured that there would be no false positives errors), so that weak or spurious associations are not used in the naïve Bayesian analysis
circularity	testing of the Bayesian analysis using the same data that was used determine the prevalence and build the likelihood ratios in the first place
comorbidity	the coexistence of more than one eye condition or disease in the same individual
decision matrix	is a table of rows and columns containing data that allows the analysis of the data and determine their relationships. In this study, only 2x2 decision matrices have been used, that is 2 rows by 2 columns. Decision matrices are also known as contingency tables
decision tree analysis	a method of multivariate analysis that is used to classify statistical data in hierarchical manner, with the ability to handle both discrete and continuous variables.
diagnosis	an eye disease or condition
frequentist statistical theory	the drawing of conclusions from sample data with emphasis on the frequencies of the data, using hypothesis testing and confidence intervals to see how well the sample represents the population.
Gibbs sampling	a form of Monte Carlo sampling based on Markov chains
heuristic	enabling a person to discover or learn for themselves
kappa	is an agreement coefficient statistic used to measure the agreement between tests which has been corrected for agreements achieved by chance. Maximum value of kappa is 1 which achieved when there is perfect agreement.
kappa - weighted, (k_r)	is where cells in the decision matrix are weighted according to their importance after which the kappa coefficient is applied. When the weighting "r" =1 , this gives kappa value for a screening test, that is sensitivity is maximised. For r=0, kappa value is for a diagnostic test, that is specificity is maximised; when r=0.5, both sensitivity and specificity are given equal importance and false positives and false negatives are minimised.

Glossary –Cont.	
likelihood ratio - positive	the ratio of sensitivity to the false positive error rate ***LR+ = sensitivity/ (1-specificity)***
likelihood ratio- negative	the ratio of the false negative error to specificity ***LR- = (1-sensitivity)/ specificity***
Markov Chain	a stochastic process in which future states are independent of past states given the present state
MCMC	A combination of Markov chains and Monte Carlo methods
Monte Carlo	a computer generated simulation of a probability distribution by taking random samples
naïve Bayesian analysis	Bayesian analysis which assumes that all tests items are independent of each other, and test outcomes are independent of other test outcomes.
naïve Bayesian sequential analysis	in the context of this study, naïve Bayesian analysis with a continuous input of data, altering outcomes with the input of new data where the order of new data being applied to the analysis is not of significance
odds - post test	the odds of the condition being present after a test outcome. If the test outcome is positive, then ***Post-test odds = Pre-test odds x positive likelihood ratio .*** If however the test outcome is negative, then ***Post-test odds = Pre-test odds x negative likelihood ratio***
odds - pre-test	the odds of the condition being present prior to testing. ***Pre-test odds=pre-test probability/ (1- pre-test probability)***
ophthalmic procedure	examination of a particular structure (e.g. cornea or optic disc) or measurement (e.g. intraocular pressure or fixation disparity)
predictive value - negative	proportion of people who with a negative test outcome who do not have the condition
predictive value - positive	proportion of people with a positive test outcome who actually have the condition
presentation variation	the difference between observed clinical data and expected "textbook" data
prevalence	the ratio of the number cases showing that particular diagnosis to the total number of cases.
priors - equal	equal probability is assigned as the pre-test probability as previous data to relating this diagnosis is vague and inconclusive
priors - subjective	where the pre-test probability is based on a persons belief as opposed to being based on previous data
priors - unequal	unequal probability is assigned to the pre-test probability when there is previous informative data relating to the diagnosis, that is, the prevalence.
priors -objective	where the pre-test probability is based on previous data
probabilities (post-test)	is the probabilty of the diagnosis after the application of the test
probabilities (pre-test)	is the prevalence of the diagnosis in practice

Glossary - Cont.	
QROC	is a graphical plot of the weighted kappa k_1 to k_0. The weighted kappa represent "quality indices" in an ROC curve (therefore QROC)
receiver operator characteristic (ROC) curve	a graphical plot of the true positive rate against the false positive rate that is ***sensitivity against (1-specificity)*** for different pass/fail criteria. ROC curves are used to determine optimal models for decision making
recommended tests	following history and symptoms, the post-test odds determine the 6 most likely diagnosis. Recommended tests are those which would then confirm the diagnosis, and these are those test items that have the highest positive likelihood ratios indicating a strong test item/ condition association.
sampling	using a finite number of observations to learn about a much larger population
sensitivity	proportion of people with a condition who will have a positive test outcome
sequential analysis	a continuous analysis of data; essentially analysing new data as it arrives whilst taking into account previous data.
specificiity	proportion of people without the condition who will have a negative test outcome
statistical inference	drawing conclusions from data using statistical methods
stochastic	of or pertaining to a process involving a randomly determined sequence of observations each of which is considered as a sample of one element from a probability distribution.
test item	any clinical recording that can provide information to aid/improve diagnosis (such as ocular or medical history demographics, presenting symptoms, clinical signs)
test outcome	the test outcome is whether or not a test item is present or absent

1. Introduction

This chapter briefly sets the stage for the research presented in this thesis before describing the scope of the study and an outline of what each chapter covers.

1.1 Global Action Plan 2014-2019

In May 2013, the World Health Assembly approved the Global Action Plan (GAP) for the Prevention of Avoidable Blindness and Visual Impairment 2014-2019 (3). This supersedes the Vision 2020, the Right to Sight (4) (5), public health initiative of the World Health Organisation (WHO) and the members of the International Agency for the Prevention of Blindness (IAPB). The goal of the GAP is to "reduce avoidable visual impairment and to secure rehabilitation services for the visually impaired". The GAP purpose is to achieve this goal by improving access to comprehensive eye care services which are a part of a general health system. This will be achieved by

- Collecting epidemiological evidence
- Training more eye health professionals
- Provide comprehensive eye care

Figure 1 From Vision 2020- The Right to Sight to the Global Action Plan 2014-2019

According to the WHO and IAPB (4) (5) about 285 million people in the world today are estimated to be visually impaired. Of these, 39 million are estimated to be blind. Of all the people that are visually impaired, it is estimated that 80% are due to preventable causes. The United Republic of Tanzania became a signatory in 2003 and is currently running its 2nd Strategic Plan (6). The summary of the GAP is shown in Figure 2

Figure 2 A summary of the Global action Plan 2014 -2019: Aims and Objectives

1.2 The need for effective differential diagnosis

Eye disease is most likely to be detected first in the primary care setting during routine eye examinations. Failure to detect early eye disease leads, at best, to the need for more sophisticated and costly secondary care treatments and, at worst, to avoidable loss of vision that brings with it other socioeconomic costs (7). On the other hand, unnecessary referrals lead to overburdening of secondary eye care services (8). This potentially leads to delivery of suboptimal treatment in overstretched hospitals. Effective and timely referral between eye care professionals can ensure the best treatment possible, and efficient use of costly secondary care resources (9).

Primary care optometrists are therefore an essential part of the strategy to eliminate avoidable blindness and the consequences of sight threatening conditions can be limited if optometrists make a timely diagnosis and manage eye disease appropriately.

There are currently only about 190 registered optometrists in Tanzania (10), serving a population of 43 million people (11), (12). Efforts are being made to increase the manpower and the level of clinical training. However, one of the consequences of this shortage in man power is that Tanzanian optometrists are frequently called upon to make clinical decisions that extend beyond their training.

In the UK, there is a different kind of pressure. Commercial pressure to reduce chair time (13) puts the optometrist at risk of not detecting sight threatening conditions in time due to a lack of clinical vigilance. Clinical vigilance guided by the recognition of symptoms and signs is important for making accurate diagnoses and the most appropriate referrals (14).

Poor record keeping and missed diagnoses can be a potential minefield for any optometrist, with possibly severe legal implications. Computerised record keeping systems ensure comprehensive and legible records, but currently lack artificial intelligence.

1.3 Scope of the present study

This thesis explores the factors influencing the accuracy of a modified form of naïve Bayesian sequential analysis as a means of providing artificial intelligence for making differential diagnoses and selecting diagnostic tests. As far as its author is aware, the investigations described in this study have never appeared in the ophthalmic literature before now.

Chapter 2 introduces Bayes' theorem from an historical perspective. Chapter 3 provides worked examples to demonstrate the simplicity of naïve Bayesian calculations and the theoretical influence

on its accuracy of factors such as prevalence, comorbidity, presentation variation and circularity. The use of the Laplacian correction, Chi-square filtration and sequential analysis are also discussed in chapter 3. Chapter 4 covers the methodology of the preliminary and main studies of this thesis. Chapter 5 presents the findings of a preliminary study designed to determine the influence of circularity, prevalence, comorbidity, presentation variation and Chi-square filtering on the accuracy of naïve Bayesian sequential analysis applied to a small group of diagnoses, each of which had at least one definitive diagnostic test outcome. Chapter 6 shows the findings of the main study designed to re-examine the influence of prevalence, comorbidity, presentation variation and Chi-square filtering on the accuracy of naïve Bayesian sequential analysis applied to a larger group of diagnoses, some of which did not necessarily have definitive diagnostic tests outcomes. The use of positive likelihood ratios to identify diagnostic tests was also explored in chapter 6. Chapter 7 summarises the findings of this thesis and makes recommendations for future study.

2. Bayes' Theorem

This chapter introduces the 250 year old history of Bayes' theorem and presents a brief overview of its applications to date in ophthalmic research.

2.1 What is Bayes' theorem?

Thomas Bayes', a mathematician and theologian who lived in the 18th century, proposed a mathematical basis for the change in probability of an event following new evidence (15) (16) . The concept of Bayes' theorem is simple: An initial belief (referred to as a prior probability or just a prior or, in the clinical context, a pre-test probability) is altered by new evidence (that may be expressed in terms of a likelihood ratio) to give a new and improved belief (posterior probability or, in the clinical context, a post-test probability).

Systems that apply Bayes' theorem are not only relatively simple but can self-learn, leading closer and closer to certainty as more data is entered (2) (17). Bayesian models also have the ability to self-correct; that is, when the inputted data changes, the model changes to accommodate the new data (18).

2.2 Early History of Bayes' Theorem

Bayes' Theorem has had a turbulent history since its first emergence in 1763. Thomas Bayes was very much involved with the issues of the early 18th century, and although he only published one paper during his lifetime, defending and explaining Newton's theory of calculus (19) (20), he was an active member of an informal group that peer reviewed other mathematicians' papers. Bayes' work was known to have been influenced by Richard Price (who edited and posthumously published the work of Thomas Bayes (15)), Jakob Bernoulli, Abraham de Moivre and possibly also the philosopher David Hume (21). Bayes provided a reasoned argument for inference from observed frequencies to unknown probabilities (that is, from effect to cause), something also, independently, worked on by Pierre Simon Laplace (22). This was the inverse of the Bernoulli theorem and de Moivre's theory (from cause to effect (23)). Bayes assumed equal priors, which would be modified on collection of further data.

Across the English Channel in France, the prediction of the return of Halley's Comet by French scientists of the time convinced Laplace that natural events could be revealed by mathematics.

Laplace was, like Bayes, influenced by de Moivre's work on probability. In the mid-18th century, probability was mainly applied to gambling, assessing situations of commercial risk, and philosophical questions such as the existence of God. Laplace presented his paper "Memoire on the Probability of Causes given Events" in 1773 (22). This was the very first expression of Bayesian statistical inference. Initially Laplace assigned equal priors to his theory but later amended it to include unequal priors. Laplace then used this theory in demography and judicial reform before applying it to studies of astronomy, the tides and barometric pressure.

During Laplace's lifetime, and immediately afterwards, the thirst for facts grew, whether they were birth statistics, the number of crimes in a city, the number of undelivered letters in the Paris post office, or the number of cholera cases. These were then used to make decisions in government and other institutions. As the number of facts grew the need for Bayesian analysis, (which worked well in the face of uncertainty), diminished and frequency analysis (which worked better when hard facts were available) took over in the latter half of the 19th Century.

2.3 The Early 20th Century

Joseph Bertrand (24) used Bayes' theorem to assist the French army artillery to improve their performance. Bertrand advocated the strict use of equal priors only when they were truly equal or when there was actually no prior information. Bertrand's textbooks and methods were used by both the French and Russian military. At that time, all the judiciary attended military school and had studied Bayes' theorem from Bertrand's textbooks. When Alfred Dreyfus, a Jewish French army officer, was convicted of spying for the Germans his defence lawyer requested the assistance of Henri Poincare who used Bayes' theorem to dismiss a letter presented by the prosecution as forgery thus exonerating Dreyfus (25).

In 1907, Bell Telephones Systems was facing financial collapse. A financial plan to make the company more economically viable, based on Bayes' theorem by Edward Molina, helped Bell to convince a banking consortium, led by the House of Morgan, to give financial support crucial for Bell's survival (26) (27).

In another important development, in the early 20th century, legislative changes were introduced in the United States which required employers to provide employees with occupational injury and illness insurance. Thus insurance premiums had to be formulated for a wide variety of circumstances. Isaac Rubinow, working for the American Medical Association, collated insurance claims from all over

the world to give the prior information in order to set premiums (28) (29) (30). Whitney in 1918 formulated the credibility theory for the insurance industry based on Bayes' theorem.

Although Bayes' theorem was being used by the military and many other disciplines, statisticians were being strongly influenced by the theories of Karl Pearson and Ronald Fisher in the early 1920's. Pearson and Fisher promoted the frequentist theories for statistics, condemning Bayes' theorem, and introduced sampling theory, tests of significance, analysis of variance and experimental design methods. Fisher redefined uncertainty, not by probability but by relative frequency (31) (32) (33) (34). Egon Pearson (the son of Karl Pearson) used both Bayes' theorem and Fisher's frequentist statistical theories to advance statistical theoretical work (35). Egon Pearson then teamed with Jerzy Neyman to develop the Neyman-Pearson theory of hypothesis testing (36). Fisher and Neyman were staunch anti-Bayesians and their influence prevented the use and the advancement of Bayes' theorem in statistics for almost 50 years. Their objection to Bayes' theorem was mainly due to the fact that the priors being used were subjective, that is that the prior probability was determined according to an individual's "personal belief" and was subject to great variation. However, they found it acceptable to use Bayes' theorem when prior probabilities were "objective" that is, the prior probabilities were based on previous data, or on a collective rather than individual belief.

Around the same time, in the 1920's, three people (Emile Borel, Frank Ramsey and Bruno de Finetti) independently came to the conclusion that a person's subjective belief could be quantified by the amount that he was willing to bet (as in a horse in a race, for example). This was an important step forward for Bayes' theorem, in that the controversial subjective prior could now be quantified. De Finetti is particularly recognised as having put the use of subjective priors on a firm mathematical basis (37) (38) (39) (40). However, de Finetti's work was not recognised for a long time in the predominantly English speaking Bayesian circles.

Harold Jeffreys, a contemporary of Fisher, also made a major contribution to the development of Bayes' theorem after Laplace. Jeffreys studied earthquakes and, using this theorem, was able to determine their epicentres based on the tsunamis that followed the earthquakes (41)

2.4 Bayes' applications during World War II

Alan Mathison Turing is credited with the modern revival of Bayes' theorem. Turing's primary contribution was the decoding of German "Enigma" messages during World War II (42); especially that of the U boats. Turing quantified information in terms of a "ban" which is similar to today's byte

(used in information technology); Irving John Good, Turing's statistical assistant, wrote about 900 articles on Bayes' theorem, most of which were published.

Another good friend of Turing's, Claude Shannon (also a committed Bayesian) developed information theory (43), programmed a machine to play chess (44), performed cryptography analysis (45) and developed communications theory (46).

Also during the Second World War, Andrei Kolmogorov used Bertrand's tables to testing artillery for the Russian army. His theoretical work showed that probability can legitimately be used in both frequentist and Bayes' theory (47) (48). Abraham Wald tested the quality of ammunitions which led him to develop sequential analysis, that is, the continuous analysis of data, as it arrives, whilst taking into account previous data (49).

2.5 The Cold War Years

At the height of the Cold War, search techniques were developed using Bayes' theorem by Bernard Osgood Koopman (50), (51) (52). This helped in locating lost Hydrogen-bombs as well as submarines. Crews were rescued by the Coast Guards after their boats had capsized using Bayesian and Monte Carlo Methods (53). Monte Carlo methods use repeated sampling to determine the properties of a phenomenon.

In civilian life, Bayes' theorem had found many uses. The Essen-Moller index was a probability index based on Bayes' theorem and used to determine paternity in lawsuits until DNA profiling became available. Mickey et al (54) showed that the index is quite reliable, provided a realistic prior is used. The Credibility Index, used by the insurance industry and developed by Whitney in 1918, was further updated by Arthur Bailey. Bailey discusses the fact that the insurance business uses Bayes' theorem where prior knowledge is combined with current knowledge where the current knowledge is also weighted in order to make decisions in the face of uncertainty of future events (27). Bailey's son, Robert used Bayes' theorem to give drivers merit when they had no previous insurance claims (55) (56).

Jimmie Savage's book, "***The Foundations of Statistics***" in 1954 (57), gave further impetus to the revival of Bayes' theorem and Savage himself led many statisticians on the road to accepting this theorem.

In medicine, Jerome Cornfield of the National Institute of Health (NIH) applied Bayes' theorem to epidemiology (58) and showed a positive link between lung cancer and smoking (59). He also showed the link between cardiovascular disease and smoking, cholesterol, heart abnormalities and blood pressure (60). Homer Warner created the first computerised programme for detecting disease in 1961. Basing it on 1,000 children with various congenital heart diseases, Warner showed that Bayes' theorem could identify their underlying problems quite accurately (61). Warner had a total of 54 tests but this could be reduced to 7 or 8 tests to give the proper diagnosis (62).

At the Harvard Business School, John Pratt, Howard Raiffa and Robert Schlaifer applied Bayes' theorem to business decision making. Here the theorem flourished, because often decisions needed to be made fast on limited information (63).

Allen Birnbaum introduced likelihood theory (64) derived from observed data only. This followed on from George Barnard's work on statistical inference (65).

One of the largest Bayesian projects undertaken before the advent of personal computers was the determination of authorship of the "Federalist papers" by Fred Mosteller and David Wallace (66). This project took more than ten years to complete with Wallace at times working full time, together with students and statistical assistants. Mosteller also found that Bayesian models proved to be more accurate than predictions based solely on prior probabilities.

John Tukey used Bayes' theorem to predict the Presidential election win for John F Kennedy in 1960; six hours before the official declaration. He, like Turing, had also worked on Enigma and decoding Soviet messages. Tukey also worked at Bell Labs where his work included the development of the cathode ray tube, development of anti-aircraft guided missiles and improving statistical sampling (67). Recommendations on improving statistical analysis became a priority after criticism of the statistical methods used in the Rasmussen Report on nuclear safety (68) and the Kinsey study on the sexual behaviour of the human male (69).

Adrian Raftery discovered that with Bayesian analysis he could discover how abrupt changes can affect data. Raftery was analysing the change in fatalities in coal mining in the later 1880's and early 1890's and discovered that the fall in fatalities was associated with the establishment of the Coal Worker's Union, whose main concern was safety (70).

2.6 Bayes' theorem comes of age

The full force of the Bayesian revolution was felt in the 1980s. Considerably more powerful and easily accessible computers together with advances in sampling theory (71) (72) (73) (74) made Bayesian inference more applicable to a variety of different situations. Sampling allows Bayesian analysis to draw inferences even when data is missing or incomplete. The development of the BUGS (Bayesian inference using Gibbs sampling) software by the Medical Research Council Biostatistics Unit, Cambridge in 1989 followed by WinBUGS developed together with Imperial College School of Medicine London (75) allowed the fitting of complex statistical models using Markov chain Monte Carlo (MCMC) methods. Further development was also built on recently declassified information relating to the use of Bayes' theorem during World War II. Such advances allowed Bayesian analysis to use increasingly larger datasets and ask increasingly complex questions. (76) (77)

Thus, so far, Bayesian analysis has had many applications, ranging from analysing bird and animal distributions, mining, crop production, space programs, medical diagnosis and research (78) (79), engineering, decision analysis, artificial intelligence, email filtering, theoretical mathematical analysis (80) (81) (82) (83), child abuse (84) and prediction of the occurrence of epidemic meningitis (85)

Bayesian analysis has also been used to differentiate the different levels of risk for disease treatment in medical negligence cases (86). This demonstrates the use of evidence based medicine and best practice in medico-legal cases and could be applied in the field of optometry.

The search for missing Air France plane Flight 447 from Rio de Janerio to Paris in 2009 was conducted using Bayes' theorem. An early assumption that the black box or the cockpit voice recorder would be emitting a signal led to delays. However, when the assumption was changed (that is the black box or the cockpit voice recorder was not emitting a signal), the search area was significantly reduced and the wreckage of the plane was found (87).

2.7 The future of Bayes' theorem …..

So where does Bayes' go from here? Every day data is being collected as part of routine tasks: institutions such banks, social media (Facebook and Twitter, for example), healthcare and sales. Incredible computer power today can use this data to make predictions based on Bayesian networks (88). Computers learn from accumulated data (the prior) to make predictions about the future. Predictions can be made in healthcare (89), risk analysis (insurance), marketing, crime and fraud, safety and efficiency, politics, government, education and human resources (90) (91). Such vast

quantities of data represent human experience on an unprecedented scale, which Bayes' theorem can transform into powerful decision making tools.

2.8 Previous applications of Bayes' theorem in ophthalmic research

Aspinall & Hill (92) described how Bayes' theorem could be applied to making rational clinical decisions in optometry. Since Aspinall & Hill's work in 1983, applications in optometry and ophthalmology have included detection of ophthalmic disease from visual acuity and contrast sensitivity (93) and from colour vision (94) (95), the prediction of visual acuity following post- cataract surgery (96) (97), the evaluation of visual field defects in patients with and without glaucoma (98) (99) (100) (101), prediction of retino-neural function following Nd-YAG laser in patients with posterior sub-capsular opacification (102) and prediction of childhood myopia (103).

Most of the studies in eyecare, considered above, have been applications of naïve Bayesian analysis. Hand and Yu (104) have shown that naïve analysis works well in most cases and that the accuracy of naïve Bayesian analysis improves as the number of variables (tests) increases. Naïve Bayesian analysis supports heuristic solutions to disease diagnosis especially when more than one possible diagnosis is present (comorbidity) (105). The advantage of naïve Bayesian analysis is its mathematical simplicity and for this reason it was applied in the present study. Naïve Bayesian analysis makes the following assumptions:

- Each diagnosis is independent of any other diagnosis (e.g. dry eye is often associated with blepharitis; naïve Bayesian analysis would assume no relationship between the two)
- Independence of tests outcomes (e.g. a patient with corneal abrasion might present with two test outcomes - photophobia and cornea Fluorescein staining; Bayesian analysis would assume that these two test outcomes were independent of each other (106).

2.9 Summary

Relatively recent publications that have introduced the idea of applying Bayes' theorem to help optometrists and ophthalmologist make clinical decisions (92), (101), (107), (108) often refer to frequentist statistical methods as traditional when, in actual fact, Bayesian statistical methods fell out of favour and have now returned. The use of Bayes' theorem is now widespread. Though it has seen some application in ophthalmic research, the full potential of naïve Bayes' theorem for use in day-to-day primary care optometry has barely been touched upon. The next chapter demonstrates the

simplicity of naïve Bayesian calculations and introduces some of the factors that might influence its accuracy.

3. Naïve Bayesian analysis

This chapter uses worked examples to demonstrate the simplicity of naïve Bayesian calculations and the influence on diagnostic accuracy of various factors later studied in this thesis.

3.1 Decision Matrices

With conventional probability notation, the probability of the presence of an eye condition/disease is stated as p(D+). Following a positive test result, the probability of an eye condition/ disease = p(D+|T+) and is expressed as:

$$p(D+|T+) = \frac{p(T+|D+) \times p(D+)}{p(T+)}$$

That is, the probability of a positive diagnosis given a positive test result is equal to the probability of a positive test result given a positive diagnosis multiplied by the probability of the presence of the diagnosis (that is, its prevalence) all divided by the probability of a positive test result.(101)

The form of naïve Bayesian analysis carried out in this study essentially calculates the post-test probability of a number of eye conditions or diseases by modifying unequal objective priors (pre-test probability or prevalence) of those conditions or diseases using likelihood ratios (relating to a number of test outcomes).

Sequential analysis refers to the use of more than one test outcome to calculate post-test probabilities for each eye condition or disease irrespective of the order of the tests.

Throughout this thesis a simpler form of probability notation (ref) is applied to decision matrices in order to make the understanding of Bayes' theorem more intuitive to the clinician, and demonstrate the ease with which Bayes' theorem can be applied to routine clinical practice. The number of decision matrices required is the product of the number of eye conditions or diseases and the number of test outcomes considered.

Table 1 shows a 2x2 decision matrix (also known as a contingency table), that is used for Bayesian analyses (as described by Aspinall and Hill (92)). In the definitions below, both the simpler and conventional probability notation is given.

	Condition/Disease		
Test	Present	Absent	**Row Total**
Positive	a	b	a+b
Negative	c	d	c+d
Column Total	a+c	b+d	**a+b+c+d**

Table 1 Decision Matrix (contingency table) (109), (107)

Definitions of the elements of the decision matrix shown in Table 1 are as follows:

- **a+b+c+d** = the total sample size (i.e. the total number of cases seen)
- **a+c** = the total number of cases seen with the diagnosis
- **b+d** = the total number of cases seen without the diagnosis
- **a+b** = the total number of cases testing positive
- **c+d** = the total number of cases testing negative
- **a**= total number of cases with a positive diagnosis and positive test result (observed true positives)
- **b**= total number of cases with a negative diagnosis and positive test result (observed false positives)
- **c**= total number of cases with a positive diagnosis and negative test result (observed false negatives)
- **d**= total number of cases with a negative diagnosis and negative test result (observed true negatives)

From the above elements the following quantities can be calculated

- **Prevalence or pre-test probability =** (a+c)/ (a+b+c+d) using simpler notation and **p(D+)** using conventional notation.
- **Pre-test odds = prevalence / (1 - prevalence)** using simpler notation and **p(D+)/(1-p(D+))** using conventional notation.
- **Sensitivity =** a / (a + c) using simpler notation and **p(T+|D+)** using conventional notation. This is the ratio of observed true positives to all those with a positive diagnosis.
- **Specificity =** d / (b + d) using simpler notation and **p(T-|D-)** using conventional notation. This is the ratio of observed true negatives to all those with negative diagnosis.
- **Positive likelihood ratio (LR+) =** sensitivity / (1 - specificity) using simpler notation and **p(T+|D+)/ p(T+|D-)** using conventional notation. This ratio is used when a positive test result occurs.
- **Negative likelihood ratio (LR-)** = (1 - sensitivity) / specificity using simpler notation and **p(T-|D+)/p(T-|D-)** using conventional notation. This ratio is used when a negative test result occurs. This is shown for completion only. Negative likelihood ratios were not used in the present study because only positive test findings were recorded (see 4.3 Data Collection).
- **Post-test odds** = pre-test odds x LR (positive or negative likelihood ratio is used, depending on the test outcome)

- **Post-test probability** = post-test odds / (1 + post-test odds). This value is equal to the predictive values described by Hill and Aspinall (92), and Parikh et al (107), which can be calculated more directly (that is, without the use of likelihood ratios). However, the use of likelihood ratios has the advantage of enabling sequential analysis (3.7 Naïve Bayesian sequential analysis) and even diagnostic test selection (see 6.2 The use of positive likelihood ratios to select diagnostic tests)
- **Positive Predictive Value** = a/a+b using simpler notation and **p(D+|T+)** using conventional notation. This is the ratio of true positives to all positive and is equal to the probability of a positive diagnosis when the test is positive. A high positive predictive value (that is close to 1) indicates a test that is as good as a "gold standard" test.
- **Negative Predictive Value** = d/c+d using simpler notation and **p(D-|T-)** using conventional notation. This is the ratio of all the true negatives to all negatives and is equal to the probability of the absence of a diagnosis when the test is negative. Again, as with the positive predictive value, if the negative predictive value is close to 1 then the test is as good as a "gold standard" test. The probability of a positive diagnosis following a negative test result is = **p(D+|T-)** and is equal to **1-Negative predictive value.**

LR+ and LR- vary from 0 to infinite. Ratios of greater than 1 raise the post-test probability of a diagnosis. Ratios of less than 1 lower the post-test probability of a diagnosis. Ratios of 1 indicate that a test is not very useful for the diagnosis in question as post-test probability remains unaltered regardless of the test result (110)**,** (111)**.**

Note how likelihood ratios are dependent on the sensitivity and specificity of the test under consideration. That is, they say something about the test but not the probability of disease in the patient who has just been tested. Likelihood ratios have the advantage of portraying the amount by which a test result alters the probability of an eye condition or disease, not just whether a test is positive or negative (112). The fact that likelihood ratios are not influenced by prevalence (108), makes them very useful as, for example, ratios derived from data collected in the secondary (hospital) eye care setting (in which the prevalence of eye disease is relatively high) can be applied in the primary eye care setting (in which the prevalence of eye disease is relatively low).

The positive predictive value and negative predictive value are dependent on prevalence. Where prevalence is low, the number of false positives will be high and therefore the number of individuals with healthy eyes that are mistakenly diagnosed with the condition will be high, leading to unnecessary treatment. Thus predictive values are not constant over different clinical settings (being dependent on prevalence), unlike the likelihood ratio (being independent of prevalence) which can, therefore, be obtained from one clinical setting and used in another. (113), (107).

3.2 Multiple levels of test outcomes

Naïve Bayesian analysis assumes that all tests are independent of each other (see 2.8 Previous applications of Bayes' theorem in ophthalmic research)

Optometric clinical reality is not quite so simple. Certain measurements have multiple values, such as age, cup-to-disc ratios, and intraocular pressure measurements, and the clinician has to decide at which point a certain measurement assumes diagnostic importance. An incorrect decision would result in a missed diagnosis, or unnecessary referral. Therefore choosing the correct value is of immense diagnostic importance. In this study, approved guidelines for "best practice" were used to determine test/ diagnostic levels of importance such as those from the National Institute for Health & Clinical Excellence (NICE) and College of Optometrists Clinical Management Guidelines (114), (115) as applied by the author in daily practice.

3 Worked examples of naïve Bayesian analysis

Table 2 shows a worked example of how the information in a decision matrix is used to calculate the post-test probability a diagnosis being present. Here, the test is whether or not reduced vision is reported and the diagnosis is cataract.

Observed

Reduced Vision (test)	Cataract (diagnosis)		Total
	Present	Absent	
Positive	36	217	253
Negative	25	433	458
Total	61	650	711

Here:

Prior probability =	**0.0858**
Sensitivity =	0.5902
Specificity =	0.6662
LR+ =	1.7678
LR- =	0.6152
If a test positive occurs ...	
Pre-test odds =	0.0938
Post-test odds =	0.1659
Post-test probability =	**0.1423**
If a negative test occurs …	
Pre-test odds=	0.0938
Post-test odds =	0.0577
Post-test probability=	**0.0546**

Table 2 Worked example of naïve Bayesian analysis carried out on a test (reduced vision) in relation to a specific diagnosis (cataract).

The pre-test probability, from the prevalence for this sample, is 9%. A positive test result (that is, the reporting of reduced vision) increases the probability of cataract from 9% to 14%, whereas a negative test result (that is, the patient does not report a reduction in vision) reduces the probability of cataract from 9% to 5%.

Although the present study did not use negative test findings and therefore the negative likelihood ratio, the usefulness of a negative test outcome is demonstrated here and in the following examples.

3.4 Factors affecting Bayesian Analysis

Various factors potentially influence the accuracy of Bayesian analysis. These include prevalence (116), comorbidity and variation in presentation of symptoms and signs

3.4.1 The effect of Prevalence

Table 3and Table 4 show how likelihood ratios calculated for a specific test give rise to very different post-test probabilities in primary and secondary eye care settings in which the prevalence of eye disease is likely to be, respectively, relatively low and high. For example, the prevalence of glaucoma for a population seen in primary care is thought to be between 1- 4% (117) (118). In a secondary care setting, the prevalence may be much higher. Using the figures cited by Parikh et al (108), Table 3 shows the post-test probability of glaucoma when a test (high intraocular pressure > 24 mmHg, sensitivity 50%, specificity 92% (119)) is used in the primary care setting in which the prevalence of glaucoma (prevalence 2.5%) is low.

Diagnosis with low prevalence

Intraocular pressure > 24 (test)	Open Angle Glaucoma		
	Present	Absent	Total
Positive	125	780	905
Negative	125	8970	9095
Total	250	9750	10000

Here:

Prior probability =	**0.0250**
Sensitivity =	0.5000
Specificity =	0.9200
LR+ =	6.2500
LR- =	0.5435

If a test positive occurs ...

Pre-test odds =	0.0256
Post-test odds =	0.1603
Post-test probability =	**0.1381**

If a negative test occurs …

Pre-test odds=	0.0256
Post-test odds =	0.0139
Post-test probability=	**0.0137**

Table 3 Worked example of naïve Bayesian analysis carried out on a test (intraocular pressure > 24 mmHg) in relation to a specific diagnosis (glaucoma) in the primary care setting in which the prevalence of glaucoma is taken to be 2.5%.

Table 4 shows the post-test probability of glaucoma that arises when the same test is used in secondary care, where perhaps the prevalence of glaucoma rises to 40%.

Diagnosis with high prevalence

Intraocular pressure > 24 (test)	Open Angle Glaucoma		
	Present	Absent	Total
Positive	2000	480	2480
Negative	2000	5520	7520
Total	4000	6000	10000

Here:

Prior probability =	**0.4000**
Sensitivity =	0.5000
Specificity =	0.9200
LR+ =	6.2500
LR- =	0.5435

If a test positive occurs ...

Pre-test odds =	0.6667
Post-test odds =	4.1667
Post-test probability =	**0.8065**

If a negative test occurs …

Pre-test odds=	0.6667
Post-test odds =	0.3623
Post-test probability=	**0.2660**

Table 4 Worked example of naïve Bayesian analysis carried out on a test (intraocular pressure > 24 mmHg) in relation to a specific diagnosis (glaucoma) in a secondary care setting in which the prevalence of glaucoma is taken to be 40%.

As the same test (intraocular pressure > 24 mmHg) was used in both the primary (Table 3) and secondary (Table 4) care settings, the same sensitivity (50%) and specificity (92%) were used in both sets of calculations. Given that positive and negative likelihood ratios are calculated using sensitivity and specificity, then it follows that that these values also remain the same (6.25 and 0.54, respectively) in both settings. However, the post-test probability for a positive test result differs dramatically in the primary (14%) and secondary (81%) care settings. So exactly the same test carried out in two different settings gives rise to very different levels of suspicion that glaucoma exists. It follows that with low prevalence a higher portion of positive test results will be false positives (113). The importance of the clinical setting has been recognised in previous studies (120). In qualitative terms, a positive test result in the primary care setting may only slightly raise the suspicion of eye disease, whereas a positive test in a secondary care setting may raise suspicion significantly (121). It has thus been recognised that when prevalence is not taken into account in interpreting positive test results, unnecessary diagnostic errors can occur (122), (123), (124), (125). Accounting for prevalence is a key feature of Bayesian analysis that sets it apart from other forms of statistical analysis (e.g. t-

tests, analyses of variance, that are all frequentist in nature (see 2.3 The Early 20th Century) that are more used in clinical research (92).

3.4.2 The effect of Comorbidity

Comorbidity refers to the coexistence of eye conditions or diseases in the same individual (126). This could confound the linkage between test findings and specific diagnoses. This factor was considered by a research group that applied Bayesian analysis to the diagnosis of dementia (127), (128). Their findings indicated that Bayesian analysis could be successfully performed even in the presence of comorbidity. Table 5 and Table 6 illustrate confounding arising from comorbidity together with the additional influence of prevalence and suggest that Chi-square filtering is a possible remedy.

Table 5 reveals, as might be expected, that uncorrected ametropia has a high prevalence (81%) in the primary care setting. One would expect this condition to be associated with reported reduced vision and, taken as a test for uncorrected ametropia; this symptom has a sensitivity of 40% and a specificity of 82% with corresponding positive and negative likelihood ratios of 2.20 and 0.74, respectively.

Now consider pinguecula (Table 6). This would not normally be associated with reported reduced vision. However, given the high prevalence of uncorrected ametropia, it is to be expected that some individuals with uncorrected ametropia will also have pinguecula. Pinguecula is also relatively uncommon (prevalence = 2%). So a rather high proportion of cases with pinguecula might also have uncorrected ametropia and reported reduced vision. Consequently, the decision matrix shown in Table 6 shows that reported reduced vision, as a test for pinguecula, has an unexpectedly high sensitivity (56%) and specificity (65%) with positive and negative likelihood ratios (1.63 and 0.66, respectively) of approximately the same magnitude as found for uncorrected ametropia.

Observed

Reduced Vision (test)	uncorrected Ametropia (diagnosis)		
	Present	Absent	Total
Positive	228	25	253
Negative	345	113	458
Total	573	138	711

Here:

Prior probability =	**0.8059**
Sensitivity =	0.3979
Specificity =	0.8188
LR+ =	2.1964
LR- =	0.7353

If a test positive occurs ...

Pre-test odds =	4.1522
Post-test odds =	9.1200
Post-test probability =	**0.9012**

If a negative test occurs …

Pre-test odds=	4.1522
Post-test odds =	3.0531
Post-test probability=	**0.7533**

Chi-square with Yates' correction =	21.86
df =	1
p-value =	<0.001

Table 5 Worked example of naïve Bayesian analysis carried out on a test (reported reduced vision) in relation to a specific diagnosis (uncorrected ametropia). Chi-square with Yates' correction has been applied to the diagnostic matrix and reveals a strong association (p-value < 0.001) between the test and the diagnosis being considered.

Observed

Reduced Vision (test)	Pinguecula		Total
	Present	Absent	
Positive	8	245	253
Negative	6	452	458
Total	14	697	711

Here:

Prior probability =	**0.0197**
Sensitivity =	0.5714
Specificity =	0.6485
LR+ =	1.6257
LR- =	0.6609

If a test positive occurs ...

Pre-test odds =	0.0201
Post-test odds =	0.0327
Post-test probability =	**0.0316**

If a negative test occurs …

Pre-test odds=	0.0201
Post-test odds =	0.0133
Post-test probability=	**0.0131**

Chi-square with Yates' correction =	2.02
df=	1
p-value=	>0.05

Table 6 Worked example of naïve Bayesian analysis carried out on a test (reported reduced vision) in relation to a specific diagnosis (pinguecula). Chi-square with Yates' correction has been applied to the diagnostic matrix and reveals a weak association (p-value >0.05) between the test and the diagnosis being considered.

It is proposed that using a Chi-square filter might be a remedy. Here, Chi-square is used to indicate the strength of the association between any test and diagnosis. Weakly associated test-diagnosis combinations could be disregarded. Calculating Chi-square involves determining expected counts in each cell of the corresponding decision matrix and, in instances where expected counts drop below 10, Yates' correction is advised (129). For computational simplicity Yates' correction could be carried out on all diagnostic matrices. Table 5 shows that the association between reported reduced vision and uncorrected ametropia is strong (Chi-square with Yates' correction = 21.9, df = 1, p-value <0.001). Table 6 on the other hand, shows that the association between reported reduced vision and

pinguecula is weak (Chi-square with Yates' correction = 2.0, df = 1, p-value > 0.05). Note how Chi-square alters a great deal in both instances despite the likelihood ratios being very similar.

The proposed use of Chi-square to filter out weakly associated test-diagnosis combinations may be unnecessary. Table 5 shows that, because uncorrected ametropia is common (that is, its prevalence is high), a positive test outcome (i.e. reduced vision is reported) gives rise to a post-test probability of 90%. Table 6, however, shows that, because pinguecula is relatively rare (that is, its prevalence is low), a positive test outcome gives rise to a post-test probability of only 3%. So spurious associations between tests and diagnoses that arise due to comorbidity may have only a very small impact on accuracy. Nevertheless, Gill et al. (130) pointed out that small elevations of post-test probability can accumulate when large numbers of tests are used. As the question of whether or not to use Chi-square filtering remains equivocal, the effect on accuracy of using such a filter was tested in the present study.

3.4.3 The Effect of Presentation Variation

It is likely that variations in presentation of various eye conditions and diseases will alter accuracy in a similar manner to that described above for comorbidity. For example, not all people with uncorrected ametropia will report reduced vision. This was also tested in the present study.

3.5 Circularity

Circularity, that is, the testing of the Bayesian analysis using the same data that was used determine the prevalence and build the likelihood ratios in the first place could, theoretically, overestimate the performance of the analysis. Therefore, this also was tested in the present study.

3.6 Use of the Laplacian Correction

Problems arise if any cell in a diagnostic matrix contains zero counts. This could result in likelihood ratios of zero and, in turn, post-test probabilities of zero would occur at any point in during naïve Bayesian sequential analysis (see section 3.1). A post-test probability of zero would absolutely rule out a diagnosis even if subsequent tests strongly indicated that the diagnosis was actually present. A remedy for this is to add a very small increment to cell counts a-d in each decision matrix Table 8. This is called a Laplacian correction (131). Commonly "1" is added to each cell and is known as "add 1

smoothing". However to ensure that the increment added to the cells made the least difference to the original cell counts, an addition of 0.001 to the count in each cell was considered to be sufficient, and thus ensured that no diagnosis can be entirely ruled out (after all, it is not possible to be 100% certain of the presence or absence of any diagnosis). The probability of such a diagnosis being present just becomes very small instead. Table 7 and Table 8 show a worked example demonstrating the effect of the Laplacian correction adopted in the present study. Table 9 shows the effect of using the commonly used "add 1 smoothing" and as can be seen by comparing Table 8 and Table 9, the larger the increment added to each cell, the greater the effect on the calculated prevalence and the post-test probability.

Although there are many different methods of smoothing and eliminating the effects of a zero count in any of the cells, the investigation of these was considered to be beyond the scope of this study.

Hypothetical

Reduced Vision (test)	Cataract (diagnosis)		
	Present	Absent	Total
Positive	61	531	592
Negative	0	119	119
Total	61	650	711

Here:

Prior probability =	**0.0857947**
Sensitivity =	1.0000000
Specificity =	0.1830769
LR+ =	1.2241055
LR- =	0.0000000

If a test positive occurs ...

Pre-test odds =	0.0938462
Post-test odds =	0.1148776
Post-test probability =	**0.1030405**

If a negative test occurs …

Pre-test odds=	0.0938462
Post-test odds =	0.0000000
Post-test probability=	**0.0000000**

Table 7 Worked example of the calculation of prevalence, likelihood ratios and post-test probability before the application of the Laplacian correction. Here, if a negative test occurs the negative likelihood ratio would be applied and this effectively reduces the post-test probability to zero. In a sequential analysis, this would rule out any further diagnosis, even if the likelihood ratio applied later was of a very high magnitude.

Hypothetical With Laplacian Correction

Reduced Vision (test)	Cataract (diagnosis)		
	Present	Absent	Total
Positive	61.001	531.001	592.002
Negative	0.001	119.001	119.002
Total	61.002	650.002	711.004

Here:

Prior probability =	**0.0857970**
Sensitivity =	0.9999836
Specificity =	0.1830779
LR+ =	1.2240869
LR- =	0.0000895

If a test positive occurs ...

Pre-test odds =	0.0938489
Post-test odds =	0.1148793
Post-test probability =	**0.1030419**

If a negative test occurs …

Pre-test odds=	0.0938489
Post-test odds =	0.0000084
Post-test probability=	**0.0000084**

Table 8 Worked example of the calculation of prevalence, likelihood ratios and post-test probability after the application of the Laplacian correction. Note how addition of a correction of 0.001 to each cell of a diagnostic matrix makes little difference to calculations apart from removing the possibility of post-test probabilities of zero.

Hypothetical With Laplace smoothng Correction "add 1"

Reduced Vision (test)	Cataract (diagnosis)		
	Present	Absent	Total
Positive	62	532	594
Negative	1	120	121
Total	63	652	715

Here:

Prior probability =	**0.0881119**
Sensitivity =	0.9841270
Specificity =	0.1840491
LR+ =	1.2061105
LR- =	0.0862434

If a test positive occurs ...

Pre-test odds =	0.0966258
Post-test odds =	0.1165414
Post-test probability =	**0.1043771**

If a negative test occurs …

Pre-test odds=	0.0966258
Post-test odds =	0.0083333
Post-test probability=	**0.0082645**

Table 9 Worked example of the calculation of prevalence, likelihood ratios and post-test probability after the application of the Laplace smoothing add 1 correction. Note how addition of a correction of 1 to each cell of a diagnostic matrix makes a greater difference to the prevalence, and the post-test probability of the diagnosis being present following a negative test result compared to the example in Table 8.

3.7 Naïve Bayesian sequential analysis

Primary eye care does not involve carrying out one isolated test to indicate the presence of just one diagnosis. In fact, a battery of screening tests is carried out, often leading to more than one tentative diagnosis.

To apply Bayesian analysis to such a situation, the post-test odds of the first test becomes the pre-test odds for the second. In turn, the post-test odds of the second test become the pre-test odds for the third and so on (130). In this sequential manner both positive and negative likelihood ratios may be used to raise or lower the final probabilities of each possible diagnosis depending on the outcome of each test (though in the present study, only positive likelihood ratios were used, see 4.3 Data Collection). Wald and Wolfowitz (49) had used the term “sequential analyses” as long ago as 1950.

For the purposes of this study, naïve Bayesian sequential analysis refers to a continuous analysis of data; essentially analysing new data as it arrives whilst taking into account previous data, where the order of new data being applied to the analysis is not relevant.

For example, the diagnosis of primary open angle glaucoma (POAG) is usually made after several tests which inform the decision making process. In the following example, 3 tests (intraocular pressure [IOP] >= 24mmHg; cup to disc ratio [CDR] >= 0.7; If ISNT rule is true or false) are used to change the post-test probability of a positive diagnosis of POAG. Table 10 shows the three tests being used and their sensitivity, specificity and associated likelihood ratios.

	Sensitivity	Specificity	LR+	LR-
IOP >=24mmHg	50	92	6.2500	0.5435
ISNT	72	79	3.4286	0.3544
CDR >= 0.7	20	99	20.0000	0.8081

Table 10 Three tests used to inform decision making about POAG with their sensitivity, specificity and associated likelihood ratios

Taking the prevalence of POAG to be 2%, the pre-test odds are calculated. With a positive test result, the pre-test odds are multiplied by the LR+ to give the post-test odds. This, in turn becomes the pre-test odds for the next test. So,

$\text{Post-test odds}_{IOP} = \text{pre-test odds} \times LR+_{IOP}$

$\text{Post-test odds}_{ISNT\ \&\ IOP} = \text{post-test odds}_{IOP} \times LR+_{ISNT} = \text{pre-test odds} \times LR+_{IOP} \times LR+_{ISNT}$

$\text{Post-test odds}_{CDR\&\ ISNT\ \&\ IOP} = \text{Post-test odds}_{ISNT\ \&\ IOP} \times LR+_{CDR} = \text{pre-test odds} \times LR+_{IOP} \times LR+_{ISNT} \times LR+_{CDR}$

Thus , for n number of tests

$\text{Post-test odds}_n = \text{Pre-test odds} \times LR+_1 \times LR+_2 \times LR+_3 \ldots\ldots\ldots LR+_n$

And from this is can be seen that the order of the tests is not critical to the analysis.

Using the three tests (that is IOP, ISNT rule and CDR measurement) the post-test probabilities can be calculated at each stage.

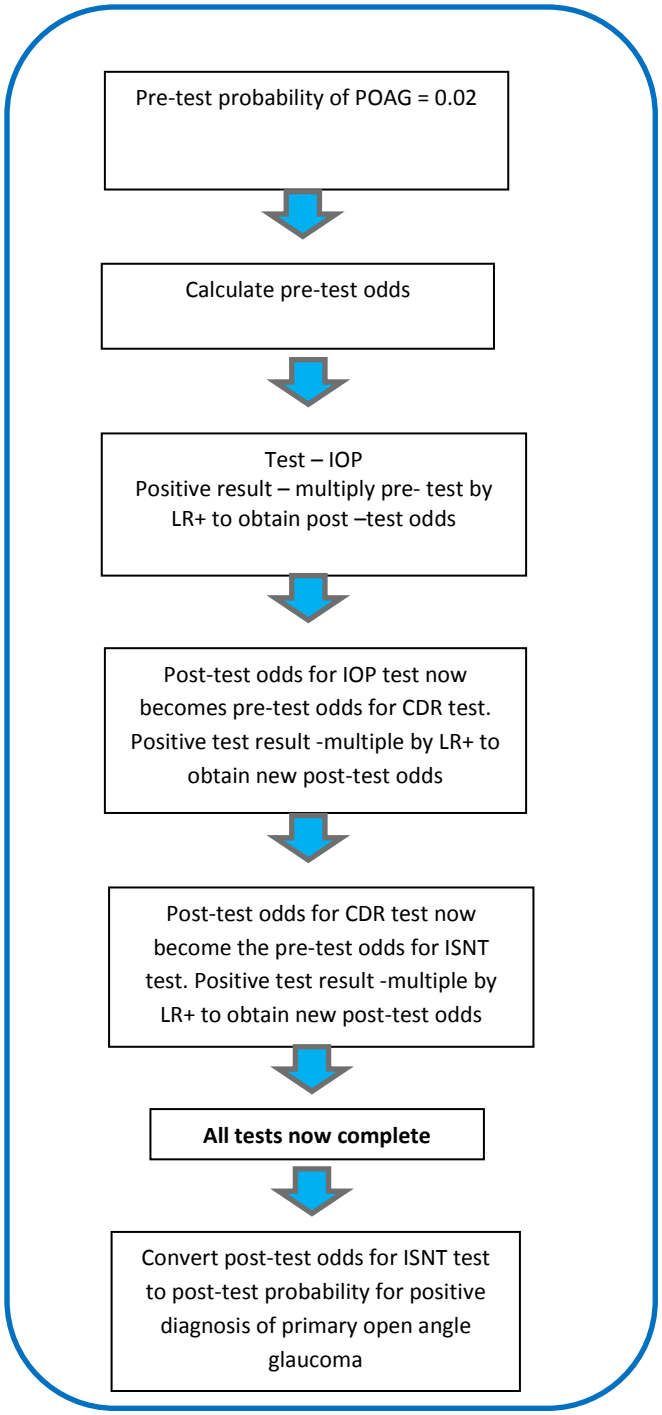


Figure 3 shows more than one test is used to give a final post-test probability. Here IOP, CDR measurement and presence or absence of ISNT rule determines the final probability of a positive diagnosis of POAG.

	Pre-test Prevalence	IOP	CDR	ISNT
No test	0.0200	0.0200	0.0200	0.0200
IOP only	0.0200	0.1131	0.1131	0.1131
IOP and CDR	0.0200	0.1131	0.7184	0.7184
IOP, CDR &ISNT	0.0200	0.1131	0.7184	0.8974

Table 11 Prevalence of primary open angle glaucoma is set at 2%. Note how the post-test probability of a positive diagnosis of primary open angle glaucoma increases with each successive positive test result.

To take this example further, consider the following four ocular surface conditions, dry eye, allergic conjunctivitis, bacterial conjunctivitis and viral conjunctivitis. Using 4 symptoms as reported by the patient and 3 signs as noted by the optometrist, Bayesian analysis can give the possible differential diagnosis between the four ocular conditions. Test outcomes are represented by 1=positive test result, 0= negative test result, blank = test not carried out.

Table 12, Table 13, and Table 14, show how a test or a combination of tests can improve the differential diagnosis achieved. Table 14 also shows the advantage of recording a negative sign in improving the diagnosis.

		presenting symptoms				signs			
	Pre-test probability	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae	Final Post-test probability (%)
test outcome? >>		1							
dry	0.1525	0.3726	0.3726	0.3726	0.3726	0.3726	0.3726	0.3726	37.26
allergic	0.0654	0.4949	0.4949	0.4949	0.4949	0.4949	0.4949	0.4949	49.49
bacterial	0.0196	0.0291	0.0291	0.0291	0.0291	0.0291	0.0291	0.0291	2.91
viral	0.0040	0.0068	0.0068	0.0068	0.0068	0.0068	0.0068	0.0068	0.68

Table 12 Pre-test probabilities for the four conditions are shown on the left. The probabilities are changed with the application of a positive test result for itchy eye. Two conditions, ocular allergy and dry eye have significantly raised post-test probabilities.

		presenting symptoms				signs			
	Pre-test probability	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae	Final Post-test probability (%)
test outcome? >>		1						1	
dry	0.1525	0.3726	0.3726	0.3726	0.3726	0.3726	0.3726	0.5024	50.24
allergic	0.0654	0.4949	0.4949	0.4949	0.4949	0.4949	0.4949	0.9920	99.20
bacterial	0.0196	0.0291	0.0291	0.0291	0.0291	0.0291	0.0291	0.0206	2.06
viral	0.0040	0.0068	0.0068	0.0068	0.0068	0.0068	0.0068	0.0000	0.00

Table 13 Carrying out a second test, the presence of palpebral papillae, the post-test probabilities change again and now a diagnosis of allergy becomes significantly more likely than a diagnosis of dry eye. Note how, with a positive test result for palpebral papillae, the post-test probability for a viral eye condition actually is reduced.

		presenting symptoms				signs			
	Pre-test probability	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae	Final Post-test probability (%)
test outcome? >>		1	1			1		0	
dry	0.1525	0.3726	0.8330	0.8330	0.8330	0.9920	0.9920	0.9920	99.20
allergic	0.0654	0.4949	0.8045	0.8045	0.8045	0.8045	0.8045	0.2916	29.16
bacterial	0.0196	0.0291	0.0431	0.0431	0.0431	0.0000	0.0000	0.0000	0.00
viral	0.0040	0.0068	0.0219	0.0219	0.0219	0.0511	0.0511	0.0511	5.11

Table 14 In this example, there are two presenting symptoms, itchy eye that has a high association with ocular allergy and burning which is highly associated with dry eye. To distinguish between the two conditions, two signs are looked for and a positive result for fluorescein staining and a negative result for palpebral papillae make dry eye the more likely of the two conditions. Note, that prior to determining the presence of palpebral papillae, but after noting the presence of fluorescein staining, the probability for dry eye is 99% and ocular allergy is 80%; after determining the absence of palpebral papillae the probability of dry eye remains unchanged, but the probability of ocular allergy is reduced to 29%. This demonstrates the importance of recording negative findings during the eye examination.

In this manner, a list of alternative diagnoses can be obtained and, by ranking them in order of ascending probability, the most likely diagnoses can be identified. Thus, Bayesian analysis has the potential to assist in increasing the accuracy of eye examinations carried out by community optometrists.

3.8 Summary

Recall, from chapter 2, that one of the historical objections to earlier Bayesian analyses related to the use of equal and subjective priors. Less objectionable unequal objective priors are used in this thesis. These come in the form of the prevalence of eye conditions or diseases. Prevalence is converted to pre-test odds that are altered using likelihood ratios generated for each test outcome (representing new evidence) to give post-test odds and, ultimately, probability. Likelihood ratios have a number of advantages and one of them is to enable sequential analysis of many test outcomes, including those with multiple levels. Naïve Bayesian sequential analyses tested in this thesis include Laplacian correction. In addition, Chi-square filtering is explored, for the first time, as a means of overcoming some of the factors (prevalence, comorbidity and presentation variation) that might influence the

accuracy of naïve Bayesian sequential analysis. Circularity is another factor that is explored for the first time. The next chapter provides a description of the methodological aspects of this study.

4. Methodology

This chapter covers the methodological aspects of this study.

4.1 Introduction

The purpose of the present study was to explore factors influencing the accuracy of naïve Bayesian sequential analysis when applied to clinical data collected in the primary eye care setting. As such, it is important for the reader to understand that neither the provision of eye care nor the range of eye diseases encountered in this part of Tanzania were the primary focus of the study.

The study was retrospective, i.e. it posed a question and looked back at data to find the answers, as opposed to a prospective study, which would pose a question and then design a study to find the answers. (132)

Confidential clinical data (such as patient age, sex, ethnicity, symptoms, signs and diagnoses) are being collected in primary care practices all over the world every day. This wealth of data is very rarely used for epidemiological studies, and yet can inform public health policies as well as informing clinical decision-making by individual eye care practitioners.

4.2 Ethical Approval

The clinical data needed to be collected without prior consent as any refusal to participate in this study would corrupt estimates of prevalence (i.e. pre-test probability) that are essential for Bayesian analysis. Attempts to gain consent early on in the study raised subtle issues of trust. For example, patients may wonder why an optometrist would need to carry out such research. Might it be that optometrists doubt their ability to make clinical decisions without the assistance of computers? If that is the case, then are optometrists to be trusted? Fortunately, there is precedent for unconsented retrospective analysis of fully anonymised clinical data (effectively a clinical audit (133)). This case was made to the research ethics committees of Aston University and Tanzania's National Institute of Medical Research (NIMR) and ethical approval for the study was granted. A research certificate was also obtained from the Commission for Science and Technology (COSTECH) in Tanzania. Clearances from both NIMR and COSTECH ensured that the project was undertaken in accordance with Tanzanian laws and protocols (see pages 111 & 113)

4.3 Data Collection

The clinical data for this analysis was collected from Eyeline Optometrists (see Figure 4), a family owned community practice, in Dar es Salaam city centre. The data constituted clinical records of routine eye examinations [134] [135].

Figure 4 Staff of Eyeline Optometrists, Dar es Salaam

The patient base at Eyeline Optometrists is varied, but not typical of Tanzania. This is partly due to the location of the practice. The language used in the practice varies almost equally between English, Kiswahili (the local African language) and Gujarati (spoken by most Indians). Hindi and French speaking patients are occasionally seen.

The time period chosen for data collection was one year, from October 2010 to October 2011. Initially, data was recorded as necessary for the purposes of primary care practice. Practice data needs to inform the practitioner's decisions, whether at that particular examination or at an examination sometime in the future [136]. Here, positive and negative test findings have to be recorded as part of the practitioner's legal defence in case of a complaint. For the purposes of this study, cases that presented for follow-up visits, where diagnoses had been previously made, were excluded from the analysis.

Prior to any analysis, all cases were anonymised by the removal of any personal details that could lead to identification of a particular patient. Of the 1524 cases seen during the designated time period, 1422 met the criteria for analysis. In these 1422 cases, 199 test outcomes were found to be associated with 57 diagnoses. Preliminary examination of this data revealed that further refinement was necessary in the interests of consistency (for example, where several reported visual complaints actually represented the same symptom or where the exact nature of a symptom, for example its laterality or onset, were not recorded because they did not aid diagnosis at the time of the eye examination). This refinement process was extremely time consuming and leads to the recommendation for future research of this type that data collection should be prospective and should use standardised test outcomes and diagnoses.

Only positive test outcome were included in the analysis, as the data contained positive outcomes only of tests carried out. While positive test outcome had an unambiguous meaning, the absence of a positive sign was open to the following interpretation:

- Either the test was carried out and was negative
- Or the test was not carried out at all.

Because of this ambiguity, it was not considered valid to assume that the absence of a positive test sign could be interpreted as a negative test result and thus prevented the use of negative likelihood ratios in the naïve Bayesian sequential analysis. The consequence of this decision was that a modified form of naïve Bayesian analysis was performed that only included positive likelihood ratios.

4.4 Methods

4.4.1 Splitting the database

The refined data was split into two equal datasets of 711 cases each (datasets A and B). These represented the first 711 cases seen (dataset A) and then the second 711 cases seen (dataset B). Creation of the two datasets allowed the accuracy of Bayesian analysis to be tested with and without circularity.

Each dataset was checked against the other to ensure that the same diagnoses and tests occurred in both datasets. This left the two datasets with 105 test items and 35 diagnoses. Bayesian analysis on this data required the construction of 3,675 decision matrices (that is, 105 tests x 35 diagnoses).

4.4.1.1 Preliminary analysis

A preliminary investigation was carried out on 10 diagnoses and 15 test items. This reduced the analysis to 150 decision matrices (that is, 15 tests x 10 diagnoses), and enabled an initial assessment of the effects on accuracy of circularity, prevalence, comorbidity, presentation variation and Chi-square filtering..

4.4.1.2 Main analysis

The process was then repeated with the full dataset, involving 105 test items and 35 diagnoses and the construction and analysis of 3.675 decision matrices.

4.4.2 Comorbidity

The comorbidity for each case was based on the original complete data (that which included all 57 original diagnoses) as some of the test outcomes could have resulted from diagnoses that were eventually not included but might, nevertheless, have influenced the accuracy of diagnoses arising from Bayesian analysis.

4.4.3 Prevalence

Prevalence was calculated as described in section 3.1. If comorbidity was present, then the median prevalence was calculated from all diagnoses made for that case.

4.4.4 Presentation variation

Datasets A and B showed test outcomes as they naturally occurred (observed data). That is, it was not always the case that "textbook" relationships between diagnoses and test outcomes arose. The impact on the accuracy of Bayesian analysis of natural variations in test outcomes associated with each diagnosis was investigated by creating two further datasets with all presentation variation removed. In the preliminary study and the main study, this was carried out in two very different ways. The presentation variation was the number of mismatches found between the observed dataset and the "textbook" dataset.

4.4.4.1 Presentation Variation in Preliminary Analysis

The calculation of presentation variation required reconstruction of new datasets (A and B) for which all presentation variation was removed. For the preliminary analysis, this was a relatively simple process as only 15 test outcomes were matched to 10 diagnoses; the 15 test outcomes were selected (5.1 Selection of tests and diagnoses) partly because of their unambiguous linkage to each diagnosis. The required presentation variation was then equal to the number of mismatches found, case by case, when comparing the test outcomes for the observed and "textbook" data in both datasets

4.4.4.2 Presentation Variation in Main Analysis

The main study consisted of 3,675 decision matrices representing 105 test outcome and 35 diagnosis combinations. It was not feasible to search the literature for "textbook" associations for all these combination. The supervisor of this project had also previously carried out literature searches of this sort, as part of a series of dissertations completed by final year optometry students, which indicated that the required information was lacking.

The lengthy process of removing presentation variation from 105 test outcomes in relation to 35 diagnoses had to be carried out in two stages. The first stage involved re-combining both datasets (A and B) and calculating Chi-square filtered positive likelihood ratios for all 3675 test outcome/diagnosis combinations. The second stage involved selection of "textbook" test outcomes for each diagnosis by including only those for which the Chi-square filtered positive likelihood ratio was greater than 1 (a value of 1 indicated a test of no diagnostic value) and, in the professional opinion of the author, "made sense". By this means, the 35 diagnoses were typically assigned 2 definitive test outcomes but up to 9 "textbook" test outcomes were noted. Finally a "textbook" dataset B (that for which Bayesian analysis was applied), with presentation variation removed, was constructed by taking the observed diagnoses for each case and replacing the observed test outcomes for each diagnosis with "textbook" test outcomes. . The required presentation variation was then equal to the number of mismatches found, case by case, when comparing the test outcomes for the observed and "textbook" B datasets.

4.4.5 Decision matrices with Laplacian correction

Decision matrices incorporating Laplacian correction were constructed (as shown in section 3.6 Use of the Laplacian Correction) for each test outcome and diagnosis combination. As mentioned earlier, this involved calculation of 150 matrices for the preliminary analysis and 3,675 matrices for the main analysis (see 4.4.1 Splitting the database). These matrices were used to calculate prevalence and positive likelihood ratios (as shown in section 3.1 Decision Matrices)

4.4.6 Chi-Square filtering of likelihood ratios

Chi-square with Yates' correction was used to filter out positive likelihood ratios that represented weakly associated test outcome/diagnosis combinations. Chi-square values were calculated as shown in 3.4.2 The effect of Comorbidity. A critical minimum Chi-square then had to be determined, below which Likelihood ratios would be filtered out. The minimum Chi-square value was determined using a Probability Distribution Calculator (127)

4.4.6.1 Chi-square filtering in the preliminary study

For the preliminary study, 150 decision matrices were constructed (4.4.1.1 Preliminary analysis) In order to avoid even 1 statistical false positive error (that is, to have a probability of a statistical false

positive error of less than 1 in 150 = 0.006667). The Probability Distribution Calculator indicated that the minimum Chi-square value required was 7.36.

4.4.6.2 Chi-square filtering in the main study

Similarly, for the main analysis, 3,675 decision matrices were constructed (see section 4.4.1.2 Main analysis). To avoid a statistical false positive error (i.e. to have a probability of a false positive error of less than 1 in 3,675 = 0.000272) the Probability Distribution calculator indicated the minimum Chi- square value required was 13.25.

4.4.7 Naïve Bayesian sequential analysis

This was performed case by case. All test outcomes were entered into the analysis. The pre-test odds for all diagnosis were multiplied by the Chi-square filtered positive likelihood ratios of each positive test outcome. The final post-test odds for all diagnoses were then converted to post-test probabilities. Diagnoses were then ranked in order or descending post-test probability.

4.4.7.1 Naïve Bayesian sequential analysis in the preliminary study

As described in 4.4.7 Naïve Bayesian sequential analysis, positive likelihood ratios were used to convert pre-test probability into post-test probability (Figure 5).

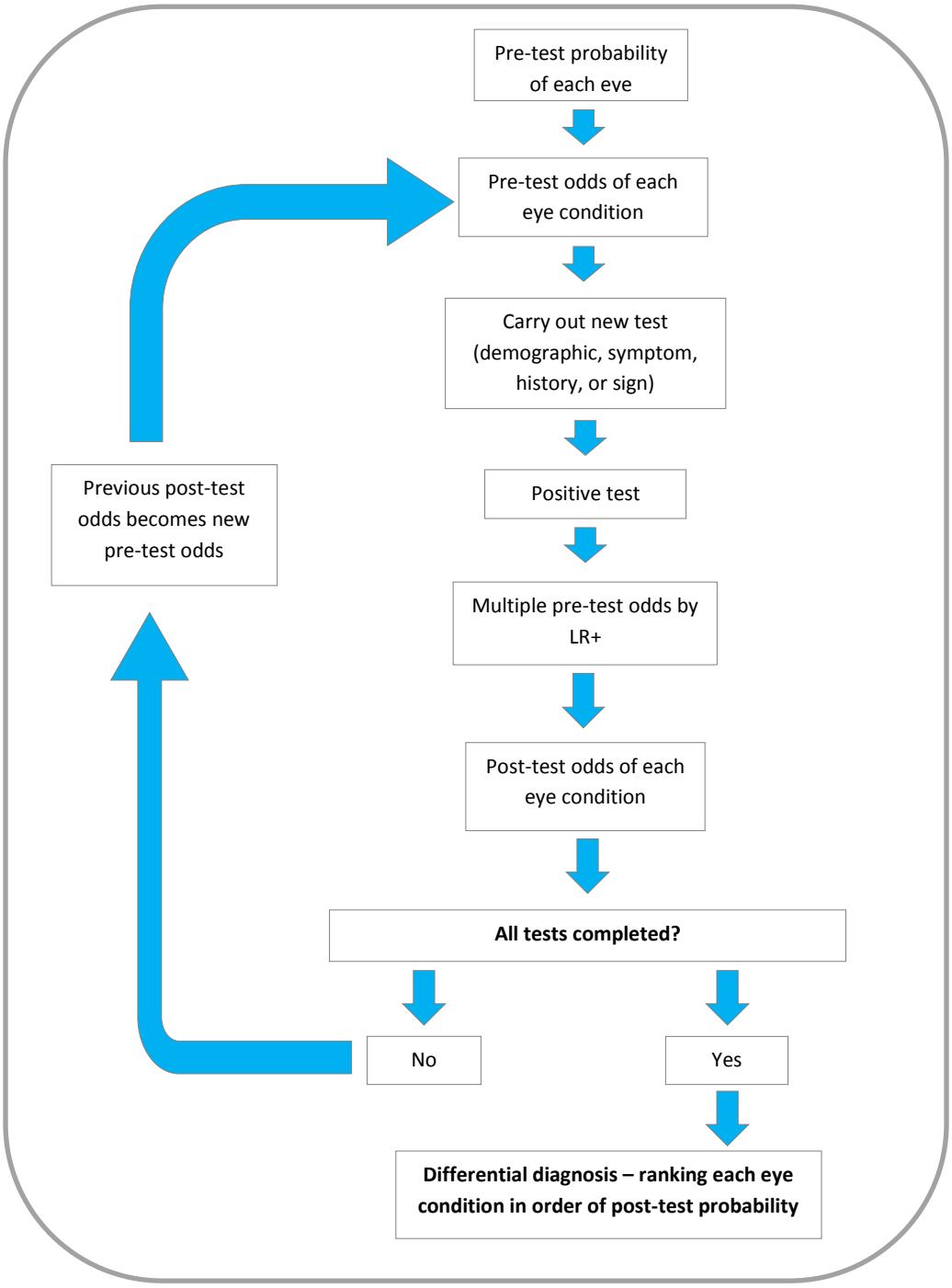


Figure 5 Flow chart showing the process of naïve Bayesian sequential analysis using only LR+. In the main study, the analysis was extended to include tests recommended by the use of LR+.

4.4.7.2 Naïve Bayesian analysis in the main study

As described in 4.4.7 Naïve Bayesian sequential analysis, pre-test probabilities were converted to post-test probabilities using the likelihood ratios. However an extension of the use of likelihood ratios was incorporated to recommend supplementary tests to confirm diagnoses.

4.4.8 Accuracy

Accuracy was expressed as the percentage of cases for which the diagnoses made by the clinician appeared at the top of a ranked list generated by naïve Bayesian sequential analysis.

4.4.9 Circularity

In the preliminary analysis, percentage accuracy was compared with and without circularity. Circularity was present if accuracy was investigated by using the pre-test odds and likelihood ratios generated from dataset A to make diagnoses on the test outcomes of dataset A. On the other hand, circularity was absent if the accuracy was investigated by using the pre-test odds and likelihood ratios generated from dataset A to make diagnoses on the test outcomes of dataset B. The main analysis was carried out in the absence of circularity.

4.4.10 Use of Microsoft® Excel® spreadsheets`

Microsoft® Excel® spreadsheets were used to perform all the calculations necessary for all decision matrices (See example given in Table 15 Tables). The data used in all the calculations to determine prevalence and the likelihood ratios were taken from dataset A.

Table 15 Tables 15.1 to 15.21 serve to illustrate how Microsoft® Excel® spread sheets were used to calculate filtered likelihood ratios. Note that the example shown only considers 4 diagnoses and 7 tests extracted from the main study data (chapter 6). The 4 diagnoses shown are all causes of conjunctivitis (dry eye, allergic, bacterial and viral). The 7 tests shown include 4 symptoms reported by patients (itchy eye, burning eye, watery discharge and sticky discharge) and 3 signs (fluorescein staining, palpebral redness, palpebral papillae). All tables make reference to cells a, b, c and d of the decision matrix described in Table 1. Tables 15.1 to 15.8 create the decision matrices for each test/diagnosis combination. Tables 15.9 to 15.17 lead to the calculation of Chi-square for cells a, b, c and d. Tables 15.18 to 15.20 show the calculation of positive likelihood ratios from sensitivity and specificity. Table 15.21 show filtered likelihood ratios. The critical value of 13.25 shown in Table 15.21 was that used in the main study (see section 4.4.6.2). The prevalence of each diagnosis was calculated by dividing the number of cases seen with the diagnosis (Table 15.2) by the total number of cases seen (Table 15.1).

Table 15.1 The total number of cases seen. This gives cell (a+b+c+d) in each diagnostic matrix. The Laplacian correction of 0.001 is added (a value of 0.004 appears as the total number of cases seen represents 4 cells in each diagnostic matrix).

(a+b+c+d)	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	711.004	711.004	711.004	711.004	711.004	711.004	711.004
allergic	711.004	711.004	711.004	711.004	711.004	711.004	711.004
bacterial	711.004	711.004	711.004	711.004	711.004	711.004	711.004
viral	711.004	711.004	711.004	711.004	711.004	711.004	711.004

Table 15.2 shows the total number of cases seen with the diagnosis. This represents the (a+c) cell of the decision matrix in Table 1. The Laplacian correction appears as 0.002 as the total number of cases with the diagnosis represents 2 cells of the diagnostic matrix.

(a+c)	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	110.002	110.002	110.002	110.002	110.002	110.002	110.002
allergic	44.002	44.002	44.002	44.002	44.002	44.002	44.002
bacterial	11.002	11.002	11.002	11.002	11.002	11.002	11.002
viral	3.002	3.002	3.002	3.002	3.002	3.002	3.002

Table 15.3 shows the number of cases seen without the diagnosis. This is calculated from the above two tables, that is by subtracting the values of Table 15.2 from Table 15.1. This is gives the value of cell (b+d) of the decision matrix.

(b+d)	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	601.002	601.002	601.002	601.002	601.002	601.002	601.002
allergic	667.002	667.002	667.002	667.002	667.002	667.002	667.002
bacterial	700.002	700.002	700.002	700.002	700.002	700.002	700.002
viral	708.002	708.002	708.002	708.002	708.002	708.002	708.002

Table 15.4 shows the number of cases seen that had a positive test result. This represents (a+b) of the decision matrix

(a+b)	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	48.002	49.002	59.002	8.002	39.002	5.002	43.002
allergic	48.002	49.002	59.002	8.002	39.002	5.002	43.002
bacterial	48.002	49.002	59.002	8.002	39.002	5.002	43.002
viral	48.002	49.002	59.002	8.002	39.002	5.002	43.002

Table 15.5 Observed true positives, that is, cases that tested positive with a positive diagnosis. These represent the (a) cell of the decision matrix.

(a)	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	18.001	30.001	29.001	4.001	32.001	1.001	6.001
allergic	21.001	11.001	7.001	3.001	1.001	0.001	38.001
bacterial	2.001	1.001	1.001	3.001	0.001	5.001	1.001
viral	0.001	0.001	0.001	0.001	0.001	1.001	0.001

Table 15.6 Observed false positives, that is, cases that had a positive test result but a negative diagnosis. These represent the (b) cell of the decision matrix. These are calculated by subtracting values of Table 15.5 (a) from Table 15.4 (a+b)

(b)	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	30.001	19.001	30.001	4.001	7.001	4.001	37.001
allergic	27.001	38.001	52.001	5.001	38.001	5.001	5.001
bacterial	46.001	48.001	58.001	5.001	39.001	0.001	42.001
viral	48.001	49.001	59.001	8.001	39.001	4.001	43.001

Table 15.7 Observed false negatives , that is, cases that had a negative test result but a positive diagnosis. These represent the (c) cell of the decision matrix. These are calculates by subtracting the values of Table 15.5 (a) from Table 15.2 (a+c)

(c)	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	92.001	80.001	81.001	106.001	78.001	109.001	104.001
allergic	23.001	33.001	37.001	41.001	43.001	44.001	6.001
bacterial	9.001	10.001	10.001	8.001	11.001	6.001	10.001
viral	3.001	3.001	3.001	3.001	3.001	2.001	3.001

Table 15.8 Observed true negatives, that is, cases that had a negative test result and a negative diagnosis. These represent the (d) cell of the decision matrix and are calculated by subtracting the values of Table 15.6 (b) from Table 15.3 (b+d)

(d)	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	571.001	582.001	571.001	597.001	594.001	597.001	564.001
allergic	640.001	629.001	615.001	662.001	629.001	662.001	662.001
bacterial	654.001	652.001	642.001	695.001	661.001	700.001	658.001
viral	660.001	659.001	649.001	700.001	669.001	704.001	665.001

Table 15.9 Expected true positives that is expected (a), = [(total cases with the diagnosis) *(total cases testing positive)]/(total number of cases). Referring to the decision matrix in Table 1, this is [(a+c)*(a+b)]/(a+b+c+d). The values in this table are calculated from Tables 15.2, 15.4 and 15.1

Expected a	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	7.427	7.581	9.128	1.238	6.034	0.774	6.653
allergic	2.971	3.033	3.651	0.495	2.414	0.310	2.661
bacterial	0.743	0.758	0.913	0.124	0.604	0.077	0.665
viral	0.203	0.207	0.249	0.034	0.165	0.021	0.182

Table 15.10 Expected false positives that is, expected (b) = [(total cases without the diagnosis) *(total cases testing positive)]/(total number of cases). Referring to the decision matrix in Table 1, this is [(b+d)*(a+b)]/(a+b+c+d). The values in this table are calculated from Tables 15.3, 15.4 and 15.1

Expected b	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	40.575	41.421	49.874	6.764	32.968	4.228	36.349
allergic	45.031	45.969	55.351	7.507	36.588	4.692	40.341
bacterial	47.259	48.244	58.089	7.878	38.398	4.925	42.337
viral	47.799	48.795	58.753	7.968	38.837	4.981	42.820

Table 15.11 Expected false negatives, that is, expected (c) = [(total cases with the diagnosis) *(total cases testing negative)]/(total number of cases). Referring to the decision matrix in Table 1, this is [(a+c)*(c+d)]/(a+b+c+d). The values in this table are calculated from Tables 15.2, 15.7,15.8 and 15.1

Expected c	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	102.575	102.421	100.874	108.764	103.968	109.228	103.349
allergic	41.031	40.969	40.351	43.507	41.588	43.692	41.341
bacterial	10.259	10.244	10.089	10.878	10.398	10.925	10.337
viral	2.799	2.795	2.753	2.968	2.837	2.981	2.820

Table 15.12 Expected true negatives, that is, expected (d) = [(total cases without the diagnosis) *(total cases testing negative)]/(total number of cases). Referring to the decision matrix in Table 1, this is [(b+d)*(c+d)]/(a+b+c+d). The values in this table are calculated from Tables 15.3, 15.7, 15.8 and 15.1

Expected d	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	560.427	559.581	551.128	594.238	568.034	596.774	564.653
allergic	621.971	621.033	611.651	659.495	630.414	662.310	626.661
bacterial	652.743	651.758	641.913	692.124	661.604	695.077	657.665
viral	660.203	659.207	649.249	700.034	669.165	703.021	665.182

Table 15.13 is the calculated chi square value for (a) incorporating Yates' correction. Chi square = [(expected value-observed value)^2 -0.5]/ expected value. From the tables above this is calculated by using tables 15.5 and 15.9

Chi-square for a	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	13.666	63.376	41.113	4.137	107.482	0.096	0.003
allergic	103.447	18.393	2.224	8.124	0.345	0.118	456.100
bacterial	0.774	0.087	0.186	45.638	0.017	252.818	0.041
viral	0.439	0.418	0.255	6.461	0.687	10.904	0.562

Table 15.14 is the calculated chi square value for (b) incorporating Yates' correction. From the tables above this is calculated by using tables 15.6 and 15.10.

Chi-square for b	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	2.501	11.600	7.525	0.757	19.673	0.018	0.001
allergic	6.824	1.213	0.147	0.536	0.023	0.008	30.089
bacterial	0.012	0.001	0.003	0.717	0.000	3.974	0.001
viral	0.002	0.002	0.001	0.027	0.003	0.046	0.002

Table 15.15 is the calculated chi square value for (c) incorporating Yates' correction. From the tables above this is calculated by using tables 15.7 and 15.11.

Chi-square for c	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	0.989	4.691	3.720	0.047	6.238	0.001	0.000
allergic	7.490	1.361	0.201	0.092	0.020	0.001	29.361
bacterial	0.056	0.006	0.017	0.519	0.001	1.791	0.003
viral	0.032	0.031	0.023	0.074	0.040	0.077	0.036

Table 15.16 is the calculated chi square value for (d) incorporating Yates' correction. From the tables above this is calculated by using tables 15.8 and 15.12.

Chi-square for d	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	0.181	0.859	0.681	0.009	1.142	0.000	0.000
allergic	0.494	0.090	0.013	0.006	0.001	0.000	1.937
bacterial	0.001	0.000	0.000	0.008	0.000	0.028	0.000
viral	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 15.17 is the chi square value for a+b+c+d and is calculated as the sum of tables 15.13, 15.14, 15.15 and 15.16

Chi-square for a+b+c+d	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	17.338	80.526	53.039	4.949	134.534	0.115	0.004
allergic	118.255	21.057	2.585	8.758	0.389	0.127	517.487
bacterial	0.843	0.095	0.206	46.883	0.019	258.611	0.044
viral	0.473	0.451	0.279	6.562	0.730	11.028	0.601

Table 15.18 Sensitivity is calculated as the ratio of true positives to all positive diagnosis. Referring to the decision matrix, this is the ratio of the cells (a)/ (a+c) and is calculated from Tables 15.2 and 15.5

Sensitivity (a/(a+c))	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	0.164	0.273	0.264	0.036	0.291	0.009	0.055
allergic	0.477	0.250	0.159	0.068	0.023	0.000	0.864
bacterial	0.182	0.091	0.091	0.273	0.000	0.455	0.091
viral	0.000	0.000	0.000	0.000	0.000	0.333	0.000

Table 15.19 Specificity is calculated as the ratio of true negatives to all negative diagnosis. Referring to the decision matrix, this is the ratio of the cells (d)/ (b+d) and calculated from tables 15.3 and 15.8

Specificity (d/(b+d))	presenting symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	0.950	0.968	0.950	0.993	0.988	0.993	0.938
allergic	0.960	0.943	0.922	0.993	0.943	0.993	0.993
bacterial	0.934	0.931	0.917	0.993	0.944	1.000	0.940
viral	0.932	0.931	0.917	0.989	0.945	0.994	0.939

Table 15.20 The positive likelihood ratio is the ratio sensitivity/(1-specificity) and is calculated using tables 15.18 and 15.19

Positive likelihood ratio	symptoms				signs		
	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
Dry	3.278	8.626	5.281	5.464	24.973	1.367	0.886
allergic	11.790	4.388	2.041	9.096	0.399	0.003	115.184
bacterial	2.768	1.327	1.098	38.180	0.002	318189	1.516
Viral	0.005	0.005	0.004	0.029	0.006	59.005	0.005

Table 15.21 Filtered likelihood ratio. Using the chi square value determined for all the cells in Table 15.16, if the calculated chi square value is greater than 13.25 (the critical chi square value , see section 4.4.6.2 Chi-square filtering in the main study) then the corresponding likelihood ratio is used (filtered likelihood ratio) , otherwise the likelihood ratio is set to 1 (this would have the least effect of the final post-test probability).

Positive Likelihood ratio	symptoms				signs		
filter LR+ if (chi-square > critical chi-square, LR+, 1)	itchy eye	burning eye	watery discharge	sticky discharge	fluorescein staining	palpebral redness	palpebral papillae
dry	3.278	8.626	5.281	1.000	24.973	1.000	1.000
allergic	11.790	4.388	1.000	1.000	1.000	1.000	115.184
bacterial	1.000	1.000	1.000	38.180	1.000	318189	1.000
viral	1.000	1.000	1.000	1.000	1.000	1.000	1.000

4.4.10.1 Use of Microsoft® Excel® spread sheets: Preliminary study

Using the data from both datasets A and B, two new datasets were created without presentation variation. A "dashboard" was created which allowed the selection of the dataset, (circular or non-circular and with or without presentation variation) and selection of the value of Chi-square filtration, to be used to test the analysis. As these are relational spread sheets, using formulae relating to particular cells, calculations could be carried out very quickly. Results from each analysis were displayed within the spread sheet and could easily then be copied into the IBM® SPSS® 21 statistical package to carry out the decision tree analysis.

4.4.10.2 Use of Microsoft® Excel® spread sheet: Main study

As with the preliminary analysis, data from dataset A was used to determine prevalence and likelihood ratios.

Dataset B (non-circular with presentation variation) was used to test the naïve Bayesian sequential analysis in two different ways:

- All test items: naïve Bayesian sequential analysis was carried out using all demographic factors, symptoms and signs recorded by the clinician.
- Recommended tests: Using only symptoms and history, the best tests to confirm the diagnoses using the positive likelihood ratio were entered into the analysis.

Accuracy levels were recorded in a separate spread sheet to facilitate the transfer of data into the IBM® SPSS 21 statistical package for further analysis.

A dashboard was also created allowing easy data entry, a summary of the pertinent data relating to that case (that is, prevalence, comorbidity and presentation variation) and the summary of the accuracy levels for the two analyses. A graphical display comparing the accuracy achieved by naïve Bayesian sequential analysis to the diagnoses made by the clinician was also displayed on the dashboard.

Figure 6 shows the dashboard where naïve Bayesian sequential analysis identified the same diagnoses as the clinician.

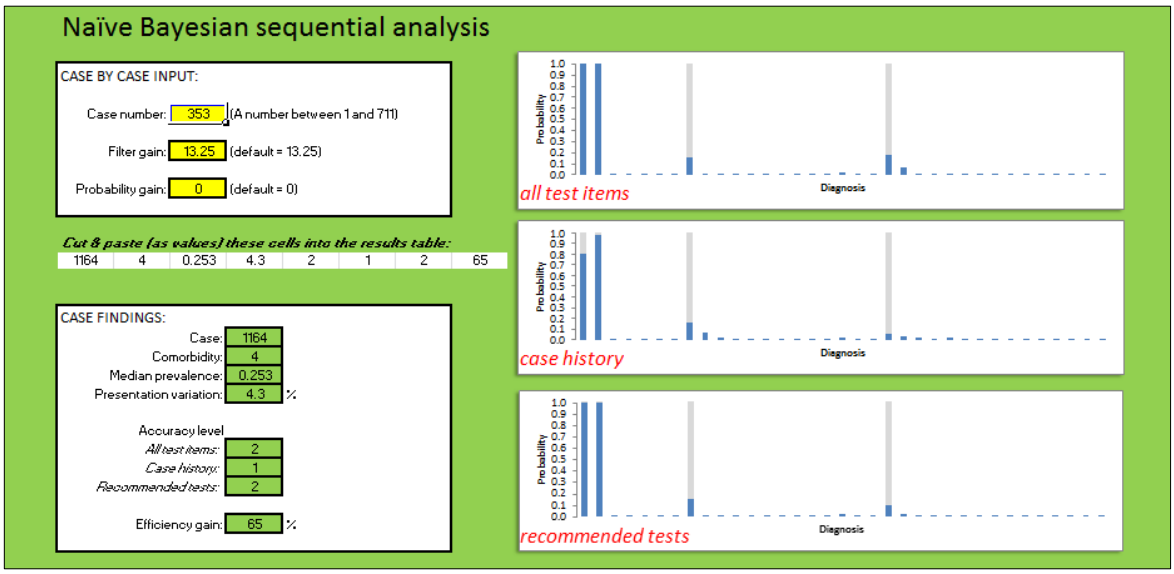


Figure 6 Dashboard of Excel spreadsheet showing how naive Bayesian sequential analysis achieves clinical accuracy by correctly identifying the same diagnoses as the clinician. Light grey bars represent the diagnoses made by the clinician and the blue bars represent the diagnosis made by naïve Bayesian sequential analysis.

Note how the clinician has noted 4 eye conditions and these are shown as the light grey bars. The naïve Bayesian sequential analysis has calculated the post-test probabilities and has correctly identified the same 4 conditions.

Each of the 711 cases had to be entered into the analysis and a macro (that is, a set of instructions) was created that enabled the naïve Bayesian sequential analysis to enter each case and record the data automatically. In this manner all 711 cases of dataset B could be analysed in about 15 minutes.

4.4.11 Decision tree analysis (DTA)

Decision tree analysis (DTA) is a method of multivariate analysis that is used to classify statistical data in a hierarchical manner. In optometry, DTA has previously been applied to the:

- Identification of refractive error (137);
- Interpretation of data from videokeratography in a both quantative and objective manner (138);
- Classification of keratoconus (139);
- Identification of factors which may contribute to the loss of visual and optical performance of in myopes wearing silicone hydrogel lenses (140);
- Assessment of the cost effectiveness of school based screening and primary eye care in relation to the supply of spectacles to correct refractive error in rural and urban areas in India. (141);
- Assessment of the factors influencing habits and attitudes to retinoscopy (142).

For the present study, DTA was considered suitable as it can handle both discrete and continuous variables. DTA can also be used for primary classification of data, where data is analysed and leads to a decision being made for further classification. However, this study used DTA to determine the level of influence of the three independent factors (prevalence, comorbidity and presentation variation) on the final accuracy achieved. Each independent factor was categorised to enable a preliminary analysis of the influence of each factor (see 5.2.1 Distribution and categorization of prevalence, comorbidity and presentation variation)

Previous studies have used Regression tree analysis and algorithms such as C4.5 and ID3 to carry out decision tree analysis (143). More recently, Dunstone (142) used the same software as the current study (IBM® SPSS® 21). This software was used to perform DTA with the Chi-squared Automatic Interaction Detection (CHAID) tree growing method which enabled investigation of the influences of prevalence, presentation variation and comorbidity (each being treated as independent variables) on the accuracy of naïve Bayesian sequential analysis (the dependent variable). CHAID was able to indicate the relative importance of each independent variable. (144)

To simplify interpretation of the decision trees, the dependent and independent variables were assigned categories. That is, accuracy was categorised as “1” (meaning not accurate) or “2” (meaning accurate), prevalence as “rare” or “common”, comorbidity as “absent” or “present” and presentation variation as “high” or “low”. By assigning these categories to the variables, all the data was treated as being discrete.

An alternative form of analysis (stepwise multiple regression) was considered not to be suitable as it required all variables to be continuous (142).

DTA gives rise to decision trees containing parent nodes from which child nodes grow. For the analyses presented in this study, the default settings were used, that is, parent nodes had to contain at least 100 cases while child nodes had to contain at least 50 cases. This was possible as the sample was large enough to allow at least 2 levels of tree growing. This allowed for the influence of the 3 independent variable to be displayed in the tree.

The following CHAID statistical criteria were also kept at the default levels:

- Significance level p=0.05
- Chi-square statistic – Pearson

Using Pearson (which is the default statistic) enabled faster calculations within the software, and the sample was deemed to be large enough to allow this. SPSS® recommends that, for small samples, the likelihood ratio is preferred but calculations may take longer (145)

The Bonferroni correction (although default) was not used (its use is sometimes recommended in order to avoid a Type II error; that is, accepting a hypothesis in error). Armstrong (146) suggests that use of the Bonferroni correction in this type of decision tree analysis reduces the power of the statistical test

4.5 Summary

Having outlined the methods that are common to all analyses carried out in this study, Chapter 5 presents the findings of the preliminary analysis while chapter 6 presents those of the main analysis.

5. Findings of the preliminary analysis

This chapter shows the findings of the preliminary study on 15 test outcomes and 10 diagnoses. The objectives of the preliminary study were to determine the influence of circularity, prevalence, comorbidity, presentation variation and Chi-square on the accuracy of naïve Bayesian sequential analysis.

5.1 Selection of tests and diagnoses

Five non-definitive test outcomes were selected, including age (over 50 years) and 4 symptoms. These represented the occurrence of test outcomes that may be common to several diagnoses and might, therefore, challenge Bayesian analysis. The "textbook" relationships between these 5 non-definitive test outcomes and the 10 chosen diagnoses are shown in Table 16

The 10 diagnoses were selected using two criteria. The first criterion was that each diagnosis was required to have one definitive test. This ensured that the accuracy of Bayesian analysis was not underestimated through failure to make a diagnosis for which there were no definitive signs. The second criterion was that the selected diagnoses had to represent the full range of observed prevalence and that the prevalence of each selected disease was approximately equal in datasets A and B (see Figure 7). This ensured that dataset A (used to generate pre-test odds and likelihood ratios prior to performing Bayesian analyses) contained the same distribution of diagnoses as dataset B (upon which Bayesian analyses were performed).

	Age	Vision	Strain	Comfort	Red eye	Refraction-uncorrected	Near add determination-uncorrected	Fixation disparity	TBUT<=10s	Conjunctival small papillae	Conjunctival lump/nodule	Crystalline lens opacity	Retinal haemorrhage	Macular hole	IOP >21mmHg
Uncorrected ametropia		*	*			*									
Uncorrected presbyopia	*	*	*				*								
Uncompensated heterophoria			*					*							
Dry eye				*	*				*						
Allergic Conjuctivitis				*	*					*					
Pinguecula	*										*				
Cataract		*										*			
Diabetic retinopathy	*												*		
Macula hole	*	*												*	
Primary open angle glaucoma	*														*

Table 16 The 10 diagnoses selected for analysis together with their associated definitive signs and associated reported symptoms

Prevalence of the 10 selected diagnoses on datasets A (blue bars) and B (red bars) are shown in Figure 5-1. No statistically significant difference was found between prevalence of each of the 10 selected diagnoses in datasets A and B (Chi square = 7.83, degrees of freedom = 9, P > 0.05).

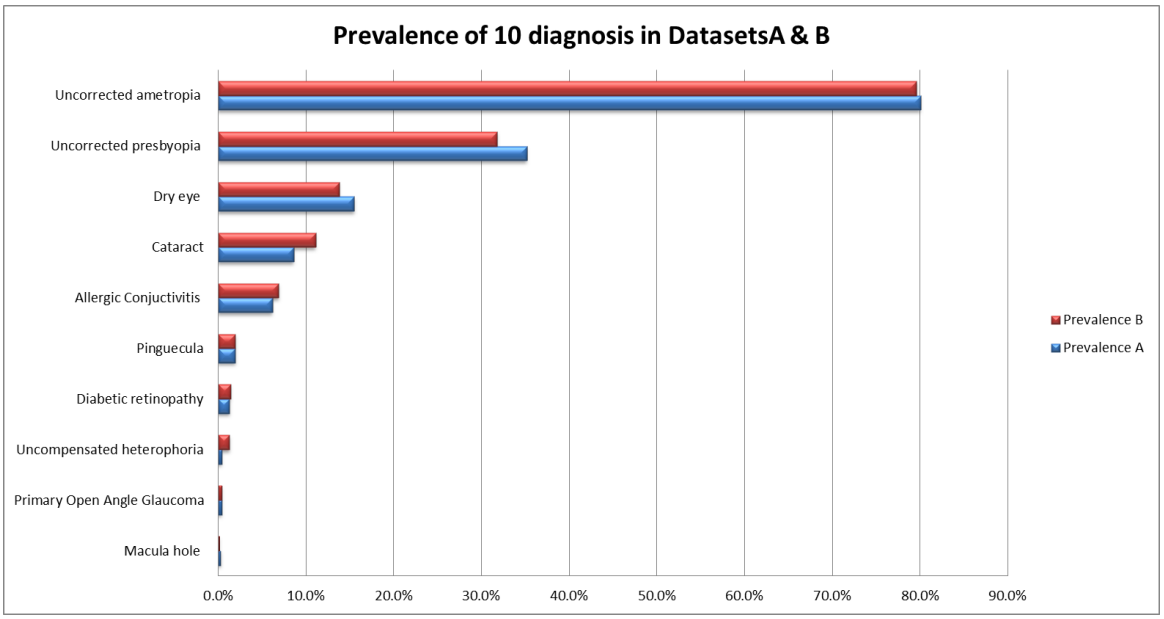


Figure 7 Prevalence in Datasets A & B

5.2 Results

5.2.1 Distribution and categorization of prevalence, comorbidity and presentation variation

Prevalence (or median prevalence) ranged from 0.001 to 0.80 with a median of 0.56. Comorbidity ranged from 0 (no eye conditions present) to 6 (6 eye conditions present) with a median of between 1 (no comorbidity as only 1 eye condition was present) and 2. This indicated that approximately 50% of the sample showed some degree of comorbidity. Presentation variation ranged from 0 to 7 items with a median of 1 item.

For the purposes of decision tree analyses, prevalence, comorbidity and presentation variation were categorised as shown in Table 17

Prevalence	> 0.56 (common)
	<= 0.56 (rare)
Comorbidity	<=1 (absent)
	>1 (present)
Presentation variation	<=2 (low)
	>2 (high)

Table 17 Categories of prevalence, comorbidity and presentation variation used for decision tree analyses.

5.2.2 Overall accuracy of naïve Bayesian sequential analysis

Circularity was present in the analyses denoted AA (Table 18) as accuracy was investigated by using the pre-test odds and likelihood ratios generated from dataset A to make diagnoses on the test outcomes of dataset A. Circularity was absent in analyses denoted AB (Table 18) as accuracy was investigated by using the pre-test odds and likelihood ratios generated from dataset A to make diagnoses on the test outcomes of dataset B. Presentation variation was absent in analyses performed on datasets A and B that contained cases with "textbook" test outcomes for all diagnoses. Presentation variation was present in analyses performed on datasets A and B that contained cases with observed test outcomes for all diagnoses. Percentage accuracy represented the percentage of cases for which the diagnoses made by the clinician (the author) appeared at the top of a list generated by Bayesian analysis.

Analysis	**Percentage accuracy**
AA (circularity present, presentation variation absent)	100.0%
AB (circularity absent, presentation variation absent)	100.0%
AA (circularity present, presentation variation present)	94.5%
AB (circularity absent, presentation variation present)	94.0%

Table 18 The overall percentage accuracy of naïve Bayesian sequential analysis achieved with and without circularity and presentation variation. .

Table 18 shows the overall percentage accuracy of naïve Bayesian sequential analysis with and without circularity and presentation variation. Accuracy of 100% was achieved when presentation variation was absent. This, in a sense, proved the basic concept of applying this form of naïve Bayesian analysis to the sort of clinical data collected in the setting of primary care optometry.

Accuracy fell be up to 6% when presentation variation was present. The presence of circularity with presentation variation artificially increased the accuracy by 0.5%.

Counter intuitively, removal of the Chi-square filter did not influence the accuracy of Bayesian analysis when presentation variation was absent and increased accuracy by 0.4% when presentation variation was present.

5.2.3 Decision Trees Analysis

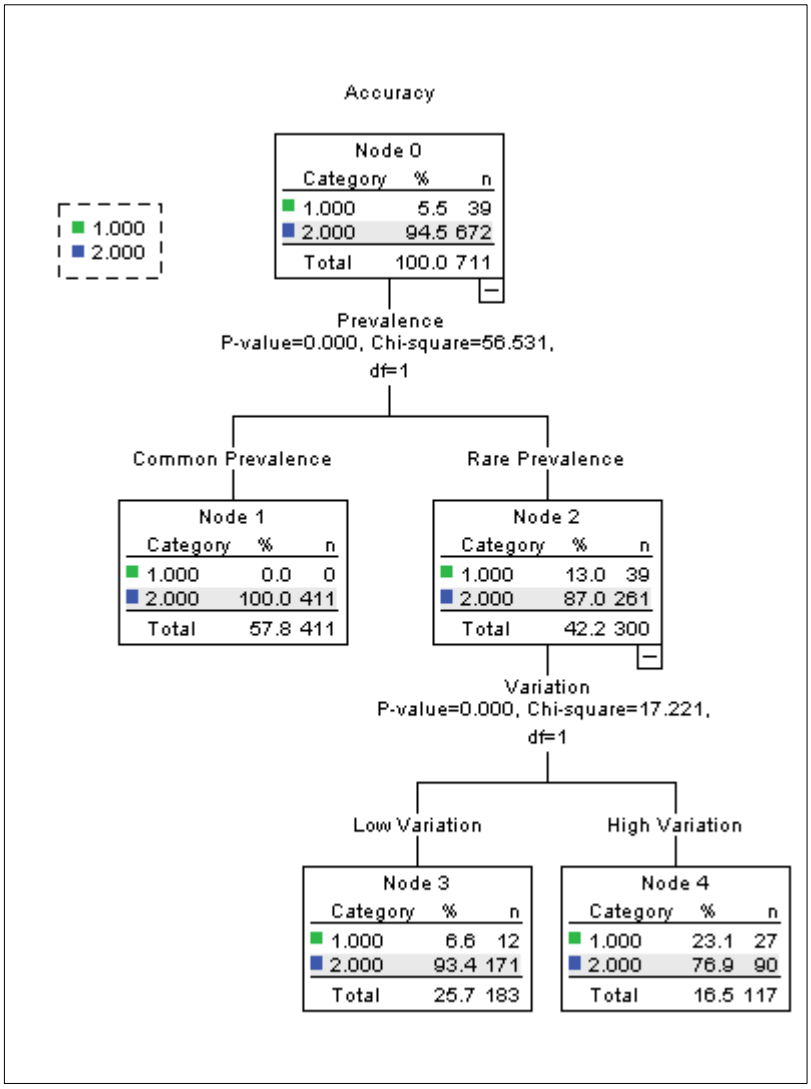


Figure 8 Decision tree analysis for analysis AA (circular analysis with presentation variation)

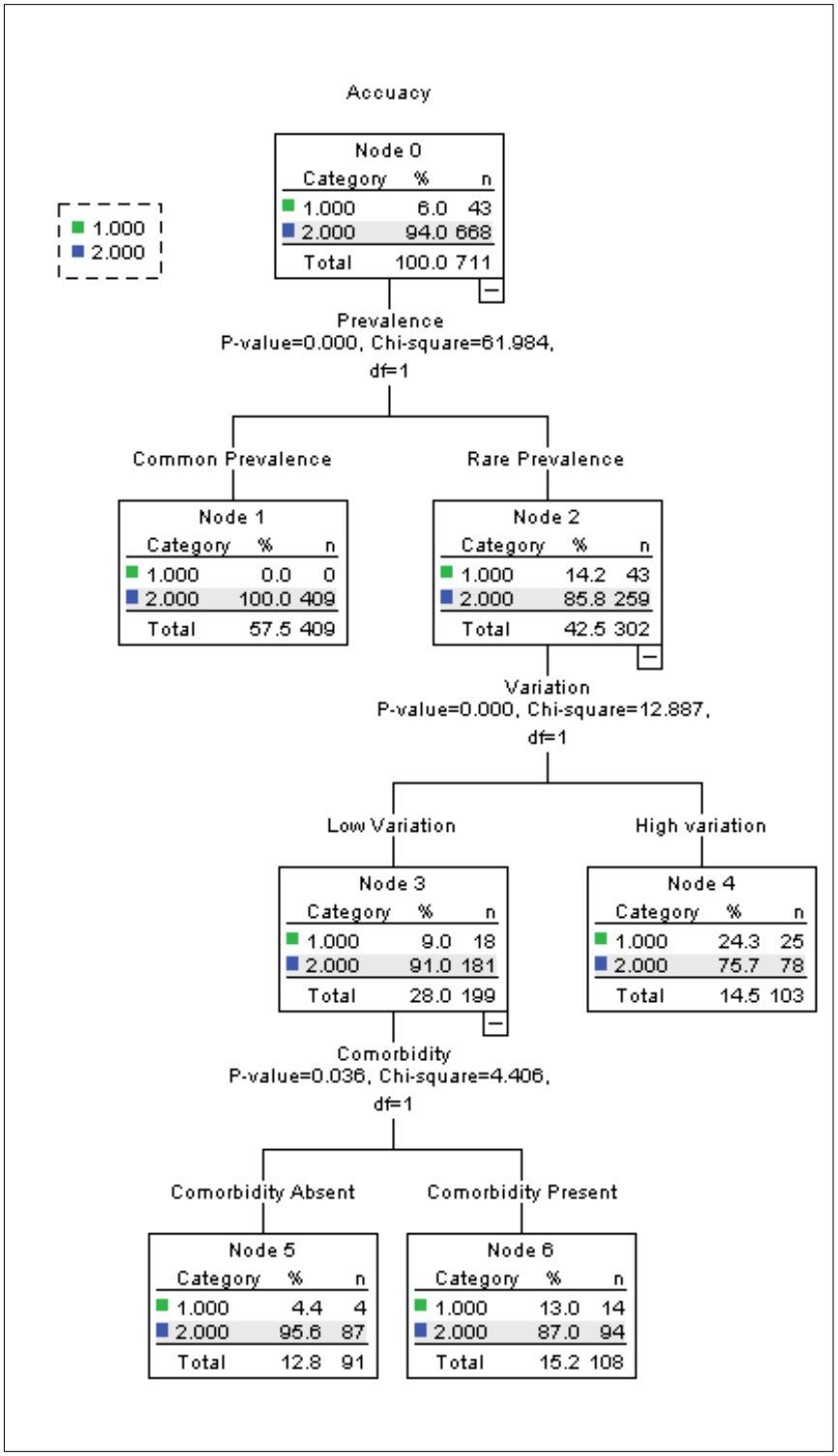


Figure 9 Decision tree analysis for analysis AB (non- circular analysis with presentation variation)

Separate decision tree analyses were performed for naïve Bayesian sequential analyses carried out with circularity present (AA- Figure 8) and absent (AB-Figure 9) in the presence of presentation variation. The overall accuracy of naïve Bayesian sequential analysis was found to be 94.5% for analysis AA and 94% for analysis AB. Decision trees exhibited up to 3 levels of branching (denoted as the 1st, 2nd and 3rd level of influence). The hierarchical influences of prevalence, comorbidity and presentation are indicated by the level of influence in which they are found. The percentage of cases represented in each level is shown. The percentage accuracy achieved by Bayesian analysis is also shown at each level.

In both analyses, prevalence influenced accuracy the most, followed by presentation variation and, only when circularity was absent, comorbidity.

The finding that comorbidity did not influence the accuracy of Bayesian analysis when circularity was present is consistent with previous research carried out on patients with dementia (127).The present study is, however, as far as the author is aware, the first to have carried out Bayesian analysis when circularity is absent and to have also considered the influence of prevalence and presentation variation.

In the presence of circularity, 100% accuracy was achieved for 57.8% of all cases. All of these exhibited eye conditions of common prevalence. Accuracy of 93.4% was achieved in 25.7% of all cases, all of which exhibited eye conditions of rare prevalence in combination with low presentation variation. Accuracy of 76.9% was achieved in 16.5% of cases, all of which exhibited eye conditions of rare prevalence in combination with high presentation variation. This finding was very encouraging as it showed that accuracy of over 76% was achievable even for the most challenging cases; those with atypical presentations of rare eye conditions.

However, Bayesian analyses lacking circularity were of most interest in the context of the future application of this form of artificial intelligence to new cases. Under these conditions, Bayesian analysis achieved 100% accuracy for cases exhibiting eye conditions of common prevalence and this occurred for 57.5% of all cases. Accuracy of 95.6% was achieved in 12.8% of all cases, all of which exhibited eye conditions of rare prevalence in combination with low presentation variation and no comorbidity. Accuracy of 87% was achieved in 15.2% of all cases, all of which exhibited eye conditions of rare prevalence in combination with low presentation variation and comorbidity. Accuracy of 75.7% was achieved in 14.5% of all cases, all of which exhibited eye conditions of rare prevalence in

combination with high presentation variation but with no comorbidity. Therefore, even with circularity present, accuracy of over 75% was achievable in the most challenging cases.

5.3 Summary

The preliminary study, carried out on 10 diagnoses and 15 test outcomes, showed that naïve Bayesian sequential analysis could make diagnoses without error in the absence of presentation variation. Accuracy dropped by only 6% when presentation variation existed. Circularity artificially elevated accuracy by 0.5%. Surprisingly, removal of Chi-square filtering increased accuracy by 0.4%. Decision tree analysis showed that accuracy was influenced primarily by prevalence followed by presentation variation and comorbidity. Chapter 6 presents the findings of the main study that was carried out to explore whether these encouraging findings also arise when naïve Bayesian sequential analysis is applied to diagnoses that may not necessarily have definitive tests. Chapter 6 also presents an investigation carried out to determine whether positive likelihood ratios could be used to identify diagnostic tests.

6. Findings of the main study

This chapter shows the findings of the main study on 105 tests and 35 diagnoses. The objectives of this study were (1) to re-examine the influence of prevalence, comorbidity, presentation variation and Chi-square filtration on the accuracy of naïve Bayesian sequential analysis for diagnoses that may or may not have definitive tests and (2) to explore the use of positive likelihood ratios to identify diagnostic tests.

6.1 Selection of tests and diagnoses

The 35 diagnoses represented all diagnoses that were present in both datasets (A and B). The requirement in selection 5.1 Selection of tests and diagnoses to have one definitive test associated with each diagnosis was removed in this analysis to reflect clinical reality. Likewise, the 105 tests represented all tests that were present in both datasets (A and B).

6.2 The use of positive likelihood ratios to select diagnostic tests

Recall that the magnitude of each positive likelihood ratios gives the amount by which the post-test odds of any given diagnosis rises or falls after a positive test result (section 3.1 Decision Matrices). Logically, then, the test outcome with the largest positive likelihood ratio associated with a given diagnosis is that which has greatest diagnostic value for that diagnosis. Previous workers have hinted that this would work (92), or have actually tried it out (62). As far as the author is aware, this has not been tested in the field of optometry. Using Chi-square filtered positive likelihood ratios for this purpose could help optometrists select the most appropriate diagnostic tests. Exploring the potential for this was one of the key objectives of the main analysis.

An overview of the scheme adopted now follows. For each case, naïve Bayesian sequential analysis was performed until all positive test outcomes for demographic items (ethnicity and age), symptoms and history items (ocular and medical) had been entered. In other words, the history and symptoms part of the eye examination had been completed. This included 54 of the 105 test outcomes. At this point, diagnostic tests (all clinical signs) were recommended based on those that had the highest Chi-square filtered positive likelihood ratios associated with the top 6 diagnoses (these diagnoses being ranked according to their post-test probability at that point). Only 6 diagnoses were shown as the

maximum comorbidity was known to be equal to 6 (see 5.2.1 Distribution and categorization of prevalence, comorbidity and presentation variation). Logically, this would only lead to the recommendation of 6 of the remaining 51 diagnostic tests that represented signs. Naïve Bayesian sequential analyses were then resumed until all positive outcomes of these 6 diagnostic tests had been entered. At this point, the end of the eye examination, all diagnoses were again ranked according to their post-test probability. The accuracy of the diagnoses was calculated (as described in 4.4.8 Accuracy) and compared to the accuracy achieved when positive test findings for all 105 test outcomes had been entered.

This approach would only work in optometric practice if optometrist had the autonomy to select diagnostic tests on their diagnostic value alone, as medical practitioners do. However, in reality, optometrists have to perform certain tests to meet medico-legal obligations. Nevertheless, if it could be demonstrated that positive likelihood ratios could identify fewer but more diagnostically valuable tests without a loss in diagnostic accuracy, then the approach tested in this study might, in the fullness of time, be considered medico-legally acceptable.

This approach has another potential advantage – reduced "chair-time". The 51 test outcomes, representing clinical signs, could actually be grouped into just 17 ophthalmic procedures. Appendix III (Ophthalmic Procedures) shows the ophthalmic procedures in the first column and the signs associated with these procedures that were used in the analysis in the second column. Each ophthalmic procedure included between 1 and 13 of the 51 test items. For example one of the test procedures was fundus examination. This included 3 test items: haemorrhages, cotton wool spots and exudates. The thinking here was that if an optometrist invested the "chair time" to examine the fundus then they would very likely detect any of the 3 test items. So, if 3 of the 6 recommended diagnostic tests fell under fundus examination then the "chair time" invested in those 3 recommended diagnostic tests would only amount to that involved with carrying out 1 ophthalmic procedure. Now, if the 6 recommended diagnostic tests represented 1 test item from each of 6 different ophthalmic procedures then the total reduction of "chair time" would be equal to ((17 – 6) / 17) x 100 = 65%. On the other hand, if all 6 recommended diagnostic tests fell under just one ophthalmic procedure (this would be theoretically possible for ophthalmic procedures involving examination of the conjunctiva, cornea and optic disc) then the total reduction of "chair time" would be equal to ((17 – 1) / 17) x 100 = 94%. So, adopting the approach of recommending just 6 out of 17 diagnostic tests could reduce "chair time" by anything between 65% and 94%,

Given what has been said above, it becomes clear that if this approach reduces “chair time” without compromising diagnostic accuracy (that is, clinical vigilance), then this might be of great interest to practices that rely on high volume eye examinations to remain economically viable.

Dataset A was used to generate the Chi-square filtered positive likelihood ratios used to identify recommended diagnostic tests for each diagnosis. Two criteria were used for recommended test selection. The first was, as mentioned earlier, that the recommended test had to exhibit the largest positive likelihood ratio for the diagnosis in question. The second criterion was that the recommended test had to be the sole test item with the largest positive likelihood ratio for that diagnosis. That is, there should be no ties between test items exhibiting maximum positive likelihood ratios. These criteria were met for 34 of the 35 diagnoses; the diagnosis of hypertensive retinopathy lacked a recommended diagnostic test. It is worth considering the loss of accuracy that this might cause. Given that Bayesian analysis was carried out on the cases in the dataset B, it follows that the potential loss of accuracy, when basing diagnoses on recommended tests, amounts to the number of cases in dataset B that had hypertensive retinopathy. This amounted to 2 (0.3% of dataset B). Therefore, little loss in accuracy was anticipated. Nevertheless, the possibility that this type of accuracy loss can occur should be noted when considering this approach to making diagnoses in the future.

6.3 Results

All naïve Bayesian sequential analyses carried out in the main study were made without circularity (that is, Bayesian analyses were carried out using initial pre-test odds and positive likelihood ratios from dataset A in order to make diagnoses based on positive test outcomes from dataset B) and included natural presentation variations observed in dataset B.

6.3.1 Overall accuracy of naïve Bayesian sequential analysis on all test items

The overall percentage accuracy of Bayesian analysis, when diagnoses were based on positive test outcomes from all 105 test items, was 75% when Chi-square filtering was used and 74% when Chi-square filtering was removed. As found in the preliminary analysis, Chi-square filtering had little impact on accuracy.

Recall that the accuracy found under the same conditions (circularity absent, presentation variation present) in the preliminary study was 94% (Table 18). The question arose as to what might have caused this 20% drop in accuracy.

6.3.1.1 Possible causes of reduced accuracy: lack of definitive tests

It has already been suggested that this was due to some of the 35 diagnoses lacking definitive tests. The definitive test selected by naïve Bayesian sequential analysis had the highest positive likelihood ratio for that particular test/diagnosis combination. Of the 35 diagnoses, 34 diagnoses had definitive tests associated with them. However on closer inspection, some of these test/diagnosis associations were spurious and in normal clinical practice would not be associated with the diagnosis (147). Of the 35 diagnoses only 29 were found to have a test that would be associated in clinical practice; that is 17% of the diagnoses lacked definitive tests or tests that were equivocal at best.

Figure 10 shows how the diagnosis of hypertensive retinopathy is missed by the naïve Bayesian sequential analysis because there were no associated definitive signs. Similarly, nuclear cataract had equivocal signs associated with the diagnosis (hazy view of fundus, small retinal haemorrhages), so again this diagnosis is missed.

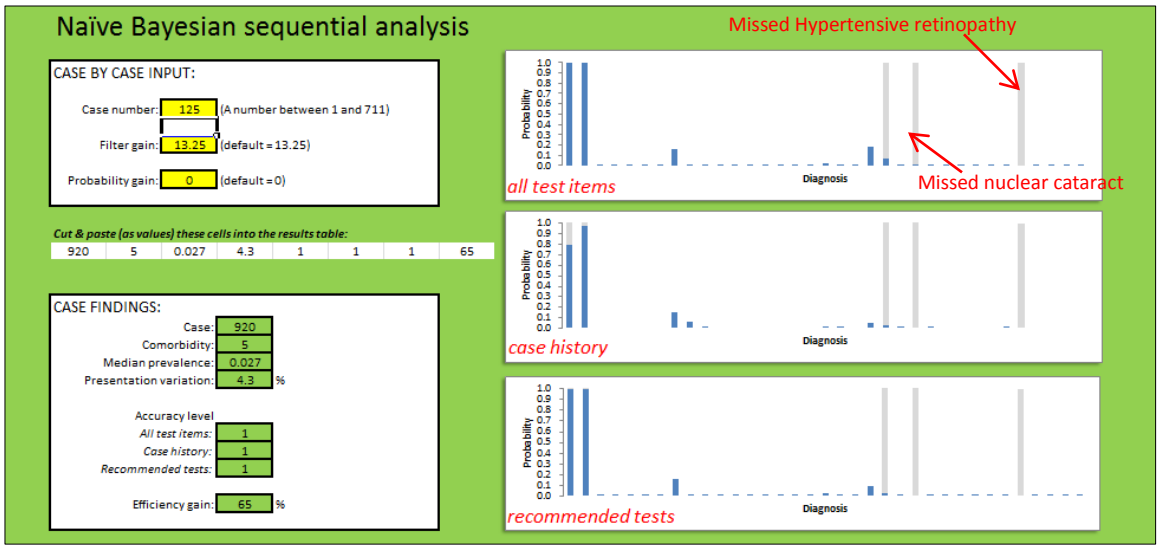


Figure 10 Due to a lack of a definitive sign, the diagnosis of hypertensive retinopathy is missed completely. Similarly, the diagnosis of nuclear cataract is missed as the signs associated with this diagnosis (hazy view of fundus, small retinal haemorrhages) are equivocal at best

The diagnoses lacking a definitive test identified by the use of likelihood ratios were:

- Allergic dermatitis (0.2%)
- Bacterial Conjunctivitis (1.6%)
- Corneal abrasion (1.7%)
- Cataract – cortical (1.3%)
- Primary open angle glaucoma (0.8%)
- Hypertensive retinopathy (0.3%)

Prevalence of each condition for dataset B is given in brackets. Thus, diagnoses with rare prevalence, which lack a definitive sign, may account for part of the drop in accuracy.

6.3.1.2 Possible causes of reduced accuracy: reduced potency of definitive tests

There is, however, another possible explanation. Might the positive likelihood ratios of the definitive tests of the preliminary analysis have been much higher in value that those of the main study? In other words, might the definitive test of the preliminary study have had greater diagnostic power (or potency) than those of the main study.

This was investigated by taking the average of all the LR+ associated with a diagnosis and multiplying this by the prevalence (thus giving the weighted average). The average of all the weighted averages was then compared between the two studies. The preliminary study average was found to be 1,740 and the main study average was found to be 82. This suggested that there was more than 20 times difference in the potency of the tests used in the preliminary study and the main study. It is pertinent to note that in the preliminary study the tests were "chosen" by the author, whereas in the main study the tests were "chosen" by using the Bayesian analysis (that is, identifying the best test/diagnosis associations by using the LR+). The comparison of the positive likelihood ratios for the preliminary study and the main study are shown in Appendix V -Comparison of LR+ between the preliminary study and the main study.

6.3.1.3 Possible causes of reduced accuracy: lack of recorded signs

Loss of accuracy (that is, where naïve Bayesian sequential analysis did not concur with that of the clinician) occurred in 26% percent of cases. Individual cases were further examined to elicit the

possible causes of the drop in accuracy. One such loss of accuracy was found to be due to the lack of a recorded sign associated with the diagnosis.

Figure 11 shows a diagnosis of pinguecula that is missed completely as there were no recorded associated signs. Three eye conditions were recorded in this case and the other 2 were correctly identified.

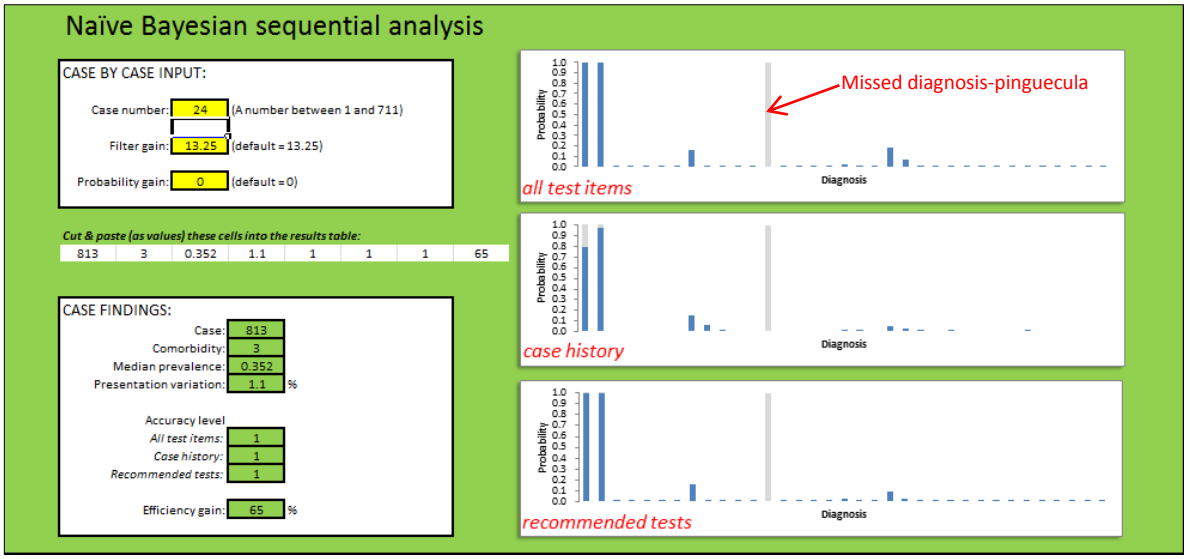


Figure 11 Naïve Bayesian sequential analysis misses a diagnosis as there is no recorded sign.

Other eye conditions that were diagnosed but were lacking in recorded signs by the clinician were corneal arcus, pterygium, and corneal abrasion.

Further, if incomplete signs are recorded, then the naïve Bayesian sequential analysis identifies the clinician's diagnoses, but included other diagnosis as well. (Figure 12)

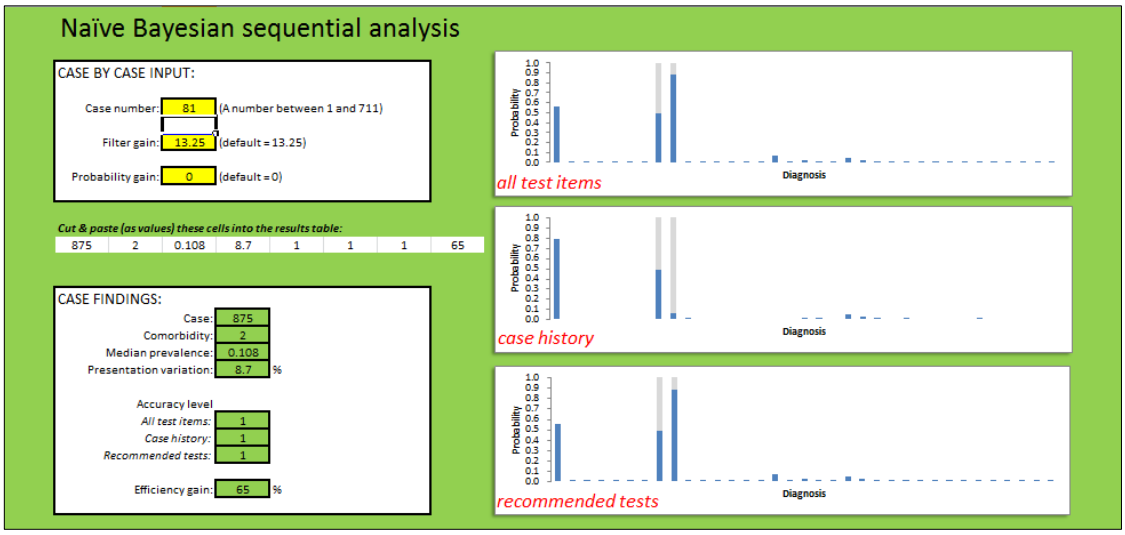


Figure 12 Incomplete recording of signs achieves partially accurate diagnoses, but also achieves high probability for a spurious diagnosis. In this case, the two diagnoses recorded by the clinician were "dry eye" and "allergic conjunctivitis". However, the only sign that was recorded was "conjunctival small papillae". The diagnosis for the dry eye has been made on the history and symptoms. A high diagnostic probability for "uncorrected ametropia" is also seen, which was not recorded by the clinician; thus in this case although the original diagnoses were confirmed by naïve Bayesian sequential analysis, accuracy is lost by the introduction of a spurious diagnosis.

6.3.1.4 Possible causes of reduced accuracy: influence of small likelihood ratios

Could a large number of small positive likelihood ratios have been the cause of the reduced accuracy? Likelihood ratios less than 1 diminish the probability of a diagnosis. However, where the positive likelihood ratio is greater than 1, the effect of a large number of such small likelihood ratios can have a cumulative effect raising the probabilities of a diagnosis. (130)

Sixty eight of the test/diagnosis combinations had a positive likelihood ratio of greater than 1 and of these 59 had a positive likelihood ratio of greater than 5. It would, therefore, seem very unlikely that a preponderance of smaller positive likelihood ratios (of between 1 and 2) could have a significant effect on the final diagnosis.

6.3.2 Overall accuracy of naïve Bayesian sequential analysis on recommended test items

The overall percentage accuracy of Bayesian analysis, when diagnoses were based on just 6 recommended tests, was 73% when Chi-square filtering was used and 70% when Chi-square filtering was removed. This was an encouraging finding as it indicated that basing diagnoses on just 6 recommended tests (which, as mentioned earlier, could reduce "chair-time" by between 65% and 94%) also bought about a 3% increase in accuracy when Chi-square filtering was used.

6.3.2.1 The quality of recommended test items

The quality of recommended tests for each of the 35 diagnoses was investigated. The author used her clinical judgement to rate each recommended test using a Likert scale of 1 to 5; in which 1 represented strong disagreement and 5 represented strong agreement. The maximum score possible was 175 (that is, maximum score of 5 per test multiplied by the total number of tests, 35). The score for the analysis with Chi-square filtering was 150 and without the Chi-square filtering was 137. Expressed as a percentage of the highest possible score, the tests with Chi-square filtering achieved 85.7% and without Chi-square filtering achieved 78.3%. These findings suggested that Chi-square filtering improved recommended test selection – albeit by a small degree. The recommended tests selected by using the maximum likelihood ratios are shown on page 122.

6.4 Decision tree analysis

Decision tree analyses was carried out to further explore the hierarchical influences of prevalence, comorbidity and presentation variation on the accuracy of naïve Bayesian sequential analyses carried out in the main study. Separate decision tree analyses were performed for Bayesian analyses carried out with all test items and with recommended tests. Logically, the percentages of cases found at each level in the decision tree were identical in both analyses (after all, both analyses were carried out on dataset B). The decision tree structure was also identical in both analyses showing that prevalence influenced accuracy the most, followed by comorbidity and then by presentation variation. Compared to the decision tree analyses performed in the preliminary study (section 5.2.3 Decision Trees Analysis), comorbidity had increased its hierarchical position from last to second.

The decision tree analyses based on all test items and recommended tests are shown in Figure 13 and Figure 14, respectively.

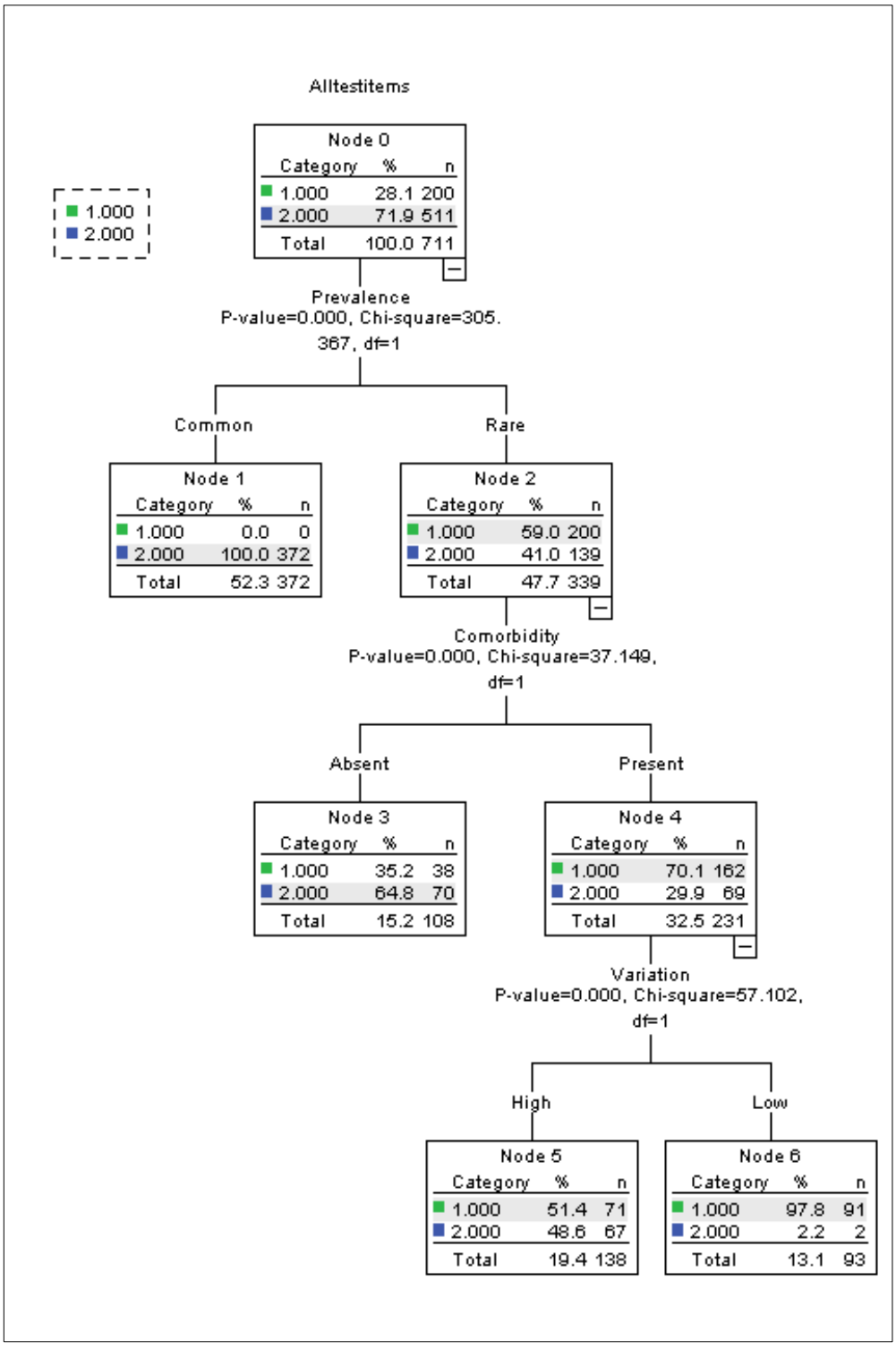


Figure 13 Decision tree analysis for naive Bayesian sequential analysis using all 105 test items comprising of non-circular data with presentation variation

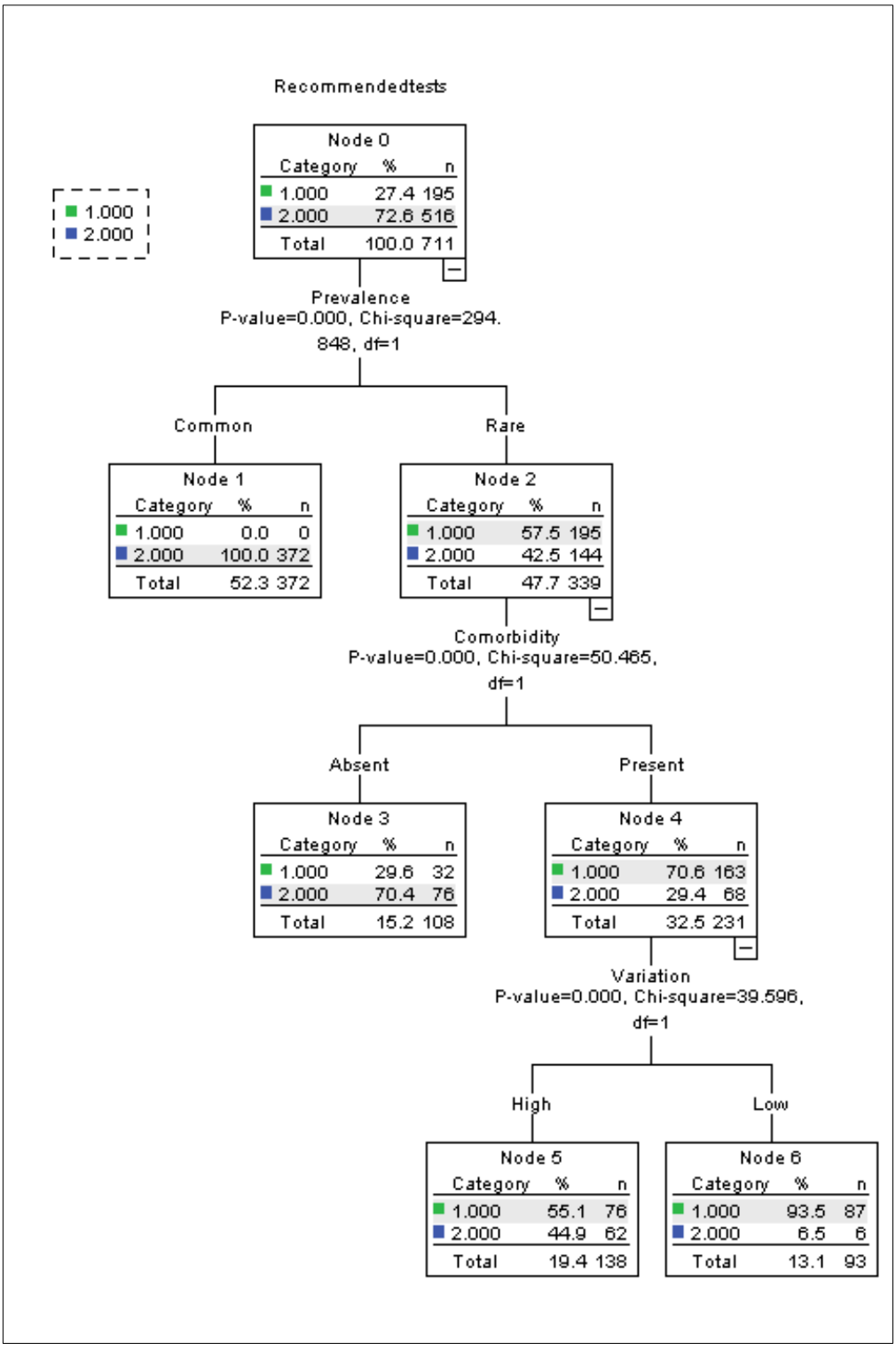


Figure 14 Decision tree analysis based on only recommended tests following history and symptoms comprising of non-circular data with presentation variation

The only differences in both analyses were the accuracy levels achieved. Even so, the differences observed were only subtle so that the accuracy trends were identical. In cases exhibiting eye conditions of common prevalence (52.3% of all cases), both analyses achieved 100% accuracy. In cases exhibiting eye conditions of rare prevalence without comorbidity (15.2% of all cases), analyses based on all test item and recommended tests achieved 64.8% and 70.4% accuracy, respectively. In cases exhibiting eye conditions of rare prevalence with comorbidity and high presentation variation (19.4% of all cases), analyses based on all test item and recommended tests achieved 48.6% and 44.9% accuracy, respectively. In cases exhibiting eye conditions of rare prevalence with comorbidity and low presentation variation (13.1% of all cases), analyses based on all test item and recommended tests achieved 2.2% and 6.5% accuracy, respectively.

The reason for the paradoxical elevation of accuracy in cases with high presentation variation compared to low presentation variation remains obscure. On closer examination of individual cases, lack of accuracy appeared again due to lack of definitive tests for some diagnoses and missing diagnostic signs (see 6.3.1.1 Possible causes of reduced accuracy: lack of definitive tests; 6.3.1.3 Possible causes of reduced accuracy: lack of recorded signs).

6.5 Summary

The main study, on 35 diagnoses and 105 test outcomes exhibiting observed presentation variations, showed that naïve Bayesian sequential analysis, without circularity, achieved accuracy of 72% when all clinical signs were entered. So the encouraging findings of the preliminary analysis (chapter 5) were not sustained. Removal of Chi-square filtering made little difference. The 20% drop in accuracy, compared to the preliminary analysis, was attributed to the fact that some diagnoses lacked strong diagnostic signs, a lack of potency of diagnostic tests used or a lack of recorded diagnostic signs.

The main study also explored the use of Chi-square filtered positive likelihood ratios, calculated after history and symptoms, to recommend diagnostic signs to look for. This approach reduced “chair time” by between 65% and 94% while maintaining clinical vigilance. In fact, when compared to the entry of all positive test findings, the accuracy increased by 1% when only recommended signs were entered. Chi-square filtering improved recommended test selection.

Decision tree analysis showed that accuracy was influenced primarily by prevalence, followed by comorbidity and presentation variation. The influence of these factors was also greater than that observed in the preliminary analysis. Accuracy fell dramatically in 13% of the cases exhibiting eye

conditions of rare prevalence compounded with comorbidity and low presentation variation (13.1% of all cases). Here, analyses based on all test item and recommended tests achieved 2.2% and 6.5% accuracy, respectively.

The above analyses were all carried out using only positive likelihood ratios as only positive test outcomes had been recorded. As a result of these findings, it is strongly recommended that, in future studies, Bayesian analysis should be based on positive and negative test findings.

Chapter 7 provides a review of the thesis findings and makes recommendations for further studies.

7. Discussion

This study was designed to investigate the influences of prevalence, comorbidity, presentation variation and circularity on the application of naïve Bayesian sequential analysis to make differential diagnoses in the optometric primary care setting. This study also investigated the extent to which Chi-square filtering maximised accuracy. Given the widespread use of Bayesian analysis in many other areas of our daily lives, it is surprising that this is the first investigation of its kind in the field of optometry. Even where similar investigations have been carried out in other medical fields, none have explored the potential sources of error investigated in the present study.

One of the key features of Bayesian analysis, in this context, is that it accounts for the prevalence of each diagnosis. Reliable estimates of prevalence are, therefore, essential and were only possible if all eligible data was included in the analysis. It soon became apparent that asking for patient consent raised subtle trust issues. Therefore, it became necessary to proceed without patient consent and this became a major ethical issue standing in the way of the continuation of this study. Risks to patients would be the accidental disclosure of clinical data and yet the advantage of using Bayesian analysis to improve the accuracy of differential diagnosis had vast potential in preventing avoidable blindness.

Fortunately, precedent for the use of completely anonymised clinical data without prior consent appears in Section 3.3 of the Guidelines for Researcher and Research Ethics Committees on Psychiatric Research involving Human Participants (133), which states that…"Individual consent should not be necessary for group analysis of anonymised data but the ethics committee should ensure that the required anonymisation has been achieved before the data are made available" and …"the ethics committee should have specifically agreed to exempt the research from the general requirement for individual consent from each research subject. In that respect, our recommendation on research using records and archived samples follow the same principles as those that apply to clinical audit". A not insignificant breakthrough occurred when the research ethics committees of Aston University and the Tanzanian National Institute of Medical Research made this study possible by accepting this guidance.

The findings of this study indicate that naïve Bayesian sequential analysis works without error when variations in the presentation of various eye conditions or diseases are removed. The preliminary study served as a proof of concept in that prevalence and comorbidity did not render the application of naïve Bayesian analysis to primary care clinical data a "non-starter". The main study showed that naïve Bayesian analysis can be applied to a wealth of clinical data with some accuracy.

The sample of 1422 cases examined in this study seems adequate for the purposes of evaluating the factors affecting naïve Bayesian analysis, given that previous studies of ocular conditions in primary eye care have been carried out on samples of between 1438 and 5308 cases (148) (149) (150) (151) (152).

In the preliminary study, 100% accuracy was achieved when there was no presentation variation in the data. A surprisingly small fall in accuracy (6%) did, however, occur when observed presentation variations were included in the analysis. Though some inconsistency was evident in the findings, analyses indicated that prevalence, comorbidity and presentation variation reduced accuracy. Prevalence was consistently highest in the hierarchy of these influencing factors. There were initial concerns about circularity but even this exerted a surprisingly small (0.5%) artificial elevation of accuracy.

Using all the available data, naïve Bayesian sequential analysis achieved 72% accuracy in data with low presentation variation. A similar study carried out on the detection of pancreatic cancer using sequential Bayesian analysis involving clinical, laboratory and image data achieved 67% accuracy (153). Using Chi-square filtered positive likelihood ratios to recommend between 65% and 94% fewer confirmatory tests, no fall in the accuracy of naïve Bayesian sequential analysis (73%) arose, thus opening the way for this form of artificial intelligence to direct problem-orientated eye examinations. This has potential in an environment in which commercial pressures demand reduced “chair-time”. Recall that Warner, in his early computerised diagnostic program for congenital heart conditions, had a total of 54 tests but this could be reduced to 7 or 8 tests to give the proper diagnosis (61).

In the preliminary study, the tests chosen were such that they had definitive links with each of the diagnoses. In the main study however, not all the tests were definitive and this may have led to a 20% reduction in accuracy.

Accuracy could be further improved by recording negative test findings, thereby allowing the use of negative likelihood ratios.

No attempt was made in this study to calculate confidence limits for sensitivity, specificity and likelihood ratios as this did not assist Bayesian analysis. In the future, however, these quantities could be published with confidence limits as an indicator of their reliability.

Attempts made to remove circularity (3.5 Circularity) involved splitting the data into subsamples A (containing the first half of the data collected) and B (containing the second half of the data

collected). The first sample was used to calculate prevalence and likelihood ratios. The second sample was used to test the accuracy of Bayesian analyses carried out using the calculated prevalence and likelihood ratios. An alternative approach would have been to randomly pick the data used to calculate prevalence and likelihood ratios and then to test accuracy on the remaining data. As these were both samples from the same population, future studies could calculate prevalence and likelihood ratios from one clinic and test the accuracy of Bayesian analyses, based on these quantities, in another clinic.

Decision tree analysis (DTA) was applied to broadly categorised data; for example, prevalence was categorised as being high and low. This led to uncomplicated decision trees that could be interpreted easily. Initial analysis using DTA on independent variables that had not been broadly categorised resulted in confusing results. However, future work could be directed towards narrower categorization of the independent variables. For example, to understand at what level of prevalence does the accuracy drop below 100%? (see Figure 8, Figure 9, Figure 13 and Figure 14).

Accuracy may also be improved by the greater use of tests with multiple levels, such as for cup-to-disc ratios, and intraocular pressure measurements. Different likelihood ratios for the different levels of a test (108), can be calculated, which may also provide useful information about an individual patient and thereby making the diagnosis more customised for each patient.

This is shown in the example below, derived from the present study, where age was recorded in multiple levels and positive likelihood ratios are shown for uncorrected presbyopia (see Table 19)

Positive Likelihood ratio	**Age**						
	0-10	**11-20**	**21-30**	**31-40**	**41-50**	**51-60**	**61-70**
n	43	112	111	117	128	94	67
LR+ for uncorrected presbyopia	0.00	0.00	0.00	0.13	12.18	2.47	1.46

Table 19 The Variation of positive likelihood ratios for uncorrected presbyopia with age.

Uncorrected presbyopia (or the need for a near vision correction) is not generally seen before the age of 40. So, positive likelihood ratios of below 1 are seen up to 40 years of age. Between the ages of 41-50, most patients need a new reading prescription, or a change to a current one and this is reflected by the high positive likelihood ratio recorded for that age group. The falling positive likelihood ratios after 50 years of age reflect the fact that progressively fewer patients require further change to their reading prescription. Table 19 therefore, illustrates how multiple levels of test outcomes can generate useful information about a patient's individual characteristics, such as age.

Looking at the challenges for eye care today in terms of the aging demographic throughout the world, the main conditions that have to be addressed in order to prevent avoidable blindness are cataract, diabetic retinopathy, glaucoma and age related macular degeneration (3). Table 20 shows how positive likelihood ratios reveal that all of these conditions are more likely to occur with advancing age. The positive likelihood ratios for diabetic retinopathy are reduced to zero at the age group 81-90 and this may be due to the fact that there is a poor survival rate at this age with diabetes. Similarly, the positive likelihood ratio for age-related macular degeneration also drops to zero. This may be due to the fact that few people were actually seen in the practice of this age group in the data that was analysed (only 2 patients of this age group were seen) , or that they may be attending a specialist low vision clinic and may not require the services of a community optometrist.

Positive likelihood ratio (LR+)	Age								
	0-10	**11-20**	**21-30**	**31-40**	**41-50**	**51-60**	**61-70**	**71-80**	**81-90**
n	43	112	111	117	128	94	67	31	8
Cataract - nuclear	0.00	0.00	0.00	0.00	1.60	1.48	0.97	6.13	37.66
Primary open angle glaucoma	0.00	0.00	0.00	0.40	1.37	3.00	2.96	1.95	9.76
Diabetic retinopathy	0.00	0.00	0.00	0.00	1.54	1.88	3.36	7.03	0.00
Age-related macular degeneration	0.00	0.00	0.00	0.00	0.00	0.00	7.80	15.78	0.00

Table 20 The variation of positive likelihood ratios with age for cataract, primary open angle glaucoma, diabetic retinopathy and age-related macular degeneration.

It is pertinent to note that a completely novel method of evaluating test/diagnosis association was considered in this study- Chi-square filtering. In the past, several other methods have been used to evaluate diagnostic tests. Aspinall & Hill (154) used receiver operator characteristic curves (ROC) to determine the best cut-off values for a positive or negative test outcome (ideally maximising true positives and minimising false negatives), whereas Gilchrist (100) used a weighted kappa and QROC curves.

Figure 15 shows an ROC curve with the dashed line representing the indecision line. That is where the diagnostic information is so poor, that cases with or without the condition or disease cannot be differentiated except by chance. The top left hand corner of the graph is where sensitivity and specificity is at its best. Each curve represents a test with the sensitivity and specificity for a certain value. A point on the curve that is furthest away from this line and closest to the top left hand corner of the graph represents the best criteria for a positive or negative test outcome. This point ensures

maximum true positives and the minimum false negatives. Curve B represents a test that is more likely to identify true positives and true negatives than the test represented by Curve A.

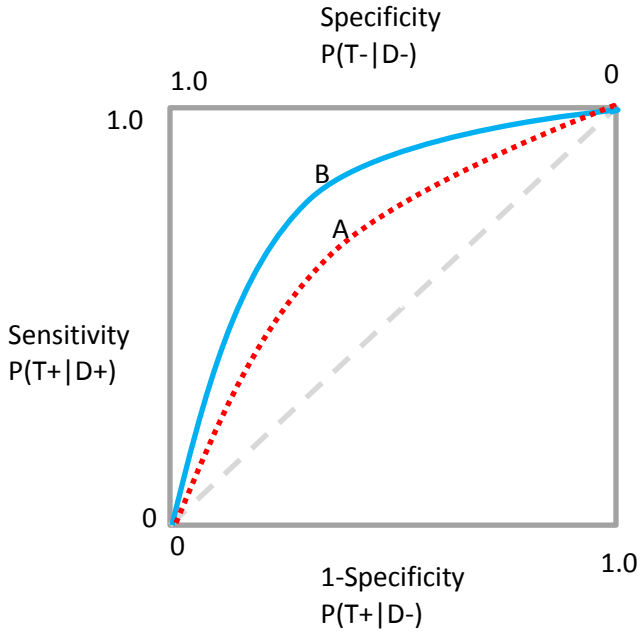


Figure 15 ROC curves (after Aspinal and Hill)

Using the data from Crick and Daubs (155) using intraocular pressure for detecting glaucomatous visual field loss, Aspinall & Hill used ROC curves to determine the intraocular pressure that predicts visual field loss caused by primary open angle glaucoma. The ROC curve thus obtained was shown to be very close to the indecision line, suggesting that intraocular pressure measurement, as a stand-alone test, cannot be a strong indicator for glaucomatous field loss.

In the present study, intraocular pressure > 21 mmHg was considered to be a definitive sign for primary angle glaucoma. In the preliminary study, the positive likelihood ratio that was generated between the test "intraocular pressure >21mmHg" and the diagnosis of "primary open angle glaucoma" was 0.88. (see page 120). Likelihood ratios that are close to 1 have minimal effect on post-test probability (see 3.1 Decision Matrices) and therefore this shows that the test is of very little diagnostic value, concurring with the findings of Aspinall & Hill (154).

ROC curves are useful for a single test/diagnosis combination as shown above. However, the current study deals with test outcomes in relation to many alternative diagnoses. For this, each test/diagnoses combination would require the calculation of separate ROC curves. This was thought to be an unwieldy process.

Gilchrist (100) applied a weighted coefficient “kappa” which is a measure of association between the test and the diagnosis and this gives an optimal test criterion as well as estimating the test quality. By adjusting the weighting “r”, kappa can be used to find the best screening method (when r=0) or the best diagnostic test (when r=1). However, to compare to the methods used in the present study, the weighting that gave the least false positive and least false negatives r=0.5 can be used.

The weighted kappa was worked out (see Table 21) for the examples shown in section 3.4.2 The effect of Comorbidity, where Chi-square filtering was used to test the association between reduced vision and ametropia (Table 5) and between reduced vision and pinguecula (Table 6),

Reduced Vision/Uncorrected Ametropia

k(0)	0.65
k(0.5)	0.15
k (1)	0.08

Reduced vision/ Pinguecula

k(0)	0.01
k(0.5)	0.02
k(1)	0.33

Table 21 Weighted kappa was worked out for two test/diagnosis combinations (reduced vision/uncorrected ametropia and reduced vision/pinguecula) for different values of "r".

Note how reduced vision is a good test for screening for uncorrected ametropia, but not a good diagnostic test. For pinguecula, reduced vision is not a good screening test, yet appears to have a fair association as a diagnostic test. Recall that the positive likelihood ratios were similar for both tests (Reduced vision/Uncorrected ametropia LR+ = 2.2; Reduced vision/pinguecula LR+ = 1.6) suggesting that reduced vision was of equal diagnostic value for detecting uncorrected ametropia and pinguecula. So both likeliood ratios on their own and kappa can make spurious associations.

However, in this study, more than one test/diagnosis association is being tested for the strength of that association. Kappa allows for only one test at a time and whether that test is suitable for screening or diagnosis, whereas the filtered Chi-square removes spurious association for a number of tests at a time.

QROC curves (similar to ROC curves) determine the strength of a test/diagnosis association at different values of the test, by using the different sensitivity and specificity at different values to produce the QROC curve. In this respect, all the data has to be analysed to find the best cut-off value

for each test. Like the ROC curves this process was felt to be too cumbersome for the purposes of the current study.

Advantages of using the Chi-square filtering are

- It can be incorporated into the naïve Bayesian sequential analysis, and can work in the "background" within the analysis,
- It allows the use of multiple test criteria (such as age, the severity of symptoms and signs) which allows a "tailor-made" diagnosis for each person.

Yet, the use of Chi-square filtering appears to be questionable when applied to all available data, and does not appear to justify the computational effort involved. However, the use of Chi-square filtering does appear to make a worthwhile contribution when it comes to identifying "sensible" recommended tests to confirm diagnosis. Further work is required, especially when both positive and negative likelihoods ratios are used and where diagnoses without definitive tests occur, to see if Chi-square filtering is a worthwhile addition to Bayesian analysis.

The clinical records used for the analysis were created during routine eye examinations. Although only positive outcomes were assessed in this study, it is important to note that the clinical records did contain "negative" outcomes such as "anterior segment– normal". Unfortunately, this type of recording of negative findings was deemed not to be specific enough for the purposes of this study. Ideally, each test item should have a positive or negative test outcome or otherwise an indicator that the test had not been carried out. This study shows the importance of recording negative findings as a part of best practice.

In clinical practice, a practitioner may well recognise a condition/disease due to experience and having done so, may not note all the signs leading to the diagnosis. In such cases, where information is incomplete, loss of accuracy may occur.

Sadatsafavi et al (156) created a mathematical model to assess inter-observer agreement when using Bayesian analysis to assist in clinical diagnosis. 3 main sources of errors were cited in interpreting diagnostic findings using Bayes' theorem:

- pre-test probabilities (prevalences in the present study);
- the misclassification of the test outcomes;
- prior knowledge of the observer/clinician which can affect the decision making process. (156).

Although these findings were not applied to clinical practice, a similar finding was made in the present study. That is, where the clinician had prior knowledge (or experience), diagnoses were often made without recording the associated signs (see 6.3.1.3 Possible causes of reduced accuracy: lack of recorded signs), leading to a loss of accuracy.

If sufficient further accuracy can be achieved with the above suggestions, naïve Bayesian analysis could be incorporated into practice software systems. This would create an element of artificial intelligence, assisting in clinical decision making and diagnostic test selection. Global use of such an "intelligent" system could base prevalence on entered details of the practice location and setting (i.e. primary or secondary care), as prevalence is dependent on the clinic location and setting (157). However, likelihood ratios would apply to any setting (as they are independent of prevalence). This "intelligent" system would update itself with continued input of clinical data (becoming a circular, sequential analysis). The potential here is vast as this would enable epidemiological surveillance on an unprecedented scale.

In addition, an "intelligent" system within practice software could help to safeguard against missed diagnosis (that is, by suggesting the most appropriate tests to be carried out). In this manner, such a system would have legal value in that it could provide an evidence base for "best or current practice" that would support legal defence in cases of malpractice. The resulting reduction of legal actions might, in turn, reduce legal costs and the cost of professional indemnity insurance.

Bayes' theorem has been used within the courts of law (25), (86). However, Bayesian principles may be difficult to understand for the layman and, recently, in an appeal against a murder conviction in the UK, a judge ruled against the use of Bayesian arguments "unless the underlying statistics are 'firm'" (158). Bayesian analysis has always been a method of drawing inference from limited data; drawing conclusions from the effects about the cause (see 2.2 Early History of Bayes' Theorem). Thus it appears that the judge has ruled against what may be the best method of obtaining statistical inference from limited data, not necessarily because the statistical methods themselves are flawed, but because it is difficult for the layman to understand the principles involved in Bayesian analysis.

Another potential advantage of "intelligent" systems is the possibility that they could increase efficiency through better use of optometric assistants. Here, "intelligent" systems could be used by optical assistants to elicit the presenting symptoms from which the most appropriate diagnostic tests would automatically be identified. The optometrist would then concentrate on the recommended

diagnostic tests. This would save time by providing the most efficient problem-orientated eye examination directed to the resolution of the presenting symptoms.

Would such an “intelligent” system lead to “deskilling” of the optometrist? A comparable situation was experienced when soft contact lenses improved in design and quality. As they became more popular, many optometrists thought they would lose this potentially lucrative market to unskilled personnel. However, the reality showed that optometrists were facing new challenges, and therefore had to learn new skills to deal with the complications posed by the extensive use of soft contact lenses (159) (160) (161). Thus, the optometrist should not fear deskilling but look to different challenges and to a new level of optometry.

Such “intelligent” systems also have the potential of helping new optometrists make the best possible decisions based on the symptoms and signs that they have elicited from the patient. Similarly such a tool can help more seasoned optometrists make decisions more confidently and provide the evidence base to facilitate the training of younger colleagues. Bayesian analysis is a powerful decision-making tool allowing the clinician to consider treatments based on rigorous evidence tailored to the patient’s individual requirements (162). The use of Bayesian analysis can also inform problem-based learning, which tends to encourage an inquisitive style of learning as opposed to the more traditional, short term, rote memorisation. This encourages the learner (be it a student or a seasoned practitioner) to become a pro-active, independent thinker, with good problem solving skills. (163). This would harmonise well with the current General Optical Council continuing professional requirements for registration to practise in the UK (164). A further consideration would be to develop problem orientated clinical records in which Bayesian decision making is part of the record and can help to identify when a test should be carried out and when treatment should be started (165). Such examination of probabilities can rule out unnecessary tests, making medical care more cost and time efficient (166). Clinical records should be user friendly, allow data analysis for the purposes of self-evaluation and learning, enable the formulation of public health policies and research, with strict parameters for patient confidentiality and comprehensive data recording (167).

Bayesian analysis in medicine has advanced considerably since Jerome Cornfield’s discovery linking lung cancer and smoking (58) and, especially in the last 30 years, allowing modelling where flexibility and innovation are desired, where models can be continually changed according to data available (168) (169). Bayesian analysis applied to optometry can assist in decision-making especially where data may be incomplete or equivocal.

Adoption of Bayesian thinking in optometry can allow the eye examination to become more problem orientated and more efficient. Using Bayesian analysis with electronic record keeping can assist in meeting the challenges of the Global Action Plan for Universal Eye Health (3), by assisting collection and analysis of epidemiological data (170), formulating public health policies for resource and infrastructure management, supporting training of eye care personnel (171) and providing an evidence based clinical support guidelines where there may be lack of human and infrastructure resources.

This study has shown the value of data collection from primary eye care in formulating an evidence base, in order to guide the profession and the practitioner alike.

8. References

1. Laplace PS. Philosophical essay on Probablilties (AI Dale, Trans.) New York: Springer; 1995.

2. McGrayne S. Sharon Bertsch McGrayne Answers Questions about "The Theory that would not Die". Chance. 2012 February 7; 25(1): p. 24-29.

3. International Agency for the Prevention of Blindness. IAPB. [Online].; 2014 [cited 2014 May 11. Available from: http://www.iapb.org/advocacy/who-action-plan.

4. World Health Organisation. World Health Organisition. [Online].; 2013 [cited 2013 June 3. Available from: http://www.who.int/blindness/partnerships/vision2020/en/.

5. International Agency for the Prevention of Blindness. International Agency for the Prevention of Blindness. [Online].; 2012 [cited 2012 June 5. Available from: http://www.iapb.org/vision-2020/global-facts.

6. Saguti G. Personal communication. 2013. Dr Saguti- National Eyecare co-ordinator.

7. Murthy G, Raman U. Perspectives on Primary Eye Care. Community Eye Health Journal. 2009; 22(69): p. 10-11.

8. Clarke M. NHS sight tests include unevaluated screening examinations that lead to waste. BMJ. 2104 March; 348:g2084.

9. Blakeney. The College of Optometrists. [Online].; 2010 [cited 2010 August 23. Available from: http://www.college-optometrists.org/coo/download.cfm?uuid=729FE39F-3048-216E-804AA6284C06A348&type=framework.

10. Mshana J. 2013. Deputy Registrar, Optometry Council, United Republic of Tanzania.

11. CIA. The World Factbook. [Online].; 2012 [cited 2012 May 13. Available from: https://www.cia.gov/library/publications/the-world-factbook/geos/tz.html.

12. Community Eye Health Journal. The Role of Optometry in Vision 2020. Community Eye Health Journal. 2002; 15(43): p. 33-36.

13. Shah R, Edgar D, Rabbetts R, Blakeney S, Charlesworth P, Harle D, et al. The content of optometric eye examinations for a young myope with headaches. Ophthal. Physiol. Opt. 2008; 28: p. 404-421.

14. Thampy R, Kwartz J. Clinical decision-making for optometrists in the management of the red eye. Optometry in Practice. 2010; 11(2): p. 51-56.

15. Bayes T. An Essay towards solving a problem in the doctrine of chances. Philos Trans Roy Soc London. 1764; 53: p. 370-418.

16. Encyclopaedia Britannica. Encyclopaedia Britannica. [Online].; 2013 [cited 2013 March 9. Available from: http://www.britannica.com/EBchecked/topic/56807/Thomas-Bayes.

17. McGrayne S. The Theory that Would Not Die: How Bayes' Rule Cracked the Engima Code hunted Down Russian Submarines & Emerged Triumphant from Two Centuries of Controversy. 1st ed. Boston: Yale University Press; 2012.

18. CNET. CNET. [Online].; 2003 [cited 2013 June 12. Available from: http://news.cnet.com/2009-1001-984695.html.

19. Bayes T. An Introduction to the Doctrine of Fluxions and a defence of the Mathematicians against the objections of the Author of the Analyst. 1736. Printed by John Noon (Dr Williams Library).

20. Holland J. The Reverend Thomas Bayes, FRS (1702-61). Journal of the Royal Statistical Society. Series A (General). 1962;: p. 451-461.

21. Gillies D. Was Bayes a Bayesian? Historia Mathematica. 1987; 14: p. 325-346.

22. Stigler S. Laplace's 1774 Memoir on Inverse Proabability. Statistical Science. 1986; 1(3): p. 359-378.

23. Dale A. On Bayes'Theorem and the Inverse Bernoulli Theorem. Historica Mathematica. 1988; 15: p. 348-360.

24. Bru B, Jongmans F. Joseph Bertrand. In Statisticians of the Centurie.: Springer; 2001.

25. Kaye D. Revisiting "Dreyfus": A More Complete Account of a Trial by Mathematics. Minnesota Law Review. 2007 Feb; 91(3): p. 825-835.

26. Molina E. Some Fundamental curves for the solution of sampling problems. The Annals of Mathematical Statistics. 1946;: p. 325-335.

27. Bailey A. Credibility Procedures Laplace's Generalistion of Bayes Rule and the combination of Collateral Knowledge with Obseerved Data. New York: New York State , Insurance Department; 1950.

28. Rubinow I. Social Insurance. ; 1916.

29. Rubinow I. Scientific Methods for Computing Compensation Rates. ; 1914.

30. Rubinow I. A Standard Accident Table as a Basis for Compensation Rates. American Statistical

Association. 1915 March; 14(109): p. 358-415.

31. Fisher R. On the Mathematical Foundations of Theoretical Statistics. Phil. Trans. R. Soc. Lond. 1922 April 19; 222.

32. Fisher R. Frequency Distributions of the Values of the Correlation Coeffiecient of Samples from an Indefinately Large Population. Biometrika. 1915 May; 10(4): p. 507-521.

33. Fisher R. On the Interpretation of χ2 from Contingency Tables and the Calculation of P. Journal of the Royal Statistical Society. 1922 Jan 87-94; 85(1).

34. Fisher R, Bennett J. Statistical methods. experimental design and scientific inference: Oxford University Press; 1973.

35. Pearson E. Bayes' Theorem, examined in the light of experimental sampling. Biometrika. 1925; 17(3-4): p. 388-442.

36. Neyman J, Pearson E. On the use and intepretation of certain test criteria for the purposes of statistical inference part i. Biometrika. 1928; 20(1-2): p. 175-240.

37. De Finetti B. Methods for discriminating levels of partial knowledge concerning test item. British Journal of Mathemathical and Statistical Psychology. 1965; 18(1): p. 87-123.

38. De Finetti B. Bayesianism: its unifying role for both the foundations and application of statisitics. International Statistical Review. 1974; 42(2): p. 117-130.

39. De Finetti B. Logical foundations and measurements of subjective probabilty. Acta Psychologica. 1970; 34: p. 129-145.

40. De Finetti B. The Bayesian approach to the rejection of outliers. In Berkeley UoC. Proceedings of the fourth Berkeley Symposium on Probability and Statisitics.: University of California Press Berkeley; 1961. p. 199-210.

41. Jeffreys H. The theory of Probability: Oxford University Press ; 1998.

42. Good I. Studies in the History of Probability and Statistics XXXVII. AM Turing's Statistical Work in World War II. Biometrika. 1979 Aug; 66(2): p. 393-396.

43. Shannon CE. Key papers in the development of information theory. Bell Syst. Tech. J. 1948; 27: p. 623-656.

44. Shannon CE. XXII. Programming a computer for playing chess. Philosophical magazine. 1950; 41(314): p. 256-275.

45. Shannon CE. Communication Theory of Secrecy Systems. Bell System Technical Journal. 1949;

28(4): p. 656-715.

46. Shannon CE. A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review. 2001; 5(1): p. 3-55.

47. Kolmogorov AN. English translations of Kolmogorov's 1941 papers. Proceedings of the Royal Society of London, Series A. 1991; 434: p. 9-17.

48. Kolmogorov AN. Foundations of the theory of probability Oxford,England: Chelsea Publishing Co; 1950.

49. Wald A, Wolfowitz J. Bayes solutions of sequential decision problems. The Annals of Mathematical Statistics. 1950;: p. 82-99.

50. Koopman BO. The theory of search. I. Kinematic bases. Operations research. 1956; 4(3): p. 324-346.

51. Koopman BO. An operational critique of detection laws. Operations Research. 1979; 27(1): p. 115-133.

52. Morse PM. In Memoriam: Bernard Osgood Koopman, 1900-1981. Operations Research. 1982 May-June; 30(3): p. viii+417-427.

53. Stone L. Theory of Optimal Search. 1975..

54. Mickey M, Gjertson D, Terasaki P. Empirical Validation of the Essen-Moller Proability of Paternity. Am J Hum Genet. 1986; 39: p. 123-132.

55. Bailey R. Insurance rates with minimum bias. In ; 1963. p. 4-11.

56. Bailey R, Simon L. Two studies in automobile insurance ratemaking. Astin Bulletin. 1960; 1(4): p. 192-217.

57. Savage LJ. The Foundations of Statistics. 2nd ed. New York : Wiley; 1954.

58. Cornfield J. The Bayesian Outlook and its applications (with discussion). Biometrics. 1969; 25: p. 617-657.

59. Cornfield J, Haenszel W, Hammond EC, Lilienfeld A, Shimkin M, Wynder E. Smoking and Lung Cancer: Recent evidence and a discussion of some questions. Internatoinal Journal of Epidemiology. 2009; 38: p. 1175-1191.

60. Truett J, Cornfield J, Kannel W. A multivariate analysis of the risk of coronary heart disease in Framingham. Journal of chronic diseases. 1967; 20(7): p. 511-524.

61. Warner H, Toronto A, Veasy L. Experience with Bayes' Theorem for Computer Diagnosis of Congenital Heart Disease. Annals of the New York Academy of Sciences. 1964; 115: p. 558.

62. Gorry G, Barnett G. Experience with a Model of Sequential Diagnosis. Working Paper. Cambridge, Massachusetts: Massachusetts Institute of Technology, Alfred P Sloan School of Management; 1968.

63. Pratt J, Raiffa H, Schlaifer R. The foundations of decision under uncertainty: An elementary exposition. Journal of the American Statistical Association. 1964; 59(306): p. 353-375.

64. Birnbaum A. On the Foundations of Statistical Inference (with discussion). Journal of the American Statistical Association. 1962; 57(298): p. 269-326.

65. Barnard G, Jenkins , Winsten C. Likelihood inference and time series. Journal of the Royal Statistical Society. Series A (General). 1962: p. 321-372.

66. Mosteller F, Wallace D. Inference in an Authorship Problem: A comparative study of discrimination methods applied to the authorship of the disputed "Federalist" papers. Journal of the American Statistical Association. 1963 June; 58(302): p. 275-309.

67. Brillinger D. John W Tukey: His Life and Professional Contributions. The Annals of Statistics. 2002; 30(6): p. 1535-1575.

68. Rasmussen N, Levine S. Reactor Safety Study - An assessment of accident risks in US commercial nuclear power plants. Technical Report. US Nuclear Regulatory Commission; 1975.

69. Kinsey A, Pomeroy W, Martin C. Sexual behavior in the human male. 1948 Dec. Published by Saunders, Philadelphia.

70. Raftery A, Akman V. Bayesian Analysis of a Poisson Process with a Change-point. Biometrika. 1986 April; 73(1): p. 85-89.

71. Gelfand A, Smith A. Sampling-based approaches to calculating marginal densities. Journal of the American statistical association. 1990; 85(410): p. 398-409.

72. Gelfand A, Hills S, Racine-Poon A, Smith A. Illustration of Bayesian Inference in Normal Data Models using Gibbs Sampling. Technical report no 421. Stanford California: Stanford University, Dept of Statistics; 1989.

73. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1984; 6: p. 721-741.

74. Resnik P, Hardisty E. Gibbs Sampling for the Uninitiated. 2010. Notes from University of Maryland.

75. Lunn D, Thomas A, Best N, Spiegelhalter D. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing. 2000; 10: p. 325-337.

76. Fienberg SE. When Did Bayesian Inference Become "Bayesian". Bayesian Analysis. 2003; 1: p. 1-41.

77. Fraser D. Is Bayes Posterior just Quick and Dirty Confidence. Statistical Science. 2011; 26(3): p. 299-316.

78. Hall G. The Clinical Application of Bayes' Theorem. The Lancet. 1967 Sept; 2(7515): p. 555-557.

79. Okeh U, Ugwu A. Bayes' theorem: A paradigm research tool in biomedical sciences. African Journal of Biotechnology. 2008 December; 7(25): p. 4807-4818.

80. Sahami M, Dumais S, Heckerman D, Horvitz E. A Bayesian Approach to Filtering Junk E-mail. 1998.

81. Heckerman D, Mamdani A, Wellman M. Real-world applications of Bayesian Networks. Communications of the ACM. 1995 March: p. 24-26.

82. Tucker K, Rushton S, Sanderson R, Martin E, Blaiklock J. Modelling bird distributions — a combined GIS and bayesian rule-based approach. Landscape Ecology. 1997 April; 12(2): p. 77-93.

83. Aspinall R. An inductive modelling procedure based on Bayes' theorem for analysis of pattern in spatial data. International Journal of Geographical Information Systems. 1992 Oct; 6(2).

84. Best N, Ashby D, Dunstan F, Foreman D, McIntosh N. A Bayesian approach to complex clinical diagnoses:a case-study in child abuse. J. R. Statist. Soc. 2013; 176, Part 1 : p. 53-96.

85. Zeng G, Thacker S, Hu Z, Lai X, Wu G. An assessment of the use of Bayes' theorem for the forecasting in public health: the casse of Epidemic Meningitis in China. International Journal of Epidemiology. 1988; 17(3): p. 673-679.

86. Fenton N, Neil M. Comparing risks of alternative medical diagnosis using Bayesian arguments. Journal of Biomedical Informatics. 2010; 43: p. 485-495.

87. BBC. BBC. [Online].; 2014 [cited 2014 April 11. Available from: http://www.bbc.com/news/magazine-26680633.

88. Charniak E. Bayesian Networks without Tears. AI Magazine. 1991: p. 50-63.

89. Spiegelhalter MJA. Bayesian methods in health technology assessment: a review. Health Technology Assessment. 2000; 4(38).

90. Siegal E. Predictive Analysis. 1st ed. Hoboken : John Wiley &Sons; 2013.

91. Berger J. Bayesian analysis: A look at Today and Thoughts of Tomorrow. Journal of the American Statistical Assoc. 2000 Dec; 95(452): p. 1269-1276.

92. Aspinall P, Hill A. Clinical Inferences and decisions - I. Diagnosis & Bayes Theorem. Ophthal. Physiol. Opt. 1983; 3(3): p. 295-304.

93. Woods R, Tregear S, Mitchell R. Screening for ophthalmic disease in older subjects using visual acuity and contrast sensitivity. Ophthalmology. 1998 December; 105(12): p. 2318-2326.

94. Hill A, Aspinall P, Verriest G. Principles of Colour Vision Test Battery Selection. Documenta Ophthalmologica Proceeding Series. 1984; 39: p. 181-187.

95. Heron G, Erskine N, Farquharson E, Moore A, White H. Colour vision screening in glaucoma: the Tritan Album and other simple tests. Ophthalmic Physiol Opt. 1994 July; 14(3): p. 233-238.

96. Odom J, Chao GM, Weinstein G. Preoperative prediction of postoperative visual acuity in patients with cataracts: A quantitative review. Documanta Ophthalmologica. 1988 Sept; 70(1): p. 5-17.

97. Odom J, Hobson R, Coldren J, Chao GM, Weinstein G. 10-Hz flash visual evoked potentials predict post-cataract extraction visual acuity. Documenta Ophthalmologica. 1987 Aug; 66(4): p. 291-299.

98. Cox M, Wood I. Measurement of Friedmann Visual Field Analyzer Tests in Primary Open Angle Glaucoma. Optometry and Vision Science. 1993; 70(5): p. 421-432.

99. Jansonius N. Bayes theorem applied to perimetric progression detection in glaucoma:from specificity to positive predictive value. Graefe's Arch Clin Exp Ophthalmol. 2005; 243: p. 433-437.

100. Gilchrist J. QROC curves and kappa functions: new methods for evaluting the quality of clinical decision making. Ophthal Physio Opt. 1992 July; 12(3): p. 350-360.

101. Thomas R, Mengersen K, Parikh R, Walland M, Muliyil J. Enter the reverend: introduction to and application of Bayes' theorem in clinical ophthalmology. Clinical and Experimental Ophthalmology. 2011; 39: p. 865-870.

102. McGraw P, Brosnahan D, Winn B, Whitaker D. Assessment of Retinal-Neural Function Before Neodymium:YAG Laser Capsulotomy. Investigative Ophthalmology & Visual Science. 1995 May; 36(5).

103. Mutti D, Zadnik K. The Utility of Three Predictors for Childhood Myopia: a Bayesian Approach. Vision Res. 1995; 35(9): p. 1345-1352.

104. Hand D, Yu K. Idiot's Bayes-Not So Stupid After All? International Statiscal Review. 2001 March;

69(3): p. 385-398.

105. Charniak E. The Bayesian basis of common sense medical diagnosis. In AAAI; 1983. p. 70-73.

106. van Zadelhoff B. Towards Large Diagnostic Bayesian Network Models. Bachelor Thesis. Radboud University, The Netherlands, Computer Science; 2013.

107. Parikh R, Mathai A, Parikh S, Chandra Shekar G, Thomas R. Understanding and using sensitivity, specificity and predicitive values. Indian J Ophthalmol. 2008 Jan - Feb; 56(1): p. 45-50.

108. Parikh R, Parikh S, Arun E, Thomas R. Likelihood ratios: Clinical application in day-to-day practice. Indian J Ophthalmol. 2009 May-June; 57(3): p. 217-221.

109. Chang A. StatTools. [Online].; Oct 2012 [cited 2013 March 9. Available from: http://www.stattools.net/Contingency2x2_Exp.php.

110. McGee S. Simplifying Likelihood Ratios. J Gen Intern Med. 2002 August; 17(8): p. 647-65.

111. Wikimedia Foundation, Inc. Wikipedia. [Online].; 2013 [cited 2013 March 11. Available from: http://en.wikipedia.org/wiki/Likelihood_ratios_in_diagnostic_testing.

112. Sonis J. How to Use and Interpret Interval Likelihood Ratios. Fam Med. 1999; 31(5): p. 432-7.

113. Altman D, Bland J. Diagnostic tests 2: predictive tests. BMJ. 1994 July; 309: p. 102.

114. NHS National Institute for Health and Clinical Excellence. Glaucoma -Diagnosis and management of chronic open angle glaucoma and ocular hypertension. NICE clinical guidelines. London: National Health Service UK, National Collaborating Centre for Acute Care ; 2009. Report No.: ISBN 1-84629-949-7.

115. College of Optometrists. The College of Optometrists. [Online].; 2014 [cited 2104 May 28. Available from: http://www.college-optometrists.org/en/utilities/document-summary.cfm/docid/D97A25AB-E034-4056-A462894D52D1F5C8.

116. Kass R, Raftery A. Bayes Factors. Journal of the American Statistical Association. 1995 June; 90(430): p. 773-795.

117. Rudnicka A, Owen C. Epidemiology of primary open angle glaucoma. In Edgar D, Rudnicka A. Glaucoma - Identification & Co-management. Philadelphia: Elsevier; 2007. p. 1-16.

118. Minassian D, Reidy A, Coffey M, Minassian A. Utility of predictive equations for estimating the prevalence and incidence of primary open angle glaucoma in the UK. Br J Ophthalmol. 2000; 84: p. 1159-1161.

119. Tielsch J, Katz J, Singh K, Quigley H, Gottsch J, Javitt J, et al. A Population- based evaluation of

Glaucoma Screening: The Baltimaore Eye Survey. Am J Epidemiology. 1991; 134(10): p. 1102-1110.

120. Summerton N. Making a diagnosis in primary care:Symptoms and context. British Journal of General Practice. 2004 August;: p. 570-571.

121. Medow M, Lucey C. A qualitative approach to Bayes' theorem. Evidence-Based Medicine December 2011 | volume 16 | number 6. 2011 Dec; 16(6): p. 163-167.

122. Agoritsas T, Courvoisier D, Combescure C, Deom M, Perneger T. Does Prevalence Matter to Physicians in Estimating Post-test Probability of Disease? A Randomized Trial. J Gen Intern Med. 2010 Nov; 26(4): p. 373-378.

123. Summerton N. The medical history as a diagnostic technology. British Journal of General Practice. 2008 April; 58: p. 273-276.

124. Grimes D, Schulz K. Epidemiology 3: Refining clinical diagnosis with likelihood ratios. Lancet. 2005; 365: p. 1500-1505.

125. Hayden S, Brown M. Likelihood ratio- A Powerful Tool for Incorporating the Results of a Diagnostic Test into Clinical Decisionmaking. Ann Emerg Med. 1999 May; 33: p. 575-580.

126. Valderas J, Starfield B, Sibbald B, Salisbury C, Roland M. Defining Comorbidity: Implications for Understanding Health and Health Services. Ann Fam Med. 2009 Juky; 7(4): p. 357-363.

127. Mitnitski A, Graham J, Mogilner A, Rockwood K. Vector Diagnostics in Dementia Derived from Bayes Theorem. Am. J. Epidemiol. 1997 July; 146(8): p. 665-671.

128. Graham J, Mitnitski A, Mogilner A, Gauvreau D, Rockwood K. An Algorithmic Approach to the Differential Diagnosis of Dementia. Dementia and Geriatric Cognitive Disorders. 1996; 7(6): p. 324-330.

129. Finney D. The Fisher-Yates Test of Significance in 2 ×2 Contingency Tables. Biometrika. 1948 May; 35(1/2): p. 145-156.

130. Gill C, Sabin L, Schmid C. Why clinicians are natural Bayesians. BMJ. 2005 May; 330: p. 1080-3.

131. Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. [online]. 317117th ed. London: Elsevier Science & Technology; 2011.

132. Institute of Work and Health. Institute of Work and Health. [Online].; 2010 [cited 2013 March 20. Available from: http://www.iwh.on.ca/wrmb/retrospective-vs-prospective-studies.

133. PARTICIPANTS, ROYAL COLLEGE OF PSYCHIATRISTS' WORKING PARTY. GUIDELINES FOR RESEACHERS AND FOR RESEARCH ETHICS COMMITTEES ON PSYCHRIATRIC RESEARH INVOLVING

HUMAN PARTICIPANTS. Guideline. London: ROYAL COLLEGE OF PSYCHIATRISTS; 2000.

134. Church R. The Effective use of Secondary Data. Learning and Motivation. 2001; 33: p. 32-45.

135. Rakes J. The medical record. Clinical Eye and Vision Care. 1995; 7: p. 49-55.

136. Warburton T. Staying out of Trouble- will your record keeping help or hinder? Optometry Today. 2008 November : p. 38-40.

137. Chandra Shekar D, Sesha Srinivas V. Clinical Data Mining- An Approach for identification of Refractive Errors. In Proceedings of the International MultiConference of Engineers and Computer Scientists; 2008.

138. Twa M, Parthasarathy S, Roberts C, Mahmoud A, Raasch T, Bullimore M. Automated Decision Tree Classification of Corneal Shape. Optom Vis Sci. 2005 Dec; 82(12): p. 1038-1046.

139. Twa M, Parthasarthy S, Bullimore M, Roberts C, Mahmoud A, Raasch T, et al. Automated Decision Tree classification of Keratoconus from videokeratography. [Online]. [cited 2013 Dec 7 [College of Optometry, The Ohio State University]. Available from: http://www.optikon.com/it/articles/keratron_036/media/Twa_1.pdf.

140. Keir N, Simpson T, Fonn D. Visual and optical performance of silicone hydrogel contact lenses for moderate myopia. J Optometry. 2010; 3(3): p. 149-157.

141. Frick K, Riva-Clement L, Shankar M. Screening for refractive error and fitting with spectacles in rural and urban India: Cost-effectiveness. Ophthalmic Epidemiology. 2009; 16(6): p. 378-387.

142. Dunstone D, Armstrong R, Dunne M. Survey of habits and attitudes to retinoscopy by optometrists in the UK. Optometry in Practice. 2013; 14(2): p. 45-54.

143. Wikimedia Foundation Inc. Wikipedia. [Online].; 2014 [cited 2014 June 5. Available from: http://en.wikipedia.org/wiki/C4.5_algorithm.

144. IBM. IBM SPSS Statistics. [Online].; 2014 [cited 2014 June 5 [User guide with IBM SPSS software]. Available from: http://user-pc:17661/help/index.jsp?topic=/com.ibm.spss.statistics.help/overvw_auto_0.htm.

145. IBM. IBM SPSS Decision Trees 20. 2011. User guide to IBM SPSS.

146. Armstrong R. When to use the Bonferroni correction. Ophthalmic Physiol Opt. 2014 March.

147. Gerstenblith A, Rabinowitz M, editors. The Wills Eye Manual. 6th ed. Philadelphia: Lippincott, Williams & Wilkins; 2012.

148. The Eye Diseases Prevalence Research Group. The Prevalence of Refractive Errors Among Adults

in the United States, Western Europe, and Australia. Archives of Ophthalmology. 2004; 122: p. 495-505.

149. Katz J, Tielsch J, Sommer A. Prevalence and risk factors for refractive errors in an adult inner city population. Investigative Ophthalmology & Visual Science. 1997; 38: p. 334-340.

150. Wu SY, Nemesure B, Leske M, Group BES. Refractive Errors in a Black Adult Population: The Barbados Eye Study. Investigative Ophthalmology and Visual Science. 1999; 40: p. 2179-2184.

151. Wedner S, Ross D, Balira R, Kaji L, Foster A. Prevalence of eye diseases in primary school children in a rural area of Tanzania. Br J Ophthalmol , 84:1291-1297. 2000; 84: p. 1291-1297.

152. Wedner S, Ross D, Todd J, Anemona A, Balira R, Foster A. Myopia in secondary school students in Mwanza City, Tanzania: the need for a national screening programme. Br J Ophthalmol. ; 66: p. 1200-1206.

153. Icaza Ed, Lopez-Cervantes M, Arredondo A, Robles-Diaz G. Likelihood ratios of clinical, laboratory and image data of pancreatic cancer: a Bayesian approach. Journal of Evaluation in Clinical Practice. 2009; 15: p. 62-68.

154. Aspinal P, Hill A. Clinical Inferences and Decisions II-Decision Trees, Receiver Operator Curves and Subjective Probability. Ophthal.Physio.Opt. 1984; 4(1): p. 31-38.

155. Crick R, Daubs J. Epidemiological aids to clinical decision making in primary open angle glaucoma. Int Ophthal. 1980; 3(1): p. 37-41.

156. Sadatsafavi M, Moayyeri A, Bahrami H, Soltani A. The Value of Bayes Theorem in the Interpretation of Subjective Diagnostic Findings: What Can We Learn from Agreement Studies? Med Decis Making. 2007; 27: p. 735-743.

157. Reid M, Lane D, Feinstein A. Academic Calculations versus Clinical Judgements: Practicing Physicians use of Quantitative Measures of Test Accuracy. Am J Medicine. 1998 April; 104(4): p. 374-380.

158. The Guardian. The Guardian. [Online].; 2014 [cited 2014 May 11. Available from: http://www.theguardian.com/law/2011/oct/02/formula-justice-bayes-theorem-miscarriage.

159. Efron N. Contact Lens Complication. 3rd ed.: Elsevier Health Sciences ; 2012.

160. Burnett-Hodd C. The Future of RGP lenses? Larger-diameter and hybrid lenses offer all the benefits of RGP lenses with the comfort of a soft lens. Optometry in Practice. 2011; 12(2): p. 49-56.

161. Morgan P, Efron N. A decade of contact lens prescribing trends in the United Kingdom (1996-

2005). Contact Lens & Anterior Eyee. 2006; 29: p. 59-68.

162. Ashby D, Smith A. Evidence-based Medicine as Bayesian decision-making. Statist. Med. 2000; 19: p. 3291-3305.

163. Schmidt H, Dauphinee W, Patel V. Comparing the Effects of Problem-based Curricula and Traditional Curricula in an International Sample. Journal of Medical Education. 1987 April; 62: p. 305-314.

164. General Optical Council. General Optical Council. [Online]. London; 2012 [cited 2014 April 14 [Outline of new CET requirements]. Available from: http://www.optical.org/en/Education/CET/index.cfm.

165. Realdi G, Previato L, Vitturi N. Selection of diagnostistic tests for clinical decision making and translation to a problem orientated medical record. Clinica Chimica Acta. 2008; 393: p. 37-43.

166. Herrle S, Corbett E, Fagan M, Moore C, Elnicki D. Bayes' Theorem and the Physical Examination: Probability Assessment and Diagnostic Decision-making. Acad Med.. 2011 May; 86(5): p. 618-627.

167. Gillum R. From papyrus to the electronic tablet: A brief history of the clinical medical record with lessons for the digital age. The American Journal of Medicine. 2013 Oct; 126: p. 853-857.

168. Ashby D. Bayesian Statistics in medicine : A 25 year review. Statist. Med. 2006 August; 26: p. 3589-3631.

169. Wertli M, Ruchti K, Steurer J, Held U. Diagnostic indicators of non-cardiovascular chest pain:a systemic review and meta-analysis. BMC Medicine. 2013; 11: p. 239.

170. Dunson D. Commentary: Practical Advantages of Bayesian Analysis of Epidemiologic Data. Am J Epidemiol. 2001; 153: p. 1222-1226.

171. Andriamanjato H, Mathenge W, Kalua K, Courtright P, Lewallen ,S. Task shifting in primary eye care: how sensitive and specific are common signs and symptoms to predict conditions requiring referral to specialist eye personnel. Human Resources for Health. 2014; 12(S1:S3).

172. Vision 2020- The Right to Sight. Vision 2020- the Right to Sight. [Online].; 2011 [cited 2013 Jan 30. Available from: http://www.vision2020.org/main.cfm?type=WHATVISION2020.

173. StatSoft. StatSoft. [Online].; 2013 [cited 2013 Oct 14. Available from: www.statsoft.com/.

174. Nielsen R, Ulrichsen S. The New Co-Morbidity Index. 2010. Masters thesis paper at Aalborg University, Denmark.

175. Vollmer L, Schechtman D, Woods A, Pizzimenti J. Use of multifocal ERG and OCT for diagnosing Stargardt's disease. Clin Exp Optom. 2011; 94(3): p. 309-313.

176. Emmett K. Nonspecific and atypical presentation of disease in the older patient. Geriatrics. 1998; 53(2): p. 50-52; 58-60.

177. Lemp M. Advances in Understanding and Managing Dry Eye Disease. Am J Ophthal. 2008 Sepy; 146(3): p. 350-356.

178. Fetterolf D, Tucker T. Assessment of Medical Management Outcomes in Small Populations. Population Health Management. 2008 OCt; 11(5): p. 233-239.

179. Breese J, Heckerman D, Kadie C. Empirical Analysis of Predicitve Algorithmsfor Collaborative Filtering. In Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence; 1998. p. 43-52.

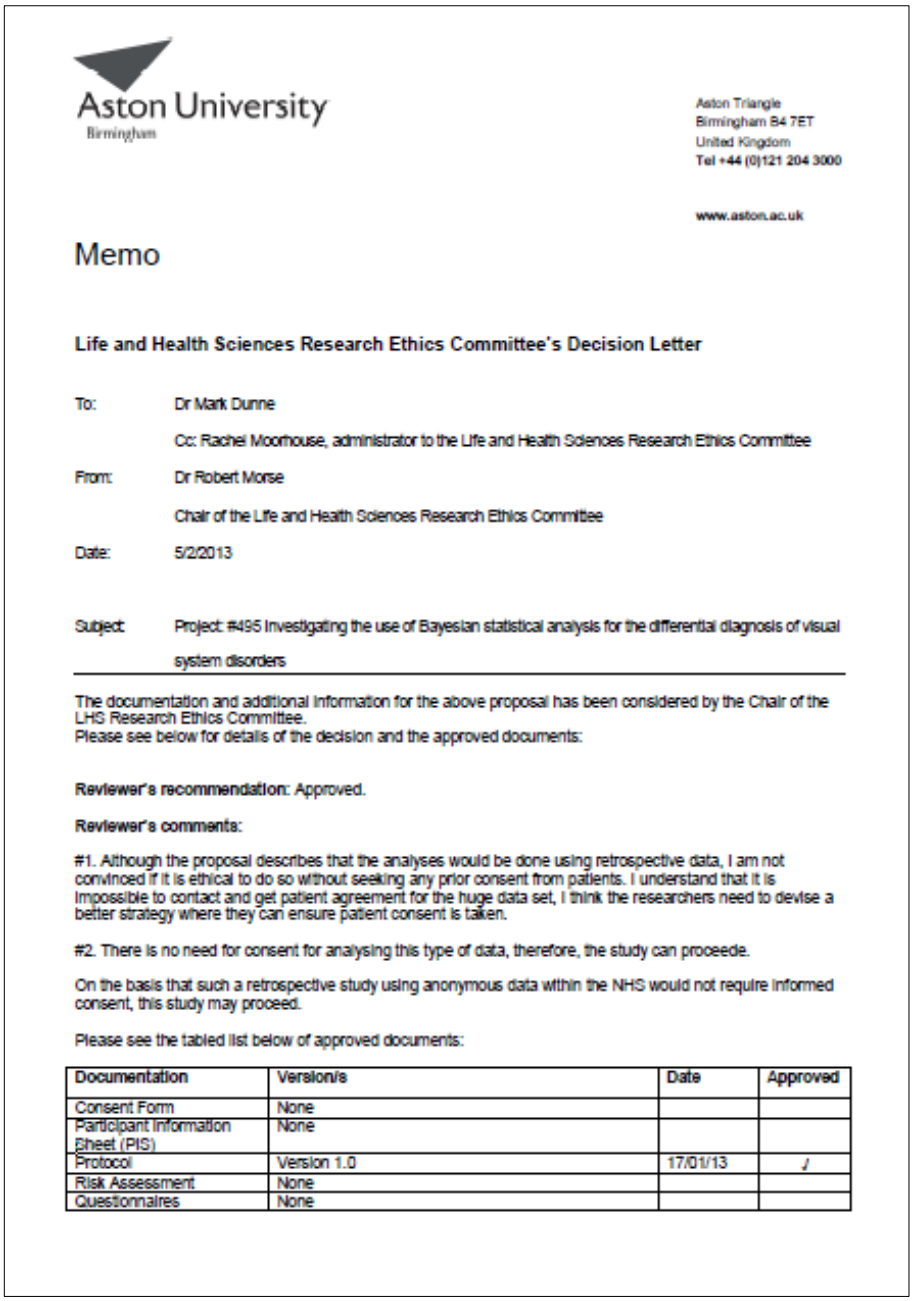
180. Aston University. Aston University. [Online].; 2013 [cited 2013 March 21. Available from: http://www1.aston.ac.uk/study/postgraduate/taught-programmes/school/life-health-sciences/doctor-of-optometry/.

181. IBM. IBM SPSS Decision Trees 20. 2011. User guide with IBM SPSS software.

182. Yates F. Contingency Tables Involving Small Numbers and the χ2 test. Supplement to the Journal of the Royal Statistical Society. 1934; 1(2): p. 217-235.

Appendix I

Aston Ethical Clearance



Aston University
Birmingham

Aston Triangle
Birmingham B4 7ET
United Kingdom
Tel +44 (0)121 204 3000

www.aston.ac.uk

Memo

Life and Health Sciences Research Ethics Committee's Decision Letter

To: Dr Mark Dunne

Cc: Rachel Moorhouse, administrator to the Life and Health Sciences Research Ethics Committee

From: Dr Robert Morse

Chair of the Life and Health Sciences Research Ethics Committee

Date: 5/2/2013

Subject: Project: #495 Investigating the use of Bayesian statistical analysis for the differential diagnosis of visual system disorders

The documentation and additional information for the above proposal has been considered by the Chair of the LHS Research Ethics Committee.
Please see below for details of the decision and the approved documents:

Reviewer's recommendation: Approved.

Reviewer's comments:

#1. Although the proposal describes that the analyses would be done using retrospective data, I am not convinced if it is ethical to do so without seeking any prior consent from patients. I understand that it is impossible to contact and get patient agreement for the huge data set, I think the researchers need to devise a better strategy where they can ensure patient consent is taken.

#2. There is no need for consent for analysing this type of data, therefore, the study can proceede.

On the basis that such a retrospective study using anonymous data within the NHS would not require informed consent, this study may proceed.

Please see the tabled list below of approved documents:

Documentation	Version/s	Date	Approved
Consent Form	None		
Participant Information Sheet (PIS)	None		
Protocol	Version 1.0	17/01/13	✓
Risk Assessment	None		
Questionnaires	None		

Appendix I (continued)

Other (please detail)			

After starting your research please notify the LHS Research Ethics Committee of any of the following:

Substantial amendments. Any amendment should be sent as a Word document, with the amendment highlighted. The amendment request must be accompanied by all amended documents, e.g. protocols, participant information sheets, consent forms etc. Please include a version number and amended date to the file name of any amended documentation (e.g. "Ethics Application #100 Protocol v2 amended 17/02/12.doc").

New investigators

The end of the study

Please email all notifications and reports to lhs_ethics@aston.ac.uk and quote the original project reference number with all correspondence.

Ethics documents can be downloaded from: http://www.ethics.aston.ac.uk/documents-all. Please note that these documents can ONLY be opened using Mozilla Firefox or the latest Internet Explorer version (IE9).

Statement of Compliance

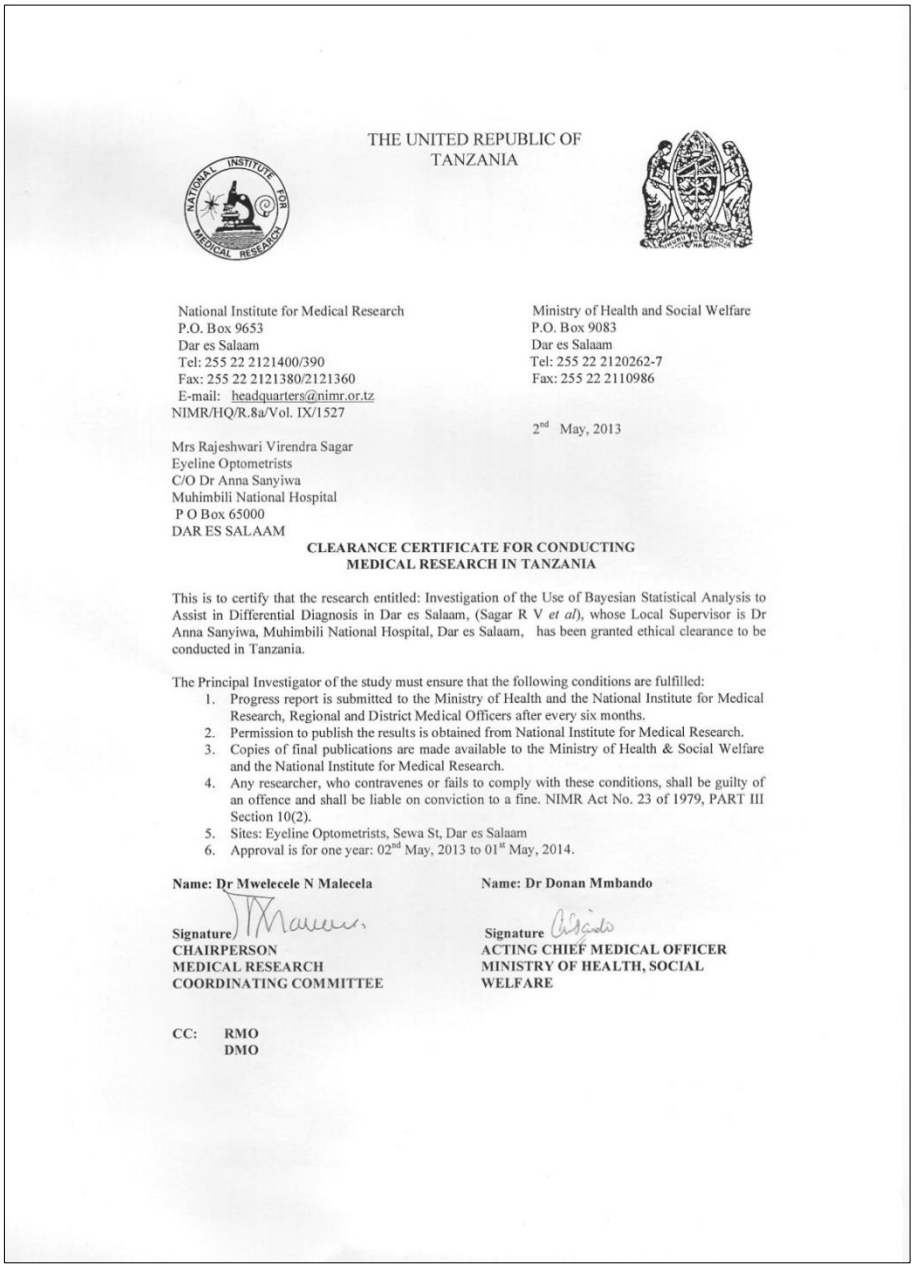
The Committee is constituted in accordance with the Government Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK. In accord with University Regulation REG/11/203(2), this application was considered to have low potential risk and was reviewed by three appropriately qualified members, including the Chair of the Life and Health Sciences Research Ethics Committee.

Yours sincerely

Dr Robert Morse
Chair of the LHS Research Ethics Committee

Appendix II

Tanzanian Ethical Clearance



THE UNITED REPUBLIC OF TANZANIA

National Institute for Medical Research
P.O. Box 9653
Dar es Salaam
Tel: 255 22 2121400/390
Fax: 255 22 2121380/2121360
E-mail: headquarters@nimr.or.tz

NIMR/HQ/R.8a/Vol. IX/1527

Ministry of Health and Social Welfare
P.O. Box 9083
Dar es Salaam
Tel: 255 22 2120262-7
Fax: 255 22 2110986

2nd May, 2013

Mrs Rajeshwari Virendra Sagar
Eyeline Optometrists
C/O Dr Anna Sanyiwa
Muhimbili National Hospital
P O Box 65000
DAR ES SALAAM

CLEARANCE CERTIFICATE FOR CONDUCTING MEDICAL RESEARCH IN TANZANIA

This is to certify that the research entitled: Investigation of the Use of Bayesian Statistical Analysis to Assist in Differential Diagnosis in Dar es Salaam, (Sagar R V *et al*), whose Local Supervisor is Dr Anna Sanyiwa, Muhimbili National Hospital, Dar es Salaam, has been granted ethical clearance to be conducted in Tanzania.

The Principal Investigator of the study must ensure that the following conditions are fulfilled:

1. Progress report is submitted to the Ministry of Health and the National Institute for Medical Research, Regional and District Medical Officers after every six months.
2. Permission to publish the results is obtained from National Institute for Medical Research.
3. Copies of final publications are made available to the Ministry of Health & Social Welfare and the National Institute for Medical Research.
4. Any researcher, who contravenes or fails to comply with these conditions, shall be guilty of an offence and shall be liable on conviction to a fine. NIMR Act No. 23 of 1979, PART III Section 10(2).
5. Sites: Eyeline Optometrists, Sewa St, Dar es Salaam
6. Approval is for one year: 02nd May, 2013 to 01st May, 2014.

Name: Dr Mwelecele N Malecela
Signature
CHAIRPERSON
MEDICAL RESEARCH
COORDINATING COMMITTEE

Name: Dr Donan Mmbando
Signature
ACTING CHIEF MEDICAL OFFICER
MINISTRY OF HEALTH, SOCIAL WELFARE

CC: RMO
DMO

Appendix II (continued)

NATIONAL INSTITUTE FOR MEDICAL RESEARCH

National Institute for Medical Research
RECEIVED
27 FEB 2013
P. O. Box 9653
DAR ES SALAAM

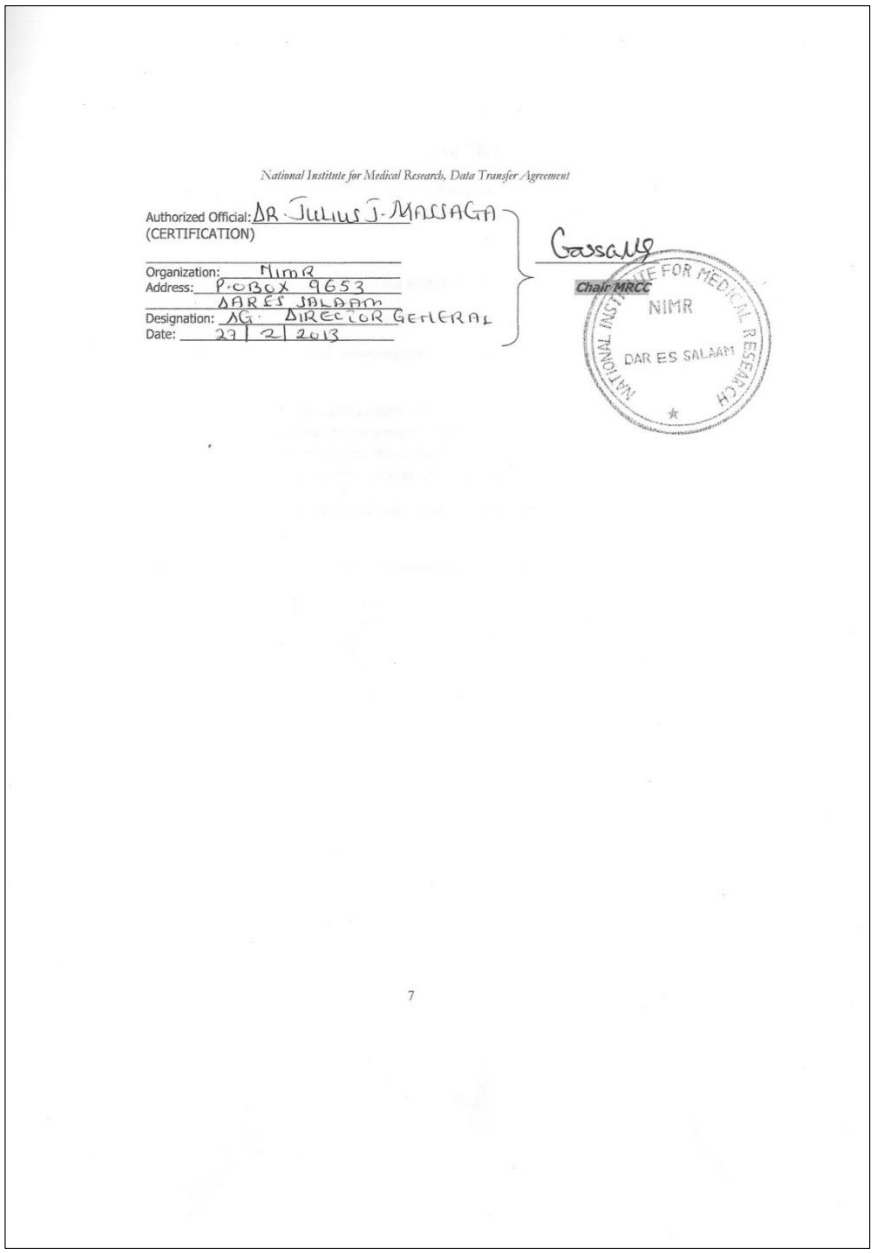
DATA TRANSFER AGREEMENT

DATED this 14 day of Feb, 2013

The United Republic of Tanzania Parliamentary Act No. 23 of 1979 mandates the National Institute for Medical Research to:

1) *to establish a system of the registration of, and to register the findings of medical research carried out within Tanzania, and promote the practical application of those findings for the purpose of improving or advancing the health and general welfare of the people of Tanzania; and*
2) *to establish and operate systems of documentation and dissemination of information on any aspect of the medical research carried out by or on behalf of the institute*

Appendix II (continued)



National Institute for Medical Research, Data Transfer Agreement

Authorized Official: DR. JULIUS J. MASSAGA
(CERTIFICATION)

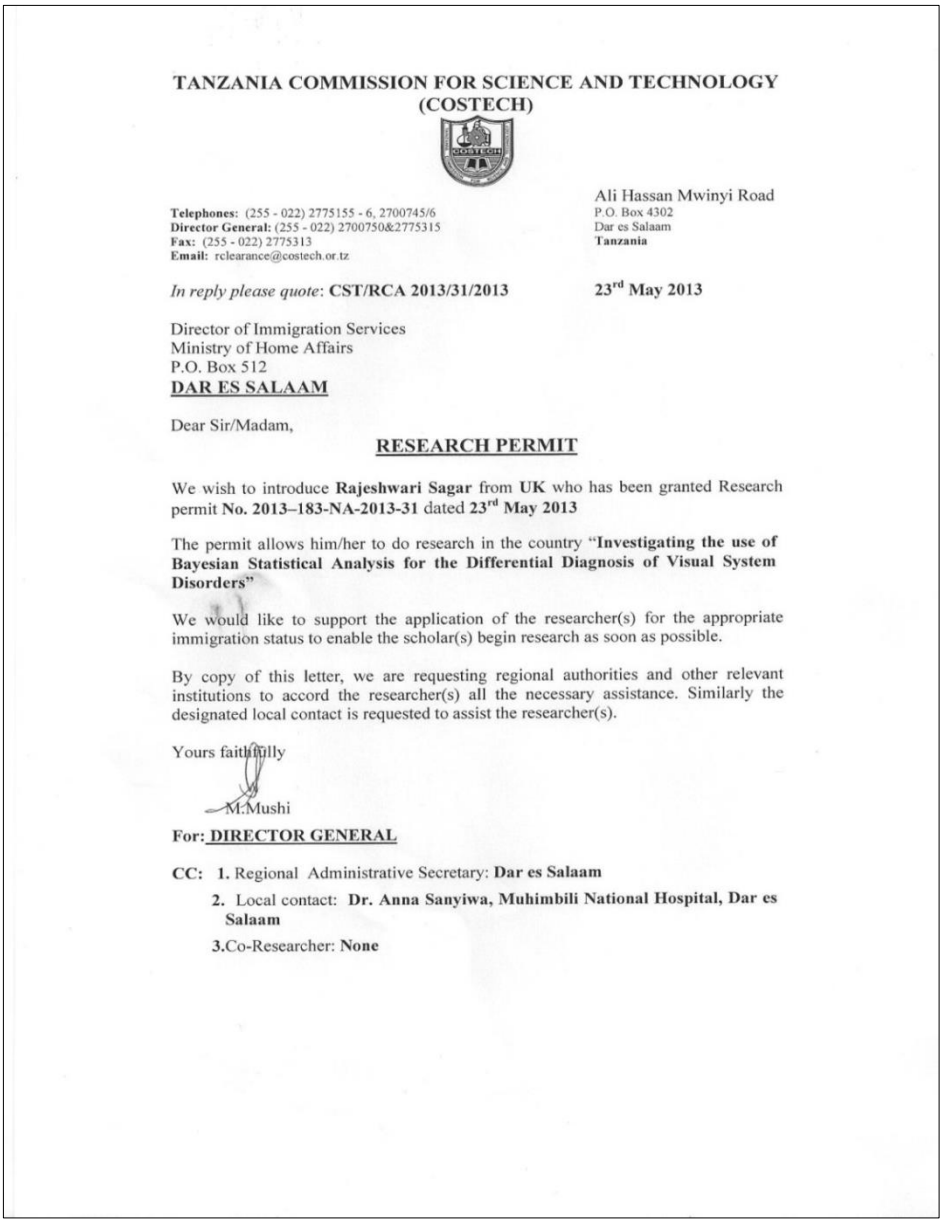
Organization: NIMR
Address: P.O.BOX 9653
DARES SALAAM
Designation: AG. DIRECTOR GENERAL
Date: 23 | 2 | 2013

Gassaga
Chair MRCC

NATIONAL INSTITUTE FOR MEDICAL RESEARCH
NIMR
DAR ES SALAAM

7

Appendix II (continued)



TANZANIA COMMISSION FOR SCIENCE AND TECHNOLOGY (COSTECH)

Telephones: (255 - 022) 2775155 - 6, 2700745/6
Director General: (255 - 022) 2700750&2775315
Fax: (255 - 022) 2775313
Email: rclearance@costech.or.tz

Ali Hassan Mwinyi Road
P.O. Box 4302
Dar es Salaam
Tanzania

In reply please quote: **CST/RCA 2013/31/2013** **23rd May 2013**

Director of Immigration Services
Ministry of Home Affairs
P.O. Box 512
DAR ES SALAAM

Dear Sir/Madam,

RESEARCH PERMIT

We wish to introduce **Rajeshwari Sagar** from **UK** who has been granted Research permit **No. 2013–183-NA-2013-31** dated **23rd May 2013**

The permit allows him/her to do research in the country "**Investigating the use of Bayesian Statistical Analysis for the Differential Diagnosis of Visual System Disorders**"

We would like to support the application of the researcher(s) for the appropriate immigration status to enable the scholar(s) begin research as soon as possible.

By copy of this letter, we are requesting regional authorities and other relevant institutions to accord the researcher(s) all the necessary assistance. Similarly the designated local contact is requested to assist the researcher(s).

Yours faithfully

M.Mushi

For: DIRECTOR GENERAL

CC: **1.** Regional Administrative Secretary: **Dar es Salaam**
2. Local contact: **Dr. Anna Sanyiwa, Muhimbili National Hospital, Dar es Salaam**
3.Co-Researcher: **None**

Appendix II (continued)

TANZANIA COMMISSION FOR SCIENCE AND TECHNOLOGY (COSTECH)

Telephones: (255 - 022) 2775155 - 6, 2700745/6
Director General: (255 - 022) 2700750&2775315
Fax: (255 - 022) 2775313
Email: rclearance@costech.or.tz

Ali Hassan Mwinyi Road
P.O. Box 4302
Dar es Salaam
Tanzania

RESEARCH PERMIT

No. 2013–183-NA-2013-31 **23rd May 2013**

1. Name : **Rajeshwari Sagar**
2. Nationality : **British**
3. Title : **"Investigating the use of Bayesian Statistical Analysis for the Differential Diagnosis of Visual System Disorders"**
4. Research shall be confined to the following region(s): **Dar es Salaam**
5. Permit validity **23rd May 2013 to 22nd May 2014**
6. Contact /Collaborator: **Dr. Anna Sanyiwa, Muhimbili National Hospital, Dar es Salaam**
7. Researcher is required to submit progress report on quarterly basis and submit all Publications made after research.

M. Mushi

for:DIRECTOR GENERAL

Appendix III

Ophthalmic Procedures

Ophthalmic Procedure	Associated Signs used in analysis
1: VA	Reduced VA on presentation
2: CT	Heterotropia Heterophoria
3: FD	Fixation disparity
4: Pupils	Not reactive to light
5: Lids	External (eyelid) - lump/swelling
6: Tears	External (lacrimal apparatus) - TBUT ≤10s
7: Conjunctiva	External (conjunctiva) - lump/nodule External (conjunctiva) - winged mass External (conjunctiva) - redness - bulbar External (conjunctiva) - redness - bulbar - diffuse External (conjunctiva) - redness - bulbar - diffuse - mild External (conjunctiva) - redness - bulbar - diffuse - severe External (conjunctiva) - redness - bulbar - sectoral External (conjunctiva) - redness - palpebral External (conjunctiva) - discharge - clear watery External (conjunctiva) - discharge - yellow pus like External (conjunctiva) - papillae - small External (conjunctiva) - papillae - cobblestone External (conjunctival) - foreign body
8: Cornea	External (cornea) - foreign body External (cornea) - fluorescein staining - punctate External (cornea) - fluorescein staining - diffuse External (cornea) - fluorescein staining - foreign body tracks External (cornea) - limbal ring opacity External (cornea) - haze External (cornea) - lesion External (cornea) - scar
9: Lens	Internal (media) - crystalline lens - cloudy/yellowing Internal (media) - crystalline lens - small white flake like opacities Internal (media) - pseudophakia - opaque capsule
10: Vitreous	Internal (media) - vitreous - floaters Internal (media) - hazy view of fundus
11: Optic disc	Internal (optic disc) - nasal displacement of blood vessels Internal (optic disc) - tilted Internal (optic disc) - CDR>0.7 Internal (optic disc) -intraocular difference of CDR >= 0.2 Internal (optic disc) - well defined margins Internal (optic disc) - pallor Internal (optic disc) - temporal pallor
12:Retinal Vasculature	Internal (Retinal B/Vs) - AV nipping
13: Fundus	Internal(fundus) - haemorrhage - small Internal(fundus) - cotton wool spots

	Internal(fundus) - exudates
14: Macula	Internal(macula) - pigment clumping Internal(macula) - hole Internal(macula) - oedema Internal(macula) - scar
15: Rx	Refraction - uncorrected
16. Add	Near add determination - uncorrected
17: IOP	Tonometry - IOP >21mmHg

Appendix IV

Comparison of LR+ between the preliminary study and the main study

Preliminary Study	min LR+	max LR+	average LR+	weighted average (= prevalence x average LR+)
Uncorrected ametropia	0.093233283	137,038.40	9136.89815	7,273.54
Uncorrected presbyopia	0.087807903	449,803.38	29988.3724	9,532.17
Dry eye	0.816783297	27.32	3.12663961	0.43
Cataract	1	47.54	5.19562989	0.58
Allergic Conjuctivitis	0.076248892	115.18	8.57181606	0.59
Pinguecula	1	174.19	12.6469705	0.25
Diabetic retinopathy	1	175.44	12.751062	0.18
Uncompensated heterophoria	0.078588283	707,766.16	47185.2823	597.28
New Primary Open Angle Glaucoma	0.078588283	1.00	0.8771451	0.00
Macula hole	0.118009615	707.94	48.0705282	0.07
			Average =	1,740.51

Appendix IV continued

Main study	min LR+	Max LR+	Average LR+	weighted average (= prevalence x average LR+)
Uncorrected ametropia	0.04	142,001.75	1,353.34	1,083.05
Uncorrected presbyopia	0.00	461,000.16	4,391.73	1,544.21
Uncompensated heterophoria	0.06	707,766.16	6,748.00	28.47
Heterotropia	0.09	702,914.15	7,532.82	84.76
Allergic Dermatitis	0.06	177.16	2.56	0.01
Blepharitis	0.00	140.95	4.05	0.03
Stye	0.06	176.66	4.17	0.02
Dry eye	0.20	27.32	1.67	0.26
Allergic Conjuctivitis	0.12	115.18	2.39	0.15
Bacterial Conjunctivitis	0.06	318,188.51	3,033.49	46.93
Viral Conjunctivitis	0.06	117.98	3.27	0.01
Pterygium	0.23	586.74	7.23	0.20
Pinguecula	0.00	174.19	2.67	0.05
Subconjunctival haemorrhage	0.06	75.73	1.83	0.01
Foreign body on conjunctiva	0.06	709,293.42	6,755.71	9.50
Contact lens associated GPC	0.06	70.69	1.86	0.01
Contact lens associated red eye	0.06	44.29	1.88	0.01
Corneal arcus	1.00	99.41	3.24	0.07
Corneal abrasion	0.08	78,060.88	744.43	9.42
Foreign body on cornea	0.06	709,293.42	6,755.71	9.50
Cataract - early	1.00	21.24	1.29	0.06
Cataract - posterior pole	1.00	36.42	2.11	0.06
Cataract - cortical	0.07	70,156.27	1,337.56	18.81
Cataract - nuclear	0.06	106.01	2.32	0.01
Primary open angle glaucoma	0.07	140,242.45	1,337.70	18.81
Ocular hypertension	0.06	64.42	1.18	0.00
Normal tension glaucoma	0.07	141.14	2.13	0.01
Anterior Ischaemic optic neuropathy	0.06	709,293.42	6,757.40	9.50
Optic neuropathy (nutritional)	0.06	708,647.85	6,751.85	18.99
Diabetic retinopathy	0.08	311.70	5.64	0.07
Hypertensive retinopathy	0.06	177.28	2.23	0.00
Retinitis pigmentosa	0.06	354.15	3.95	0.01
Diabetic maculopathy	0.07	564.15	7.12	0.05
Macula hole	0.06	707.94	7.45	0.02
ARMD	0.06	708.58	9.54	0.01
			Average=	82.38

Appendix V

Comparison of recommended tests with and without Chi-square

Diagnosis	Best recommended test as identified by LR+	Score	Best recommended test as identified by LR+ without use of Chi-square filtration	Score
Uncorrected ametropia	Refraction - uncorrected	5	Refraction - uncorrected	5
Uncorrected presbyopia	Near add determination - uncorrected	5	Near add determination - uncorrected	5
Uncompensated heterophoria	Fixation disparity	5	Fixation disparity	5
Heterotropia	Cover Test - heterotropia	5	Cover Test - heterotropia	5
Allergic Dermatitis	External (conjunctiva) - discharge - yellow pus like	1	External (conjunctiva) - discharge - yellow pus like	1
Blepharitis	External (eyelid) - lump/swelling	4	External (eyelid) - lump/swelling	4
Stye	External (eyelid) - lump/swelling	5	External (eyelid) - lump/swelling	5
Dry eye	External (lacrimal apparatus) - TBUT ≤10s	5		1
Allergic Conjuctivitis	External (conjunctiva) - papillae - small	5	External (conjunctiva) - papillae - small	5
Bacterial Conjunctivitis	External (conjunctiva) - redness - palpebral	4	External (conjunctiva) - redness - palpebral	4
Viral Conjunctivitis	External (conjunctiva) - redness - bulbar - diffuse - severe	5	External (conjunctiva) - redness - bulbar - diffuse - severe	3
Pterygium	External (conjunctiva) - winged mass	5	External (conjunctiva) - winged mass	5
Pinguecula	External (conjunctiva) - lump/nodule	5	External (conjunctiva) - lump/nodule	5
Subconjunctival haemorrhage	External (conjunctiva) - redness - bulbar - sectoral	5	External (conjunctiva) - redness - bulbar - sectoral	5
Foreign body on conjunctiva	External (conjunctival) - foreign body	5	External (conjunctival) - foreign body	5
Contact lens associated Giant papillary conjunctivitis	External (conjunctiva) - papillae - cobblestone	5	External (conjunctiva) - papillae - cobblestone	5
Contact lens associated red eye	External (conjunctiva) - redness - bulbar - diffuse - mild	4	External (conjunctiva) - redness - bulbar - diffuse - mild	4
Corneal arcus	External (cornea) - limbal ring opacity	5	External (cornea) - limbal ring opacity	5
Corneal abrasion	External (cornea) - foreign body	3	External (cornea) - foreign body	3
Foreign body on cornea	External (cornea) - foreign body	5	External (cornea) - foreign body	5
Cataract - early	Internal (media) - crystalline lens - cloudy/yellowing	5	Internal (media) - crystalline lens - cloudy/yellowing	5
Cataract - posterior pole	Internal (media) - hazy view of fundus	4	Internal(fundus) - cotton wool spots	1
Cataract - cortical	External (conjunctiva) - discharge - clear watery	1	External (conjunctiva) - discharge - clear watery	1
Cataract - nuclear	Internal (media) - hazy view of fundus	4	Internal (media) - hazy view of fundus	4
Primary open angle glaucoma	Pupils - not reactive to light	1	Pupils - not reactive to light	1
Ocular hypertension	Tonometry - IOP >21mmHg	5	Tonometry - IOP >21mmHg	5
Normal tension glaucoma	Internal (optic disc) - nasal displacement of blood vessels	4	Internal (optic disc) - nasal displacement of blood vessels	4
Anterior Ischaemic optic neuropathy	Internal (optic disc) - well defined margins	5	Internal (optic disc) - well defined margins	5
Optic neuropathy (nutritional)	Internal (optic disc) - temporal pallor	5	Internal (optic disc) - temporal pallor	5
Diabetic retinopathy	Internal(fundus) - exudates	4	Internal(fundus) - exudates	4
Hypertensive retinopathy		1	Near add determination - uncorrected	1
Retinitis pigmentosa	Internal(macula) - pigment clumping	5	Internal(macula) - pigment clumping	5
Diabetic maculopathy	Internal(fundus) - exudates	5	Internal(fundus) - exudates	5
Macula hole	Internal(macula) - hole	5	Internal(macula) - hole	5
Age-related macular degeneration - dry	Internal(macula) - pigment clumping	5		1