THE APPLICATION OF A BAYESIAN MACHINE LEARNING PLATFORM TO THE CLINICAL ACTIVITY OF INDEPENDENT PRESCRIBING OPTOMETRISTS

Farkhandah Raqib

Doctor of Optometry

Aston University

February 2022

Aston University

The Application of a Bayesian Machine Learning Platform to the Clinical Activity of Independent Prescribing Optometrists

Farkhandah Raqib

Doctor of Optometry

2022

Abstract

Background:
The NHS Long Term Plan has laid the foundations for healthcare transformation. Artificial intelligence, digitally enabled care, and decision support are mentioned in the NHS Long Term Plan and the Topol Review as enablers for transformation.

Independent Prescribing (IP) optometrists practising in isolation lack "live" peer support, seeking guidance from clinical guidelines and literature reviews instead. This sacrifices the currency of evidence for its quality and contributes to an evidence-to-practice gap.

Aims of study:
- to apply machine learning (ML, a subset of artificial intelligence) to the clinical activity of IP optometrists
- to develop the concept of an interactive and evolving "live" evidence-based support system for IP optometrists and those in training

Methods: Over a year, 1351 first patient consultations were collected by the Acute Primary Care Ophthalmology Service in West Kent (APCOS), a service delivered by IP optometrists.
A digital learning platform was developed (MyDLP) to apply supervised machine learning (naïve Bayes') to the data. A combined "intelligent" electronic patient record and virtual patient tool (iEPR/iVPT) within MyDLP provides decision support and automated grading. MyDLP also evaluates the performance of ML (accuracy, informedness and markedness) using cross validation and learning efficiency curves. The data in MyDLP can be manipulated to promote an understanding of ML concepts amongst clinicians.

Results: A 'proof-of-concept' was demonstrated using the diagnoses and prescribing decisions for keratoconjunctivitis sicca (KCS) and uveitis. Maximum learning efficiency was reached, meaning more data would not have improved model performance. The study findings indicate that Bayes' ML results in good replication for diagnoses and prescribing decisions.

Conclusion: ML can be used to power "live", "white box" decision support tools, useful to both qualified IP optometrists and those in training. As far as the author is aware this was the first time ML was applied to the clinical activity of IP optometrists.

Keywords: Machine learning, Bayes' theorem, IP optometrists, decision replication, support system

For the ones who started this journey with me, and who I lost along the way. My beautiful and intelligent mother and my brave and hardworking father. I hope you are both watching me with pride from heaven up high. It is only your prayers that keep my boat afloat.

To my three amazing and patient children who have sacrificed mama-time for me to fulfil my dreams. Ishaaq, Ahmad, and baby Maryam, my life would be empty without your laughter. I do this and everything for you.

Acknowledgements

First and foremost my thanks and gratitude are directed to my supervisor, Dr Mark Dunne. I could not have wished for someone as knowledgeable, sincere, kind and encouraging. Mark has provided not only academic support but personal too, for which I will be eternally grateful.

Thanks to Dr Preeti Bhogal-Bhamra my associate supervisor, for her reassurances regarding my ideas and writing, and her kind manner.

Thanks to the most amazing and innovative group of specialist optometrists I have ever met, Dr John Gurney, Dr Deacon Harle, Dr Nicola Sabokbar and Mrs Thurka Sivapalan. Your data-collection for this study alongside your everyday practice has been invaluable. You have taught me the vastness of optometric practice.

A special thanks to Mr Amar Alwitry, consultant ophthalmic surgeon, author, educator and my 'big brother', your kindness and support has been unparalleled.

Contents

List of Figures

List of Tables

List of Abbreviations

A/C	Anterior Chamber
Acc	Accuracy
APCOS	Acute Primary Care Ophthalmology Service
Bayes'	Naïve Bayes' Theorem
bTLC	Bayes' Translational Learning Concept
CL	Confidence limits
CMG	Clinical Management Guidelines
CMO	Cystoid Macula Oedema
EPR	Electronic Patient Record
FB	Foreign Body
HSK	Herpes Simplex Keratitis
HZO	Herpes Zoster Ophthalmicus
iEPR	"intelligent" Electronic Patient Record
Inform	Informedness
IP	Independent Prescribing
iVPT	"intelligent" Virtual Patient Tool
KCS	Keratoconjunctivitis Sicca
LR	Likelihood Ratio
Marked	Markedness
MGD	Meibomian Gland Dysfunction
MyDLP	My Digital Learning Platform
NaFl	Sodium Fluorescein
NHS	National Health Service
NPV	Negative Predictive Value
NSAID	Non-Steroidal Anti-inflammatory Drug
PPV	Positive Predictive Value
PTP	Post-test probabilities
KT	Knowledge Translation
AI	Artificial Intelligence
MECS	Minor Eye Conditions Service
CUES	Covid-19 Urgent Eye Care Service
GOS	General Ophthalmic Services
GOC	General Optical Council
CPD	Continuing Professional Development
CET	Continuing Education and Training
DOCET	The Directorate of Optometric Continuing Education and Training
DoH	Department of Health
NICE	National Institute for Health and Care Excellence
SAC	Seasonal Allergic Conjunctivitis

Chapter 1: Introduction and Background

1.1 Introduction

In a world boasting rapid and unparalleled technological development in every field of science, adaptation is no longer a choice but a necessity. The professions, the gatekeepers to specialist knowledge and its application, must have an awareness of, and reflect these developments in practice. All professions, including optometry are transforming at an unprecedented rate (2). Optometry has historically embraced technological changes. The past few decades have witnessed both the widespread use of automated machinery and an increase in scope of the clinical practice of optometrists. This latter activity has been particularly evident during the recent Covid-19 crisis (3,4). With optometrists largely left to practise in isolation, supporting the workforce through this technological and clinical revolution is of paramount importance.

The purpose of the research presented in this thesis was to explore the role of artificial intelligence driven clinical decision support for independent prescribing optometrists. This chapter provides the background prior to outlining the scope of the remaining thesis.

1.1.1 Background

A background covering the evolving role of the optometrist is now presented. This section is split into three main areas as shown in figure 1.1. The section culminates in a discussion on the limited on-going support for the now increasingly broadened role of the optometrist.

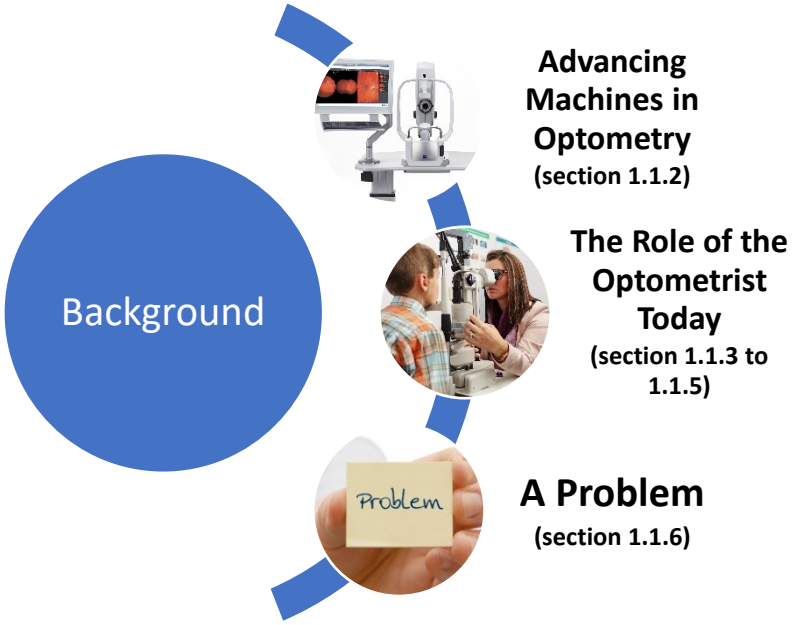


Figure 1.1 Three stages in discussing the role of optometrists

1.1.2 Advancing Machines in Optometry

1.1.2.1 Clinical Automation

Many aspects of an optometrists' examination have been automated. This includes the use of autorefractors, tonometry and topography devices, OCTs, and visual field machines. The routine evaluation of a visual field, for example, no longer necessitates a Bjerrum screen and the manual plotting of points, nor the unsupported interpretation of results. Instead, it invites a combination of state-of-the-art hardware nourished with algorithms dictating the reliability, pattern interpretation and even the likelihood and degree of future progression in the face of background noise (5,6). Increasingly, automated phoropters are being used to allow accurate refractions within a few minutes and with minimal exertion. More recently, automated refraction machines (objective and subjective) such as the Visionix® Eye Refract (Visionix USA, Luneau Technology USA Inc.) and the Chronos® (Topcon Corporation) enable a complete refraction with the acquisition of corneal curvature readings and pupillary distance measures. Although few of the latter are currently available in community practice (7), the Foresight Report, commissioned by The Optical Confederation and The College of Optometrists in 2016, predicts such automation will soon become commonplace (8).

1.1.2.2 Electronic Patient Records

The use of software in the form of patient records is also mainstream, with a 2014 study showing 80% of practices engaged (9). Patient records have historically been identified as a rich source of data (10). These data can be used to not only analyse the prevalence of disease and the effectiveness of treatments and protocols, but also to inform public health policy. Electronic patient records go a step towards making data accessible, but this data can be difficult to analyse. A recent report commissioned by The College of Optometrists identified the weakness of current electronic patient records as a lack of standardisation, i.e. most of the data is available in free text rather than organised menus and is therefore difficult to analyse (11). Consequently, the data held within patient records in optometry is largely being underutilised. Virtual (web-based) platforms such as, OptoManager® (Cegedim RX® Ltd, www.cegedimrx.co.uk) and OPERA® (www.optom-referrals.org) have been developed in an effort to streamline the connectivity between primary and secondary eye services. These platforms also automatically extract audit data for patients presenting to optometrists offering commissioned examinations. Examinations include Covid urgent eyecare services (CUES), minor eye conditions services (MECS), pre and post-cataract assessments and glaucoma referral refinement services (12). Although the data generated by these systems is mostly standardised and therefore can be analysed with ease, it is restricted in the amount of clinical information it contains. These platforms are largely used to demonstrate the financial viability of schemes to the commissioners with a basic insight to clinical investigation in the form of diagnoses and management decisions. They have been successful in doing so, and studies showing the effectiveness of the MECS and CUES schemes have been published (4,13,14).

1.1.3 The Role of the Optometrist Today and Tomorrow

Optometrists have evolved greatly since statutory recognition in the form of the first Opticians Act in 1958. Before this legislation they were viewed as refractionists (15). Appendix 1 shows a timeline of the pivotal events in the development of the profession. Readers interested in the history of optometry are directed to publications by Margaret Mitchell (15) and for a brief overview, Jeffery Stanton (16).

Gains in the professionalism and competence of optometrists have been bolstered by an increase in research and academia. The earlier detection of ocular diseases and a growing battery of non-invasive testing methods and successful interventions also bore favourably. Graduate optometrists now boast a knowledge encompassing not only refraction, methods of vision correction and

binocular vision function, but also ocular health and therapeutics. Since 1999 optometrists were allowed to manage ocular conditions within their competency and refer only where necessary (17).

1.1.3.1 The Optometrist Today

The identification of refractive error and case-finding for eye disease, following the dictate of the sight test, remain the mainstay of optometric practice. However, optometrists are increasingly involved in hospital and community-based patient pathways for the delivery of extended eyecare services (12,18,19). Community-based services are often delivered in partnership with the Hospital Eye Service (HES) and funded by Clinical Commissioning Groups (CCGs). Although many patient pathways are consultant-led, they are delivered by optometrists. Such pathways are designed to ease the burden of an aging population with more chronic visual problems necessitating ongoing long-term follow-up. An aging population, the higher sensitivity of testing methods, the availability of successful treatments to avoid ocular morbidity and of course the Covid-19 pandemic have contributed to a significant increase in the demand for eye services over the last decade (20). At over 7.9 million out-patient appointments, ophthalmology constituted the highest volume speciality in the UK during 2019/20, second only to allied health professional episodes (21). Delays to follow-up are a significant problem, the Healthcare Safety Investigation Branch (HSIB – a government funded body dedicated to improving patient safety through independent investigations into NHS-funded care across England) reported 22 patients per month suffer severe or permanent sight loss as a result of delays to follow-up (22). Capacity is becoming an increasing concern especially since it is projected that the number of people with reduced vision is set to double by 2050 in the UK (largely due to the increase in patients with chronic visual problems necessitating long-term follow-up), driving up the demand for specialist eyecare services (20). The traditional hospital-based eyecare delivery system needs to be re-evaluated.

1.1.3.2 The NHS Long Term Plan

The NHS Long Term Plan was published by the Department of Health and Social Care in January 2019. This important document outlined the plans for healthcare transformation (including that of eyecare) over the next decade. It aimed to reduce out-patient appointments by a third using measures including the transfer of care to community settings such as optical practices (23). Chapter 5 of the publication set out the process by which this is envisaged to happen, mentioning “digitally enabled care” and more specifically an NHS “Where clinicians can access and interact with patient records and care plans wherever they are, with ready access to decision support and AI, and without the administrative hassle of today. Where predictive techniques support local Integrated Care Systems to plan and optimise care for their populations. And where secure linked clinical, genomic

and other data support new medical breakthroughs and consistent quality of care" (23). The integration of AI is therefore intrinsic to the Long Term Plan. Evaluating the scope of AI, its safety, and the preparation of the workforce to understand and embrace it, will enable the aims of the Long Term Plan to materialise. That is, the long-term sustainability of the NHS.

Given the size of the NHS and local differences in service delivery, digitising existing healthcare systems is a huge task. Thus far, NHSX, an organisation dedicated to "digitise services, connect them to support integration and, through these foundations, enable service transformation" (24) has been at the forefront of the dissemination of guidance and support for Integrated Care Systems (ICS). The work has focussed on creating the digital footprint required for AI implementation and connectivity. NHSX published implementation guidance (25) complete with funding sources (26) in 2021 to facilitate the changes described in the Long Term Plan. It is expected that local ICS teams can implement the digital strategy (including developing connectivity between the providers of healthcare, including optical practices, through the electronic referral system - eERS) at a local level, servicing local needs. The final piece of guidance for implementation from NHSX clarifying "Who Does What" is expected this year (2022) (27).

1.1.3.3 The Topol Review

The Secretary of State for Health and Social Care commissioned The Topol Review which was also published in 2019. This review was led by Dr Eric Topol, a consultant cardiologist, geneticist, and digital medicine researcher. The aim was to advise the government on how to prepare the healthcare workforce to deliver a digital future. The review made recommendations that would "enable NHS staff to make the most of innovative technologies such as genomics, digital medicine, artificial intelligence and robotics to improve services" (28). The review concluded that the aforementioned technologies would not replace the clinician, but rather enhance (augment) them resulting in a higher quality personalised service, freeing up more time to "care" for the patient. The report predicted that within 20 years, 90% of all jobs in the NHS would require some element of digital skills. Since the report, and as recommended, educational institutions have started to deliver courses on data science, specifically for eyecare professionals including optometrists (29,30).

1.1.3.4 The Ophthalmology Elective Care Handbook

A further publication, namely The Ophthalmology Elective Care Handbook, was also published in 2019, as part of a series of Elective Care Speciality Handbooks. The aim was to aid a rapid transformation of services in response to the increase in demand for eyecare (20). The recommendations clearly state that the loss of sight due to delayed follow-up is a problem

necessitating innovation. This loss equated to 22 patients per month losing vision due to health service initiated delays, this being prior to the increased backlog created by Covid lockdowns (31). There is no doubt that this number is expected to increase post Covid as the eye departments struggle to catch-up with patient follow-ups (32). The delivery of eyecare continues to be re-evaluated and re-structured beyond traditional patient pathways to include allied professionals, including optometrists. Indeed, optometrists continue to show that, in addition to the traditional sight test, they can monitor chronic ocular conditions such as glaucoma with clinically comparable levels of agreement to ophthalmologists and medical clinicians (12,33–35).

1.1.3.5 The Influence of Covid-19

The Covid-19 pandemic highlighted the positive contribution optometrists, and independent prescribing (IP) optometrists specifically, could make in maintaining access to urgent eyecare services when the traditional hospital settings were unavailable (3,4). High street optometry not only provided multiple practice settings conducive to social distancing and easy patient access, but also bridged the gap in care caused by HES shutdown. The scope of practice of optometrists, therefore, has hugely broadened, and is set to broaden further with plans for the transfer of care into community settings as described above. This, and digitisation are commissioning priorities described by The Clinical Council of Eye Health Commissioning following the publication of the Long Term Plan (36). Indeed work on the electronic eye care referral system (EeRS) connecting optometry to secondary care providers and GPs is current (37). Optometrists will require support for the necessary transition and increased autonomous clinical decision making. Incorporating digital technology and AI, as advised by the Topol Review, will be pivotal to the delivery of consistently efficacious eyecare.

1.1.4 The Specialist Independent Prescribing (IP) Optometrist

Since the Opticians Act 1958 and the Medicines Act in 1968, optometrists have had access to some medications to aid refraction, diagnosis, and the management of eye conditions. The NHS plan, produced by the Department of Health (DoH) in 2000, was the first endorsement of extended prescribing to certain non-medically qualified professionals, including optometrists (38). Early courses on ocular therapeutics however (including a DOCET course in 1997 and the College of Optometrists' higher diploma in therapeutics in 1998) predate this plan. This can be considered a manifestation of the desire of optometry to widen its scope of practice. It was the Crown review in 1999 (39) that was the precursor for the NHS Plan culminating in supplementary prescribing (SP) status for optometrists, realised in 2005. This increased the range of drugs available to optometrists

when working in partnership with a medical practitioner. Additional supply (AS) was proposed as an extension to the Medicines Act (1968), this involved the use of an extended list of drugs in an approved formulary. In her review, Dr June Crown recognised that extended prescribing would capitalise on the skills and placement of community-based health professionals and result in efficient access to medicines for patients.

After a joint consultation between the Medicines and Healthcare Regulations Authority (MHRA) and the DoH in 2006, optometrist independent prescribing (IP) was proposed. In June 2007 it was decided by the Commission for Human Medicines' working group that optometrist IP activity should be restricted to areas of individual practitioner competence rather than to an approved formulary (40). The latter option had been tried and tested with nurse prescribing but proved impractical as it required the constant revision of the formulary with new drug recommendations. The necessary legislative changes to the prescription-only medication (POM) order were to follow and the GOC launched the IP register in 2009 (41).

Independent prescribing status is achieved with the completion of GOC approved postgraduate training. This includes a theoretical element and a clinical placement under the overall supervision of a consultant ophthalmologist. Traditionally, the regime of drug prescribing (type, modality and duration) is left to the prescriber, resulting in much inter-ophthalmologist variation (42,43). The clinical placement is expected to provide both a means for the practical application of the theory of medicines, and a knowledge of established expert prescribing practice. Despite being a pivotal pre-requisite to independent prescribing, remaining with one clinician or within one organisation for a short period can restrict the understanding of the vast acceptable alternatives. For optometrists with no previous experience in therapeutic prescribing the expected duration is 12 days (or 24 sessions). For those with pre-existing therapeutic qualifications this requirement is halved.

Upon qualification, IP optometrists are permitted to prescribe any licenced medicine for ocular conditions, affecting the eye and adnexa. IP optometrists are also required to register with the GOC the specialisms in which they intend to practice (41).

To maintain registration, optometrists are required to keep their skills current, audit their prescribing activity annually and complete additional specialised continued education and training activity (an additional 18 specialist CET points per cycle, recently changed to CPD).

There were 1117 IP optometrists registered with the GOC as of June 2021 (appendix 2). Figure 1.2 shows the number of new IP optometrists added to the GOC register year on year since 2009. An

increase of over six times over a 9-year period can be reasonably viewed as an indication of the future direction of the profession.

NHSE (National Health Service England) has also supported the upskilling of optometrists in England (44). The organisation has provided funding for those interested in taking more specialist qualifications such as IP and higher qualifications in glaucoma. This, again, can be taken as indicative of a want to engage optometrists in extended patient care.

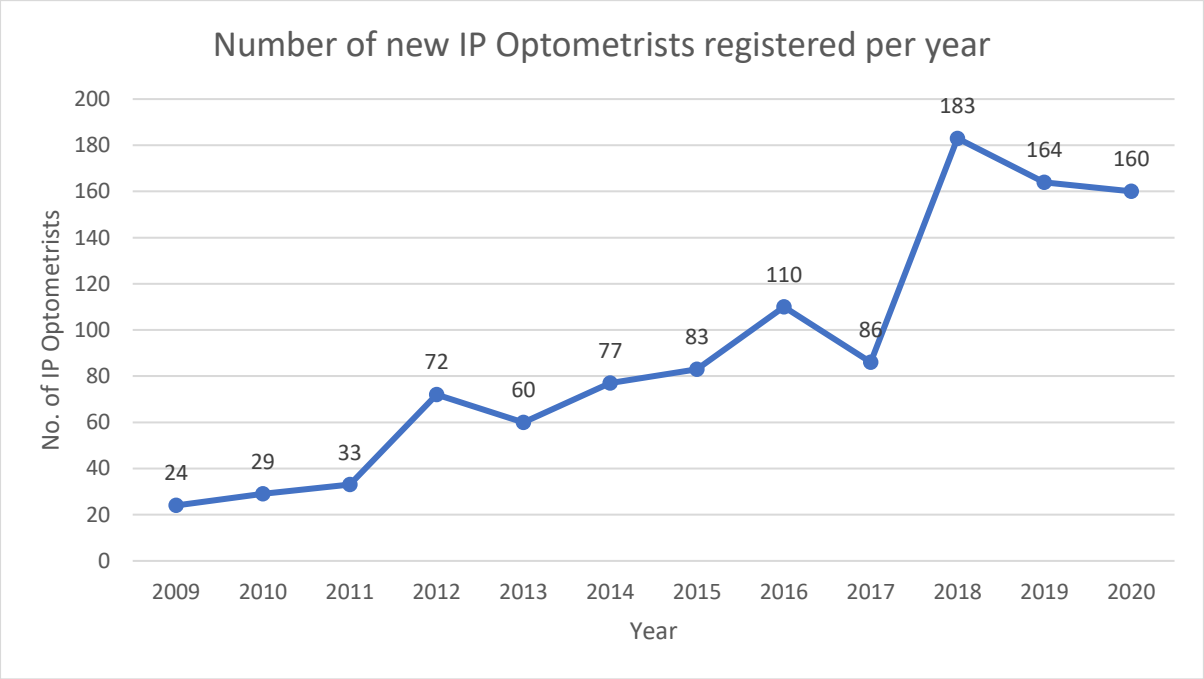


Figure 1.2 The number of new IP optometrists registered per year

1.1.5 On-going Support for IP Optometrists

Specialist IP optometrists are expected to keep abreast of the associated clinical and drug-related knowledge and current practice. In the author's experience, this can be arduous given the rapidly changing area of drug development and evidence-based treatments. The requirement of 18 additional CPD points per cycle is a means of ensuring IP optometrists engage with peers and continue to use an evidence-base when prescribing.

1.1.5.1 Prescribing Support: The Current Provision

All prescribing professionals (nurse prescribers, GPs, pharmacist, and optometrist prescribers etc.) are supported by the resources described in table 1.1.

Organisation/Provider	Support
The National Institute for Health and Care Excellence (NICE) – www.nice.org.uk	- Guidelines - best practice guidelines detailing management protocols - Clinical Knowledge Summaries (CKS) – evidence-based summaries with practical guidance - Access to the British National Formulary (BNF) via website
MIMS (A prescribing and clinical reference aimed at healthcare professionals) – www.mims.co.uk	Website includes: - Information on drugs - Drugs news - Quick reference prescribing tools - Subscription to a regular bulletin for updates regarding changes in prescribing and drug information
The Cochrane Library – www.cochranelibrary.com	- Systematic reviews - appraisals of empirical evidence forming an unbiased summary of evidence often dictating the effectiveness of treatments
Local NHS Prescribing Guidelines – available from Clinical Commissioning Groups (CCGs)	- Recommends order of medicine selection (i.e. first-line, second-line etc.) - Varied practice according to locality
Electronic Medicines Compendium (eMC) – not-for-profit organisation, funded by the pharmaceutical industry (200 subscribed companies) – www.medicines.org.uk/emc	Comprehensive and current list of UK licenced medicines available online including for each: - Summaries of Product Characteristics (SPC) - Patient information leaflets - 'Risk minimisation materials' - Medicines updates also available to browse
Medicines and Healthcare Products Regulation Authority (MHRA) – www.gov.uk/mhra	Data on medicines including: - Safety updates - Marketing authorisations - Drug-licence information - Access to Yellow Card Scheme to report adverse incidents - A Central Alerting System cascading patient alerts, important public health messages and safety critical information and guidance

Table 1.1 Resources available to support all prescribing professionals

Table 1.2 shows supporting resources specifically aimed at IP optometrists.

Organisation/Provider	Support
The College of Optometrists - www.college-optometrists.org	Provides: - Externally peer-reviewed Clinical Management Guidelines (CMGs) – include the signs, symptoms, and management options for ocular conditions seen commonly in optometric practice - College of Optometrists Formulary - IP discussion forum – encouraging peer discussions - College app which includes the CMGs and the formulary - Library, including access to online resources such as access to OpenAthens (for eye related journals only), database and reports (including Cochrane reviews) and Ebooks
Association of Optometrists – www.aop.org.uk	- Hosts therapeutics education event biannually in Manchester and London. This offers opportunities for peer review, lectures, and hands-on workshops to gain specialist CET points
Directorate of Optometric Continuing Education and Training (DOCET) – organisation to oversee the management of government funds dedicated to CET for optometrists – www.docet.info	- Produce and distribute (post and online) educational material covering topics including therapeutics - Distributed materials offer specialist CET points and news updates

Table 1.2 Supporting resources for IP optometrists specifically

1.1.5.2 The Evidence-to-Practice Gap

Evidence in the form of scientific publications has proliferated in recent decades. It is estimated that the global growth rate of scientific publications is 4% per year, and 5% for life sciences specifically. This translates to a doubling of evidence every 14 years for life sciences (45). Despite this rich and ever-growing body of evidence, optometrists (amongst other health professions) report a lack of skills, accessibility, time and resources to be able to access and assess evidence (46–48). This is despite the availability of evidence-based guidelines and systematic reviews distilling good evidence. It follows then, that there exists an evidence-to-practice gap in healthcare (49). That is, a persistent failure of best evidence to transfer to behavioural changes by clinicians. Indeed this gap has been quantified as 54.9% and 57% of patients receiving recommended care across healthcare systems in the USA and Australia respectively (50,51). Evidence relating the clinical activity of UK-based optometrists specifically against evidence-based guidelines is generally scant. An Australian study evaluating 1260 optometric patient records for compliance to evidence-based guidelines and expert

opinion across 42 optical practices, showed Australian optometrists in general fare better for the appropriateness of eyecare delivery, particularly for the management of glaucoma and diabetic eyecare (71%). Performance against guidelines, however, remained sub-optimal, especially for history taking and physical examination (52). To the authors knowledge, the activity of IP optometrists specifically has not been evaluated against guidelines, perhaps due to the relative infancy of IP practice. However it is reasonable to consider methods to minimise the evidence-to-practice gap in order to future-proof the profession.

1.1.5.3 Knowledge Translation (KT)

Knowledge translation (KT) is the process that seeks to translate best evidence into clinical practice (53). This includes producing systematic reviews and evidence-based guidelines. Implementation science is a key part of KT and examines the methods to promote the uptake of best evidence into practice (i.e. by what means can we efficiently modify clinician behaviour to mirror the evidence?). It has been established that in order for a behaviour to occur, a person needs three attributes. These are, the necessary skills (capability), motivation (will), and opportunity ("all the factors that lie outside the individual that make the behaviour possible or prompt it" (54)). These have all been identified as barriers within eyecare delivery by optometrists (48). The first attribute (the necessary skills) is addressed with education, such as the IP qualification certified by the College of Optometrists or guidelines such as the CMGs. The cost and time taken for training however is a known concern (55). For the second attribute, although optometrists appear willing to partake in further training, motivation is hindered by a lack of remuneration, a lack of community ophthalmology pathways (incorporating an NHS prescribing pad in England – FP10) and fear of litigation (55,56). The final attribute (opportunity) is particularly important as regardless of the level of training and motivation, a hostile environment stifles easy enablement, this being key to modifying clinician behaviour. Any method by which evidence is incorporated into clinical practice (perhaps as a prompt) must therefore fit seamlessly into practice without compromising the already stringent consultation times.

1.1.6 The Problems

It is evident that there are a plethora of reliable sources of information and guidance available to IP optometrists. However, in the author's opinion, there are three inherent weaknesses in those mentioned above. These relate to the implementation of the evidence.

Firstly, all the current resources require the clinician to access and work through recommendations or education. Practically, this necessitates study outside of clinic hours which is subsequently applied

to practice. This format limits the usefulness of such resources at the 'coal face'. Clinicians often lack the time during consultations to call-upon and sift through such resources despite the means of access being simple (phones, tablets etc). These resources also require a firm diagnosis to dictate treatment (such as for the CMGs), which can be problematic with atypical or ambiguous presentations, where guidance is most useful.

Secondly, many of the resources provide excellently refined research translated into recommendations that are not updated "real time". As such they force the clinician to choose between review quality and currency. Twenty-three percent of systematic reviews, for example, can require updating only 2 years post-publication, and 7% by the time they are published (57). Although there are obvious advantages of "live" systematic reviews, they too are limited to known diagnoses and are not yet mainstream (58). This means any proposed change in recommended practice incurs a delay before it can be implemented due to processing.

Lastly, although clinical guidelines, as expected, are based on highest level of evidence (clinical trials), they ignore the range of normative peer practice – i.e. what other specialist IP optometrists would do with a patient like the one being attended to (that is, clinical expertise). This last provision, although at the bottom of the pyramid of evidence, is invaluable in the practical training of prescribing professionals (in areas where higher levels of evidence is lacking), particularly as evidence-based practice is the process of combining the best available evidence with *expertise of practitioners and the individual needs of the patient* (59). Consider the training of ophthalmologists for example. After medics have completed their two foundation years, they then embark upon a further seven years of training under the supervision of a number of trained ophthalmologists (60). The vast part of training is dependent upon the interaction between the trainee (armed with theoretical knowledge) and the experienced senior (the practical application of the knowledge). The experienced ophthalmologists, although not present at individual examinations, are always within reach of the trainee as they move through the seven years. Therapeutic intervention for ambiguous signs and symptoms can therefore be checked with an experienced prescribing clinician as required.

In contrast, other than the 12 practical days required for qualification, IP optometrists lack this "real time" peer support. In the author's opinion, although the duration of the initial placement is sufficient, the key is the ongoing practical support for IP optometrists working in isolation other than reference to potentially outdated guidelines bereft of specifics (a form arguably at odds with the dictates of KT and implementation science). Many optometrists practice in isolation and a set-up such as that in the HES is impossible. Yet the work of a specialist IP optometrist can often mirror that of a trainee ophthalmologist. Upon examining the College of Optometrists' IP discussion forum (Sept

2021), a platform exclusively available to IP qualified College members, the author observed that almost 20% of the 219 posts concern IP optometrists seeking peer support in prescribing decisions for specific patients. These posts frequently call upon the more experienced IP optometrists' advice. This indicates a clear need for "real time" support.

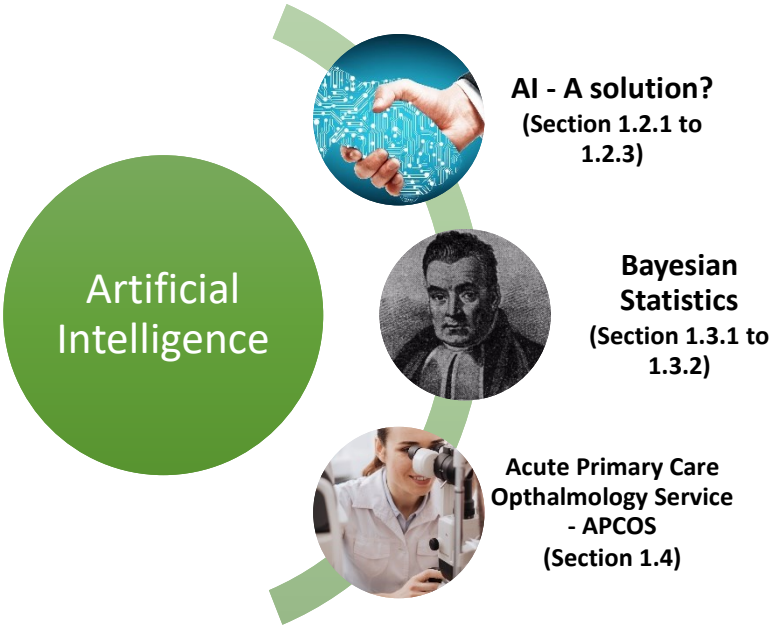


Figure 1.3 The three stages of discussing artificial intelligence and its role in eyecare

1.2 Artificial Intelligence (AI)

Figure 1.3 shows an overview of the following section. The nature of artificial intelligence is briefly discussed. Terms such as 'big data' and 'machine learning' are clarified, and the current research reviewed. An explanation of Bayesian statistics and how this has been applied in previous research is explained before finally the Acute Primary Care Ophthalmology Service (APCOS) is introduced as the data-collectors.

1.2.1 Artificial Intelligence – A solution?

The term artificial intelligence or AI often conjures up images of futuristic terminator-style robots intent on exterminating humankind. However, in reality, the precursors to AI, data acquisition and machine interpretation, have been in constant (non-threatening) use all around us. Optometrists are habitual users of machine interpretation, that is, as mentioned in section 1.1.1 in forms including IOP measurements, visual field printouts, topography maps and OCT images. These systems take raw data (for example, the amount of force needed to indent the cornea, or the reflection of light from structures within the retina), apply a rule or rules to it (algorithms – step by step instructions) and

provide an interpretation (readings or images). AI is the next logical step in refining the use of these machines. AI has been nicely defined by Kaplan and Haenlein (2019) as “a system’s ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation” (1). Figure 1.4 shows this process visually, using the example of coming to a diagnosis. The internet of things (IoT) and other big data are sources of raw data into this system. The IoT describes devices around us possessing sensors and software that enables the exchange of information between them. This applies to systems such as those used by Apple Siri® (61), Google Home® (62) and Amazon Alexa® (63). These devices use voice recognition to interpret a command and can interact with compatible devices such as electronic blinds, lighting systems or the internet to give a response. The IoT therefore generates a lot of data on individual preferences and thus is a source of big data. Patient records are also a source of big data available to optometrists. These have the potential to form a basis of AI learning applied in optometry. This concept is discussed further below and forms the crux of the current study. AI uses big data sources to identify patterns or correlations within the given population. Stacks of individual algorithms are applied to the data, simultaneously and/or in sequence, identifying common characteristics and effectively sectioning or labelling the data. The extracted patterns are then used to refine responses (machine learning). Thus, machines or computers can learn without specifically being programmed. Moreover, where there is a constant flow of new data, this translates as a continual input of information from which correlations and trends can be revised and responses refined (1,64).

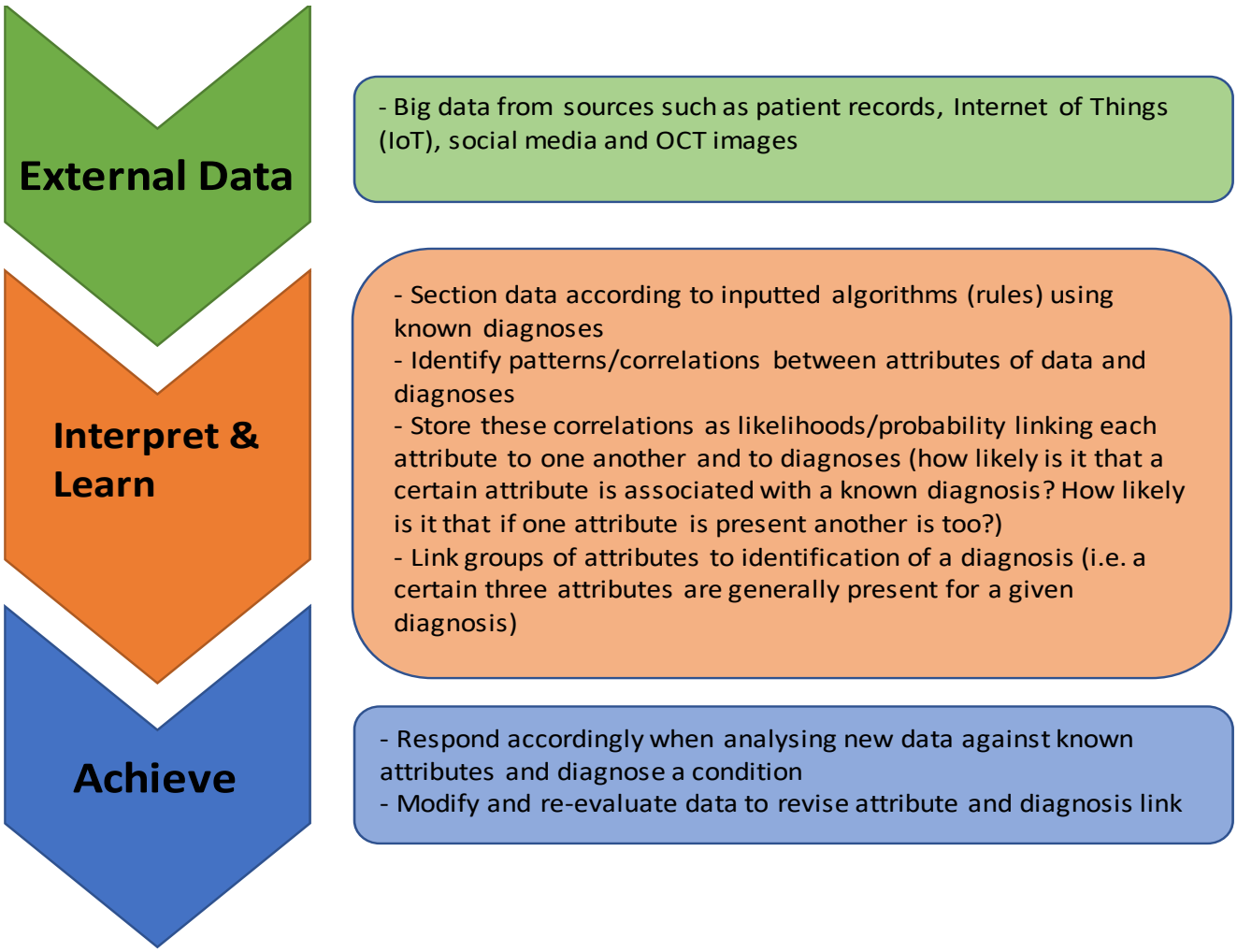


Figure 1.4 The development of AI based on the definition of Kaplan and Haenlein (2019) (1) and using the example of disease identification.

1.2.2 What is Big Data?

Many definitions for big data exist. Table 1.3 shows some of the definitions and contexts given in the literature. In summary, big data is a buzz phrase vastly used within information technology and increasingly in medical literature. Big data describes a large amount of information in various forms from which trends and correlations can be extracted. An example of this is data held within patient records. The value of data held within patient records was demonstrated by Florence Nightingale during the Crimean war and showed the beneficial effect of sanitation on death rates (10). However, due to the vast amount and unstandardized form of the data (not organised in set groups using standard terms but more free text), analysis has been historically difficult. Despite methods such as Bayesian statistics (Bayes'), which would benefit from the extracted probabilities to predict future

events, the sheer amount and variation in the form of the same data to be processed proved the main obstacle (65).

1.2.2.1 Increasing Processing Power and Data Storage

The development of central processing units (CPU) in computers changed the dynamics of data analysis. For the first time millions of calculations could be carried out per second syphoning the exertion needed to analyse great swathes of data (66). Big data, previously thought of as too large or complex for traditional processing systems to analyse, now became accessible. With the advent of advanced processing systems such as graphic processing units (GPU) and accelerators, the possible calculations rose from millions per second to billions (67). Not only this but information could now also be extracted from graphics into a computer readable format. The uptake of powerful processing systems has been impressive. Hilbert and Lopez report an increase of 83% in the world's technological capacity to compute information between 1986 and 2007 (68). To put this into context the authors note that $6.4x10^{18}$ instructions per second were carried out on general purpose computers in 2007, this being around the same as the maximum number of nerve impulses generated by the human brain. They also report an increase of 61-times in individual storage capacity during the same period. This number does not account for cloud storage, which enables the sharing of huge data files. The processing and storage capability of the world is therefore increasing exponentially in response to the ever-increasing load of available data.

Despite the increase in the processing and storage of big data, the lack of structure in patient records has remained an obstacle. Structured, or standardised patient records ensure all inputted data are conducive to computer analysis by eliminating the variation in terminology. Free text descriptors of clinical signs and diagnoses require much more sophisticated methods of data extraction prior to standardisation and analysis.

The importance of planning for a digital future has been highlighted in the Topol review (section 1.1.3.3) where it is expected that digital record keeping will form the bedrock of the NHS (69). Big data available from the digitisation of patient information will increase the potential for AI involvement moving towards a more efficient delivery of healthcare, and thus is actively encouraged. Aside from this, the increasing use of electronic patient records both in community practices and within the HES (MediSight for ophthalmology – www.medisoft.co.uk/medisight) mean that the problem of unstructured data may well be on the way to being solved. Clinicians, including optometrists must therefore understand data science and be able to allow future systems to augment their decisions. This includes being able to identify when the machines are wrong.

Source	Definition
Oxford Dictionary (70)	Extremely large data sets that may be analysed computationally to reveal patterns and associations, especially relating to human behaviour and interactions
Wong et al (2019) (71)	Data sets that are so massive or complex in nature that traditional data processing methods and methodology are inadequate to effectively analyse it… Big data usually exhibits properties of high volume (in terms of amount of data), high velocity (in terms of the speed of data in and out), high variety (in terms of the range of data types and sources) and high veracity (in terms of accuracy and correctness)
Susskind and Susskind (2017) (2)	The discipline that has emerged to specialise in the capture and analysis of information. The authors state "Big data" was initially confined to "techniques for the handling of vast bodies of data…now big data is used to refer to the use of technology to analyse much smaller bodies of information"
Gordon (2013) (72)	Data defined by some combination of the following five characteristics: 1) Volume (significantly large) 2) Variety (structured/semi-structured) 3) Velocity 4) Value (perceived or quantifiable benefit) 5) Veracity (correctness can be assessed)
Snijders at al (2014) (73)	Data sets so large and complex that they become awkward to work with using standard statistical software
De Mauro et al (2016) (74)	Big data is the information asset characterised by such a high volume, velocity, and variety to require specific technology and analytical methods for its transformation into value
Boyd and Crawford (2012) (75)	A cultural, technological, and scholarly phenomenon that rests on the interplay of: 1) Technology – Maximising computational power and algorithmic accuracy to gather, analyse, link, and compare large data sets 2) Analysis – Drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims 3) Mythology – The widespread belief that large datasets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy

Table 1.3 Definitions of big data given in the literature

1.2.3 How is Big Data, Machine Learning and AI Currently Being Researched in Eyecare?

There has been much interest and research in the application of AI and machine learning in eye care in recent years. Table 1.4 shows some of the current areas of application found in the research. AI

research covers clinical areas such as vision and refractive care, the cornea and ocular surface, the retina and image interpretation, glaucoma specifically, and neuro-ophthalmology (67). AI and big data from patient records and scans have been used with an aim to develop of decision-support tools. Current areas of research include the detection of abnormalities such as dry eye (76), congenital cataracts (77), uveitis (78) and diabetic retinopathy (79). These studies have shown that AI based decision replication (used to drive decision support) shows a high degree of agreement when compared to established specialist diagnoses. They have the potential therefore to recommend aetiologies, ancillary tests, and intervention. Although such research remains in its infancy, that is, largely limited to the demonstration of a 'proof-of-concept' using one machine learning method, the idea of a cloud based collaborative network, harnessing clinical data (thermographic images) with view to the development of computer-aided detection (CAD) of dry eye has been described (80). Perhaps one of the most publicised studies highlighting the opportunity AI presents to eyecare, was the interpretation of macula OCT scans at Moorfields Eye Hospital (81). This study, in partnership with Deepmind© (Deepmind technologies Ltd. www.deepmind.com), describes the use of a deep learning 'framework' (neural networks) able to interpret OCT scans with expert level agreement and an error rate of 5.5%. Not only does this framework deliver a diagnosis and suggest management, but also provides an 'explanation' on which the diagnosis is based, all accessible to the clinician. It does this by segmenting the scans and quantifying the retinal morphology. This produces measurements for the location and volume of pathologies such as macula oedema and a fibrovascular pigment epithelium detachment. This segmentation or explanation can not only be used to support established clinician decision-making but can also aid the training of health professionals. This system demonstrates a very advanced level of AI application dependant on specially built software trained on 14,884 scans. A number surprisingly small when compared to the 128,175 images used by Gulshan et al. (2016) for diabetic retinopathy detection. However, it firmly remains within the big data realm (79).

When combined with robotics as the output, AI has the potential to be truly revolutionary (82). Though in infancy, robotics have been developed which have undertaken the first retinal surgeries in humans (83). The surgeries undertaken were the removal of epiretinal membranes, the inner limiting membrane, or the placement of sub-retinal injections in AMD patients. Although these procedures took longer and still relied upon surgeon input (i.e. were not carried out completely independently) they represent the building blocks for future, more advanced and perhaps automated systems.

Authors	Nature of study
Situmorang et al. (2017) (84)	The establishment of a final contact lens prescription following consideration of refractive variables (including tear production, lenticular powers, and corneal curvature). This system learned from 30 pieces of sample data and displayed 100% accuracy with a testing data set.
Sudarshan et al. (2017) (76)	Machine learning is used to extract features of dry eye disease from IR thermography images of the ocular surface. Specifically, the inferior aspect only. The resultant system can aid with the diagnosis of dry eye and demonstrated mean accuracies of over 90% for each eye.
Long et at. (2017) (77)	AI deep learning and neural networks were used to identify, diagnose, and suggest treatment decisions for congenital cataracts. The system (named the CC-Cruiser) was assessed against ophthalmologist decisions and demonstrated comparable results.
González-López et al. (2016) (78)	A Bayesian network was used to identify the most probable aetiology of anterior uveitis. After learning on a set of 200 cases it was tested on 210 more. In 63.8% of cases the most probable aetiology given matched the diagnosis by a senior clinician. In 80.5% of cases the clinical diagnosis matched either the first or second most probable aetiologies given by the network.
Raghavendra et al. (2018) (85)	The development and testing of an expert system for glaucoma identification. 1000 fundus images (corrected for illumination and colour differences) were used to extract 30 significant features pertaining to glaucoma identification (machine learning from a large dataset). A maximum accuracy of 93.62%, sensitivity of 87.5% and specificity of 98.43% were achieved with 26 of these identified features when tested on a public dataset. This system is therefore useful in aiding the diagnosis of glaucoma.
Gulshan et al. (2016) (79)	The development and testing of a deep learning algorithm for detecting referable diabetic retinopathy and macula oedema in fundus images. The system was trained on 128,175 images. The system managed to demonstrate the highest sensitivity at 97.5% with a specificity at that point of 93.4%. At the highest specificity point (98.5%) the sensitivity dropped to 87%. Use of this algorithm in the clinical setting could improve patient care and outcomes, thought the authors note more research is needed.

Table 1.4 Some studies using big data, AI, and machine learning in eyecare and vision

The majority of AI related research in eyecare focusses on machine learning applied to specific areas of ocular health (table 1.4). Decision support using enhanced analysis of digital images, although extremely useful, is still limited to exactly that – image interpretation. Most studies do not collate all the information obtained during a consultation (i.e. history and symptoms, family history and medication history) to reach a diagnosis and management plan and yet in the real-world this is what optometrists do. As such current research is more ophthalmology-led than optometry-led mirroring its various sub-specialities.

1.3 Bayesian Statistics

The following section describes the concept of Bayesian statistics (Bayes), previous work in eyecare using Bayes' and how it was applied in the present study.

1.3.1 Bayes' Theorem and Eyecare

Named after its proposer the Reverend Thomas Bayes, Bayes' Theorem has a long history spanning over 250 years. In 1763 details of the theorem were first published as part of a letter submitted by his friend Richard Price, this after the death of Thomas Bayes (86). The essay addressed the problem of the unknown, that is, how can we discover the probability of an event given no prior knowledge? For Thomas Bayes the answer was to use equal priors for all the possibilities to begin with – in other words the probability of each result is initially equal. However, the crux of the rule is that the initial probabilities are updated with new information. This forms the basis of reinforced decision making. Put simply "by updating our initial belief about something with objective new information, we get a new and improved belief" (65). Pierre Simon Laplace, a well-known French mathematician became the largest contributor to the development of Thomas Bayes' original theorem, having arguably produced it independently, and spent over 40 years refining and making sense of it. His famous "Mémoire sur la probabilité des causes par les évènemens" (Memoire of the probability of causes given events) was published in 1774 and remains a seminal work for statisticians (87).

Opinion on the soundness of Bayes' theorem has a rich history of contention and conquests (88). Interested readers are directed to the publication by Sharon Bertsch-McGrayne entitled "The theory that would not die. How Bayes' rule cracked the enigma code, hunted down Russian submarines & emerged triumphant from two centuries of controversy" (65). Therewith readers can follow the history of Bayes' theorem from inception until current times detailing the reasons for contention and the opposing frequentist theory. The frequentist theory forms the basis of the types of statistical tests frequently taught in schools, colleges, and universities, and involves the much-known p-values.

Bayes' probability describes the quantification of a reasonable expectation or belief that an event may occur. These expectations or beliefs can originate from previous experiments or personal experience. As more evidence is gathered, this belief may be altered and updated. Arguably this mirrors the process we may associate with human reasoning. That is, our expectation of an ocular diagnosis depends upon our prior belief of its presence given what we know from experience and study (the gut feeling or prior). Subsequent testing and evaluation (new evidence) alters and updates our degree of belief in the diagnosis (the final post-test probability). The opposite to this is that all our beliefs are limited to 'proven facts' only. That is, the results of random sampling taken

with known properties of a population with no quantification of experience (89). This latter interpretation of probability is regarded frequentist and forms the basis of most clinical trials where p-values, α and β levels, and standard deviations determine the status of the tested hypothesis.

Bayes' theorem, expressed as the formula below, enables the necessary computation. It is based on conditional probability (i.e. the probability that arises given an observation).

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Here, the '|' is used for conditional probability estimates. So, the conditional probability P (A|B) means the probability of A (e.g. the presence of a disease) given (or conditional upon) the presence of B (e.g. a positive test result or observation). It is important to understand that conditional probability estimates ultimately confer on Bayes' the ability to contribute to decision support that suits the presentation of each individual (i.e. individualised optometry); so, one shoe does not have to fit all. The probability, - P(B|A), also conditional, means the probability of B given A (i.e. the probability of a positive test result given the disease is present). P(A) is the probability of having a disease and P(B) is the probability of a positive test result. The application of Bayes' using the data collected in the present study is explained further in chapter 3.

Aspinall and Hill provided optometrists with a description of the application of Bayes' over thirty years ago (90). Many other papers, including Thomas et al. (2011), have provided similar descriptions for ophthalmologists (91). Publications like these show Bayes' is not mathematically taxing.

Bayes', however, assumes:

- The independence of test outcomes – that is, the presence of conjunctival hyperaemia and a reduced tear break up time in a patient with dry eye are completely unrelated
- The independence of diagnoses – meaning that Bayes is not conducive to co-morbidity, even where the diagnoses are clearly linked such as blepharitis and dry eye

The assumption of test outcome independence has minimal effect on the outcomes of the complete analysis (92). The co-morbidity problem can be overcome by using heuristic methods once a Bayes' analysis is complete. This would work by ranking the post-test probability outcomes for the clinician

where more than one diagnosis exceeds a set threshold (93). An informed clinical judgement could then be made to decide (given the information) the presence of a co-morbidity or otherwise.

Bayes' is commonly used in the interpretation of clinical tests (88). It has also been used extensively in eyecare focussed research. Appendix 3 shows a list of studies in chronological order. The research ranges from computer assisted diagnosis in ophthalmology (94), the lifetime prevalence of uveal melanoma (95) and the role of cataract extraction in IOP reduction for glaucoma patients (96) .

1.3.2 Previous Work at Aston

At Aston University, Sagar (2014) investigated the accuracy of a diagnostic support system based on a modified form of Bayes' applied to differential diagnoses in primary care optometry (97). She extracted 1422 records from her practice in Dar es salaam, Tanzania. She analysed how the accuracy of this system was influenced by circularity (testing the system on the data it was trained on), co-morbidity, prevalence and presentation variation of ocular conditions and Chi-square filtering (to filter out signs/symptoms that were weakly associated with diagnoses). Her preliminary analysis of 10 diagnoses and 15 tests revealed Bayes' could be applied with 100% accuracy for cases without presentation variation (i.e. 'textbook' presentations) but dropped to 94% where cases differed in their presentation (i.e. atypical presentations). Circularity had the effect of artificially elevating accuracy by only 0.5%. Equally, Chi-square filtering only increased accuracy by 0.4%. When the system was applied to all her data (105 clinical tests and 35 diagnoses) the accuracy dropped to 72%. This drop in accuracy related most to those diagnoses where the prevalence was low and where the comorbidity and presentation variation was high. This highlighted a lack of both robust diagnostic signs and the need for sufficient amounts of training data for rare conditions. A limitation of her study was that only positive test findings were recorded during data-collection. As clinicians well know, the absence of a clinical sign can be just as diagnostic as its presence. The lack of this information could therefore have contributed to the reduced level of accuracy.

Following on from this, Gurney (2017) studied the accuracy of naïve Bayes' applied to the referral refinement decision making of specialist IP optometrists assessing for chronic open angle glaucoma (COAG) (98). He used a highly structured standard operating procedure (SOP) designed to mitigate the effects of inadequate clinical data and to increase accuracy. A 95% accuracy was achieved when 1006 patient records were analysed. The author concluded that a false discharge rate of 3.4% and a false referral rate of 3.1% meant that the naïve Bayes could not safely predict the decisions of specialist optometrists. An accuracy of 95%, though not deemed safe by Gurney (2017), is in fact comparable to the results of more sophisticated AI technologies such as that used by De Fauw et al. (2018) at Deepmind© (81). Other interesting findings included the speed at which naïve Bayes'

learned, reaching a maximum accuracy on as few as 69 cases. It has been established that the speed of learning is greatly affected by the complexity of data involved (99). That is, the simpler the data (the fewer number of variables) the quicker naïve Bayes' achieves maximum accuracy. The present study included many more variables than those used by Gurney (Chapter 2, section 2.3). It was therefore expected that significantly more cases would be required to achieve maximum accuracy. For Gurney (2017), Bayes' was least able to predict follow-up patients (as opposed to those discharged or referred). Follow-up was the rarest of all outcomes, and so Gurney's findings perhaps mirrored those of Sagar (2014), where a low prevalence contributed to a reduced accuracy (97).

Following on from these studies, the present study further explored the great potential of AI in IP optometry. Bayes' based machine learning was applied to the clinical decision-making of experienced specialist IP optometrists. The resulting AI, it is suggested, may then have the potential to disseminate best practice throughout the profession. In answer to both the Department of Health publications (The NHS Long-term Plan and the Topol Review – see section 1.1.3), a digital platform was developed (chapter 5) supporting not only the clinical decision-making of specialist IP optometrists, but also providing a teaching aid, allowing access to AI and machine learning for trainee IP optometrists. This "live" digital platform was based on the practice of experienced peers (section 1.4).

1.4 The Acute Primary Care Ophthalmology Service (APCOS)

Established in 2011 the Acute primary Care Ophthalmology Service (APCOS) describes a community optometry pathway designed, managed, and delivered by a team of specialist IP optometrists. This team manage all referrals from local optometrists, pharmacists and 62 GP surgeries concerning acute eye conditions (100). The APCOS specialists review and examine the patients referred and manage as appropriate. Management includes therapeutic intervention, minor procedures, follow-up, and referral to a community ophthalmology team, the GP or the HES (hospital eye services).

The system is unique in that it is completely optometrist led with no ophthalmological oversight or supervision. This pathway is commissioned by West Kent CCG, the largest of the seven Kent CCGs, and covers a population of 476,223 (March 2015) (101).

The advantages of the APCOS are three-fold. Firstly, the re-direction of patients to community-based specialists relieves the pressure on the HES (as described in section 1.1.2). This streamlining of referrals ensures only those who genuinely need consultant review are referred to the HES. This, in-turn, works to free up hospital capacity for those acute and follow-up patients where sight loss is a

genuine threat. Secondly, the patients avoid the need to attend the HES and instead are reviewed locally. This is particularly important as many patients are elderly and require family support to attend hospital appointments which could take hours in waiting times. The loss to the economy through lost working days, the onerousness of attending a hospital (including chronic parking issues) and the delay in receiving care are all avoided. Lastly, the APCOS team takes strides in establishing a multidisciplinary approach to patient care. Optometrists, medics, and ophthalmologists work in concert to establish adequate management plans for patients, capitalising on their individual skills. A greater understanding of the competencies of each profession enables inter-professional dialogue to flourish. For IP optometrists specifically, the ongoing practice of prescribing and the variety of cases reviewed are indispensable for professional development.

The data collection for this study was conducted by four members of the APCOS team. All these members are IP qualified optometrists and hold further qualifications such as the College diplomas in glaucoma and ocular conditions. It is therefore reasonable to conclude that the APCOS members constitute highly trained practitioners from whom meaningful data can be collected.

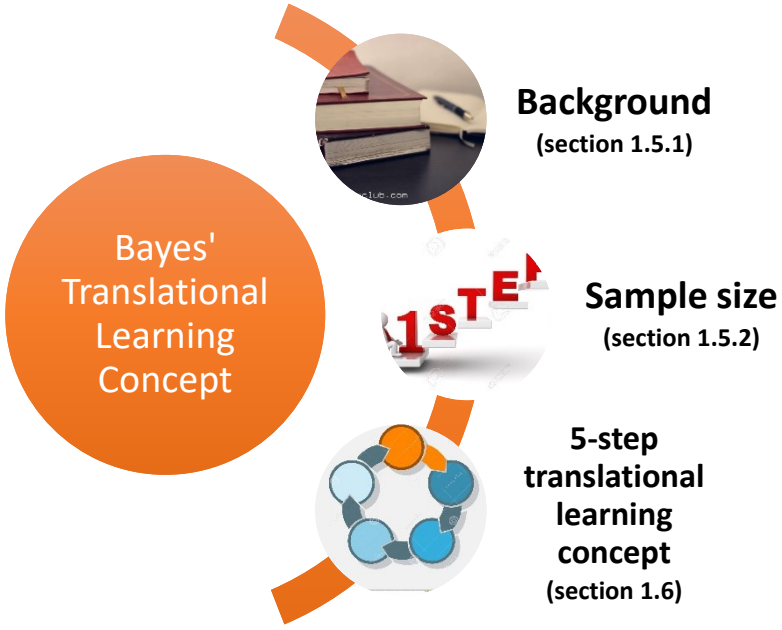


Figure 1.5 The plan for the present study is discussed in the three stages as shown

1.5 Scope and Chapter Plan of Present Study: Bayes' Translational Learning Concept (bTLC)

The following section outlines the scope of the present study. This section covers the background considerations, the preliminary step of producing a data-collection electronic patient record

followed by the introduction of the Bayes' translational learning cycle and is summarised in figure 1.5.

1.5.1 Background

The aims of the present study were to, a) apply machine learning to the clinical activity of IP optometrists and b) to develop the concept of interactive and evolving evidence-based support systems for IP optometrists and those training to become IP registered. The evidence base comprised of fully anonymised results of patient consultations (episodes). This did not involve alteration of routine clinical procedures, assessment, and management. Ethical approval was granted by the Aston Life and Health Sciences Research Ethics Committee (Project: #495 – see appendix 4).

1.5.2 Sample size

Sample size calculations determine the size of a sample of data required to make inferences which can be applied to populations (102). As such these calculations are usually a pre-requisite to defining the scope of data collection required to reveal statistically significant findings. In the present study, however, the data collected was composed of the entire population of patients examined by the APCOS members over a one-year period. A sample therefore was not taken, and this study took the form of 'big data' collection and analysis (section 1.2.2). Also, Bayesian statistical theory (i.e. conditional probability), as opposed to frequentist probability, was used to analyse the data. This meant that the alpha and beta levels required to determine whether a finding is statistically significant in frequentist probability methods, no longer applied (section 1.3.1). This is not to say that the sample size needed to demonstrate the competency of Bayes' machine learning was not evaluated. Indeed, the impact of sample size was part of the learning concept described in the following sections.

1.6 5-step Bayes' Translational Learning Concept (bTLC)

The five steps that form the bTLC are shown in figure 1.6.

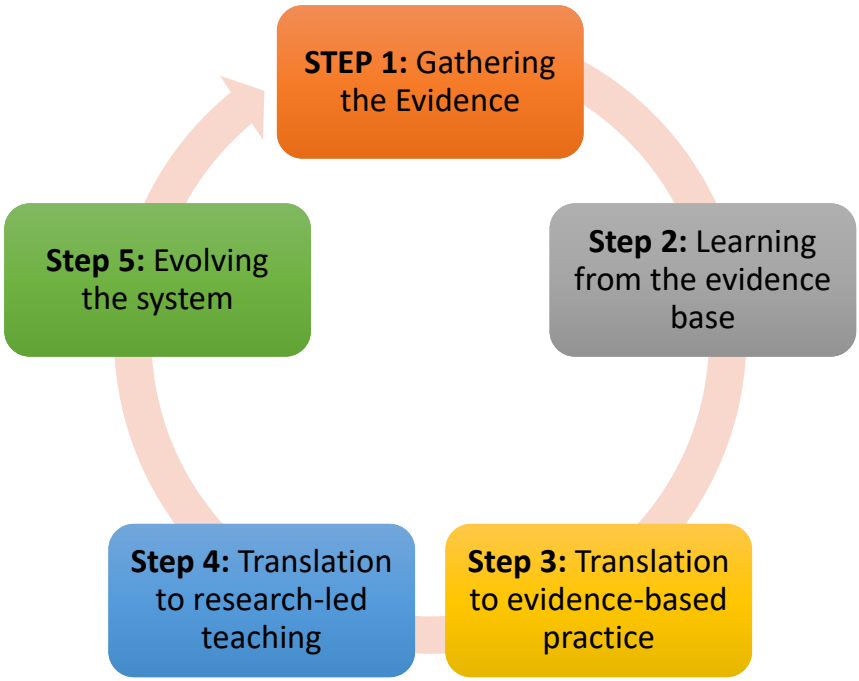


Figure 1.6 The five steps forming the bTLC

The bTLC describes the process of applying Bayes' to clinical data to drive an 'intelligent' decision support system or digital learning platform (MyDLP). This digital learning platform includes an 'intelligent' electronic patient record (iEPR) and an 'intelligent' virtual patient tool (iVPT). MyDLP supports differential diagnoses and therapeutic management decisions for IP optometrists and those in training. The translational element of the bTLC process allows the use of clinical data to inform evidence -based practice and research-led teaching through MyDLP. The bTLC also has a cyclical element that involves system evaluation. The 5 steps of the bTLC are now briefly introduced.

1.6.1 **Step 1:** Gathering the evidence base

This stage involved the collection of the evidence base by the IP optometrists serving the APCOS. These optometrists were consulted in the design of an electronic practice recording (EPR) system (chapter 2, section 2.7.2) that was largely based on the College of Optometrists' Clinical Management Guidelines (section 1.1.5.1, table 1.2) (103). The EPR took the form of a Microsoft Excel® spreadsheet with drop-down boxes designed to capture clinical data, tentative diagnoses, and therapeutic management (chapter 2, section 2.5.3). Use of the ubiquitous Microsoft Excel® database meant that future researchers could replicate bTLC for their own purposes. Another important feature of the EPR was that it converted the clinical data into an encrypted code so that data transfer could occur via email without any threat to patient confidentiality. A detailed account of the development of the EPR is provided in chapter 2.

1.6.2 **Step 2:** Learning from the evidence base

The data collected in step 1 was subjected to Bayes' with varying granulation (chapters 3, 6 and 7). Bayes' was applied in this step to learn from the evidence base established in step 1. This also made use of a Microsoft Excel® database. The encrypted EPR data was decoded and used to generate the information required for MyDLP. A key feature of this database was that it instantly and automatically updated with every new data entry. Bayes' was carried out as described by Sagar (2014) (97) and Gurney (2017) (98) with the addition of confidence limits that were new to this project. A full description of this step is provided in chapter 3.

1.6.3 **Step 3:** Translation to evidence-based practice

This step involved building a Microsoft Excel® based 'intelligent' decision support tool (iEPR – part of MyDLP) using the information generated in Step 2. The iEPR was designed to augment clinical decisions by recommending clinical tests (based on history and symptoms) before offering differential diagnoses and therapeutic management options. These recommendations were accompanied by confidence limits. In other words, the iEPR can provide live peer support with quality indicators. The iEPR can also internally judge its own competence based on the numbers of cases seen; for common diseases the confidence intervals would be small, for rare diseases, large. The entire concept of an iEPR was new to this project and this step is reported in chapter 5 and demonstrated in chapters 6 and 7.

1.6.4 **Step 4:** Translation to research-led teaching

This step was designed to demonstrate how Bayes' could be deployed in virtual patient tuition. An 'intelligent' Virtual Patient Tool (iVPT within MyDLP) allows IP students or practitioners on CET courses to explore typical presentations, the most relevant diagnostic tests and therapeutic management options for specified ocular conditions. It also allows the exploration of the relationship between clinical findings, differential diagnoses, and therapeutic management options. The entire concept of the teaching aspect of MyDLP (the iVPT) was new to this project and is reported in chapters 5, and again demonstrated in chapters 6 and 7.

1.6.5 **Step 5:** Evolving the system

This step was designed to demonstrate how MyDLP could evolve with new data. It involved evaluation of the inherent errors and learning efficiency of Bayes', as previously described by Gurney (2017). This involved the use of informedness and markedness, as described by Powers (2011), to judge the accuracy of the system (104). Step 5 provided the means of continually evaluating the performance of MyDLP as more data is added to the system. Thus, MyDLP is constantly updating and

evolving. The evaluation system is incorporated within MyDLP. The cyclical process in step 5 was also new to this project and is reported in chapters 6 and 7.

1.7 Chapter summary

This chapter has presented an overview of the development of optometry as a profession. It has highlighted a move towards increased autonomy when managing ocular conditions. Independent prescribing rights, the NHS Long Term Plan, and indeed the influence of the Covid-19 pandemic have worked to increase the scope of clinical practice for optometrists which will ease the pressures of stretched hospital eye services. Whilst this seems natural against the backdrop of an aging population with more chronic eye conditions, optometrists often find themselves practising in isolation with limited peer support. Support from experienced colleagues, such as that available to trainee ophthalmologists in the hospital eye services does not exist amongst specialist IP optometrists practising in the community. The advantages "live" peer support brings in improving clinical examination, interpretation, and patient management, especially involving therapeutics, is therefore unavailable.

Artificial intelligence, an area exhibiting unprecedented progress and availability has the potential to revolutionise the dissemination of peer support. Previous studies have shown that Bayes' based machine learning can replicate the clinical decision-making of optometrists to a high level of accuracy (section 1.3.2). The present study builds upon the previous research in this area by applying Bayes' to the clinical decisions of specialist IP optometrists. Further, it goes on to describe and develop research-based support systems applying AI and machine learning. Therapeutics is pivotal to the future development of the profession, and with these AI tools we have the potential to offer a significant leap towards professional support and development. They also align the teaching of therapeutic optometry with current practice whilst introducing future clinicians to the workings of data science.

The Bayes' Translational Learning Concept (bTLC) has been introduced as a process for the transfer of knowledge from experienced specialist IP optometrists to both peers and trainees. This is culminated with the development of MyDLP, a digital learning platform that acts as the conduit, delivering this peer support to IP optometrists and trainees. MyDLP consists of an 'intelligent' EPR and VPT alongside a 'live' system evaluation measure.

A 5-step process was described. Step 1 is to gather the evidence base. This involved the design of a data-collection EPR (chapter 2), and an evaluation of the quantity and quality of the data collected

(chapter 4). Step 2 is learning from the evidence base and is described in chapter 3. Step 3, translation to evidence-based practice, is reported in chapter 5 and demonstrated in chapters 6 and 7. Step 4, also described in chapter 5, shows the potential to extend this to research-led teaching and is demonstrated in chapters 6 and 7. The evaluation of learning efficiency (step 5) is presented alongside steps 3 and 4 in the chapters 5, 6 and 7. Finally, chapter 9 concludes the thesis by summarising the findings and makes recommendations for further areas of research.

Chapter 2: Development of an Electronic Patient Record (EPR) for Gathering the Evidence

2.1 Introduction

The Bayesian Transitional Learning Cycle (bTLC) describes the five-step evolutionary process of using clinical data to inform further clinical decision making. The steps were briefly described in chapter one. A preliminary step prior to embarking upon the bTLC involved the development of an electronic patient record (EPR). This fundamental step ensured only the most relevant data was collected, anonymised and securely transferred for analysis. After describing the rationale behind its development, the following chapter introduces the final data-collection EPR, an overview of the developmental stages, details of each stage and a final summary.

2.2 The Rationale for Development

The trove of information within medical records has been exploited for many years. Florence Nightingale has been charged with instigating this dig for data when assessing the mortality rate during the Crimean war. She subsequently went on to report on optimal hospital design in what has been described as the origin of evidence-based medicine (10). More recently, aided by advancements in information technology and artificial intelligence, a young discipline dedicated to the extraction of information from medical records has emerged. This discipline is medical informatics. Medical informatics describes the study to optimise the retrieval, storage and analysis of data contained within medical records. Secondary uses of medical data include assessing the access to, and effectiveness of new treatments as well as the analysis of patient demographics. Various subsets of informatics exist, including public health, population health, clinical, biomedical and consumer health informatics (105,106).

An obvious problem and one which remains to some degree unsolved, relates to the difficulty of extracting relevant information from descriptive or narrative patient data. This requires the pulling of data from free text, followed by a process of transformation (grouping the free text into structured data fields) and subsequent analysis. It follows therefore, that the acquisition of standardised and structured data is one way to optimise processing (107,108) and reduce cost (109). Indeed standardisation is the hallmark of electronic health records (EHRs) relied upon for pecuniary matters such as the claiming of healthcare costs (110). Beyond these matters free-text fields continue to hinder the extraction of clinical EHR data. This is confounded by the view of clinicians who deem the structuring of data fields in EHRs as limiting the qualitative value, a point disproven so

long as the EHR structure remains comprehensive and relevant (110,111). Another concern is reduction in the speed of documentation (112) which is a problem in the high volume clinics related to eyecare. However, governments recognise the importance of clinical data in continual public health surveillance and data-driven decision making. As such information technology and artificial intelligence are now actively encouraged, moreover they are deemed as the future (67,69).

Considering the above, the development of a data-collection electronic patient record must include the following requirements:

- It must be easy to use for the busy clinician during practise, working in parallel to existing practice clinical record keeping. It is vital to keep the record simple with minimal 'clicks' to ensure the retention of data collectors and the accurate recording of data (113).
- A minimum clinical data set required to reach a diagnosis, so as not to hinder or overload the processing system with redundant items. This would greatly slow processing and pattern recognition.
- A comprehensive list of anterior eye conditions likely to be encountered and therapeutically managed by the APCOS (Acute Primary Care Ophthalmology Service, the data collectors discussed in chapter 1, section 1.4).
- A comprehensive list of therapeutic options habitually used by APCOS.
- Relevant management options.
- A coding system to mask patient data as a unique code ensuring secure data transfer.
- Is developed with a decoding system which can accurately decode the transferred clinical data at the research analysis site.

2.3 The Data-collection Electronic Patient Record (EPR)

The final data-collection EPR met the requirements above and comprised of the five sections given below. A screenshot of the final EPR is shown in figure 2.1 with the actual data-collection EPR available in appendix 10.

- Patient details (Px details)
- History and Symptoms
- Clinical Observations
- Management
- Coding area

The five sections of the EPR are described in more detail below.

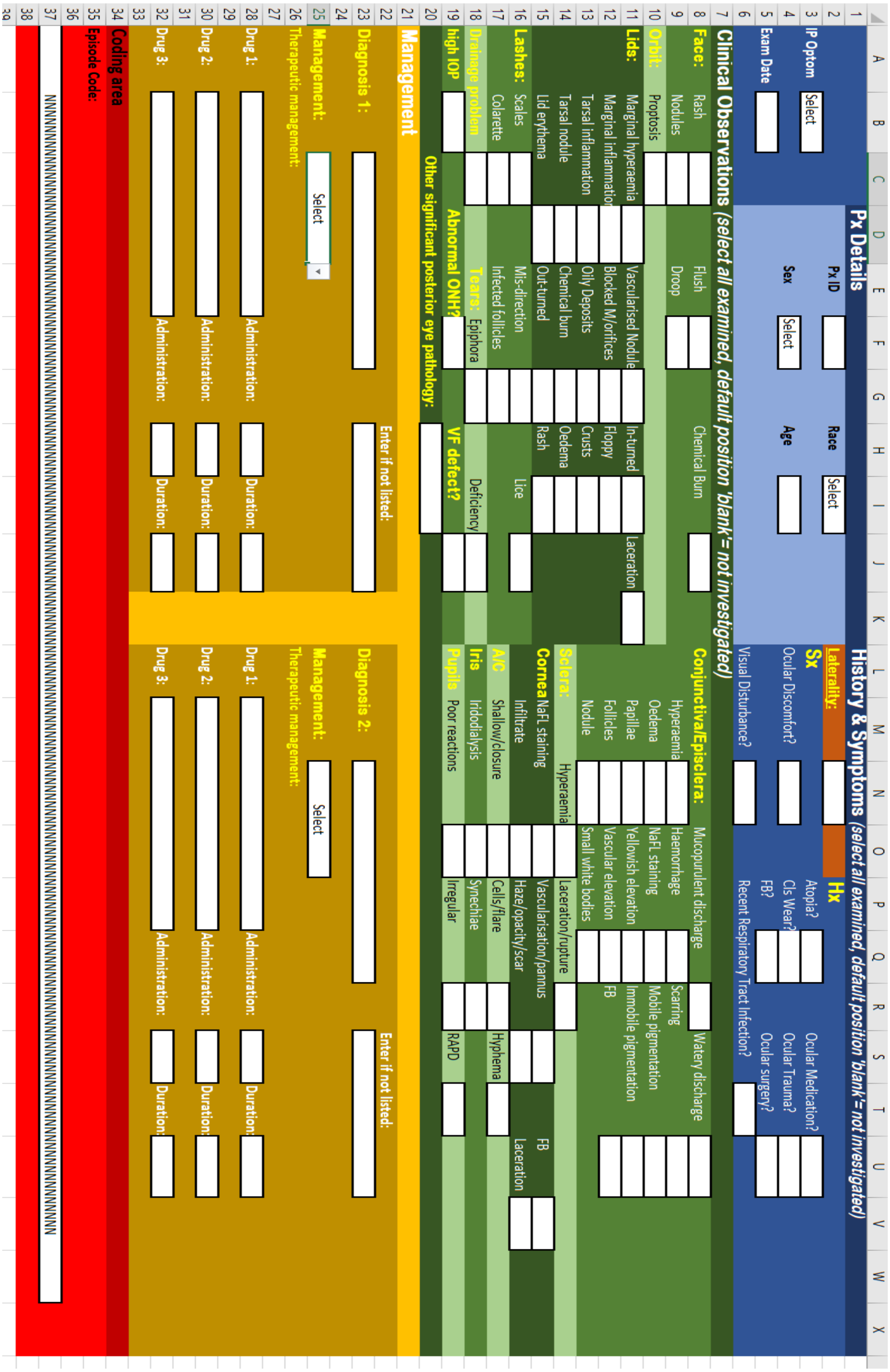


Figure 2.1 The final data-collection EPR

2.3.1 Sections of the EPR: Px Details

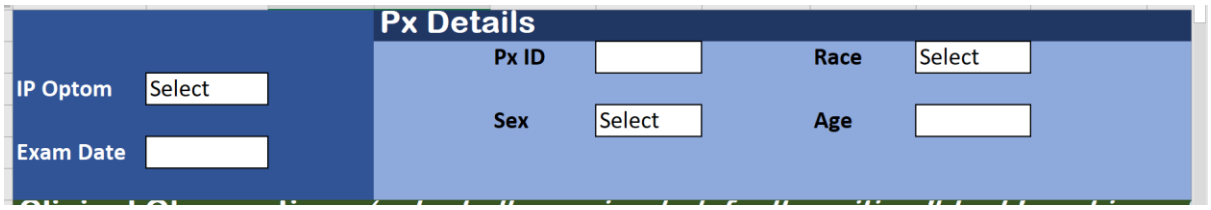


Figure 2.2 The Px details section of the EPR

This section (figure 2.2) comprised the details of the specialist IP optometrist who undertook the examination (initials) and date of the examination. Further information included a patient identification number allocated at the data-collection site and demographical information. The identification number related the anonymised record to the patient concerned. This was intended to be used as an episode identifier replacing any patient identifiable information. The data-analysis site therefore did not possess any personal details of the patient.

2.3.2 Sections of the EPR: History and Symptoms

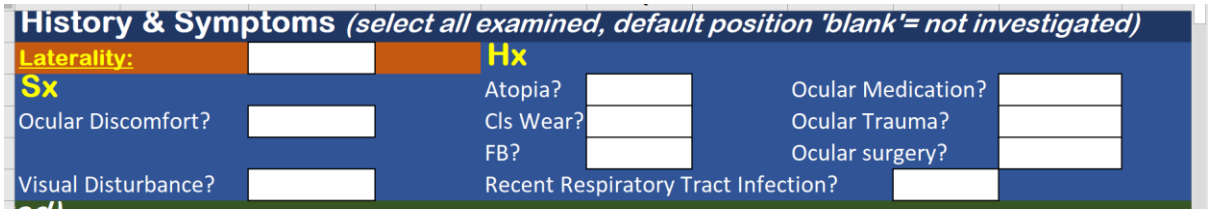


Figure 2.3 The history and symptoms section of the EPR

This section (figure 2.3) covered the pertinent patient history and symptoms extracted from conditions within the College of Optometrists' Clinical Management Guidelines (CMGs). The general term 'ocular discomfort' was used to cover all possible descriptions of discomfort (itching, burning, aching etc.). This was to reduce any subjective variation in description. The same principle was applied to 'visual disturbance' covering all the possible variations (blurring, sectorial loss, complete loss etc.). The patient history covered areas identified as the most disease differentiating, also extracted from the CMGs, and refined as described in section 2.6 (Step 2: Building the EPR Content).

The laterality was also stated here following feedback from the APCOS team (table 2.6). The EPR was to be completed for the worst affected eye, or when both eyes appeared symmetrical, the right eye.

2.3.3 Sections of the EPR: Clinical Observations

Figure 2.4 The clinical observations section of the EPR

The clinical observation section of the EPR included 16 broad anatomical areas (figure 2.4). Eleven of these were further divided into sub-basic variables covering the most pertinent clinical observations within each broad anatomical area. The drop-down lists for the clinical observations offered a choice of 'No', 'Mild' or 'Severe' leaving the blank default position to indicate 'not investigated'.

2.3.4 Sections of the EPR: Management

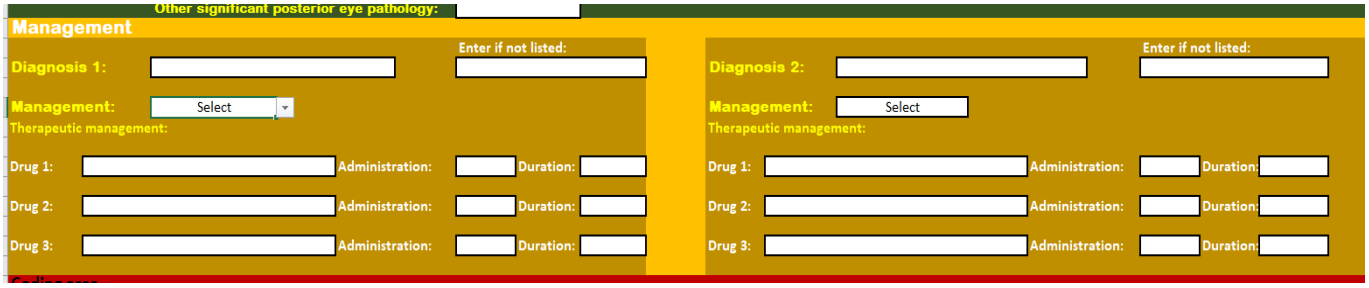


Figure 2.5 The management section of the EPR

The management section of the EPR comprised of two identical parts (figure 2.5). This represented two possible diagnoses (a co-morbidity) per patient episode and management options for each. Three drug options were given for each diagnosis with administration options (OD NOCTE, OD MANE, BDS, TDS, QDS, 5x, 6x, hourly, 2-hourly and prn) and duration options (3 days, 5 days, 1 week, 2 weeks, 3 weeks, 5 weeks, 1 month, 2 months, 3 months, 6 months, prn and on-going). The drug options were sourced from a combination of the local Clinical Commissioning Group prescribing guidelines, APCOS prescribing audits and the College of Optometrists' formulary (section 2.6.3). The diagnosis list comprised of 75 ocular conditions primarily extracted from the CMGs (section 2.5.1) with a free-text box given if the condition was not listed. The 8 management options listed reflected those available to patients seen by the APCOS. These included discharge, follow-up, rapid referral to the HES and routine referral to HES (sections 2.6.3 and 2.7.4).

2.3.5 Sections of the EPR: Coding

The final section of the data-collection EPR displayed the generated unique alpha-numeric code for each patient episode (figure 2.6). Each code was a summation of all the information selected and unselected within the EPR. The codes were copied and sent to the data-analysis site on a monthly basis. Sending data in this code format worked to ensure the secure transmission of data.

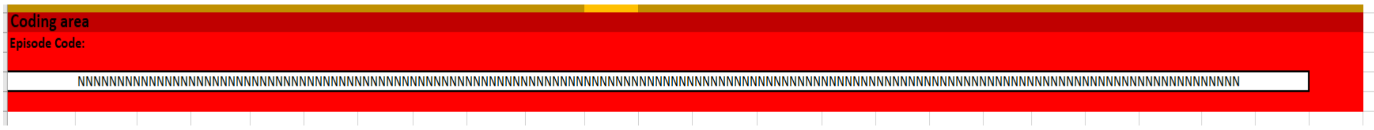


Figure 2.6 The coding section of the EPR

2.4 Developmental Overview

To achieve the aims listed above, a broad three-step process for development was followed. These three broad steps were:

1. Background research on the scope of the EPR and the technical know-how needed for the development

2. Building a database for ocular conditions and therapeutic management options dictating the EPR content

3. The actual production and refinement of the data collection EPR.

The three broad steps described were broken down into many minor stages and the whole process is described below.

2.5 **Step 1:** Background Research

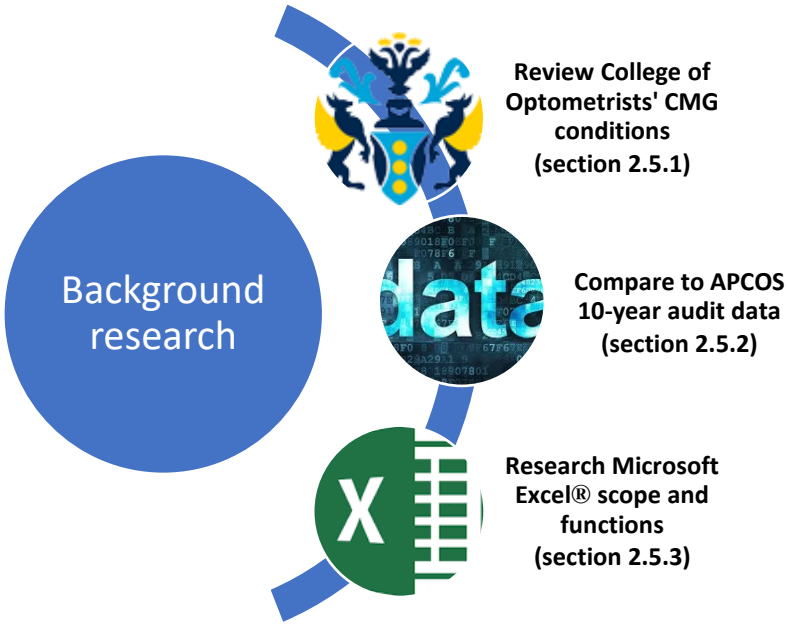


Figure 2.7 The three stages of background research

The background research consisted of three stages (figure 2.7). Firstly, the anterior-eye conditions covered by the College of Optometrists' Clinical Management Guidelines (CMGs) were reviewed (114). These conditions were then compared to the ocular conditions encountered in a 10-year APCOS audit (unpublished data). Finally, the scope and functionality of Microsoft Excel® was explored as the intended platform for data collection.

2.5.1 Review of Anterior Eye Conditions Covered in the CMGs

The CMGs produced by the College of Optometrists' are comprehensive evidence-based guidelines aimed at optometrists in general. They cover both common and rare ocular conditions likely to be seen in optometric practice. The format for each CMG is as follows:

- Aetiology
- Predisposing factors
- Symptoms
- Signs
- Differential diagnosis
- Management by optometrist
    - Non-pharmacological
    - Pharmacological

- Management category (ranging from sight-threatening conditions requiring immediate referral to an ophthalmologist without intervention, to not sight-threatening conditions that can be managed to resolution)
- Possible management by ophthalmologist

The evidence base for the recommended management complete with references are available for each condition.

The CMGs were chosen as a starting point to identify and include the most commonly presenting eye conditions to optometric practice. The guidelines are used extensively as a basis for IP optometrist training and on-going support. The terminology used within the guidelines to identify diagnoses and the differentiating signs and symptoms, was also deemed a means for standardisation of the EPR variables. This would ensure a true diagnosis was selected in place of a sign (e.g. anterior uveitis instead of cells/flare) by producing a robust standard operating procedure (SOP). Dr Gurney (2015) showed that an SOP optimised processing during the application of Bayes' theorem to glaucoma patients (115).

Other standardised healthcare terminologies with conditions represented as codes exist such as the World Health Organisation's International Classification of Disease (ICD). These were originally used to report mortality statistics by United Nations member countries (116). They are useful for the retrieval, analysis and transfer of information between regions and countries by eliminating semantic variations, a common feature of medical records. The latest version, ICD-11 was released in June 2018 to be implemented in January 2022. The major drawback of using this kind of standardised terminology is the constant need to update systems upon revision. The ICD-10 has 10 versions from the first in 2003 until the current version in 2016 (117). Regular subtle updates in the coding would create a limitation in an intended self-evolving system by necessitating manual input. An example of subtle code changes relates to the recording of septic shock in ICD-9 which changed from 785.59 to 758.52 in 2003 despite no change in the clinical term (116). Moreover, an EPR based on unchanging clinical terms from the CMGs is more relevant to practising optometrists, The College of Optometrists being a recognised and respected professional body.

The 59 indexed CMG conditions were individually viewed and a total of 80 eye conditions extracted (accounting for conditions of multiple aetiology mentioned under one heading). These conditions were refined by combining the various types of ectropion and entropion as identification of the specific cause was deemed difficult and irrelevant to practising optometrists. The management of these conditions, in their various forms, is also identical despite the aetiology (118,119). Based on the same rationale, anterior and posterior scleritis were combined in scleritis, covering both. These

two conditions necessitate an emergency referral to the HES and differentiation was not needed for the purpose of this study (120). These measures resulted in the removal of 8 conditions leaving 72 conditions forming the provisional diagnosis list for the EPR. Table 2.1 shows the combined conditions and terms chosen to replace them. A flowchart summarising the entire refinement process of the diagnosis list, with the conditions removed at each stage, can be found in figure 2.8.

Ocular Condition	Combined Term
Entropion – Involutional (age-related) Entropion – Cicatricial Entropion – Spastic Entropion – Congenital	Entropion
Ectropion – Involutional (age-related degeneration) Ectropion – Cicatricial Ectropion – Paralytic Ectropion – Mechanical Ectropion – Congenital	Ectropion
Scleritis – Anterior Scleritis – Posterior	Scleritis

Table 2.1 Combined items from original 80 condition list

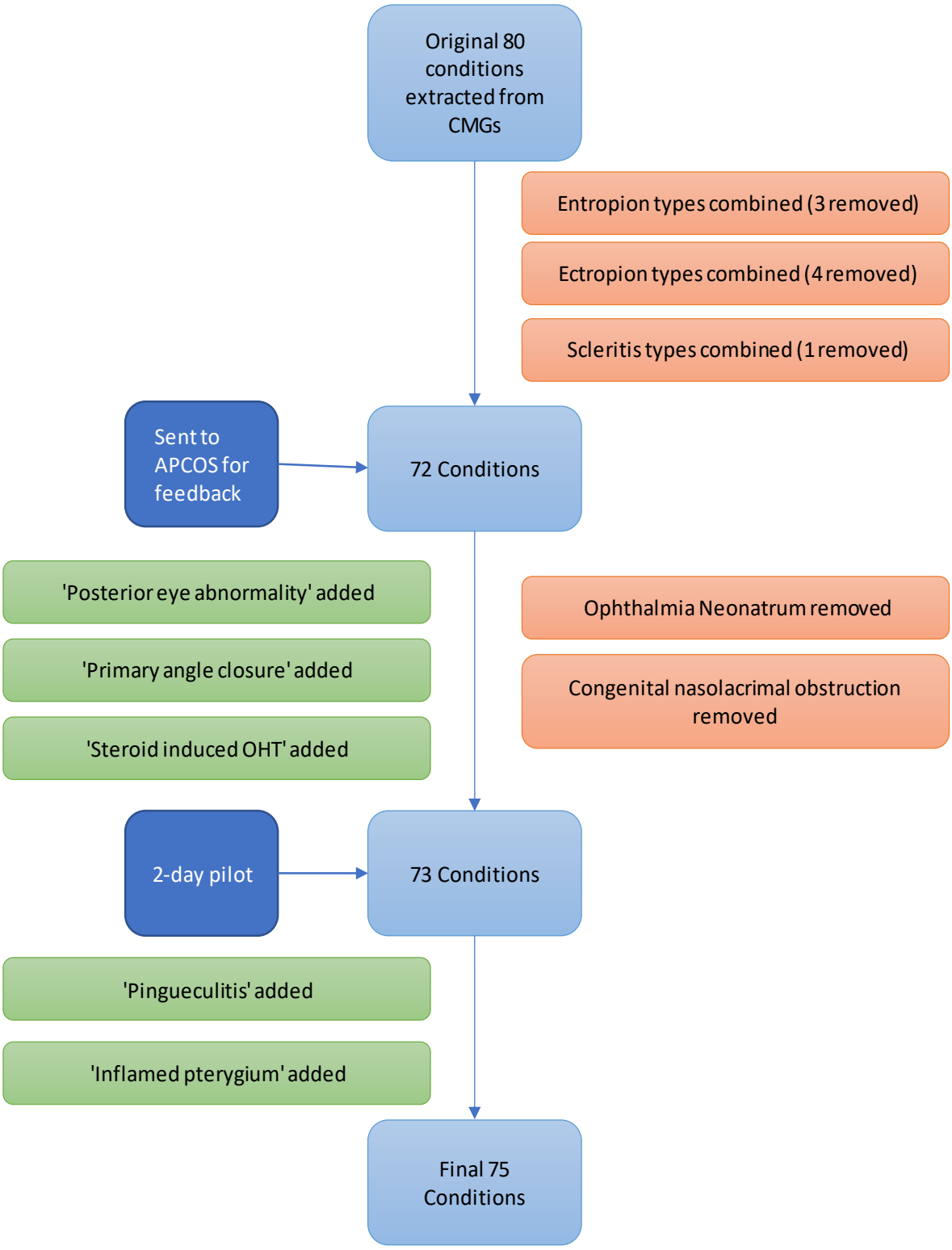


Figure 2.8 The refinement process of the diagnosis list with the conditions added and removed at each stage

2.5.2 Comparison of the CMGs to a 10-year APCOS audit

The second stage in the background research step involved comparing the diagnosis list with the results of a ten-year audit of one of the APCOS specialist IP optometrists (unpublished data). This

was done by comparing the conditions identified from the CMGs to those seen over the ten-year audit period, with an intention of enhancing the EPR diagnosis and drug lists. This task was made cumbersome by the lack of standardised diagnoses, with clinical signs given instead (e.g. visual disturbance, floaters and inward-turning eyelash), this however re-enforced the need for standardisation. Although the practice of one IP optometrist cannot be taken as representative of broad IP optometrist practice, the key findings from the audit were useful for aiding EPR development and are listed in table 2.2.

As can be seen from the results, around 59% of the episodes presented with a CMG condition, meaning 41% of episodes, were not covered by the CMGs. Of note however, is the fact that these totals include common posterior eye conditions such as PVD, a condition for which IP optometrists do not prescribe, and glaucoma, a condition commonly prescribed for but outside the scope of this study. Interestingly and perhaps most relevant to the IP-centric focus of the present study, 62.4% of the total episodes concerned the anterior segment, of which 99.2% were CMG conditions. Most of the anterior segment conditions, therefore, are adequately represented by the CMGs. Moreover, only 3.9% of these CMG conditions were referred on, indicating that most anterior segment episodes are managed within specialist IP optometry practise. Indeed, upon further analysis of therapeutic interventions, 96% of the 2277 drugs prescribed were for conditions covered by the CMGs with 95.5% of these for anterior eye conditions. For the non-CMG anterior eye conditions 86.9% were prescribed for, meaning that specialist IP optometrists are prescribing for conditions not listed in the CMGs. However, this activity is uncommon amounting to 4% of total prescribing.

From the data discussed above it can be deduced that upon considering the anterior eye, the CMGs are a reasonable basis for the development of the data collection EPR. Not only are the anterior ocular conditions listed in the CMGs reflective of most specialist IP optometrist activity, but they also represent conditions for which the vast part of prescribing takes place.

Perhaps the most telling finding becomes apparent when comparing the pharmaceutical management options given in the CMGs with the audit data. Of the total number of drugs prescribed 42% of specialist IP optometrist prescribing extended beyond the recommendations in the CMGs. The therapeutic recommendations given in the CMGs are known to be limited as they refer the reader to the optometrists' formulary for specific drug formulations. This latter resource is only available to members of the College of Optometrists and not all clinicians. Nevertheless considering the pharmacological recommendations as in the CMGs, only 38% (26 of 42) of the prescribed drugs were mentioned or linked. This highlights a considerable disconnect between theory and practice. Although the conditions listed in the CMGs are in fact the most likely to be managed by IP

optometrists, the pharmaceutical management options given in the CMGs are not equally representative.

	CMG conditions	Non-CMG conditions	Total conditions
Total number of episodes	2728 (59.1%)	1886 (40.9%)	4614
Total referrals	112 (4.1%)	363 (19.2%)	475 (10.3%)
Anterior segment episodes	2705 (99.2%)	175 (9.3%)	2880 (62.4%)
Anterior segment episodes referred on	105 (3.9%)	18 (10.3%)	123 (4.3%)
Total drugs prescribed	2193 (96%)	84 (4%)	2277
Anterior segment condition-related prescribing	2095 (95.5%)	73 (86.9%)	2168 (95.2%)

Table 2.2 Summarised results of APCOS IP optometrist audit between July 2010-March 2017

It was decided, therefore, that the data collection EPR should be based on the ocular conditions listed in the CMGs. However, the drug list should extend beyond these. The local formulary, the Optometrists' Formulary (121) and APCOS IP optometrist advice should be sought, this is described in section 2.6 (Step 2: Building the EPR content).

2.5.3 Microsoft Excel® - Scope and Functions

The third and final stage in the background research step involved a familiarisation with Microsoft Excel ®, the intended platform for both EPRs (data-collection and 'intelligent') and the 'intelligent' Virtual Patient Tools (iVPT). Produced by Microsoft ®, Excel® was chosen due to its user-friendly design and impressive processing power with 1048576 rows and 16384 columns in the latest version 16.0 (the version used)(122). Excel® allows users to automate repetitive tasks and processes, presents and analyses data in a variety of graphical ways and allows for collaborative working. This, no doubt is cause for the widespread use of Excel® across many operating systems including Windows, macOS, Android and iOS. Access to Excel® is therefore excellent over the globe, this fact being key in selecting this platform for EPR development. Artificial intelligence in optometry should be accessible to all. Through demonstrating data-collection and analysis using a 'basic' and ubiquitous spreadsheet like Excel® research participation, it is hoped, will flourish. As such "citizen data scientists" (or optometrist data scientists), as described by the director of research at Oxford University, could unlock great value from existing clinical practice (123).

Excel® available through Microsoft Office 365, comes with user tutorials accessed upon opening a new spreadsheet. The 'Take a tour' and 'Get started with formulas' tutorials can be reviewed. A

complete list of the functions and formulas deemed pertinent are shown in appendix 2. Understanding the application of these formulas is pivotal to the development and processing of data and thus EPR design. This familiarisation stage therefore represents a key element in the engagement with artificial intelligence systems.

Due to the number of ocular conditions and anticipated drugs to be listed in the data-collection EPR, a simple drop-down list (as described in appendix 2) was deemed insufficient. Data-collectors would have to search through all the items to select the one of interest. The individual functions of Excel® may however, be combined to create advanced functionality such as a searchable list. This step-by-step process has been described (124) and can also be found in appendix 3. Searchable lists were used for diagnoses and drugs options, making data collection a more time-efficient process.

2.6 **Step 2:** Building the EPR Content

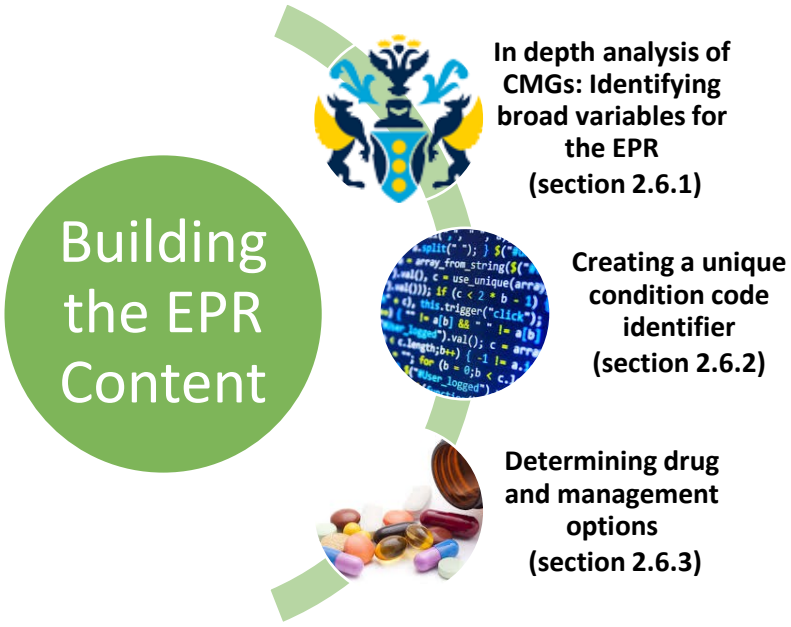


Figure 2.9 The three stages involved in the content building of the EPR

Having determined the 72 CMG conditions as a basis for the data collection EPR, compared this with the APCOS audit and becoming conversant with Microsoft Excel®, the next step was to determine the exact variables of the EPR. This step involved the three stages shown in figure 2.9.

2.6.1 In-depth Analysis of the CMGs: Identifying Broad Variables for the EPR

The first stage of step 2 was an in-depth analysis of the CMGs and the identification of broad categories required in the EPR. This was deemed necessary in order to 'flesh-up' the EPR with the

most clinically appropriate and disease-differentiating variables. Some of the recommendations for the design of electronic medical records include making them easy to use (125) with a minimum number of 'clicks' and mandatory fields (113). This, by default, implies that the number of potentially redundant data fields or variables should be minimised. In order to select the most appropriate variables (signs and symptoms) to be included in the EPR, the CMGs were distilled to produce a minimum dataset of sign and symptoms. This dataset was based upon the potential to uniquely identify the 72 ocular conditions.

In theory, upon examining this requirement mathematically, only 7 truly differentiating variables should be required to identify all 72 ocular conditions. That is, assuming each variable splits the conditions into two equal groups (one group test positive for the variable and the other negative), one variable would produce 2 groups, two variables producing 4, 3 producing 8, 4 producing 16, 5 producing 32, 6 producing 64 and 7 parameters producing 128 groups or conditions (figure 2.10). Admittedly this is a rather simplistic view and is based upon the idealistic assumption that such differentiating variables exist. Moreover, this approach ignores the implications of adding real patients to theory, which is often colourful. This approach, however, does imply that some variables are more important, or diagnostic than others, be they more than 7. The following steps were taken to produce a minimum refined list of the most pertinent clinical variables for disease identification to be included in the EPR.

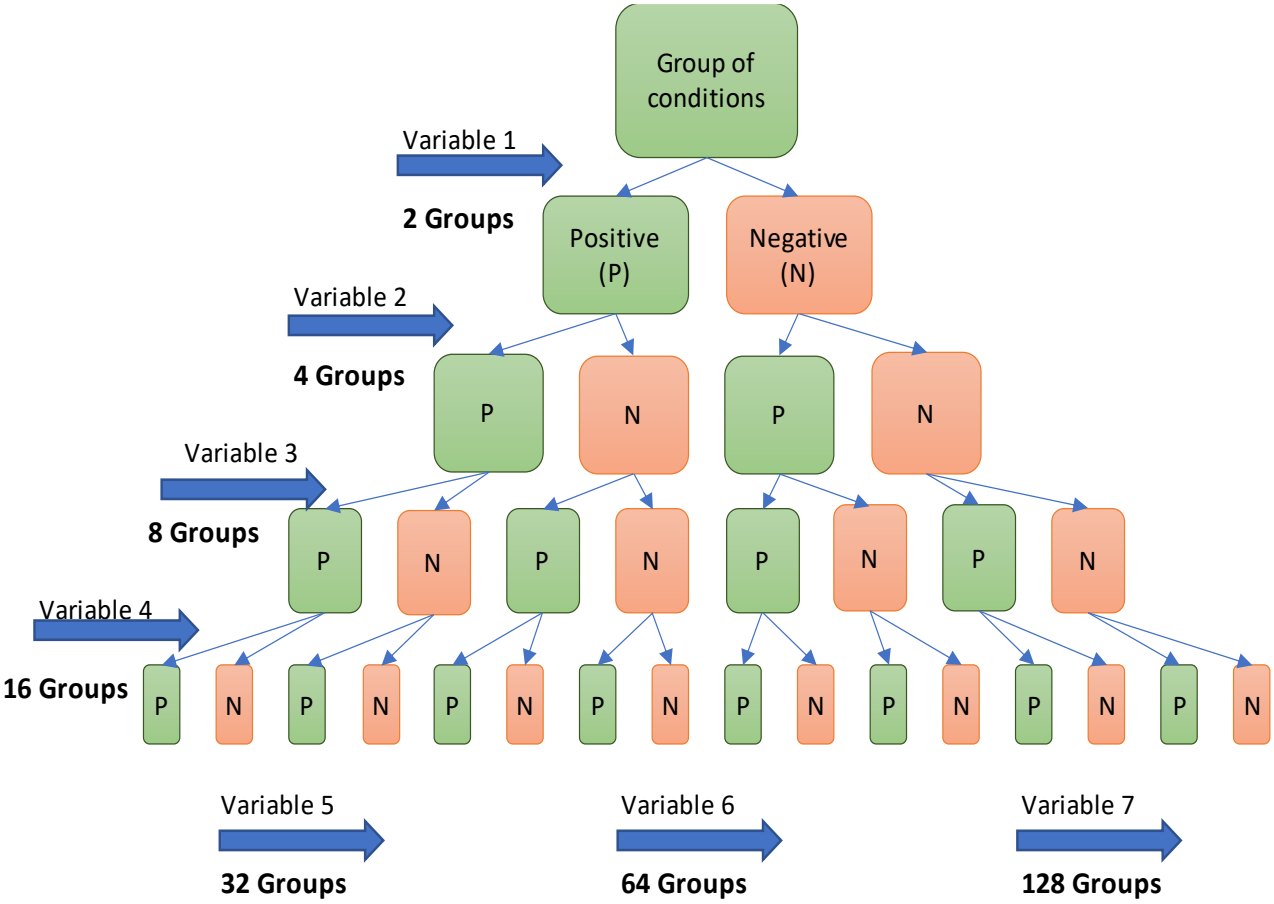


Figure 2.10 Theoretically 7 variables are needed to uniquely identify the 72 ocular conditions

First, all the signs and symptoms (variables) associated with the ocular conditions were extracted from the CMGs. This was done using an Excel® workbook. These were then grouped into broad categories representing all the anatomical areas of the eyes and visual function. The requirement of effective EPR design dictates the use of minimal variables (as described in section 2.2). The starting point was, therefore, a broad, small and robust list limited the final number by keeping-out redundant fields. The most pertinent broad categories relating to the 72 ocular conditions could then be built upon. Equally, redundant broad categories could be removed efficiently.

Ocular symptoms were grouped into two broad categories, namely 'ocular discomfort' and 'visual disturbance'. This was done to exclude subjective differences in the interpretation of the type of pain or severity of visual disturbance. To keep the EPR focussed on clinical presentation and avoid a dependency on additional information, such as family history or ethnicity, risk factors were not added at this point. Table 2.3 shows the complete list of the 25 broad categories which were chosen.

Ocular discomfort	Abnormal anterior chamber
Visual disturbance	Abnormal iris
Abnormal face	Abnormal crystalline lens
Abnormal orbit	Abnormal vitreous
Abnormal eyelids	Abnormal optic nerve
Abnormal eyelashes	Abnormal fundus
Abnormal lid eversion	Abnormal Macula
Abnormal lacrimal apparatus	Reduced visual acuity
Abnormal conjunctiva	Abnormal pupil reactions
Eye discharge	Abnormal motility
Abnormal sclera (and episclera)	Abnormal intraocular pressure
Abnormal cornea	Abnormal visual field
Abnormal colour vision	

Table 2.3 A list of the 25 broad categories relating to the signs and symptoms associated with the CMG listed conditions

2.6.2 Creating Unique Condition Codes

The CMG conditions selected in the present study were focussed on the anterior eye. This is an anatomical area that displays a variety of clinical signs in differing combinations for each condition. It was therefore expected that the 25 broad category list alone would be an insufficient minimum to uniquely identify the 72 ocular conditions. This would mean that whilst some of these categories would need to be removed as they clearly related to posterior eye conditions, others would require the addition of more specific variables (sub-basic). One approach to this would be to take an experienced guess of which categories were likely to be removed and variables added. However, a more objective approach related to the CMGs was decided upon.

A 'unique condition code identifier' was produced in Excel®. This comprised of a table with the 72 anterior ocular conditions against each broad category from table 2.3. The presence or absence of abnormality at each category for any given condition was denoted with a '0' or '1' respectively. This resulted in a row of 0's and 1's for each condition. These rows were then combined to produce a binary code for each condition relating it to all the broad categories (present and absent) from table 2.3.

A novel method of identifying any duplications in these binary codes was developed. This is described in appendix 5. The process highlighted 18 duplications, meaning that including the 25 broad categories alone was not comprehensive enough to uniquely identify each CMG condition. In order to determine which of the 25 should be expanded upon, there was a need to identify and remove the most redundant or least differentiating. This would keep the final EPR variables at a minimum. This was done using the 'mid-point proximity' value of each broad category.

The mid-point proximity describes another novel approach testing how ideal a category (or variable) is. This follows on from the theory of only requiring 7 highly differentiating variables to uniquely identify the 72 CMG conditions (figure 2.4). Ideally each of these variables must have the ability to split the group of ocular conditions in exactly half. When a variable succeeds in doing this, the mid-point proximity value becomes zero. Any other value shows how much that variable deviates from being ideal. When this method was applied to the 25 broad categories, 9 categories could be removed before the number of duplications began to increase from 18 to 21. This showed that these 9 categories made no difference in uniquely identifying the 72 ocular conditions (redundant fields) and therefore could be removed. The categories removed are listed in table 2.4. These categories would perhaps have been obvious choices considering the focus of the present study is anterior ocular conditions. However, the mid-point proximity method trades subjective decision-making for a mathematical approach. A similar concept, namely information gain exists in the realm of machine learning (126). This involves assessing the effect of an attribute (akin to our categories) on the order of a data set. In other words, the effectiveness of an attribute to 'clean up' or homogenise a disorderly set of data by sectioning. Borrowed from physics, this method involves the use of complicated formulae to quantify the pre-attribute homogeneity of data and compare it to the post attribute homogeneity. Despite this having been used in clinical research (127), this method would require an in-depth analysis of each final EPR variable for relatively little gain. As such information gain was deemed inappropriate for the purposes of this study.

Abnormal colour vision	Abnormal vitreous
Abnormal motility	Abnormal crystalline lens
Abnormal macula	Abnormal iris
Abnormal orbit	Abnormal visual field
Abnormal optic nerve	

Table 2.4 Redundant parameters removed after mid-point proximity evaluation

Following the evaluation above, 16 broad categories now remained. These, however, were still unable to uniquely identify each of the 72 CMG conditions yielding 18 duplications in coding. The CMGs were reviewed again, and the broad categories were expanded with the addition of all the sub-basic variables relating to each. These sub-basic variables also included the risk factors, namely demographic, and history information. The 'unique condition code identifier' was reproduced with the expanded variables, binary codes produced for each CMG condition and the duplication process repeated. The mid-point proximity process was also repeated to remove redundant variables. The final evaluation returned no duplications and thus each CMG condition was uniquely identifiable with the 76 variables given in table 2.5.

Hx and Sx:	**Lids:**	**Conjunctiva/episclera:**
Unilateral/bilateral	Marginal hyperaemia	Mucopurulent discharge
Ocular discomfort	Marginal inflammation	Watery discharge
Visual disturbance	Tarsal inflammation	Hyperaemia
Atopia	Tarsal nodule	Oedema
CLs wear	Lid erythema	Papillae
Foreign body	Vascularised nodule	Follicles
Ocular medication	Blocked meibomian orifices	Nodule
Ocular trauma	Oily deposits	Haemorrhage
Ocular surgery	Chemical burn	Staining
Respiratory tract infection	Out-turned	Yellowish elevation
Face:	In-turned	Vascular elevation
Rash	Floppy	Small white bodies
Nodules	Crusts	Scarring
Flush	Oedema	Mobile pigment
Droop	Rash	Immobile pigment
Chemical burn	Laceration	Foreign body
Orbit: Proptosis	**Cornea:**	**Sclera:**
Lashes:	NaFL staining	Hyperaemia
Scales	Infiltrate	Laceration/rupture
Collarette	Vascularisation/pannus	**Iris:**
Mis-direction	Haze/opacity/scar	Iridodialysis
Infected follicles	Foreign body	Synechiae
Lice	Laceration	**Pupils:**
Drainage problem:	Anterior chamber:	Poor reactions
Tears:	Shallow/closure	Irregular
Deficiency	Cells/flare	RAPD
Epiphora	Hyphema	**Abnormal ONH**
High IOP		**VF defect**

Table 2.5 The final variables identified for the data collection EPR – underlined fields represent broad categories

Ocular symptoms were kept grouped as 'ocular discomfort and 'visual disturbance'. This covered all types of discomfort (itchy, painful, mild, severe etc.) and all types of visual disturbance (transient blur, partial absolute loss, or even complete loss of vision). This did not affect the ability of the variables in table 2.5 to uniquely identify the 72 CMG conditions so a more specific interpretation was not required.

The second part of the 'Building the EPR content' step, namely the 'Creating a unique condition code identifier' formed the basis on which any additional variables or conditions were tested prior to inclusion into the data-collection EPR. The final unique condition code identifier can be found in appendix 4.

2.6.3 Determining Drug and Management Options

Lastly, having determined the diagnosis list and variables for the data-collection EPR, the final information required was a list of medications and management options to be added.

It had been established from the APCOS audit results that only 38% of the drugs prescribed were reflected in the CMGs (section 2.5.2). This discouraged the lifting of drug information from the CMGs to form a basis of a comprehensive drugs list. The APCOS audit was considered as a source but lacked the complete information on drug concentrations. Moreover, the audit covered a 10-year period ending in early 2017 and thus could potentially exclude some newer drugs. Equally so, the difficulty with taking all the eye-related medications from the BNF or eMC lists was the great number of entries for the same generic drug. For example, the hypromellose eye drop has 34 entries in the BNF(128) and 8 in the eMC (129); each varying slightly in concentration or representing a different brand. As such lifting all the drug information from the BNF and eMC would render the drugs list comprehensive but irrelevant. No search mechanism could have refined the list enough to make the EPR user friendly. A final consideration was given to local prescribing recommendations by the clinical commissioning groups in the APCOS areas (West Kent CCG, Medway CCG and Dartford, Gravesham and Swanley CCG). This approach was taken as it helped to refine the medications to those most likely to be used in the data-collection area (130).

There remained the unaddressed need to list systemic drugs which specialist IP optometrists use to treat ocular conditions such as oral acyclovir for herpes zoster ophthalmicus. These are not easily accessed through the eye sections of the BNF or CCG recommendations. Extraction of these drugs from the systemic antibiotic sections of the above sources required specialist knowledge and as such proved difficult. The systemic medications were therefore added from the College of Optometrists' formulary as this source is aimed at eyecare clinicians (131). The final 102 drugs are listed in appendix 6.

The management options (outcomes for the patient) reflected the standard examination outcome pathways available to the patient via the APCOS. These were 6, namely:

- Discharged
- Rapid access (emergency referral to Hospital Eye Service)
- Follow-up (at examining APCOS practice)
- Referral to Community Ophthalmology Team (COT) (another practitioner specialising in glaucoma services)
- Referral to wet AMD pathway
- Routine referral to HES (Hospital Eye Service)

2.7 **Step 3:** Creating and Refining the Data-Collection EPR

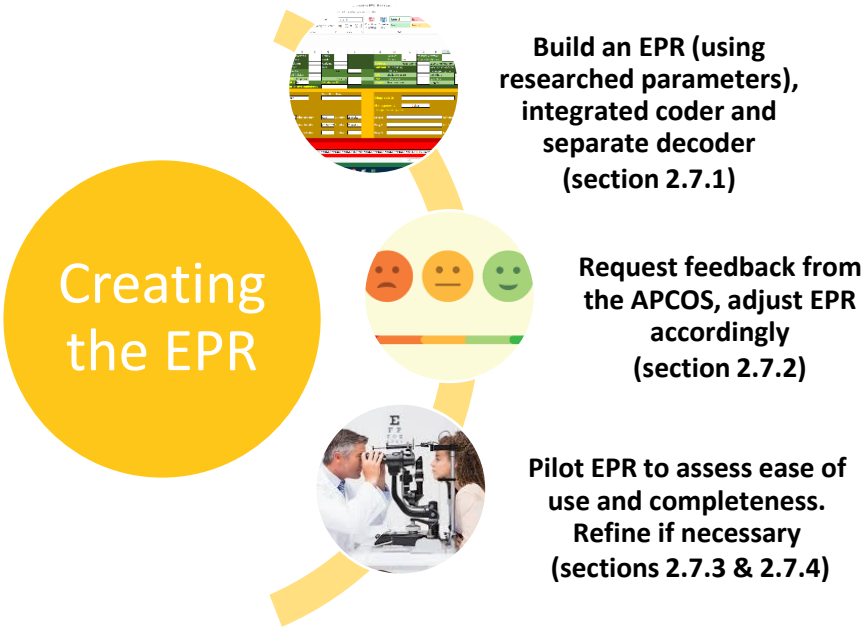


Figure 2.11 The three stages of creating the EPR

The final and on-going step of the EPR development involved the three stages shown in figure 2.11. First, the EPR was built in Excel®. An integrated coder was built translating all the clinical data of each patient episode into a unique code. These codes could then be securely transferred for analysis. Any interception would not compromise patient data as nothing was explicitly available. A separate decoder was built to enable the reverse conversion to data at the analysis site. The EPR was then sent for feedback by the APCOS and adjusted accordingly. Finally, the EPR was piloted and any suggested refinement added. Each of the stages are further discussed below.

2.7.1 Building the EPR - Including a Coder and Decoder

Figure 2.6 shows the first draft of the data collection EPR. The EPR was sectioned using colours for navigational ease. Blue represented the background and patient information, green the clinical findings, orange the management and red the coding sections. Data validation drop-down lists (appendix 2) were used for each clinical parameter showing blank as default equating to 'no'. These drop-down lists were also used for the management, drug administration and duration options. Searchable data-validation drop down lists (described in appendix 3) were used for the diagnosis and three drug options.

A means of secure, anonymised data transfer was deemed necessary owing to the distance between data collection sites and the data analysis site. This took the form of a data coder within the EPR. The

coder converted all the clinical data for each patient episode into an alpha-numeric code. This code was produced using the VLOOKUP function for each parameter relating the clinical data to individual code tables. The codes for each parameter were then combined using the CONCATENATE function to produce a final episode code (Excel® functions described in appendix 2). This method of coding the data prior to transfer ensured that in the event of a data breach during transfer no meaningful patient information would be lost. No patient identifiable information was included within the collected data EPR. The coding section is shown in red in figure 2.12.

A separate decoder was built to reverse the episode code into meaningful information at the analysis site (appendix 7). The decoder, also built in Excel®, used the MID function to separate out the individual parameters prior to analysis.

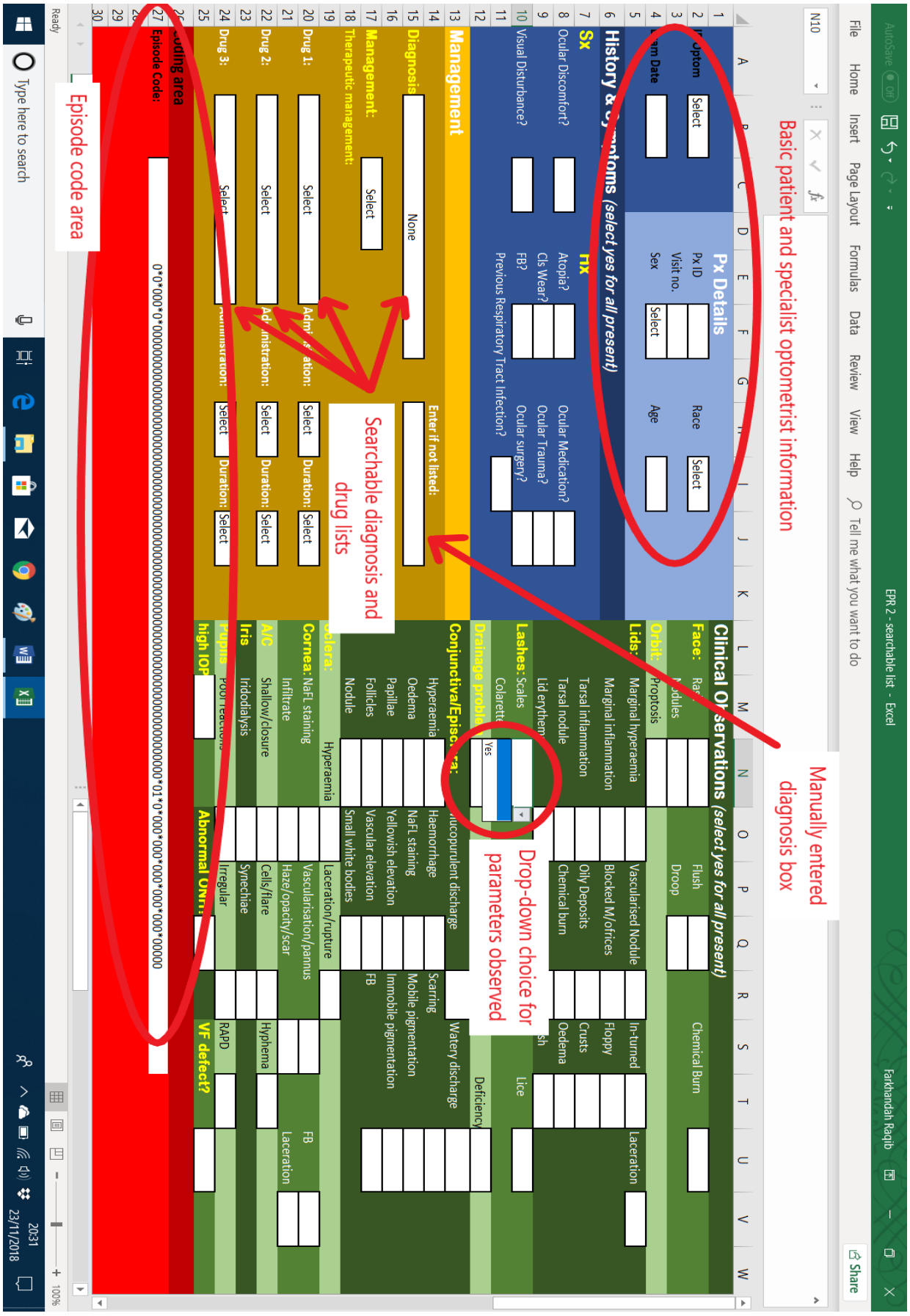


Figure 2.12 The original EPR design

2.7.2 Requesting Feedback on the EPR from the APCOS

The second stage from the list above (figure 2.11) involved seeking feedback from the APCOS regarding the completeness, general ease of use and functionality of the data-collection EPR.

The Delphi method of feedback and forecasting was considered as a means of evaluating the data-collection EPR. The Delphi method is an established process used to distil expert opinions to form a group judgement. This method and its modified versions are frequently used in many fields including those allied to medicine (132–134). It involves the use of a series of interspersed questionnaires on a subject matter sent to experts individually. Controlled individual feedback is provided showing how the group responded comparatively and giving an opportunity for the refinement of answers. Indeed, it was used recently to determine a minimum international dataset for optometry records (135). The crux of the Delphi method is the use of expert knowledge to forecast or produce projections in areas where insufficient data exists for the extrapolation of trends. This is based upon the assumption that the experts consulted are well versed in the subject matter and purpose of the research questions presented. Two problems surface on considering the use of this method in the present study. The first relates to the definition of expert, a previously documented methodological weakness of the Delphi method (134). The experts consulted with regards to the adequacy of the data-collection EPR are the specialist IP optometrists who form the APCOS. These are clinical experts bereft of expertise in optimal EPR design, a subject further complicated by the minimalization of data entry fields, which to any clinical mind seems counter-intuitive. The second relates to the purpose of the present study. A thorough knowledge of the proposed application of naïve Bayesian inference is also lacking in these experts. This translates as insufficient expertise to accurately predict the completeness of the EPR with regards to the application of Bayes'. It was enough therefore to seek feedback from the APCOS specialist IP optometrists for the clinical completeness and ease of use of the EPR. The Delphi method was deemed inappropriate.

The data-collection EPR was sent to all participants and the feedback collated. This was considered against the objective of the study. The changes made to the EPR are given in table 2.6. The full evaluation and feedback, including the suggestions declined, can be found in appendix 8.

Feedback	Adjustments made to EPR
The EPR should include negative findings separate from those areas not investigated The severity of the clinical features should be indicated in the drop-down lists for each variable	New drop-down options for each clinical feature: 'blank' (not investigated) 'no', 'mild', 'severe'. The default position for each variable will be set at blank meaning 'not investigated'.
Definition of 'High IOP' is unclear	A threshold relating to the most recent NICE CG81 guidelines will be used, namely >24mmHg. This instruction will be added as a 'pop-up' text box.
Posterior conditions necessitating referral not documented, these cases need to be added to generate accurate likelihood ratios	'Other pathology' section will be added: with options of 'lens', 'vitreous', 'macula', and 'fundus' – these will be sufficient to differentiate the referral reason.
Addition of 'Recent respiratory infection hx' to differentiate bacterial and viral conjunctivitis	Added to EPR with drop-down list comprising 'yes', 'no' and blank (not investigated).
Laterality reporting, should both eyes be documented separately?	Laterality included, data to be limited to the worst eye or in the case of symmetry, the right eye.
Glaucoma options should be limited to primary angle closure (PAC) and steroid induced as the APCOS team only reviews acute cases. Remove congenital nasolacrimal duct obstruction and ophthalmia neonatorum as the APCOS do not assess babies. Remove other chronic conditions such as choroidal nevi.	List of diagnoses adjusted
Diagnosis and therefore management options do not account for comorbidity. Two diagnoses could be managed differently.	Two diagnosis and management options added, each with three drug options (giving a total of potentially 6 medications prescribed) ǂ
Rename 'VF defect' to 'Relevant VF defect' to avoid recording of chronic unrelated defects	Adjusted drop-down list to blank (not investigated', 'no, 'relevant defect', irrelevant defect'
Chief complaint needs to be clear and whether acute or chronic in nature	Added drop-down for each Sx (visual disturbance and ocular discomfort) 'No', 'Acute', 'chronic', 'Chief acute', 'Chief chronic' – with the default blank position representing 'not investigated'.
Duration since last ocular surgery	Added drop down list to ocular surgery, default blank (not investigated), '0-3 months ago', '3-6 months ago', '6-12 months ago', '12-24 months ago', '24+ months ago'

Table 2.6 Feedback suggestions from the APCOS and adjustments made to the data collection EPR

ǂ Upon this feedback, the original data for the 10-year APCOS audit described in step 1 was re-examined to determine the maximum number of diagnoses and drug options required (4615 patient episodes). The maximum diagnoses per patient episode was 2. The maximum number of drugs prescribed per patient episode was 5 (on one occasion). Here the patient showed a comorbidity (HZO and a superficial corneal abrasion). Iritis was the second condition for which the highest number of drugs were prescribed (maximum 4 per episode including comorbidity) and accounted for 241 patient episodes. All other patient episodes required a maximum of 3 drugs. Two diagnoses and management options, and six drugs therefore was felt enough to cover the likely prescribing of the APCOS.

2.7.3 Piloting the Adjusted Data-collection EPR

Following the adjustments made above, the EPR was sent out for a second time for a 2-day pilot. The feedback was collated again, and the full evaluation can be found in appendix 9. Table 2.7 summarises the results.

Feedback	Adjustments made to EPR
Add '5x' for administration of drugs such as topical anti-viral therapy	Administration list adjusted
Duration of drug use – needs on-going option	Duration list adjusted
There is no option for manually including drugs not listed – Identified oral Co-amoxiclav as not listed. Two new drugs were also identified: Lotemax (Loteprednol 0.5%) and Softacort (hydrocortisone 3.35mg/ml)	Specialist IP optometrists are not restricted to the local CCG formulary. Therefore a 20-part drug audit provided by the APCOS was reviewed and all the missing drugs (21 in total – **see table 2.8)** added to the EPR including those suggested in the feedback. The APCOS was instructed to inform if any new drugs needed to be added during the data-collection period.
Information 'pop-up' boxes for 'ocular discomfort' and 'visual disturbance'- to clarify these broad terms cover all possibilities mild to severe	Added to relevant fields on EPR
Add pingueculitis and inflamed pterygium to diagnoses	These are not explicit diagnoses mentioned in the CMGs but are mentioned under broad headings of 'pinguecula' and 'pterygium'. The pharmaceutical management of these conditions differs from the stable type and so these diagnoses were added.

Table 2.7 Feedback from the APCOS following a 2-day pilot and the adjustments made to the EPR

Co-amoxiclav tablets (500mg/125mg)	Hylocare (Na hyaluronate)
Amoxicillin tablets (250mg)	Hyloforte (Na hyaluronate 0.2%)
Erythromycin tablets (250mg)	Hylotear (Na hyaluronate 0.1%)
Aciclovir tablets (400mg)	Blephaclean eyelid wipes
Naproxen tablets (500mg)	Hyabak (Na carboxymethylecellulose) Optive fusion
Bisoprolol fumarate tablets (5mg)	Xailin night paraffin eye ointment
Gutt. Tafluprost 15mcg/ml (Saflutan)	Theoloz duo gel 0.4%
Lacrilube ointment	Ganciclovir eye gel 0.15%
Gutt. Ketoralac 0.5% (Acular)	Diclofenac sodium 0.1%
Fuscidic Acid cream 2%	**Gutt. Loteprednol 0.5% (Lotemax)**
	Gutt. Hydrocortisone 3.35mg/ml (Softacort)

Table 2.8 Drugs extracted from a 20-part APCOS audit including two new drugs suggested (highlighted) – all were added to the EPR drug lists

The diagnoses of pingueculitis and inflamed pterygium were added to the EPR and the 'unique condition code identifier' was run again. This ensured the ocular conditions were still uniquely identifiable following the adjustments made above. The process returned no duplications showing the data-collection EPR had retained its ability to uniquely identify all the conditions from the parameters therein. Figure 2.1 shows a screenshot of the final data-collection EPR, an interactive version can be found in appendix 10.

2.7.4 The On-going Refinement of the Data-collection EPR

During the data-collection phase of the study the EPR remained under constant refinement. The APCOS feedback identified 3 issues (Table 2.9) and the EPR was adjusted accordingly. The decision to add 'tapered' as a duration option was applied to data collected at month 5 onwards. No major refinement to the clinical observations was requested, as such the theoretical ability of the variables to uniquely identify the 75 listed CMG conditions remained unaffected.

On-going feedback from APCOS	Adjustments and considerations
Aciclovir 800mg – not included in EPR, only 400mg given as option	APCOS advised they would select 400mg twice to cover this option where needed
Minor procedures not accounted for in the management options – this indicated that the patient was only therapeutically managed, not an accurate reflection of practise	Two more management options were added: 'Minor procedure and F-up' and 'Minor procedure and discharged'
Tapered doses not accounted for	The mode of tapering drug administration has no strict guidelines and thus is practitioner dependant. It was therefore deemed appropriate to add 'tapered' as an additional drug duration option .

Table 2.9 On-going APCOS feedback and adjustments made to the data-collection EPR

2.8 Chapter Summary

Chapter 2 described the rationale behind and developmental process of the data-collection EPR and decoder.

Medical records provide a recognised medium for data-collection. A data-collection EPR was therefore developed to optimise data acquisition whilst remaining user-friendly.

In order to optimise the processing of data, all data fields on the EPR were standardised. A standardisation of variables was achieved by reducing the number of 'free-text' fields. This meant including as input options, as many likely anterior ocular conditions and ocular therapeutics as possible.

Analysis of a 10-year APCOS audit revealed the following key points:

- Records lacked a standardisation of terms
- The vast majority of ocular conditions treated by specialist IP optometrists involved the anterior eye
- Most of the conditions managed by specialist IP optometrists were mentioned in the College of Optometrists' Clinical Management Guidelines (CMGs)
- The CMGs, however, did not reflect the drug prescribing practice of specialist IP optometrists (having considered the limited pharmacological guidance contained within them)

Given the points raised from the audit, the CMGs formed the basis of the 75 listed ocular conditions. The drug information however was extracted from local CCG prescribing guidelines, the Optometrists' Formulary, and a series of very recent APCOS drug audits. A total of 123 drugs were listed in the final EPR.

To ensure the secure transfer of patient data a coder was built into the EPR. This produced a unique alpha-numeric code for each patient episode. A decoder was also built at the analysis site to extract the data from this code.

Microsoft Excel® provides an effective platform for data gathering EPRs and potential decision support systems. Judicious use of the functions available therefore could enable future optometrists to gather and analyse local data aiding public eye-health surveillance and local pathway planning.

The completion of the data-collection EPR now lead to the first stage of the Bayesian Transitional Learning Cycle (bTLC), namely gathering the evidence base. This is described in chapter 3.

Chapter 3: Hierarchical Bayes'

3.1 Introduction

There were two aims of the present study. Namely, a) to apply machine learning to the clinical activity of IP optometrists, and b) to develop the concept of interactive and evolving evidence-based support systems for IP optometrists and those in training. The process to achieve these aims was conceptualised with the 5-step Bayes' Translational Learning Concept (bTLC) described in Chapter 1 (section 1.6). Step 2 of the bTLC is "learning from the evidence base" (figure 1.6, chapter 1). Having described the method used to collect the data in chapter 2, this chapter now describes the process undertaken to extract the necessary information from the data, fulfilling step 2 of the bTLC.

The data gathered was used to generate the necessary priors required for machine learning using Bayesian statistical methods (Bayes'). A hierarchical machine learning model was then developed and tested on the clinical data. This chapter describes the processes followed to produce the hierarchical model, the analyses carried out on the data using the model, and the performance of the model. Figure 3.1 shows the chapter plan.

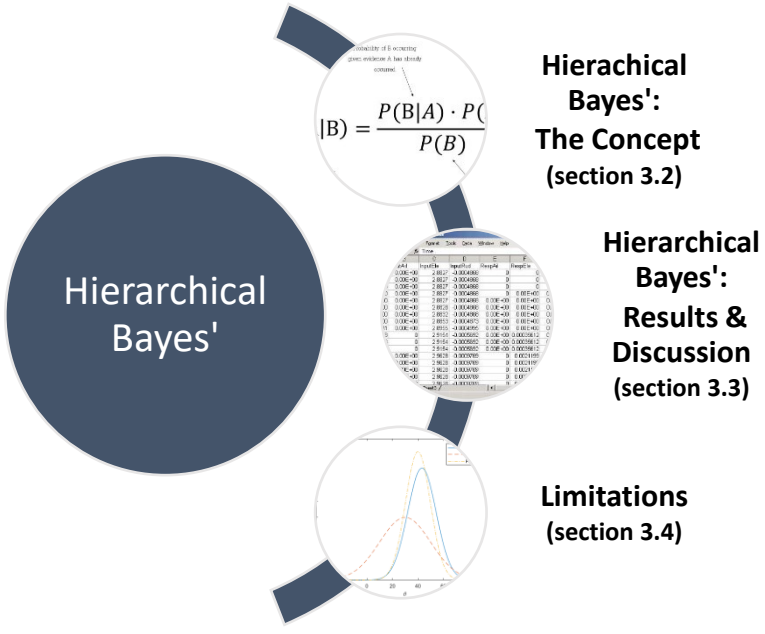


Figure 3.1 The chapter plan for Hierarchical Bayes'

3.2 Hierarchical Bayes': The Concept

Data was collected using the data-collection electronic patient record (EPR) described in chapter 2. A total of 1351 first patient episodes were collected by APCOS over a 12-month period (November 2018-October 2019). The data-collection EPR ensured highly structured data collection with information pertaining to the clinical signs and symptoms observed during each consultation (now referred to as 'tests') as well as up to 2 diagnoses and management decisions and 6 drug options (3 for each diagnosis). The drug administration and duration (drug regime) were also recorded for each drug selected. For each patient episode, the specialist IP optometrists did not undertake all the tests listed in the data collection EPR, rather only those deemed appropriate by the IP optometrist for the presenting patient.

The concept of hierarchical Bayes' (a mathematical model of supervised machine learning) was developed with the intention of using all the clinical data collected and retaining all the structured grouping (granularity) as per the data-collection EPR. This was to give the most precise results in the form of clinical decision support for IP optometrists and those in training. The following section describes machine learning, the development of Hierarchical Bayes, the mathematics involved, the planned application of the proposed model, and how the model was evaluated.

3.2.1 Supervised Machine Learning: The Basics

Machine learning is a branch of artificial intelligence (AI) that uses mathematical algorithms such as Bayes' theorem on historical data to predict the outcomes of new episodes. In the present study this meant that full clinical episodes, that is, all clinical tests (including history, symptoms, and clinical observations), and the resulting diagnoses, management and prescribing decisions were analysed using Bayes'. This resulted in a model (a set of mathematical equations or algorithms) which could predict the outcomes of future patient episodes based on the results of the clinical tests undertaken.

Machine learning can be defined in several ways including supervised and unsupervised learning (136). Supervised learning, used in this study, uses 'labelled' data. This means patient episodes for which the diagnoses, management, or prescribing decisions (the outputs) are known. Supervised learning quantifies the relationship between individual clinical tests and the known output. It is considered the easiest learning method to perform, interpret and validate (137). Moreover, it is the method of choice in most research studies due to the relatively small number of training data needed to build a model, without a significant drop in the accuracy of predictions. However, the predictive ability of a model still depends, partly, on the amount of data used to build the model.

That is, the more data provided (usually up to a plateau) the more accurate the predictive models (136).

In contrast, unsupervised learning uses 'unlabelled' data. This means patient episodes for which the output is not known. Unsupervised learning allows the detection and refinement of patterns from the results of a set of clinical tests. These results are grouped and thus lead to diagnoses (also known as clustering (138)). A major advantage of this method over supervised learning is that it may aid the discovery of previously unknown associations between tests and outcomes. A disadvantage to this method is that it requires much more data and often more sophisticated mathematical methods to build a predictive model.

3.2.1.1 Limitations of Sensitivity and Specificity – Enter Bayes'

How much does a test result affect the probability that a patient has a certain disease? This is a critical and perpetual question asked by a clinician during a patient consultation and after every test undertaken. This question asks for the post-test probability, a measure of the probability that a patient has a disease given a test result. The answer is often searched for amongst sensitivity and specificity values. Sensitivity being the number of diseased patients who will test positive (i.e., correctly identified as abnormal) and specificity being the number of normal patients (not diseased) that test negative (i.e. correctly identified as normal). The problem with looking at these values in isolation is that they only provide information about two *known* groups, that is those who are diseased and those who are normal. This tells us how good the *test* is at identifying each group, but not whether the patient being examined really has the disease (i.e. which category she falls into, and thus what the post-test probability is) (139). In order to realise this, it is important to consider how rare the disease is. For example, smallpox was eradicated in the 1970's (140). If a test for smallpox gave a positive result it is far more likely that the test was a false positive than the patient actually having smallpox. Equally, if a test for the flu virus was positive in the middle of a winter outbreak, it is far more likely that the test result was a true positive. The prevalence of the disease, often referred to as the base-rate or pre-test odds, is therefore very important in the interpretation of test results. Sensitivity and specificity values are limited by base-rate neglect (141,142). Moreover, it has been shown across disciplines that clinicians tend to overestimate the probability of disease upon a positive test result having been given both the sensitivity and specificity values (143–148). Further, even in the presence of a prevalence value, clinicians are still unable to apply this information to patient scenarios leaning again towards overestimation of the post-test probability (149). This could lead to unnecessary further investigation and stress for both the clinician and the patient (and of course the funding bodies!). Bayes' theorem, however, considers both the base-rate (pre-test odds) and the value of the test itself

(in the form of a likelihood ratio). These can be used to generate a post-test probability, thus overcoming the limitations of sensitivity and specificity. Bayes' formed the basis of the mathematical analyses in this study. The application of Bayes' is now discussed further.

3.2.1.2 The Application of Bayes'

The application of Bayes' theorem (Bayes') to quantify the effect of a clinical test result on a tentative diagnosis is not new (150). Aspinall and Hill introduced this idea to optometrists in 1983 (90). Using decision matrices, they described how using sensitivity and specificity values can generate likelihood ratios for each test result. This is then combined with the pre-test odds to generate post-test odds and probability. Clinical decision-making has been described as "fundamentally Bayesian" (151). The process of Bayesian reasoning can be regarded as a verbalisation of the subconscious analysis undertaken by clinicians even without formal training in statistics. In effect, we have an idea of the likelihood of a diagnosis and refine our thoughts with new information, that is, the results of further testing.

Continuing from the previous work carried out by Sagar (2014) and Gurney (2017), the method in which Bayes' was applied in the present study involved the generation of post-test probabilities (97,98). The components needed to generate post-test probabilities originate from decision matrices relating each test (sign/symptom or piece of demographic information, i.e. the predictor variable) to a diagnosis, management option or drug prescribed (the outcome or predicted variable) (152). A simple 2x2 decision matrix, from which the numbers required for a single test can be extracted, is shown as table 3.1. Table 3.2 shows the mathematical components needed to generate the post-test probability. The equation showing the derivation of the post-test odds reflects Bayes' rule (as discussed in chapter 1, section 1.3.1), and is highlighted in red.

Predictor	**Outcome (O+)**	**Outcome (O-)**	**Total of rows**
Positive (P+)	a	b	**a+b**
Negative (P-)	c	d	**c+d**
Total of columns	**a+c**	**b+d**	**a+b+c+d**

Table 3.1 A 2x2 decision matrix showing the possibilities of a single test and a disease (adapted from Parikh et al. (2009) (152)

Attribute	Definition
Pre-test probability (prevalence)	(a+c)/(a+b+c+d) also expressed as **p(O+)**. The prevalence of a condition or frequency of a drug prescribed within the dataset.
Pre-test odds	prevalence/(1-prevalence) also expressed as **p(O+)/(1-p(O+)**
Sensitivity	a/(a+c); This is the ratio of true positive cases to total tested positive. This is also expressed as **p(P+|O+).**
Specificity	d/(b+d); This is the ratio of true negative cases to total tested negatives. This is also expressed as **p(P-|O-)**
Positive Likelihood Ratio (LR+)	sensitivity/(1-specificity); The probability that a positive predictor result occurs for a positive outcome. This is also expressed as **p(P+|O+)/p(P+|O-)**
Negative Likelihood Ratio (LR-)	(1-sensitivity)/specificity; The probability that a negative predictor result occurs for a negative outcome. This is also expressed as **p(P-|O+)/p(P-|O-)**
Post-test odds (Bayes' Rule):	pre-test odds x LR (positive or negative LR depending on the result of the test)
Post-test probability	post-test odds/(post-test odds +1)
Positive predictive value (PPV)	a/a+b; This is the probability of a positive outcome given the predictor is positive. This is also expressed as **p(O+|P+)**. A PPV value close to 1 indicates a good predictor.
Negative predictive value (NPV)	d/c+d; This is the probability of a negative outcome given the predictor is negative. This is also expressed as **p(O-|P-)**. As with the PPV, an NPV close to 1 indicates a good predictor. NB It follows from this term that the probability of a positive outcome following a negative predictor is p(O+|P-) and = 1-NPV.

Table 3.2 The mathematical components needed to generate post-test probabilities, positive and negative predictive values, and the application of Bayes' rule (highlighted in red).

3.2.1.3 Rarity and Importance: The Two Elements of Bayes'

As mentioned in 3.2.1.2, the probability of an outcome (diagnosis, management, or drug prescription) depends upon two elements. The first is a consideration of *how rare* the outcome is, that is, its pre-test odds or base-rate (prevalence). The second element therefore is *how important* the new information is (i.e., the result of a test), that is, its likelihood ratio (LR). In clinical situations the likelihood ratio describes the number of diseased people with a test result compared to the number of well patients with the same test result (150). It follows that the most diagnostic tests will yield the highest likelihood ratios. LRs therefore assign a value to the importance of the test result,

the degree of which is of clinical importance when considering a diagnosis. LRs refine clinical judgement. It is the product of these two elements (the rareness of disease and the importance of the test results) which gives us the odds of an outcome. The post-test odds can then be converted into post-test probability.

3.2.1.4 Multi-level Likelihood Ratios (LRs) and Multiple Tests

When considering a test with two outcomes (present and not present – a binary test outcome) both positive and negative LRs exist. However, when a test has multiple outcomes, or the concept of 'not present' is impossible, positive and negative LRs do not apply in this way. Instead, each outcome carries its own LR creating a system of multi-level LRs. Consider the race of a patient for example. Here the positive LRs are divided between the known races in the study, a negative LR is impossible as a patient cannot be race-less. Multi-level LRs therefore apply here. In the present study multi-level LRs were used for the multiple parameter outcomes. This covered areas such as demographic information (e.g. age brackets) and clinical findings (e.g. no, mild, severe).

Since diagnostic and management outcomes are based on a plethora of tests and information, the product of the multi-level LRs pertaining to each of the tests undertaken gave a final combined LR. This reflected the combined importance of the information gathered during the clinical examination.

3.2.1.5 Insufficient Data and Likelihood Ratios (LR)

As mentioned, likelihood ratios (LR) represent the difference between the patients with a known disease exhibiting a particular test result and the number of patients without the disease who also exhibit that test result. It follows then that the most useful LRs are those which are either very large or very small (to support or exclude a diagnosis respectively). Higher LRs indicate the predictor as being strong and small LRs weak. A LR of 1 is the least useful as it has no predictive clinical value (it does not alter the probability of an outcome) but merely reflects the value of the pre-test odds. The value at which a LR becomes useful is ill defined. Some have taken arbitrary values of 10 and 0.1 as good indicators (153) or dismissed values of 0.5 and 2 as the least useful (154), whilst others have assigned percentages to LR values reflecting the approximate change in post-test probability (155).

The calculation of likelihood ratios depends upon the availability of sufficient data.

3.2.1.6 Laplace Correction

As described by Sagar (2014) and Gurney (2017) a Laplace correction was applied to the data counts to remove the effect of likelihood ratios of zero (97,98). This happens when a poorly represented outcome yields a zero count in the collected data for the presence of a particular sign or symptom.

This in-turn would result in a LR of zero for that indicator (sign/symptom/demographic information). Since calculating the effect of multiple clinical tests on an outcome involves taking the product of their LRs, a LR of zero can significantly distort the results. One LR of zero in a battery of tests undertaken would disregard all other test results regardless of their diagnostic strength, rendering an otherwise likely outcome impossible. The addition of a Laplace correction is important as no outcome is a true impossibility and having one of the boxes of the contingency table equating to zero reflects an insufficiency of data collection as opposed to a real lack of incidence. A Laplacian correction of 1 was used where appropriate as has been described in the literature (137).

3.2.1.7 Fractional Uncertainties

No method of measurement including mathematical computations, is perfect. Thus, all measures carry a degree of uncertainty. Uncertainties reflect the accuracy or the precision of a measurement. To appreciate the degree of uncertainty, it must be compared to the true measure (156). For example, an uncertainty of 5mmHg when measuring atmospheric pressure (around 760mmHg) translates as a relatively precise measure (i.e. 5/760 = 0.006 or 0.6% fractional uncertainty). However, the same 5mmHg of uncertainty when measured against an intraocular pressure (IOP) of 18mmHg indicates a very poor measure (i.e. 5 / 18 = 0.278 or 27.8% fractional uncertainty). It can be deduced therefore that the ratio between the true measurement and its uncertainty is of importance when considering the quality of any measure. This is fractional uncertainty and is calculated using the formula below:

$$\text{Fractional uncertainty} = \frac{\text{the uncertainty of a measurement}}{\text{true measurement (absolute value)}}$$

Note for the above formula, the true measurement is in fact an estimate itself (LR or pre-test odds for this study).

Confidence limits are commonly reported measures of error in the literature. The most common of these are the arbitrary 95% upper and lower confidence intervals. Upper and lower 95% confidence intervals indicate, with 95% certainty, that the true value falls within them. A similar approach was taken to measure the degree of error for post-test probabilities in the present study. Upper and lower fractional uncertainties of each component could be used to derive these values as described below.

Fractional uncertainties can be calculated for the generated pre-test odds (derived from the prevalence) and likelihood ratios pertaining to the prediction of diagnoses. This provides a quality measure for both the rareness of things and the importance of tests, the two components for the application of Bayes'. The compounding fractional uncertainty resulting from the pre-test odds and a battery of clinical tests can be calculated by taking the sum of all fractional uncertainties and taking the square root of the total (156). This is then converted into the upper and lower confidence limits for each post-test probability. Figure 3.2 shows lists of the equations used to get the final post-test probabilities and their confidence limits. In the present study the calculation of fractional uncertainties was left until all the Bayes' analyses were undertaken and evaluated in the first instance. The intention was to add fractional uncertainties if Bayes' had the ability to generate good decision support after performance evaluation.

Equations for the Pre-test odds and their fractional uncertainties (counts from table 5)	
Prevalence, P	a+c/(a+b+c+d)
Standard error of prevalence, P_{SE}	√ [(P x {1 - P})/(a+b+c+d)]
Upper confidence limit of prevalence, P_U	P + (1.96 x P_{SE})
Lower confidence limit of prevalence P_L	P - (1.96 x P_{SE})
Pre-test odds, Pr	P / (1 - P)
Upper confidence limit of pre-test odds, Pr_U	P_U / (1 - P_U)
Lower confidence limit of pre-test odds, Pr_L	P_L / (1 - P_L)
Squared upper fractional uncertainty of pre-test odds, Pr_{UF}^2	[(Pr_U - Pr) / Pr] ^{2}
Squared lower fractional uncertainty of pre-test odds, Pr_{LF}^2	[(Pr_L - Pr) / Pr] ^{2}

Equations for the LRs and their fractional uncertainties (counts from table 5)	
True positive rate, TPR	a / a+c
False positive rate, FPR	b / b+d
Likelihood ratio, LR	TPR / FPR
Standard Error of LR, LR_{SE}	√[(1/a)-{1/(a+b)} +{1/c)-{1/(c+d)}]
Upper confidence limit of LR, LR_U	LR x exponent (1.96 x LR_{SE})
Lower confidence limit of LR, LR_L	LR x exponent (-1.96 x LR_{SE})
Squared upper fractional uncertainty of LR, LR_{UF}^2	[(LR_U – LR) / LR]^2
Squared lower fractional uncertainty of LR, LR_{LF}^2	[(LR_L – LR) / LR]^2

Equations for the Post-test probability and its confidence limits	
Post-test odds, Po	Pre-test odds x LR (for all tests)
Upper fractional uncertainty of Po, Po_{UF}	√{Pr_{UF}^2+(sum of LR_{UF}^2 for all tests)}
Lower fractional uncertainty of Po_{LF}	√{Pr_{LF}^2+(sum of LR_{LF}^2 for all tests)}
Upper confidence limit of PO, Po_U	Po + (Po_{UF} x Po)
Lower confidence limit of Po, Po_L	Po - (Po_{LF} x Po)
Post-test probability, PTP	Po / (1 + Po)
Upper confidence limit of PTP, PTP_U	Po_U/(1+Po_U)
Lower confidence limit of PTP, PTP_L	Po_L/(1+Po_L)

Figure 3.2 Equations relating to the calculations of the post-test probabilities (PTP) and their confidence limits. The upper left table relates to the pre-test odds and its associated errors. The upper right table relates to the LRs and their associated errors. The values from these tables are combined to give the lower table generating the Post-test probability for each outcome given a combination of tests and their combined confidence limits. All equations relating to the Bayes' calculations are given in black while those used for generating the confidence limits are given in blue.

3.2.2 Hierarchical Bayes'

The development of Hierarchical Bayes' was borne of the desire to extract the fullest support from the data gathered. Four areas of clinical decision making were identified as pivotal to the management of a patient. These areas were diagnosis, management, drug selection and the administration and duration of the drug (drug regime). It follows therefore that these areas are where decision support would be the most useful. The method in which Bayes' was applied in the present study followed a hierarchical pattern covering these areas in four stages. The stages are given below:

- **Stage one** used the clinical observations (tests) to predict a probable diagnosis (from the 76 possible diagnoses) .
- **Stage two** used the clinical observations (tests) and diagnoses to predict a management decision (from the 8 possible management options described in 2.6.3 and 2.7.4).
- **Stage three** used the clinical observations (tests), diagnoses and management decisions to predict the probable drugs prescribed (from the 125 possible drug options).
- **Stage four** used the clinical observations (tests), diagnoses, management and drugs prescribed to predict the probable administration and duration of each drug prescribed (from 130 possible administration and duration options).

Each of the four stages naturally follow each other in a hierarchical fashion. This process therefore is referred to as hierarchical Bayes'.

Given the huge number of possible combinations of the outcomes of each stage of hierarchical Bayes, the number of multilevel LRs required was potentially 125,405 (18,848 for stage one, 2,696 for stage two, 43,125 for stage three, and 61,100 for stage four). Microsoft Excel® was therefore used as a processing platform for the calculations finally generating post-test probabilities.

3.2.3 Evaluating Machine Learning

Assessing the accuracy of machine learning is pivotal to evaluating the effectiveness of any model. The accuracy was assessed by allowing the model to generate outputs or decisions for each of the 4 stages of hierarchical Bayes' (i.e. diagnoses, management, drug prescribed and administration and duration decisions) using the predictors only (variables of the EPR). The generated decisions were then compared to the actual decisions taken.

3.2.3.1 Overfitting

The performance of machine learning can be overestimated where the same data is used to both train and evaluate the model. This is because the model effectively mirrors the exact patterns in the training data, including both the useful and useless information. Therefore, when presented with new data, which may vary in or lack some information, the model fails to perform as well. This is referred to as overfitting. That is, the model over-fits the training data and is not generalisable. To avoid the optimistic performance yielded by overfitting, cross validation (the 'leave-one-out' method) can be used to partition the data prior to learning and testing (137,157). Cross-validation overcomes the problem of overfitting by splitting the data into randomly generated equal cohorts (folds). One fold is withheld for testing whilst the remaining folds are used for training a model (i.e. pre-test odds and likelihood ratios are generated for each diagnostic matrix produced for every combination of clinical observation and diagnostic outcome). The accuracy of the learning is then tested on the withheld fold. This is rotated until all the folds are used for learning and testing (i.e. cross-validation). Figure 3.3 shows the process of 5-fold cross validation. A mean value for accuracy is reported. Each fold can be stratified, that is, each fold contains an equal number of the different outcomes. This ensures a more complete learning and testing process. In general, 10-fold stratified cross-validation is recommended for the resulting low bias and variance (137). Although the intention was to use cross-validation in this study, in order to get a baseline measure of the most optimistic results expected by hierarchical Bayes, cross validation was not performed in the first instance.

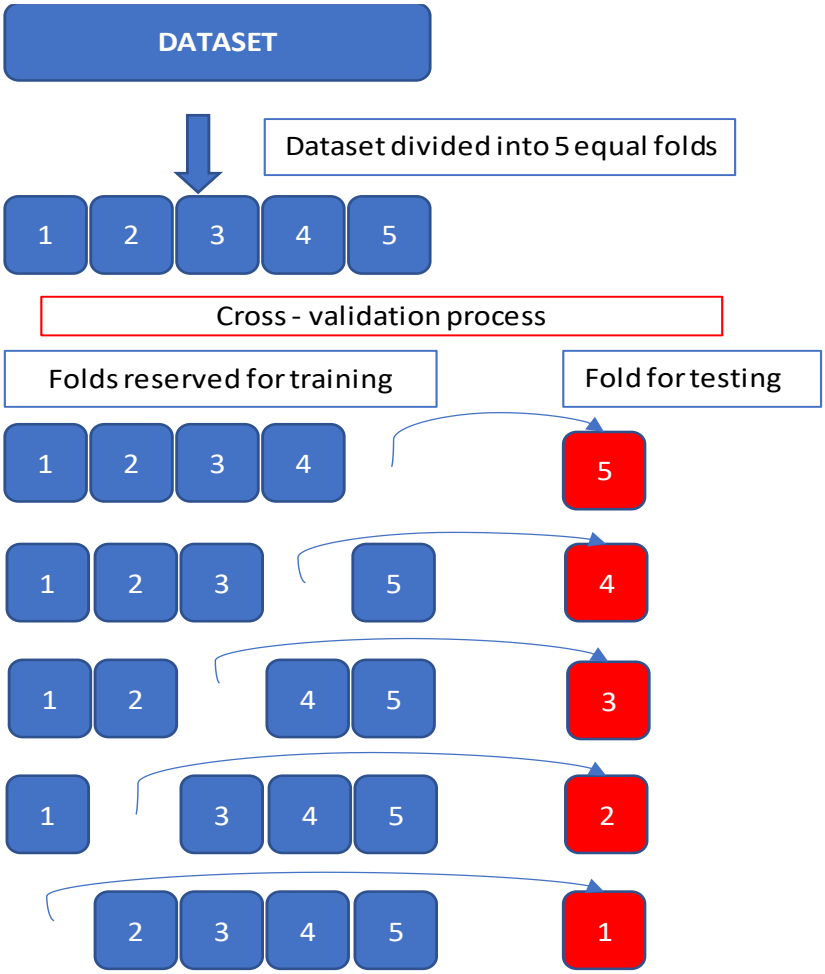


Figure 3.3 The process of 5-fold cross validation

3.2.3.2 Measuring Performance - Accuracy, Informedness and Markedness

An evaluation of diagnostic accuracy alone is insufficient as a predictor of overall system performance. This is because on considering a rare condition (X), with a prevalence of say 1%, a diagnostic accuracy of 99% would be achieved by simply excluding the condition for all cases. This is not indicative of the clinical competence of the system and would reflect chance alone. The evaluation of system accuracy therefore included estimates of weighted accuracy, informedness and markedness. Informedness is a value derived from the sensitivity and specificity, and markedness is derived from the positive and negative predictive values (table 3.3). These values are independent of the prevalence of an outcome and so are not prone to chance.

3.2.3.2.1 Informedness

Informedness quantifies how informed a predictor (e.g. a clinical test) is of a specified outcome (e.g. a diagnosis). It specifies a probability for the predictor being informed in relation to a known

outcome versus chance. Consider the prevalence of the rare condition (X) mentioned above (1%), predicting that everyone lacks this condition gave an accuracy of 99%, however the sensitivity is 0 (no cases detected) and the specificity 1 (all condition free cases detected). Following the equation in table 3.3, informedness would also be 0 (sensitivity + specificity – 1).

3.2.3.2.2 Markedness

Markedness quantifies how marked an outcome (e.g. a diagnosis) is for a specific predictor (e.g. a clinical test result). Markedness specifies a probability for the condition being marked by the predictor, again versus chance (104). For the above example the positive predictive value for rare condition X would be 0 and the negative predictive value 1. This would result in a markedness of 0 (PPV+NPV – 1), i.e. worse than chance.

Of the two measures, markedness is considered in this study to be the most relevant as it relates to the predictive value of the clinical test.

Measures of system learning	
Accuracy (counts from table 3.1)	(a + d) / (a + b + c + d)
Informedness	Sensitivity + specificity – 1
Markedness	Positive predictive value + negative predictive value - 1

Table 3.3 Equations relating to predicting system accuracy, namely accuracy, informedness and markedness.

3.2.4 The Analyses

The data was explored in the four stages of Hierarchical Bayes' as described in section 3.2.3. The pre-test odds and multilevel likelihood ratios were first generated (Appendices 1 and 5) and the resultant model was tested according to the individual analyses described below. Five analyses were carried out (including each of the 4 stages of hierarchical Bayes') to explore whether the performance of hierarchical Bayes' could be optimised. These analyses were as follows:

- **Analysis 1:**

The first analysis was designed to use all the clinical data available to create a comparative baseline for further analyses. The analysis revealed how well hierarchical Bayes' performed when trained on all 1351 patient episodes using all the predictors generated from these. As mentioned, the pre-test odds and multilevel likelihood ratios were first generated, and the resultant model was tested on the data it was trained on to give the most optimistic estimates of performance (Appendix 2). As discussed in section 3.2.3.1, circularity is a problem and can overestimate performance, however the computational demand for the performance of cross-validation prohibited its use. For the purpose of the first investigation, an optimistic estimate as a comparative baseline was felt to suffice.

- **Analysis 2:**

The second analysis attempted to streamline and optimise hierarchical Bayes' by using the most reliable or 'safe' predictors (multilevel likelihood ratios). Reliability was determined by calculating the 95% confidence limits of each multilevel likelihood ratio. As mentioned in section 3.3.1.5, a likelihood ratio of 1 has no predictive clinical value as it does not alter the probability of an outcome. Therefore, multilevel likelihood ratios with confidence limits that straddled 1 were deemed 'unsafe' due to the reduced certainty of clinical value. Only the likelihood ratios for which the confidence limits did not straddle 1, (i.e. both the upper and lower limits were either below or above 1), were labelled as "safe" and included in the analysis. Appendix 3 shows the calculations for analysis 2.

The LRs deemed as "safe" were also used to assess the learning efficiency of Bayes' as more data was collected. A gradual increase in the number of "safe" LR's would indicate continual learning of the system.

- **Analysis 3:**

The third analysis examined the generalisability of hierarchical Bayes' to different practice settings (outside APCOS practices). It is known that both the prevalence of disease (pre-test odds) and the value of clinical testing (multilevel likelihood ratios) determine the outcome (i.e. the diagnosis, management, or prescribing decision). Whilst the likelihood ratios are independent of the prevalence of disease or the decision to treat, the pre-test odds are a mirror of prevalence. Therefore, in order to investigate whether the model (hierarchical Bayes') could be applied to different practice settings, analysis 1 was repeated with the pre-test odds set to 1. This is known as using uniform or flat priors (157). That is, effectively the initial probability of any outcome was equal at 0.5. This analysis assessed whether it was sufficient to use the values of clinical testing alone to reach the correct outcome. Equivocal performance when compared to analysis 1 would indicate that hierarchical Bayes' could be applied to any clinical setting in its current form. Appendix 4 shows the calculations for analysis 3. As with the previous analyses hierarchical Bayes' was trained and tested on the same data.

- **Analysis 4:**

The fourth analysis assessed whether the performance of hierarchical Bayes' could be improved by building the model using only the data from JG, the specialist IP optometrist who contributed most of the data (68.7% of patient episodes collected). This approach was based on the rationale that consistency in the limited data used for model building could result in a more robust and therefore

better performing model than that of analysis 1. Hierarchical Bayes' was trained (appendix 5) and tested (appendix 6) on the same data.

- **Analysis 5:**

The fifth and final analysis investigated the generalisability of the model built in analysis 4. That is, the model of hierarchical Bayes' built using only the data of JG (the specialist IP optometrist who contributed most of the data – appendix 5) was tested on the data provided by the remaining three specialist IP optometrists (appendix 7). Cross-validation (section 3.2.3.1) was not performed in this study due to the complexity of the data and to yield the most optimistic results. This final analysis attempted to compensate for the lack of cross validation by testing the model on previously unseen data. Analysis 5 was expected to show the most realistic performance of the model.

Figure 3.4 shows, diagrammatically, the analyses that were performed.

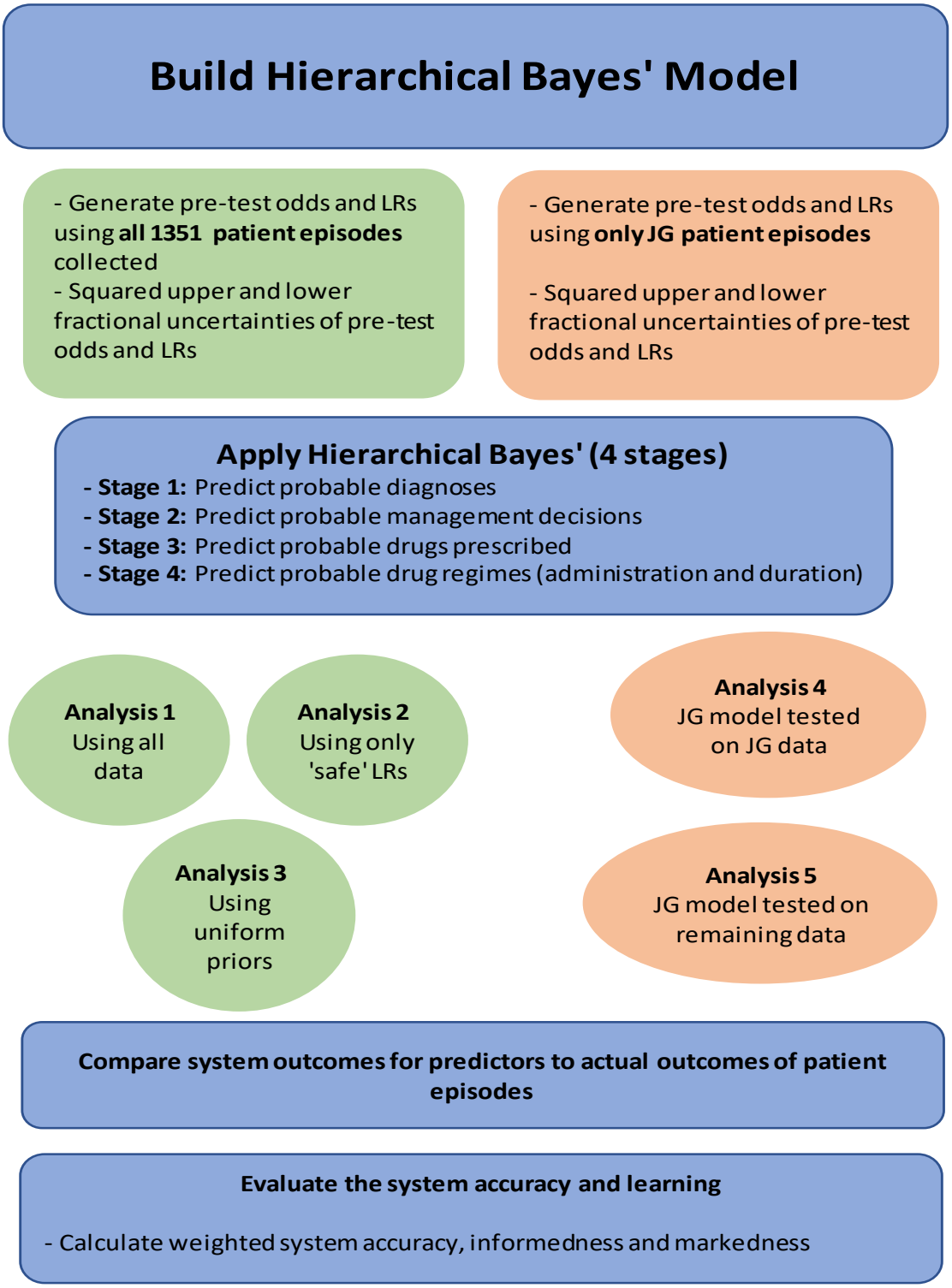


Figure 3.4 A summary of the analyses involved using hierarchical Bayes'

3.3 Hierarchical Bayes': Results and Discussion

Hierarchical Bayes' was developed to extract the fullest support from the data collected by APCOS specialist IP optometrists over the period of a year. Following the Bayes' translational learning cycle (bTLC) described in Chapter 1, the intention was to drive clinical decision support tools for both inexperienced IP optometrists and those in training. Bayes' theorem, being one of the simplest and logical mathematical methods, was used to extract the components from the data to drive such tools. As described in section 3.2.1.3 the pre-test odds (prevalence) and multi-level likelihood ratios were generated from the frequencies of each test undertaken and each clinical decision made. These values were then applied in a hierarchical fashion (section 3.2.2) to generate tentative diagnoses, management decisions, prescribing decisions (drugs), and administration and duration decisions (a drug regime). Before hierarchical Bayes' could be used to drive clinical decision support tools, it was pivotal that its performance, that is, the accuracy, informedness and markedness of the generated decisions, was evaluated against the known outcomes.

Five analyses were carried out to evaluate the performance of hierarchical Bayes'. The 1351 patient episodes collected by APCOS were used to generate the model for each analysis as described in section 3.2.4. There was no expectation that the 1351 patient episodes would be enough to maximise model performance, however a reasonably good indication of potential performance was expected.

The following section describes the results of Hierarchical Bayes' applied as the five analyses described in section 3.2 and shown in figure 3.4.

Each of the tables summarises the results for each analysis as performance measures. That is, weighted accuracy, informedness and markedness. For each stage of Hierarchical Bayes', inclusiveness is also added to the results. This shows what proportion of the data (i.e. the diagnoses, management, and prescribing options) was included at each stage of hierarchical Bayes'.

3.3.1 **Analysis 1:** Using all the data

		Weighted estimates			
Stage	**Description**	Accuracy	Informed	Marked	Inclusiveness
1	Diagnosis	0.83	0.44	0.48	0.67
2	Management	0.74	0.36	0.31	1.00
3	Drug prescribed	0.96	0.31	0.13	0.42
4	Administration & duration	0.96	0.32	0.13	0.47

Table 3.4 The results of the 4 stages of hierarchical Bayes' for analysis 1

Appendix 2 (worksheet “accuracy”) shows the full analysis from which table 3.4 has been extracted. All 1351 patient episodes were used to both generate the model of hierarchical Bayes’ and test it. The results given in table 3.4, therefore, represent the most optimistic estimates of performance. This formed the baseline to which all subsequent analyses were compared.

The weighted accuracy indicated good overall performance with values ranging from 74% for management decisions to 96% for both drugs prescribed and administration and duration. It must be noted, however, that accuracy is a limited measure in that it is not independent of chance and therefore yields the most optimistic estimates of performance (see section 3.2.3.2). The weighted informedness and markedness on the other hand are independent of chance (a value of 0% for each of these measures represents the level of chance). The results in table 3.4 show that the most optimistic estimates of the performance of hierarchical Bayes’ showed disappointingly low levels of informedness (ranging from 31% for drug prescribed, to 44% for diagnosis) and markedness (ranging from 13% for drug prescribed and administration and duration to 48% for diagnosis). These two measures gave an early indication that more data may be required to build a robust model on which any reliable clinical decision support tool could be based. Of these two measures, markedness is particularly important as the indicator used to evaluate the predictive ability of the model (as it is based upon the positive and negative predictive values). Markedness therefore will be primarily referred to when evaluating the performance of hierarchical Bayes’ for all further analyses.

The weighted markedness given in table 3.4 masks much better individual values of markedness for each clinical outcome (each diagnosis, management decision, prescribing decision, and drug regime decision). Markedness ranged from 0-100% for diagnoses, 22-50% for management decisions, 0-40% for drug prescribing decisions, and 0-50% for drug regime decisions (appendix 2, worksheet “Accuracy”). For stage 1 of hierarchical Bayes’, that is predicting of the diagnosis, 100% markedness was achieved for basal cell carcinoma, concretions, acute dacryocystitis, entropion, facial palsy, Fuch’s endothelial dystrophy, pingueculitis, steroid induced ocular hypertension and blunt trauma – all relatively rare outcomes. A good degree of markedness was also found for the more commonly presenting conditions such as KCS (71%) and uveitis (63%).

3.3.2 **Analysis 2:** Using only "Safe" Predictors

		Weighted estimates			
Stage	**Description**	Accuracy	Informed	Marked	Inclusiveness
1	Diagnosis	0.83	0.44	0.46	0.66
2	Management	0.72	0.33	0.28	1.00
3	Drug prescribed	0.95	0.26	0.10	0.42
4	Administration & duration	0.96	0.24	0.08	0.47

Table 3.5 The results of the 4 stages of hierarchical Bayes' for analysis 2

Appendix 3 (worksheet "accuracy") shows the full analysis from which table 3.5 has been extracted. The model was trained using all 1351 patient episodes and tested on the same. The results of weighted accuracy, informedness and markedness, therefore, represent the most optimistic estimates.

The second analysis attempted to streamline the calculations required in hierarchical Bayes' by using only the most reliable predictors. The multilevel LRs were deemed 'safe' if their confidence limits did not straddle 1 (as described in 3.2.4). Only these 'safe' predictors were included in the calculations. Aside from streamlining the necessary calculations, analysis 2 was also designed to determine whether using only 'safe' predictors could improve the performance of hierarchical Bayes'.

Appendix 1 ("Notes" worksheet) and table 3.6 show that only a relatively small number of the multilevel LRs were deemed 'safe' and therefore included in this analysis. For stage one of hierarchical Bayes' (the predictors of a diagnosis) only 978 (5%) of the 18848 LRs were deemed 'safe'. For stage two (the prediction of a management decision) only 274 (10%) of the 2696 LRs qualified. For stage three (predictors for drug prescribed) 1187 (3%) of the 43125 LRs were 'safe' and for stage four 1524 (2%) of the 61100 LRs were deemed 'safe'. Table 3.6 also shows how the numbers of 'safe' LRs altered with the number of patient episodes included in the analysis. This shows a very gradual increase in the number of reliable multilevel LRs across stages.

Stage		LRs Available	Usable LRs (95% CLs that do not straddle 1)					
	Patient episodes →		383		524		1351	
			Number	%	Number	%	Number	%
1	Diagnosis	18,848	590	3	649	3	978	5
2	Management	2,696	143	5	161	6	274	10
3	Drug prescribed	43,125	609	1	751	2	1187	3
4	Administration & Duration	61,100	812	1	871	1	1524	2
	Total	**125,769**	**2,154**	**1.71**	**2,432**	**1.93**	**3,963**	**3.15**

Table 3.6 The number of 'safe' multilevel likelihood ratios generated with the inclusion of more patient episodes

Table 3.5 shows the performance of hierarchical Bayes', analysis 2. The weighted accuracy, informedness and markedness were slightly reduced following the inclusion of only the 'safe' predictors. Weighted markedness fell from 48% to 46% for stage 1 (diagnoses), 31% to 28% for stage 2 (management), 13% to 10% for stage 3, and 13% to 8% for stage 4. Restricting hierarchical Bayes' to only the 'safe' predictors, therefore, did not improve performance. Of note, however, was the relatively small reduction in performance given the much-reduced number of LRs included in the analysis. Of the 125,769 LRs, only 9,963 (3%) were included, that is, a reduction of 97%. Yet, the reduction in markedness ranged from 2% to 5% across the 4 stages of hierarchical Bayes'. This represents a very small drop in performance for a very large drop in information. As such, limiting the predictors to only those deemed 'safe' could add value (by reducing the complexity of calculations and indeed the contents of the EPR) to a model of hierarchical Bayes' that initially performed better, perhaps with more data.

3.3.3 **Analysis 3:** Using Uniform Priors

Stage	Description	Weighted estimates			
		Accuracy	Informed	Marked	Inclusiveness
1	Diagnosis	0.52	0.09	0.12	0.66
2	Management	0.77	0.31	0.24	1.00
3	Drug prescribed	0.96	0.19	0.09	0.42
4	Administration & duration	0.96	0.23	0.07	0.47

Table 3.7 The results of the 4 stages of hierarchical Bayes' for analysis 3

Appendix 4 (worksheet "accuracy") shows the full analysis from which table 3.7 has been extracted. All multi-level likelihood ratios were included regardless of whether they were deemed as 'safe' or not. As with previous analyses, hierarchical Bayes' was trained and tested on the same data. The results therefore display the most optimistic estimates for weighted accuracy, informedness and markedness.

The third analysis examined the potential generalisability of hierarchical Bayes' across practice settings with differing prevalence of ocular disease and management strategies. This analysis excluded the effect of pre-test odds (or prevalence) by using uniform priors. Table 3.7 shows that using uniform priors reduced the performance of hierarchical Bayes' for all 4 stages. Weighted markedness reduced from 48% to 12% for stage 1 (diagnosis), 31% to 24% in stage 2 (management), 13% to 9% in stage 3 (drug prescribed), and 13% to 7% in stage 4 (administration and duration). Given that the greatest reduction in performance occurred at stage 1 of hierarchical Bayes' (prediction of a diagnosis), this suggests that prevalence (or pre-test odds) plays a pivotal role in determining a diagnosis and therefore the management and prescribing decisions made. This also suggests that decision support based on hierarchical Bayes' would be useful only in areas where the training data was collected. To the clinical mind, this makes perfect sense as the probability of a particular diagnosis is heavily dependent on both the results of testing but also the prevalence of the disease in the locality. In order for hierarchical Bayes' to deliver robust clinical support nationally then, training data would have to be collected from multiple sites to build the model. Hierarchical Bayes' in its current form would therefore not be generalisable.

3.3.4 **Analysis 4:** JG Model Tested on JG Data

		Weighted estimates			
Stage	**Description**	Accuracy	Informed	Marked	Inclusiveness
1	Diagnosis	0.83	0.31	0.51	0.53
2	Management	0.74	0.33	0.29	1.00
3	Drug prescribed	0.95	0.28	0.16	0.30
4	Administration & duration	0.96	0.28	0.13	0.39

Table 3.8 The results of the 4 stages of hierarchical Bayes' for analysis 4

Appendix 6 (worksheet "accuracy") shows the full analysis from which table 3.8 has been extracted. The fourth analysis evaluated whether the performance of hierarchical Bayes' could be optimised with consistency in the training data. This was achieved by training hierarchical Bayes' using data from the specialist IP optometrist who contributed the most (JG). All the predictors were used in this analysis, regardless of whether deemed 'safe' or not. The model was trained using 928 patient

episodes and it was tested on the same. The results given therefore reflect the most optimistic estimates of performance.

Table 3.8 shows a marginal improvement in markedness for stage 1 (diagnosis), increasing performance from 48% to 51%. The performance of stage 2 (management) reduced from 31% to 29%. Stage 3 (drug prescribed) showed a small improvement of 13% to 16%, whilst stage 4 (administration and duration) remained at 13%. Overall then a marginal improvement in performance was demonstrated. Table 3.7 also shows the inclusivity of the analysis, that is the proportion of the outcomes represented by the data analysed. This reduced in 3 of the 4 stages of hierarchical Bayes' (stage 1 reduced from 66% to 52%, stage 2 remained at 100%, stage 3 reduced from 42% to 30% and stage 4 reduced from 47 to 39%). A marginal improvement in performance of the model, likely due to consistency in the data, came at the expense of a reduced clinical inclusion of outcomes.

3.3.5 **Analysis 5:** JG Model Tested on Remaining Data

		Weighted estimates			
Stage	**Description**	Accuracy	Informed	Marked	Inclusive
1	Diagnosis	0.69	0.15	0.16	0.56
2	Management	0.63	0.03	0.03	1.00
3	Drug prescribed	0.93	-0.03	-0.01	0.38
4	Administration & duration	0.94	-0.01	-0.01	0.38

Table 3.9 The results of the 4 stages of hierarchical Bayes' for analysis 5

Appendix 7 (worksheet "accuracy") shows the full analysis from which table 3.9 has been extracted. Here, the data from JG (the specialist IP optometrist who contributed the most data) was used to train hierarchical Bayes'. All predictors were included regardless of whether they were deemed 'safe' or not. The resultant model was then tested on the remaining data (collected from the remaining 3 specialist IP optometrists). The model was trained on 928 patient episodes and tested on the remaining 421 patient episodes (two episodes were excluded as they did not state the identity of the examining clinician). Given the model was trained and tested on different data, the performance measures in table 3.9 give the most realistic estimates of the performance of hierarchical Bayes'.

Table 3.9 shows that weighted markedness collapsed for all 4 stages of hierarchical Bayes'. Stage 1 (diagnosis) fell from 48% to 16%, stage 2 (management) from 31% to 3%, and stages 3 and 4 from 13% to under 0% (i.e. below chance). It seemed that the model demonstrated overfitting to the

training data. Hierarchical Bayes' could therefore not be applied to other specialist IP optometrists and thus demonstrated poor generalisability at the present level of training. As in analysis 2 however, much more data from multiple practices and areas could improve the performance of hierarchical Bayes'.

3.4 Limitations

There were several limitations of hierarchical Bayes' and the presented analyses. Firstly, supervised machine learning, as demonstrated herewith, utilised Bayes' theorem to extract the mathematical components required for clinical decision replication. There are, however, many other supervised and unsupervised machine learning methods used to extract patterns from data to enable decision support. These include logistic regression, neural networks, decision trees and support vector machines (SVM) (137,138). None of these methods were compared to Bayes' and thus it is possible that these methods may have performed better given the data collected. Bayes' was chosen for simplicity of both understanding by clinicians and the application. Conditional probability is a ubiquitous concept taught in the most basic of mathematical programmes. It therefore presents a concept which clinicians are likely to be familiar with. Understanding the process by which decision support is generated is pivotal to attracting the confidence of clinicians in its use.

Secondly, the data was processed in its raw form. This study was designed to use all the clinical data provided by APCOS in an attempt to maximise clinical decision support. As such, no attempt was made to group or alter the outcomes or predictors to optimise performance. Grouping the drugs prescribed into generic forms based on the activity of the drug (for example steroids as opposed to prednisolone 1% or dexamethasone 0.1%) may have improved the performance of hierarchical Bayes' by reducing the granularity of the data. The model would therefore have less variations of outcomes to train on. This could have led to more efficient learning.

Thirdly, Bayes' theorem assumes that all predictors act independently which, for clinical observations, is clearly untrue. A result of this is that Bayes' can learn rapidly but also reach maximum learning, after which no further improvement can be observed (137). In the present study, no attempt was made to evaluate learning efficiency, it is therefore unclear whether hierarchical Bayes' reached maximal performance. Learning efficiency curves could have been plotted which aid both the assessment of learning and project how much more clinical data would be required to reach maximum learning. Plots for individual diagnoses exhibiting a high degree of markedness would have demonstrated whether maximum learning had been reached for these. In the present

study analysis 2 looked at 'safe' predictors. The gradual increase in the number of 'safe' LRs can be taken as indicative of learning (table 3.6). At all 1351 patient episodes only 3% of the LRs were deemed safe, therefore, it can be deduced that a much greater amount of data would be required to reach maximal learning and performance.

Fourthly, the patient episodes used for this study relate to first visits only. Whilst this is ideal for the prediction of the initial diagnosis, management, and drug choice, it lacks information about the success of any intervention and subsequent changes in management.

3.5 Further Work

In conclusion, the development of hierarchical Bayes' presented herewith described a first attempt (using machine learning) to extract meaningful information from the structured and unambiguous clinical records of specialist IP optometrists. Whilst the overall performance of hierarchical Bayes' was modest, a gradual increase in 'safe' LRs indicated the potential for improvement given more data. The amount of data required to optimise hierarchical Bayes' would be many times that included in this study. Therefore, it was decided that hierarchical Bayes' could not be developed any further. The calculation of fractional uncertainties (section 3.2.1.7) was deemed unnecessary at this point given the performance of hierarchical Bayes', and thus not undertaken. This was unexpected and lead to an alteration of the research objectives.

Hierarchical Bayes' performed very well in terms of markedness for some individual ocular conditions. Highly prevalent ocular conditions with a good level of markedness could be examined further with learning efficiency curves, to show whether maximal learning had been reached. Indeed, this is further explored in chapter 6.

Further, the grouping of outcomes such as drug types could also optimise the performance of hierarchical Bayes' by reducing unnecessary granulation. This is also explored in chapter 7.

The likelihood ratios generated provide a useful measure of the importance of clinical tests in relation to a diagnosis. These could be used to rank clinical testing in order to guide the clinician towards the next most appropriate test that would support (or otherwise) a diagnosis, management, or prescribing decision. This utility is further explored, in the context of clinical decision support systems, in chapter 5.

One of the ways in which the data collected from highly specialist and experienced IP optometrists can be used, is to demonstrate the scope of modern IP practice. This can be done by comparing

conditions managed and drugs prescribed to a national guideline such as the College of Optometrists' clinical management guidelines (CMGs). Indeed, the following chapter presents such an audit of the data collected.

Chapter 4: Preliminary Analyses and Assessing the Scope of IP Optometrist Practice

4.1 Introduction

Data was collected, using the EPR described in chapter 2, from the APCOS team of specialist IP optometrists in Kent over the period of one year. The four APCOS IP optometrists collected details of their new patient consultations (referred to as 'patient episodes'). This included basic demographical information, clinical findings of tests undertaken, diagnoses, management decisions and drugs prescribed (including the administration and duration of each drug or the drug regimen).

Chapter 4 presents the preliminary findings of the data collected. This is followed by a comparison of the activity of specialist IP optometrists against the recommendations set out in the College of Optometrists' Clinical Management Guidelines (CMGs). Finally, the chapter is concluded with a discussion of the findings. The chapter plan is shown in figure 4.1.

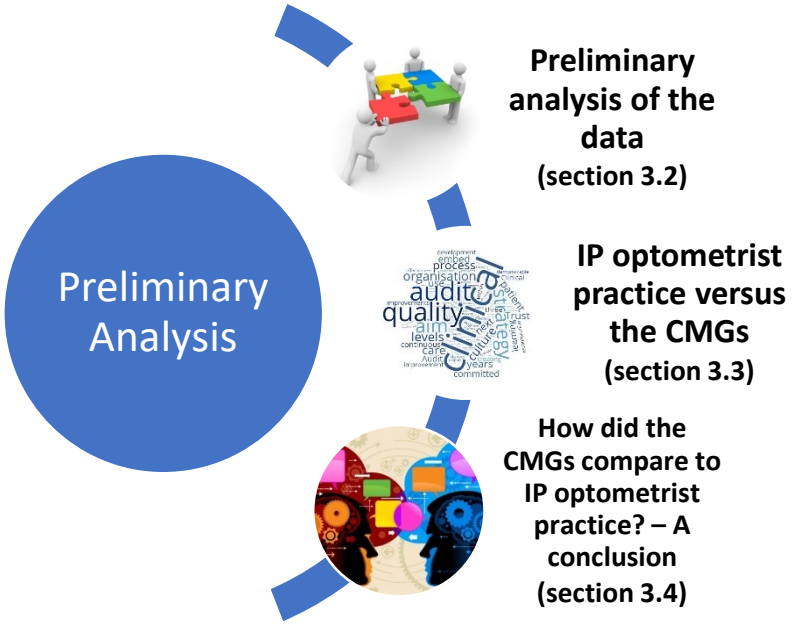


Figure 4.1 The chapter plan with relevant section numbers

4.2 Preliminary Analysis of the data

Following a 12-month period (November 2018 - October 2019) a total of 1351 new patient episodes were collected by four participating APCOS members. The patient demographics (number of patients per month, gender, age, and race distributions) can be found in appendix 1.

4.2.1 Ocular conditions

Table 4.1 shows a breakdown of the individual patient episodes sorted into the broad anatomical areas of the eye. This includes the CMG listed and manually entered conditions as well as co-morbidities. Each co-morbidity was counted separately. Two-thirds of the episodes collected involved anterior eye pathology. The work of IP optometrists therefore centres predominantly on the anterior eye. The remaining activity (34%) involved the posterior eye and other conditions (including neurological and lens-related conditions).

Activity type	Number of episodes (including co-morbidities)	Percentage of total activity
Anterior eye	887	66%
Posterior eye	300	22%
Other	167	12%

Table 4.1 A breakdown of the total activity (number of patient episodes) according to anatomical areas of the eye.

The scope of the present study was limited to anterior ocular conditions, those most likely to be treated therapeutically by IP optometrists. To this effect conditions such as cystoid macula oedema (CMO) following uncomplicated cataract surgery were excluded. However, these remain ocular conditions therapeutically treated by IP optometrists. Other than 'pigmented fundal lesions' and the various types of glaucoma (POAG, PACG and steroid induced glaucoma) no other posterior eye conditions are currently mentioned in the CMGs.

Of the 151 conditions seen over the 12-month data collection period, 80 related to anterior ocular conditions. Of the 80 conditions, 48 (60%) were CMG listed conditions, and 32 (40%) were manually entered non-CMG conditions (figure 4.2). Although this latter figure seems high at first glance, the total activity involving these conditions equated to just 65 patient episodes (4.8% of total activity). Chapter 4, appendix 2 shows the full lists of these. The manually entered anterior eye conditions are shown in table 4.2.

Manually entered anterior ocular conditions
Conjunctival retention cyst *
Limbitis *
Conjunctival concretions
Myokymia
Blepharospasm
Post-surgical inflammation *
Pseudomembranous conjunctivitis *
Post LASIK epitheliopathy *
Granuloma
Adenoviral keratitis*
Keratoconjunctivitis *
Inward turning eyelash
Infiltrative keratitis *
Post-operative corneal oedema *
Mooren's ulcer
TED (thyroid eye disease)
Post op inflammation *
Corneal compression injury
Allergy
Madarosis due to chemo
Cyst of Zeiss *
Pellucid marginal degeneration
Map dot corneal dystrophy *
Cyst of moll
Conjunctival abrasion *
Corneal ulcer
Glue from false eyelashes stuck on bulbar conjunctiva
CLARE (contact lens associated red eye)
Limbal papilloma
OCP * (Ocular cicatricial pemphigoid)
Corneal microcysts
Irregular corneal surface
Conjunctival foreign body *
Skin tag
Floppy eyelid syndrome *
Dermatochalasis
Lagophthalmos *
Squamous cell Papilloma
Canaliculitis

Table 4.2 A list of the anterior ocular conditions manually entered during data collection. Those in red indicate ocular signs as opposed to diagnoses. The asterisk shows conditions for which drugs were prescribed.

4.2.1.1 Clinical Signs versus Diagnoses

As a result of the free text manual input of ocular conditions, some clinical signs have been given in the place of diagnoses (shown in red in table 4.2). This is despite the presence of the diagnoses in the appropriate drop-down list. Clinicians can replace a given sign with the likely diagnosis when interpreting clinical records. However, using a range of names or using signs to represent a diagnosis hinders machine learning as each such "diagnosis" is considered a unique condition. Processing this data then becomes difficult as the manually entered variations in names must be sorted into their presumed clinical diagnoses before any analysis. This highlights the difficulty posed by free text inputs or the absence of standardised terms for machine learning. Davey et al. recommended a minimum data set (MDS) to promote research and data extraction from primary care optometry (135). Interestingly however, despite suggesting a comprehensive list of minimum parameters, they included many free-text fields and excluded a diagnosis, both of which are detrimental for supervised machine learning. The number of manually entered signs (instead of conditions) highlights a potential training need encouraging the standardisation of diagnostic recording by clinicians.

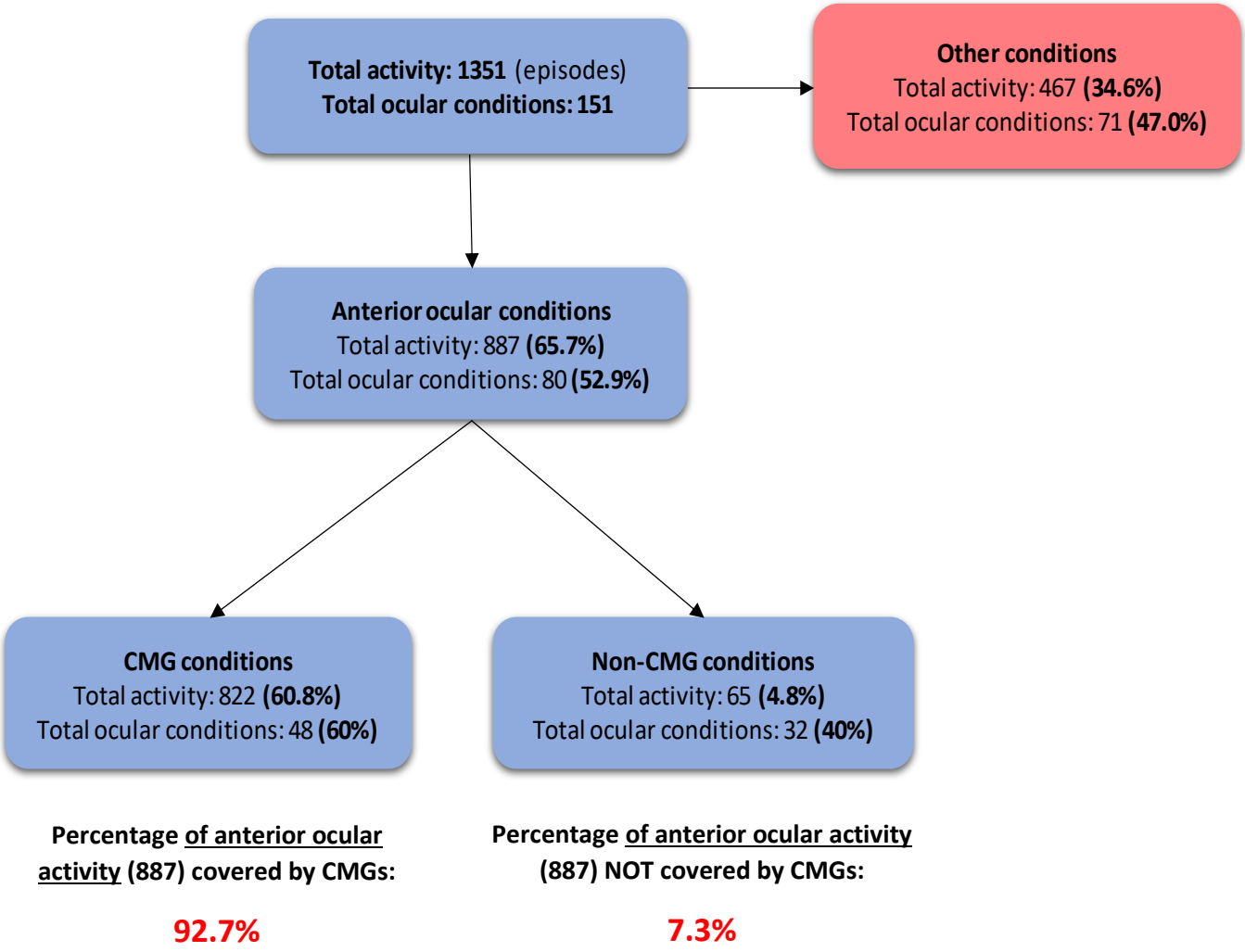


Figure 4.2 The proportions of anterior and posterior ocular conditions and activity seen over the 12-month data collection period. The lower branches divide the anterior condition numbers into CMG vs non-CMG.

Twenty-five of the 65 non-CMG anterior ocular episodes (representing 17 non-CMG conditions) resulted in therapeutic intervention (38.5%). Although uncommon, these conditions are within the scope of IP optometrist practice. These conditions are shown with an asterisk in table 4.2.

Looking at the anterior ocular activity alone, 92.7% of the patients seen fell within the CMG listed conditions (822 patient episodes). To this end the CMG conditions, though not comprehensive, reflect the vast majority of anterior ocular episodes presenting to APCOS IP optometrists. The top 20 conditions accounted for 83% of CMG-related activity (figure 4.3 - marked by the red line). These conditions therefore were the most commonly presenting anterior eye conditions to IP optometrists.

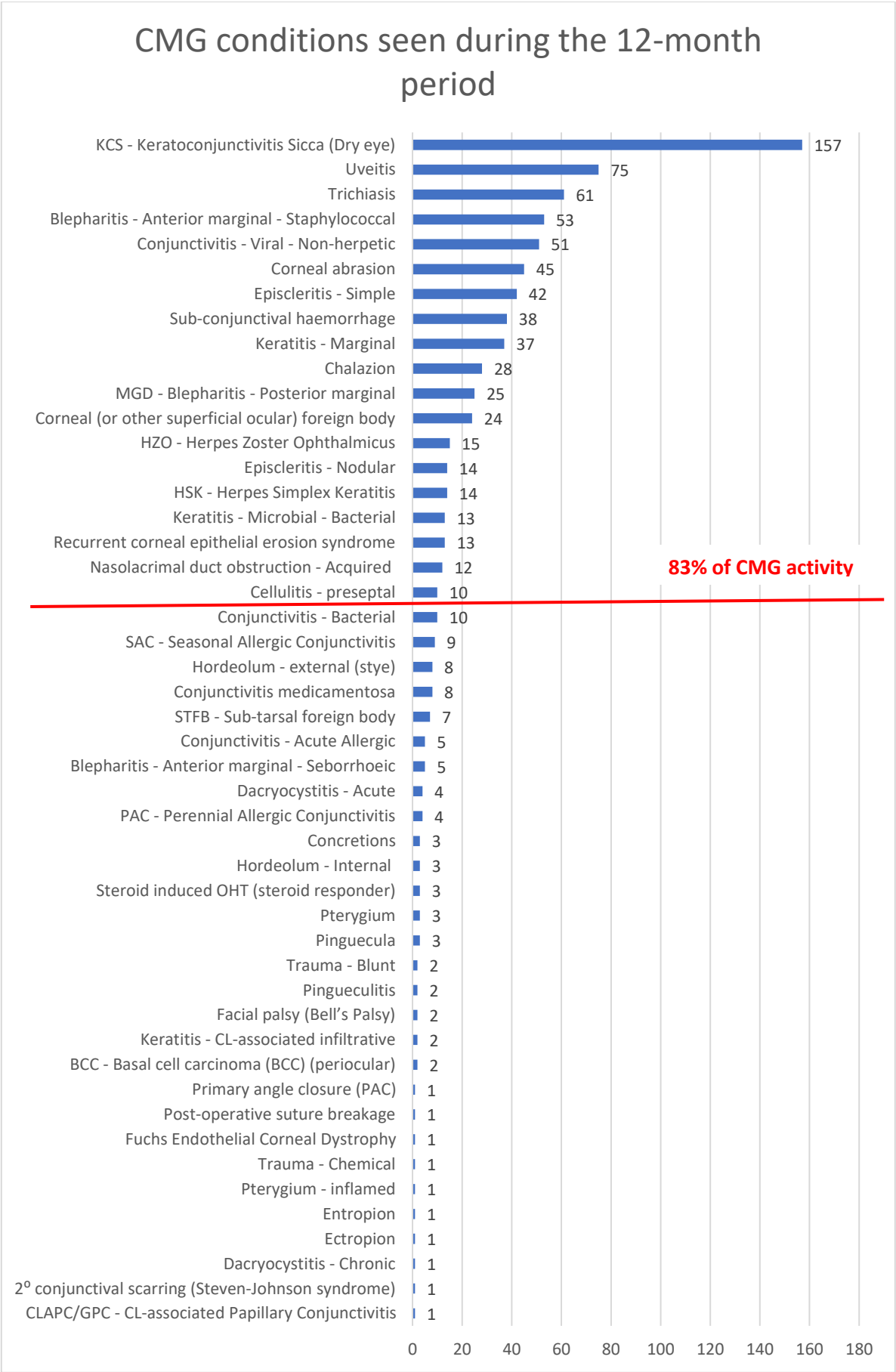
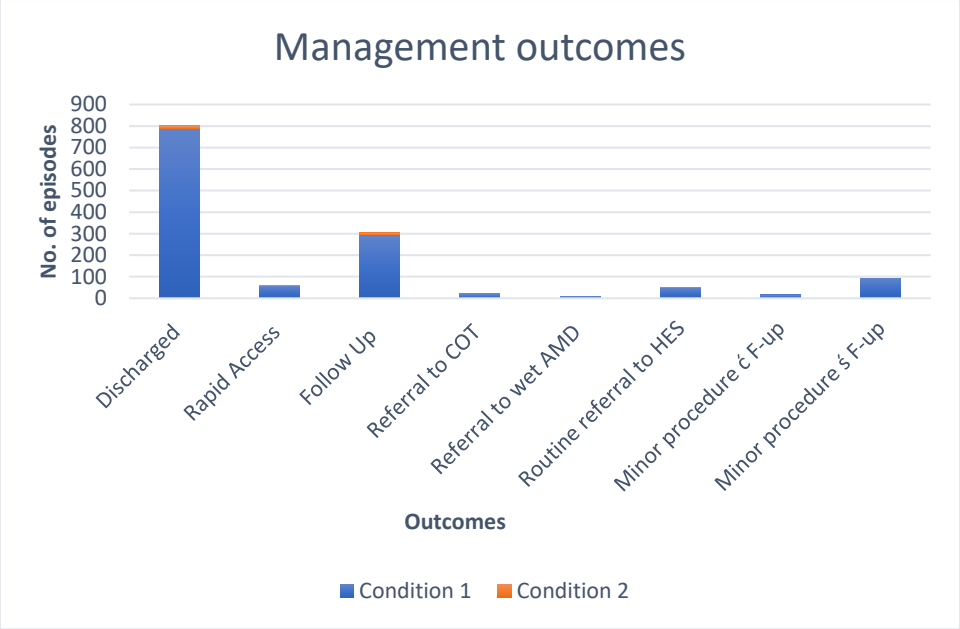


Figure 4.3 The frequency of the CMG conditions seen during the 12-month data-collection period

4.2.2 Management decisions

The patient episodes collected reported on first appointments only. Of the possible 8 outcomes (discharged, rapid access, follow-up, referral to COT (community ophthalmology team), referral to wet AMD service, routine referral to HES, minor procedure with follow-up and minor procedure without follow-up) the vast majority of patients were discharged after the first visit (65.9%). A further 23.6% were followed-up (monitored) within the APCOS service, keeping these patients out of the HES. Patients needing referral (whether routine, rapid access, to the COT or wet AMD) totalled 10.5%. Overall, 89.5% of patients were completely managed in the APCOS community setting following the first appointment (i.e. either discharged or followed-up). This number is higher than other community based schemes such as MECS/PEARS and WEHE which averaged at 80% across 14 schemes (13,158–161). However, whilst these schemes involved optometrists, the present study involved IP optometrists specifically. IP optometrists are qualified to manage a larger range of ocular conditions with greater therapeutic options. Recall that the referrals into the APCOS originate from both GPs and other optometrists. These patients would have therefore presented to the HES without this service. This represents a very effective refinement of referrals by specialist IP optometrists. A limitation of this study is that of those followed-up, it is unclear how many patients needed referral upon subsequent visits. However, given the ever-increasing demand of hospital eye services (HES)(162), an effective referral refinement or patient-triage pathway involving community optometrists could increase the provision of eyecare without the proportional increase in cost (158,163). Figure 4.4 shows the frequency of management decisions graphically followed by a table of the percentages of each broad outcome.



Total cases discharged	65.9%
Total cases follow-up	23.6%
Total Referrals	10.5%

Figure 4.4 The management outcomes of all 1351 episodes collected given as condition 1 and condition 2 (representing the primary condition and the comorbidity respectively). The total percentage of referrals represents a combination of all possible referral routes.

4.2.3 Drugs prescribed

During the data collection period 812 drug prescriptions were issued by the APCOS team. Figure 4.5 shows the frequency of prescriptions issued across categories based on the clinical action of the drugs. Ocular lubricants were the most prescribed at 37% of all prescriptions issued.

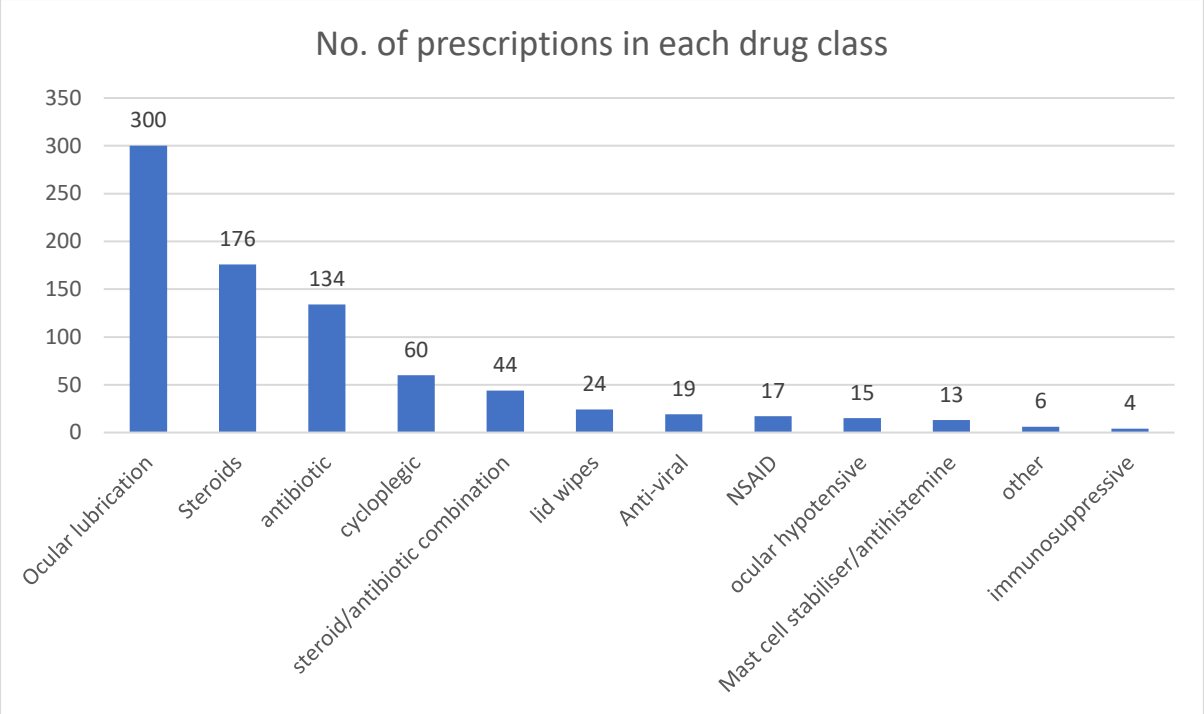


Figure 4.5 The frequency of drug prescriptions issued according to the clinical action of the drug

Of the 124 drugs listed in the data-collection EPR, 52 were prescribed (41.9%). These are given, with their individual counts, in figure 4.6. Prednisolone 1% drops (Pred Forte) accounted for the greatest number of prescriptions at 95 (11.7%). Although the most prescribed drug type was ocular lubrication, the most commonly prescribed lubricants (P/F Sodium Hyaluronate and Carbomer 0.2%) appeared lower down the individual drug count. Though the use of lubricants is recommended for conditions such as KCS (the most commonly encountered ocular condition), the effectiveness of one drop over another remains unproven (164,165). Clinicians therefore prescribe from a range of available formulations (65 mentioned in the BNF (166)). This explains the variety of ocular lubricants prescribed (17 different lubricants). Prescriptions are based upon an individual clinicians' experience or willingness to try new formulations and not on formal clinical guidance.

Chapter 4, appendix 3 shows the complete lists of ocular conditions (primary presenting and co-morbidity) and drugs prescribed along with their individual counts.

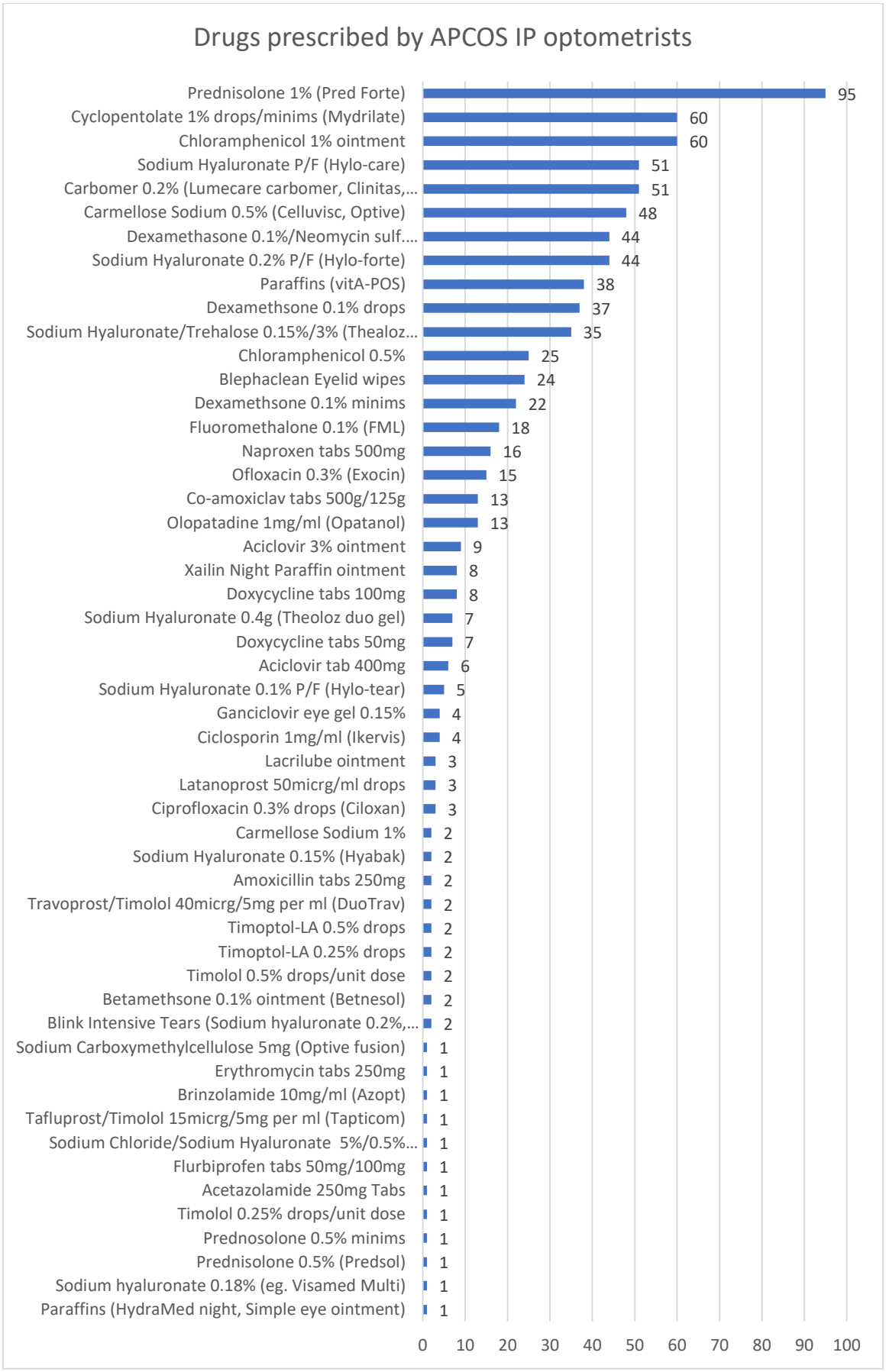


Figure 4.6 The number of prescriptions per drug issued during the 12-month data-collection period

4.3 IP optometrist practice versus the CMG recommendations

The Clinical Management Guidelines (CMGs) were created by the College of Optometrists in 2008 to coincide with the new IP status of optometrists. The Commission for Human Medicines (CHM) stated that these guidelines would support an understanding of the extent of optometrist prescribing (167). The practice of IP optometrists is restricted only by the statement "optometrist independent prescribers should be able to prescribe any licenced medicine for ocular conditions, affecting the eye and adnexa, within the recognised area of expertise and competence of the optometrist" (168). Prescribing therefore does not involve the use of a specific formulary. Instead, IP optometrists have full access to the eye section of the BNF. Systemic drugs for the management of ocular conditions (excluding controlled drugs and medicines for parenteral administration) may also be prescribed. The College of Optometrists' formulary can be used in conjunction with the CMGs for prescribing guidance. These two College resources are available to all College members (prescribing and non-prescribing). The purpose of the guidelines is to promote an equitable provision of evidence-based care, minimising clinical errors. The CMGs can be drawn upon in the event of a GOC fitness to practice or civil case to demonstrate the scope of normative practice. It is therefore of paramount importance that they are truly reflective of current IP optometrist practice in addition to being evidence-based in terms of clinical recommendations.

How reflective are the guidelines of the true and fast-evolving clinical scope of optometrists? Chapter 2 (section 2.5.2) discussed how a 10-year audit (unpublished data) suggested the CMGs were representative of the anterior eye conditions commonly seen by IP optometrists. Indeed, the data collection EPR listed the CMG conditions for this reason. Section 4.2.1 shows how the present study confirms this finding.

In the following section the clinical activity of APCOS specialist IP optometrists (over the 12-month data collection period) is compared to the CMG recommendations. Novel metrics were developed to examine compliance to CMG guidelines relating to anterior ocular episodes. The clinical elements of the examination, management decisions and therapeutic decisions were compared using these novel metrics.

4.3.1 Naïve test rates and compliance to guidelines

Generally, studies report compliance to guidelines by reporting the naïve test rates of recommended tests (variables). That is, the number of times a recommended test is carried out on a patient with a known condition. This has also been described as adherence to guidelines (169). However, there is a

fundamental flaw to this way of measuring adherence. It does not consider the number of times tests *not recommended* were undertaken for the same patient. Take for example patients presenting with acute anterior uveitis. A clinician may have performed the recommended slit lamp examination to check for anterior chamber activity in all cases. The naïve test rate for this parameter would therefore be taken as 100% or complete adherence to guidelines. The clinician may have however, also conducted 10-2 visual fields tests on all cases (not-recommended as of no diagnostic value). This departure from guidelines has no bearing on the overall recorded adherence. Therefore, so long as clinicians perform all tests available all of the time, adherence would always be excellent. Guidelines, however, promote optimal patient investigation and care. As such, at a time of limited resources, the adherence to guidelines should measure the appropriateness of eye care delivery, including when tests are undertaken without a need.

There is a need for a standardised method to measure the appropriateness of eyecare delivery against guidelines (170). Novel metrics were developed in response to this need. These metrics were used to evaluate the clinical activity of the APCOS IP optometrists against the CMGs. The naïve test rates (how many times the test was undertaken, regardless of outcome) for each of the 76 variables on the data collection-EPR were also calculated (figure 4.7).

4.3.1.1 Adherence

A novel metric, adherence, was developed. Figure 4.7 shows the components and formulae for both the naïve test rate and adherence. Adherence considers how many of the tests undertaken were as recommended by a guideline. Consider a cohort of 100 patients composed of 40 symptomatic dry eye patients and 60 normal patients. Tear function was measured in 35 of the symptomatic dry eye patients. It was also measured in 10 of the asymptomatic patients. The naïve test rate is therefore 45% (35+10/100). This rate ignores the fact that testing for tear function is recommended in dry eye patients according to the appropriate CMG (171).

In the example above, tear function was measured (as recommended) in 35 of the 40 symptomatic dry eye patients. It was not measured (again, as recommended) in all 50 of the 60 normal patients. The adherence to the CMG is therefore:

$$\text{Adherence} = \frac{(T_R + N_{R\prime})}{E}\ x\ 100$$

$$= \frac{35+50}{100}\ x\ 100 = 85\%$$

This number reflects how close clinical activity is to a guideline. Despite the naïve test rate being low, 85% of patients were tested in accordance with the CMGs.

Total number of episodes:	E
Total episodes in which the test was undertaken:	E_T
Total episodes in which test was recommended in CMG:	E_R
Total episodes in which test was NOT recommended in CMG:	$E_{R'}$
Total episodes in which test was undertaken when recommended (CMG):	T_R
Total episodes in which test was undertaken when NOT recommended (CMG):	$T_{R'}$
Total episodes in which test NOT undertaken when recommended (CMG):	N_R
Total episodes in which test NOT undertaken when NOT recommended (CMG):	$N_{R'}$
Naïve test rate (%)	$\left(\frac{E_T}{E}\right) x\ 100$
Adherence (%)	$\frac{(T_R + N_{R'})}{E} x\ 100$

Figure 4.7 The components and formulae needed to calculate naïve test rates and adherence

Whilst adherence evaluates how often clinical testing was appropriate, it does not explore non-adherence to guidelines. Non-adherence can either be the under-use of a recommended test or the over-use of a non-recommended test. Two further metrics were developed to explore these aspects of non-adherence.

4.3.1.2 Overperformed test rate

The overperformed test rate refers to tests not required by the CMGs but carried out by the IP optometrists. The overperformed test rate is calculated using the formula in figure 4.8 (the components are listed in figure 4.7). This test rate highlights the clinical variables IP optometrists may have deemed of diagnostic value despite not being mentioned in the CMGs.

4.3.1.3 Underperformed test rate

The underperformed test rate refers to those tests required by the CMGs but not carried out by the IP optometrists. This includes tests completely omitted or conducted on only some patients. The underperformed test rate is calculated using the formula shown in figure 4.8.

Over-performed test rate (%)	$\left(\frac{T_{R\prime}}{E_R + E_{R\prime}}\right) x\ 100$
Under-performed test rate (%)	$\left(\frac{N_R}{E_R + E_{R\prime}}\right) x\ 100$

Figure 4.8 The formulae needed to calculate the overperformed and underperformed test rates

Using the example of tear function testing for dry eye patients given above, the following values for over and under-performed test rates can be calculated:

Total episodes in which test was recommended in CMG (E_R): 40

Total episodes in which test was NOT recommended in CMG ($E_{R'}$): 60

Total episodes in which test was undertaken when NOT recommended ($T_{R'}$): 10

Total episodes in which test NOT undertaken when recommended (CMG) (N_R): 5

$$\text{Over-performed test rate} = \left(\frac{T_{R\prime}}{E_R + E_{R\prime}}\right) x\ 100$$

$$= \left(\frac{10}{40+60}\right) x\ 100 = 10\%$$

$$\text{Under-performed test rate} = \left(\frac{N_R}{E_R + E_{R\prime}}\right) x\ 100$$

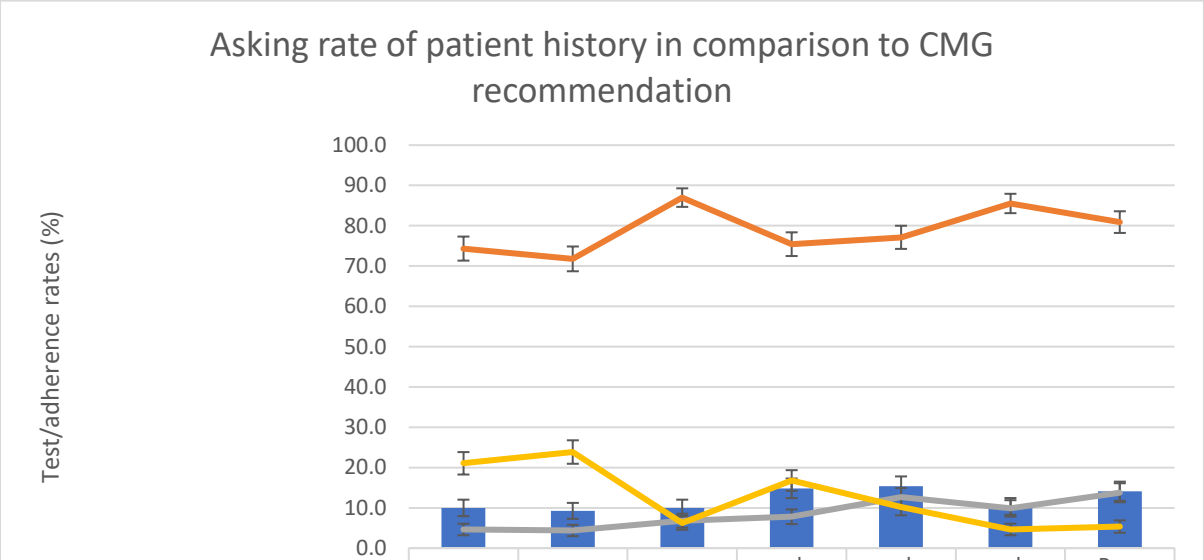
$$= \left(\frac{5}{45+60}\right) x\ 100 = 5\%$$

These values show that the testing for tear function was slightly under-performed in patients for which the test was recommended in the CMG (5%). However, in patients with no indication for the test it was over-performed by 10%. Over-performance of the test was therefore the primary reason for non-adherence to the guidelines.

The adherence along with naïve, over-performed and under-performed test rates were calculated for all 76 clinical parameters across all 74 CMG conditions. These were grouped into those covering the patient history, symptoms, and clinical signs as reported by the IP optometrists.

4.3.2 Patient history and symptoms

Figure 4.9 shows the naïve test rates of the parameters related to patient history across the CMG conditions (blue bars). Superimposed are lines to show the adherence, over and underperformed test rates when compared to the CMGs. On average, an element of patient history was asked in 12% of cases. This appears surprisingly low especially since it has been shown that a diagnosis can be reached by simply evaluating patient history (172,173). However, the occasions on which the patient history was reported show a high level of adherence (mean average: 78.9%) and low over and under-performed test rates (mean averages: 8.6% and 12.6%, respectively).

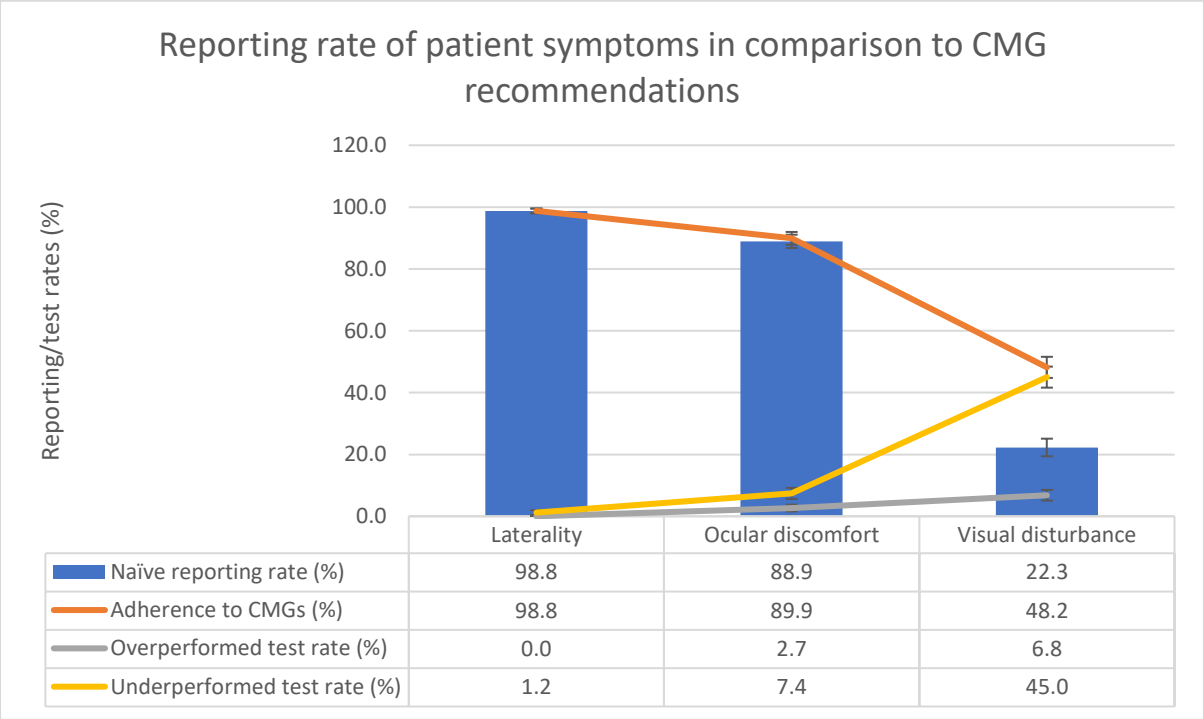


	Atopia	Cls wear	FB	ocular medicine	ocular trauma	ocular surgery	Prev Resp TI
Naïve test rate (%)	10.0	9.2	10.0	14.8	15.3	10.3	14.1
Adherence to CMGs (%)	74.3	71.8	87.0	75.4	77.1	85.5	80.9
overperformed test rate (%)	4.6	4.4	6.8	7.8	12.7	9.9	13.7
Underperformed test rate (%)	21.0	23.8	6.2	16.8	10.2	4.6	5.4

Figure 4.9 The naïve asking rates, adherence, overperformed and underperformed test rates for the elements of patient history. Abbreviations: contact lens wear (Cls wear), foreign body (FB), previous respiratory tract infection (Prev Resp TI). Error bars represent the upper and lower 95% confidence limits (174).

Figure 4.10 shows the reporting rate of patient symptoms and laterality. The IP optometrists were asked to record the clinical findings of the worst eye or the right eye for symmetrical ocular conditions. Arguably then, laterality was all but imposed upon the IP optometrists and the high reporting rate reflects this. Equally the CMGs mention laterality in every case, this is shown by an overperformed test rate of zero. Laterality is therefore of great importance in the diagnosis and management of ocular diseases presenting to IP optometrists.

The most reported primary symptom was ocular discomfort. This was investigated in accordance with the CMGs (88.9% adherence). The low underperformed test rate (7.4%) indicated that the IP optometrists largely agreed with this recommendation, and it was reflected in their practice. Visual disturbance however showed a much smaller naïve reporting rate (22.3%) with around half of these deemed appropriate (adherence 48.2%). For 45% of the patient episodes, a symptom of visual disturbance was not investigated despite being mentioned in the associated CMGs. Symptoms are subjective and whilst the CMGs report all the possible symptoms, it is expected that patients will not exhibit all those listed. However, there is no hierarchy of symptoms given in the CMGs and equal importance is placed on both ocular discomfort and visual disturbance. The data shows that a patient presenting with a CMG condition will most likely be investigated for ocular discomfort.



	Laterality	Ocular discomfort	Visual disturbance
Naïve reporting rate (%)	98.8	88.9	22.3
Adherence to CMGs (%)	98.8	89.9	48.2
Overperformed test rate (%)	0.0	2.7	6.8
Underperformed test rate (%)	1.2	7.4	45.0

Figure 4.10 The naïve reporting rates, adherence, overperformed and underperformed test rates for patient symptoms and laterality. Error bars represent the upper and lower 95% confidence limits (174).

4.3.3 Clinical signs (variables)

Chapter 4, appendix 4 shows the naïve test rates of all 66 clinical signs across the CMG episodes against their adherence, over and underperformed test rates. The naïve test rates were highly variable and ranged from 0.1% (facial chemical burn) to 63.5% (corneal staining with sodium fluorescein dye). Most of the 66 clinical tests showed an adherence of over 75% (mean adherence 93.8%, mean underperformed test rate 3.4%, mean overperformed test rate 2.8%). Most clinical

tests carried out therefore reflected the CMG recommendations for the respective ocular conditions. There were however two notable exceptions:

4.3.3.1 Conjunctival/episcleral hyperaemia

In 61.8% of cases IP optometrists examined for hyperaemia in agreement with the CMG recommendations. For just under a quarter of patients (24%) the associated CMGs recommended examining for the sign. However, this was either not done at all or was not done for all the episodes. Table 4.3 shows the ocular conditions for which there was an under-performance (<50% of cases were checked for hyperaemia). Although conjunctival/episcleral hyperaemia is a sign mentioned in the CMGs for these conditions, an under-performance suggests it may not be the most diagnostic sign and therefore of limited clinical value.

Ocular condition	Total episodes	Total tested
Pinguecula	3	0 (0%)
Pterygium - inflamed	1	0 (0%)
Trichiasis	61	3 (4.9%)
Chalazion	28	4 (14.3%)
MGD – blepharitis – posterior marginal	25	7 (28.0%)
Corneal abrasion	45	13 (28.9%)
Blepharitis – anterior marginal - staphylococcal	53	19 (35.9%)
SAC – seasonal allergic conjunctivitis	9	4 (44.4%)

Table 4.3 The ocular conditions for which there was under-testing (<50%) of conjunctival/scleral hyperaemia

In a further 14.2% of episodes the level of conjunctival/episcleral hyperaemia was examined despite not being mentioned as a sign in the associated CMG (overperformed test). Table 4.4 shows these conditions alongside their episodes and testing rates (where the testing rate exceeded 50% - or half of all episodes). Upon further analysis 2 of the CMGs relating to these conditions, although do not mention conjunctival/episcleral hyperaemia specifically, do mention other terms alluding to bulbar redness. These are given in the final column of the table. The Encyclopaedia of Ophthalmology refers to conjunctival hyperaemia as "…is a conjunctival reaction that appears as a dilatation and redness of conjunctival vessels. The pattern of hyperemia (*sic*) often appears with the greatest redness at the fornices and fades moving towards the limbus" (175). It seems, although hyperaemia is a commonly used term, it has a broad definition and perhaps confused in clinical usage. Hyperaemia is likely used to cover many different types of bulbar redness, including conjunctival/episcleral. Indeed, for uveitis the CMG refers to "hyperaemia: circumcorneal ('ciliary injection')", a description that although excludes the terms conjunctival or episcleral still denotes a form of bulbar redness. To reflect IP

optometrist activity, conjunctival/episcleral hyperaemia should be considered for addition to the guidelines of conjunctivitis medicamentosa, corneal (or other superficial) foreign body and HZO (table 4.4).

Ocular condition	Total episodes	Total tested	Episodes where sign was present	Other terms in CMG possibly confused for bulbar redness
Conjunctivitis Medicamentosa	8	7 (87.5%)	7 (87.5%)	Not mentioned
Pterygium	3	2 (66.7%)	2 (66.7%)	Scarring, thickening and distortion of bulbar conjunctiva
Hordeolum - internal	3	2 (66.7%)	1 (33.3%)	Not mentioned
Corneal (or other superficial) FB	24	13 (54.2%)	12 (50.0%)	Not mentioned
HZO – Herpes Zoster Ophthalmicus	15	8 (53.3%)	7 (46.7%)	Mucopurulent conjunctivitis

Table 4.4 The ocular conditions for which IP optometrists tested for conjunctival/episcleral hyperaemia despite this term not specifically being used in the CMGs

4.3.3.2 Epiphora

For 66.4% of patient episodes relating to epiphora, IP optometrists examined for this sign as recommended by the associated CMGs. The overperformed test rate was low at 5.1%. The underperformed test rate (the test was mentioned in the CMG but not examined by the IP optometrists) however was relatively high at 28.5%. Table 4.5 shows the conditions for which testing was low. Clearly in these cases epiphora, albeit a likely sign/symptom, is not a diagnostic feature. This may account for the lack of examination or reporting.

Ocular condition	Total episodes	Total tested
Trichiasis	61	0 (0%)
Uveitis	75	0 (0%)
Trauma - blunt	2	0 (0%)
Keratitis – microbial - bacterial	13	1 (7.7%)
Recurrent corneal epithelial erosion syndrome	13	2 (15.4%)

Table 4.5 The ocular conditions for which there was under-testing of epiphora/lacrimation

Interestingly, 9 of the 14 conditions for which epiphora was examined did not mention the sign in the associated CMG. Whilst most of these involved small numbers (1-6 episodes), one of these conditions, keratoconjunctivitis sicca (KCS), returned 28 episodes. Although this accounts for only 17.8% of total KCS-related activity, it also accounts for 47% of epiphora-related activity. It may therefore be useful to add this sign to the CMG reflecting the clinical findings (tests undertaken) by experienced IP optometrists. The concept of reflex tearing/epiphora in KCS is nothing new (176,177).

4.3.4 Clinical signs (variables) not listed in the CMGs

The ocular conditions for which there were 5 or more patient episodes were further analysed. This amounted to 24 of the 74 CMG listed conditions. For each of these conditions the clinical history, symptoms and signs were examined in more detail to identify variables that were missing from the associated CMG. The naïve test rates were evaluated, and where this was over 50% this clinical variable was deemed as important for diagnosis.

4.3.4.1 Missing variables

One clinical variable was missing in 3 conditions. These conditions were herpes zoster ophthalmicus (HZO), herpes simplex keratitis (HSK) and pre-septal cellulitis.

For HZO the missing variable was the symptom of ocular discomfort (naïve test rate 93%). Although the associated CMG mentions “pain and altered sensation of the forehead on one side” (178) it does not mention ocular discomfort specifically.

For HSK, conjunctival/episcleral hyperaemia was missing (naïve test rate 93%). This sign is however evident in the photograph given alongside the description. In 13 of the 14 cases, conjunctival hyperaemia was reported as either mild or severe. This sign was added to the CMG wording at the latest update (27/08/2021 – “redness”) and as such is reflective of the results found herewith.

Finally, for pre-septal cellulitis the single variable missing was conjunctival/episcleral hyperaemia (80% naïve test rate). This clinical sign was present in all the cases for which it was examined.

Despite the specialist IP optometrists investigating these variables regularly, they may be deemed of limited diagnostic importance for the given conditions. Perhaps, therefore, they are justified in their absence from the CMGs. However, without evidence to support the diagnostic importance of tests for given conditions, this conclusion is difficult to justify.

4.3.4.2 Seasonal Allergic Conjunctivitis (SAC)

The biggest disparity in the clinical variables was shown by seasonal allergic conjunctivitis (SAC). There were 11 variables for which the naïve test rate was ≥50%. Of these, 7 were not mentioned in the associated CMG. These are listed in table 4.6. The naïve test rates for these variables were high, indicating their potential discriminative ability for this condition. The CMGs generally list clinical variables positively associated with a given condition, that is positive test findings. Further analysis of the missing variables revealed that in most cases these investigations returned a negative result, i.e. the sign looked for was absent (table 4.6). A negative test result can be just as diagnostic as a positive. The CMG covering SAC therefore lacks these variables because they are not signs associated with the disease. However, this limits the usefulness of the CMGs for a diagnosis of exclusion such as SAC. The CMGs do list differential diagnoses. Two of the variables investigated appear to relate to the differential diagnoses given in the associated CMG. In the context of SAC, a history of contact lens wear relates to CLAPC or microbial keratitis, and a history of ocular medication to conjunctivitis medicamentosa. In the author's experience a history of previous respiratory tract infection would relate to viral conjunctivitis, tear deficiency to dry eye (KCS), A/C cells or flare to uveitis and a high IOP to primary angle closure. These conditions are not listed as differential diagnoses; however, they are clearly excluded by the IP optometrists. They could therefore be added to the CMG to reflect the thought process and practice of IP optometrists.

Hx, Sx or sign	Naïve test rate	Negative result rate ("No")
Laterality	100%	-
History: Contact lens wear	56%	56%
History: Ocular medication	67%	44%
History: Previous resp. tract infection	56%	56%
Tear deficiency	56%	22%
A/C cells or flare	56%	56%
High IOP	56%	56%

Table 4.6 The naïve test rates (≥50%) for 7 reported history, symptoms, and clinical signs associated with SAC alongside the negative results. This table shows that most of the above parameters yielded negative test findings.

4.3.4.3 Can Variables be Added to the CMGs?

In conclusion, the CMGs are representative of the most commonly encountered anterior ocular conditions by IP optometrists (92.7% of activity).

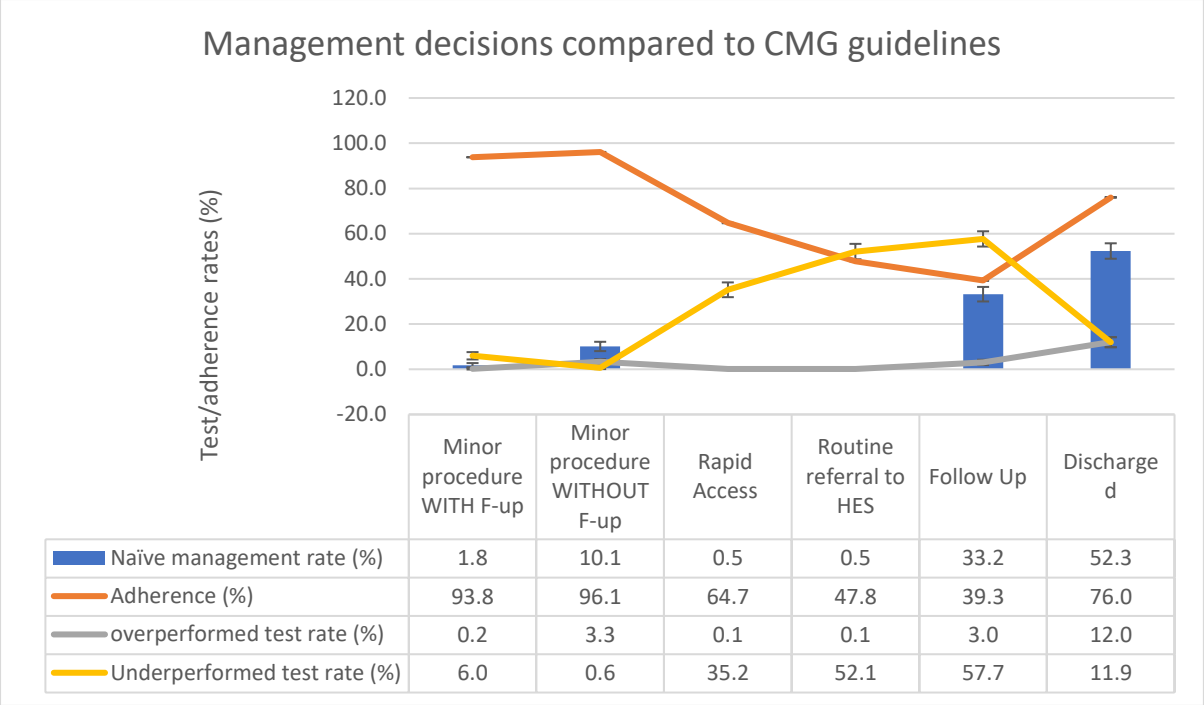
IP optometrist activity relating to the investigation of history, symptoms and clinical signs was highly appropriate when compared to the CMGs. Areas where the CMGs could be updated to further support IP optometrist practice include:

- Including a hierarchy of symptoms relating to the most encountered (ocular discomfort being the most common)
- Conjunctival/episcleral hyperaemia could be considered for addition to the guidelines of conjunctivitis medicamentosa, corneal (or other superficial) foreign body and HZO
- Epiphora could be added to the clinical signs/symptoms of KCS
- Viral conjunctivitis, uveitis and primary angle closure could be added to the differential diagnoses of SAC
- The addition of negatively associated findings for ocular conditions

4.3.5 Management decisions

Figure 4.11 shows the naïve rates for the management decisions alongside the adherence, over and underperformed rates. The management decisions shown represent 98.4% of episodes. The remaining 1.6% of patients were referred to the Community Ophthalmology Team (COT). This is a glaucoma refinement pathway unique to Kent and as such was not comparable to the management options given in the CMGs. They have therefore been removed from this analysis.

Several management options were given for each CMG condition depending on the severity. The overall adherence for management decisions ranged from 39.3% to 96.1% (figure 4.11). The primary reason for non-adherence was underperformance (range 0.6%-57.7%). Considering individual management decisions, the results showed good adherence for minor procedures with and without follow-up (93.8% and 96.1% respectively). Adherence for discharge was good at 76%. However, there were underperformances for referrals to the HES (rapid access 35.2%, routine referral 52.1%) and follow-ups within APCOS (57.7%). The most favoured management option by IP optometrists was to discharge as opposed to follow-up or refer.



	Minor procedure WITH F-up	Minor procedure WITHOUT F-up	Rapid Access	Routine referral to HES	Follow Up	Discharged
Naïve management rate (%)	1.8	10.1	0.5	0.5	33.2	52.3
Adherence (%)	93.8	96.1	64.7	47.8	39.3	76.0
overperformed test rate (%)	0.2	3.3	0.1	0.1	3.0	12.0
Underperformed test rate (%)	6.0	0.6	35.2	52.1	57.7	11.9

Figure 4.11 The naïve reporting rates, adherence, overperformed and underperformed rates for management decisions. Abbreviation: follow-up (F-up). Error bars represent the upper and lower 95% confidence limits (174).

The management decisions for ocular conditions showing more than 5 episodes were further evaluated. These constituted 24 of the 74 CMG listed conditions. A full table of management results in comparison to the CMG recommendations is given in Chapter 4, appendix 5.

Overall, the management decisions taken by the APCOS specialist IP optometrists mirrored those recommended in the CMGs. Of the 24 ocular conditions identified, 21 followed the management protocols as described (83%). One condition, namely herpes zoster ophthalmicus (HZO), with 15 cases, showed partial agreement. The associated CMG recommends “Management to resolution if co-managed with GP and keratitis limited to epithelium” (178). However, 87% of HZO cases were discharged at the first visit by the specialist IP optometrists. Although this likely implies minimal ocular involvement, it is unclear whether these patients were in fact discharged to the GP as this was not given as a management option.

The remaining two ocular conditions showed clear disagreement with the CMG recommendations. These are shown in table 4.7 and discussed individually.

Ocular condition	CMG Management Recommendation	No. of cases	APCOS management outcomes
Keratitis – Microbial (bacterial)	Emergency referral to ophthalmologist; no intervention	13	Rapid Access (HES): 15% Follow-up: **85%**
Cellulitis – pre-septal	Emergency (same day) referral to ophthalmologist; no intervention	10	Follow-up: **100%**

Table 4.7 Two ocular conditions for which management outcomes did not reflect the CMG recommendations

4.3.5.1 Keratitis – Microbial (bacterial)

The CMG associated with microbial keratitis (MK) recommends warning patients to keep contact lenses or lens cases for possible culture. It highlights that beginning any antimicrobial therapy without laboratory evaluation may delay the correct diagnosis and thus any specific treatment. Although 15% of cases of MK were referred via a rapid access pathway, 85% were followed-up within the APCOS service. This represents a significant diversion from the CMG recommendation.

Microbial keratitis is a sight threatening condition if left untreated or inadequately treated. Although microbial culture is not readily available to community optometrists, the initiation of treatment, which usually involves topical antimicrobial therapy, is within the remit of IP optometrists, particularly when concerning low risk cases. Interestingly, microbial cultures are not routinely collected in hospital settings unless either the infection is unresponsive to initial broad spectrum treatment, the infiltrates are larger than 2mm or in the visual axis or an unusual organism is suspected based on history or examination (177). Arguably then the issuing of broad-spectrum antibiotics (mono or dual therapy), a cycloplegic, ocular hypotensive treatment (if required) and oral analgesics are all within the scope of the specialist IP optometrist. Close observation would be required and prompt referral in the case of unresponsive cases. The CMG related to microbial keratitis therefore errs on the side of caution, but this may be damaging for optometrists in a fitness to practice case as it does not reflect practice.

4.3.5.2 Cellulitis – pre-septal

The CMG covering pre-septal cellulitis advises an emergency referral (same day) to an ophthalmologist without intervention. However, of the ten cases seen by the specialist IP optometrists, 100% were followed-up within the setting with no onward referral. Moreover, these

patients were all prescribed antibiotics by the specialist IP optometrists. This is another stark contrast to the guidelines.

Eye sepsis (infection of the ocular tissues) is differentiated by the position of the infection relative to the anterior boundary of the orbit, the orbital septum. Cellulitis therefore can be of two types, that anterior to the septum (pre-septal) and that posterior to the septum (post-septal/orbital). Pre-septal cellulitis is normally caused by skin trauma (insect bite or laceration) or the spread of a local infection (hordeolum or dacryocystitis) (179). It rarely extends to the orbit causing the more serious orbital cellulitis (180). The standard treatment involves broad spectrum antibiotic therapy (e.g. oral co-amoxiclav -Amoxicillin/Clavulanate). The treatment is reviewed and modified based on allergies, the response, and subsequent cultures. Intravenous intervention is considered if the oral regimen fails and can involve young children, the immunosuppressed, those showing toxicity or those not vaccinated against the offensive bacteria. Treatment, as with microbial keratitis is on an out-patient basis (HES) (177,181,182). Again, as with microbial keratitis treatment is promptly commenced prior to microbial culture analysis. The initial treatment therefore is arguably within the scope of IP optometry practice. Indeed, specialist IP optometrists appear to be treating pre-septal cellulitis. Needless to say, both with cases of MK and pre-septal cellulitis the robust rapid referral system in place in Kent pays dividends should an unresponsive patient need to be referred to an ophthalmologist.

A caveat to the above findings is the relatively small number of cases of MK and pre-septal cellulitis. Clearly, more data across multiple settings and schemes is needed to draw a generalisable conclusion that IP optometrist practice differs from the CMG recommendations. The results given herewith however show a deviation in practice for the IP optometrists at APCOS when compared to CMG recommendations.

4.3.6 Prescribing decisions

The CMGs recommend multiple therapeutic options for ocular conditions depending on the severity of the condition. The adherence for prescribing decisions (the drug type prescribed) made by the specialist IP optometrists against CMG recommendations was high (mean adherence 92.1%). The overperformed and underperformed test rates were generally low (mean overperformed rate 0.9%, mean underperformed rate 7.1%). Figure 4.12 and table 4.8 show the adherence, overperformed and underperformed test rates for prescribing decisions.

The adherence range for prescribing decisions was 61.3%-100% (figure 4.12 and table 4.8). The primary reason for non-adherence was again underperformance (range 0%-35%). The prescribing of ocular lubrication and topical steroids adhered the least to guidelines. The adherence for ocular lubrication was 61.3%, the overperformed and underperformed rates were 3.5% and 35.2% respectively. As such ocular lubrication was not prescribed as much as recommended in the CMGs. The adherence for topical steroids was 72.5%. The overperformed and underperformed rates were 3.0% and 24.5% respectively. Topical steroids were also therefore not prescribed as much as the CMGs recommended.

Figure 4.12 The naïve reporting rates, adherence, overperformed and underperformed test rates for prescribing decisions (actual figures in table 3.8). Abbreviation: Non-steroidal anti-inflammatory drug (NSAID). Error bars represent the upper and lower 95% confidence limits (174).

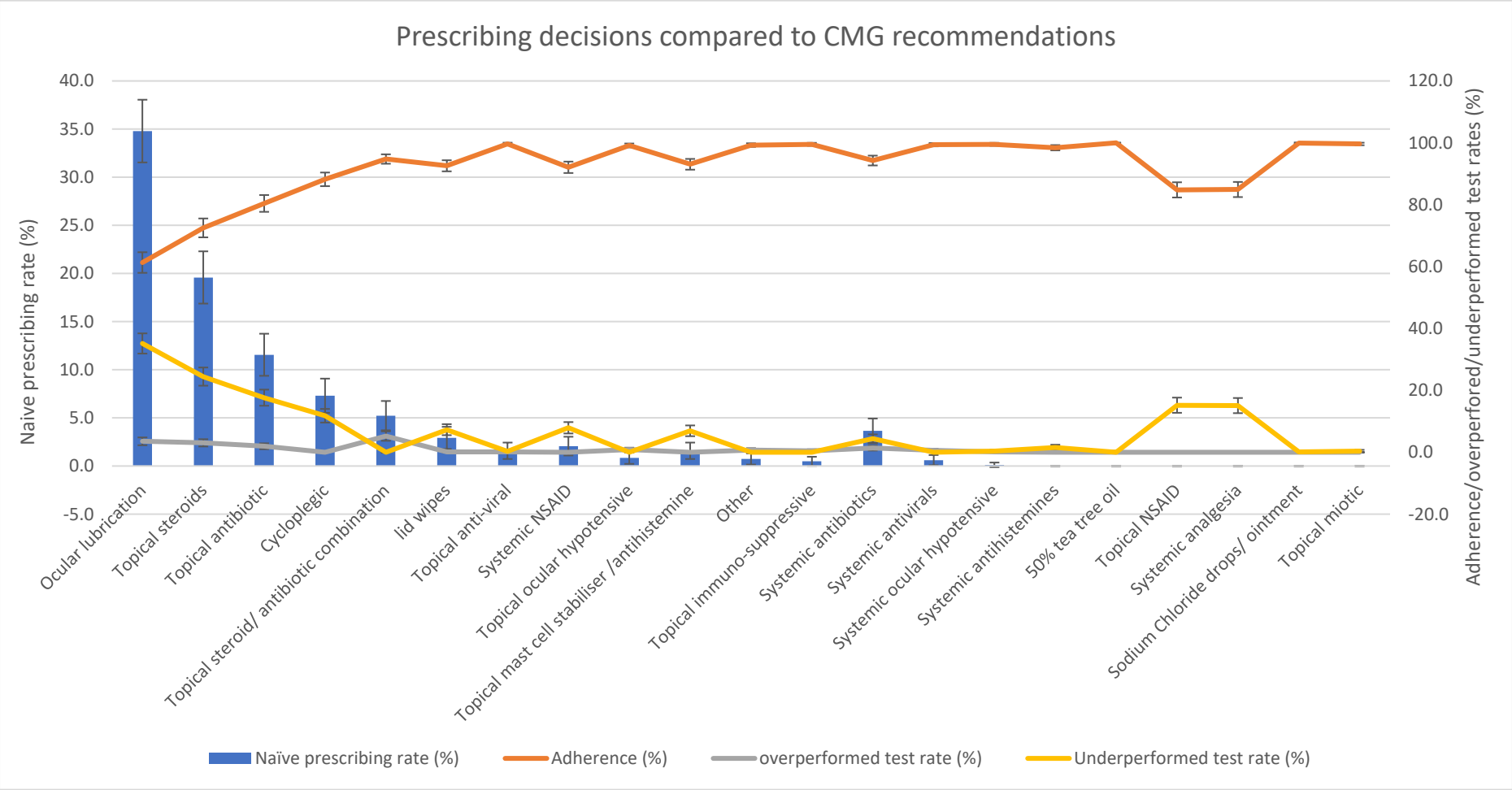


Table 4.8 The naïve reporting rates, adherence, overperformed and underperformed test rates for prescribing decisions (figure 3.12 shows these graphically). Abbreviation: Non-steroidal anti-inflammatory drug (NSAID).

	Drug types prescribed																				
Metrics	**Ocular lubrication**	**Topical Steroids**	**Topical antibiotic**	**Cycloplegic**	**Topical steroid/antibiotic combination**	**lid wipes**	**Topical anti-viral**	**Systemic NSAID**	**Topical ocular hypotensive**	**Topical mast cell stabiliser /antihistamine**	**Other**	**Topical immuno-suppressive**	**Systemic antibiotics**	**Systemic antivirals**	**Systemic ocular hypotensive**	**Systemic antihistamines**	**50% tea tree oil**	**Topical NSAID**	**Systemic analgesia**	**Sodium Chloride drops/ ointment**	**Topical miotic**
Naïve prescribing rate (%)	34.8	19.6	11.6	7.3	5.2	2.9	1.6	2.1	0.9	1.6	0.7	0.5	3.6	0.6	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Adherence (%)	61.3	72.5	80.4	88.2	94.8	92.6	99.6	92.1	99.1	93.1	99.3	99.5	94.3	99.4	99.5	98.4	100.0	84.8	84.9	99.9	99.6
Overperformed test rate (%)	3.5	3.0	1.9	0.0	5.2	0.1	0.1	0.0	0.9	0.0	0.7	0.5	1.3	0.6	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Underperformed test rate (%)	35.2	24.5	17.6	11.8	0.0	7.3	0.2	7.9	0.0	6.9	0.0	0.0	4.4	0.0	0.4	1.6	0.0	15.2	15.1	0.1	0.4

The drug regimens relating to the 24 CMG conditions with more than 5 patient episodes, were extracted and analysed further. The full counts of each prescription including varying dosages along with the CMG recommendations for these conditions are given in Chapter 4, appendix 6.

4.3.6.1 Conditions without prescriptions

Three of the 24 CMG conditions involved no therapeutic intervention, reflecting the CMG recommendations. These conditions were trichiasis, chalazion and acquired nasolacrimal duct obstruction.

4.3.6.2 Conditions with prescriptions

Of the 21 remaining CMG conditions for which prescriptions were issued, the therapeutic management of 11 reflected the CMG recommendations. The remaining 10 deviated in some aspect of therapeutic management. These 10 are now discussed in more detail.

4.3.6.2.1 KCS – Keratoconjunctivitis sicca (Dry eye)

The CMG recommends counselling the patient regarding environmental factors, lid hygiene and a diet rich in omega 3. Punctum plugs and therapeutic contact lenses are also mentioned. Pharmacological interventions involve tear supplements and weak topical steroids (FML or loteprednol) for short term use to break the inflammatory cycle associated with dry eye.

For the 157 patient episodes of KCS, a total of 196 prescriptions were issued. Table 4.9 shows the percentage of each kind of drug prescribed. The vast majority of prescriptions (86%) were for ocular lubrication (71% drops and 15% ointments), mirroring the CMG recommendation. Topical steroids composed 7% of the prescriptions (steroid or steroid and neomycin combinations), again reflecting the CMG recommendation. Anti-bacterial ointment was prescribed in 4% of cases. This was not mentioned in the CMG. Anti-bacterial ointment was likely prescribed to treat blepharitis, a common underlaying condition exacerbating KCS. An interesting observation is the prescription of ciclosporin eye drops (4 cases). According to the CMG such prescriptions fall under ‘possible management by ophthalmologist’. This may be due to the high cost of the drug (NHS tariff: £72.00 per 30 unit-dose) limiting its use to only the most severe and unresponsive cases (166). The data however shows IP

optometrists are prescribing cyclosporin for a small number of cases of KCS. Ciclosporin therefore falls within the scope of practice of specialist IP optometrists. Specialist IP optometrists all receive training in the fiscal aspect of prescribing. Indeed this is a GOC requirement of IP training (168).

Type of Drug	Percentage prescribed for KCS	Agreed/Disagreed with CMG recommendation
Eyelid wipes	0.5%	Agreed
Anti-inflammatory (Ciclosporin)	2%	Disagreed
Mast cell stabiliser/antihistamine combination	2%	Disagreed
Topical steroid/antibiotic combination	3%	Agreed
Antibacterial ointment	4%	Disagreed
Topical steroid	4%	Agreed
Lubricating ointment	15%	Agreed
Lubricating drops	71%	Agreed

Table 4.9 The distribution of drugs prescribed for KCS

The most common drug prescriptions and regimens for KCS are given in table 4.10.

Drug	Drug Type	Administration	Duration
Carmellose sodium 0.5% (Celluvisc, Optive)	Lubricating drops	QDS	On-going
Paraffins (VitA-POS)	Lubricating ointment	OD NOCTE	On-going
Sodium hyaluronate 0.1% (Hylocare P/F)	Lubricating drops	6x day	On-going
Sodium hyaluronate 0.2% (Hyloforte P/F)	Lubricating drops	TDS	On-going
Sodium hyaluronate 0.15%/Trehalose 3% (Theoloz duo)	Lubricating drops	QDS	On-going

Table 4.10 The most common drug prescriptions for KCS (descending order)

4.3.6.2.2 Uveitis

The CMG associated with uveitis recommends intensive topical steroid therapy (hourly, then tapered) alongside cycloplegic therapy to stop ocular inflammation, ciliary spasm, and the formation of posterior synechiae respectively. Urgent referral to the HES is recommended for recurrent or bilateral uveitis, severe cases, or those involving the posterior segment. Emergency referrals are indicated where there is a significant reduction in vision, severe pain or a significantly raised IOP (183). Seventy-five cases of uveitis were seen over the 12-month period of data collection. For these, 137 drug prescriptions were issued. Following the recommendations of the CMG, the most common groups of drugs prescribed were topical steroids and cycloplegics accounting for 98% of the drugs

prescribed (Table 4.11). A further 2% related to ocular hypotensive therapy (Oral Acetazolamide 250mg, gutt. Timolol 0.25% and gutt. Timolol 0.5%). This aspect of management is not mentioned in the CMG for IP optometrists. The CMG recommends 'significantly' raised IOP requires an emergency referral to the HES without intervention. The term 'significant' has not been defined, implying that a moderate increase in IOP could be treated and is within the scope of IP optometrist practice. However, ocular hypotensives have not been listed in the 'pharmacological management' section. Oral Acetazolamide is usually prescribed in cases where the IOP is very high. Although occasionally the APCOS IP optometrists are therapeutically managing a 'spike in IOP' in uveitis patients, the exact values of the IOP were not collected in the EPR. The most common drugs, with the administration and duration, are given in table 4.12.

Type of Drug	Percentage prescribed for uveitis	Agreed/Disagreed with CMG recommendation
Topical Steroid ointment	1%	Agreed
Ocular Hypotensive drugs	2%	Disagreed
Cycloplegic	42%	Agreed
Topical steroid drops	55%	Agreed

Table 4.11 The distribution of drugs prescribed for uveitis

Drug	Drug Type	Administration	Duration
Cyclopentolate 1% (Mydrilate)	Cycloplegic	TDS	1 week
Cyclopentolate 1% (Mydrilate)	Cycloplegic	TDS	2 weeks
Prednisolone 1% (Predforte)	Steroid	Hourly	Tapered
Prednisolone 1% (Predforte)	Steroid	Hourly	1 week
Prednisolone 1% (Predforte)	Steroid	6x day	Tapered

Table 4.12 The most common drug prescriptions for uveitis (descending order)

4.3.6.2.3 Blepharitis – Anterior Marginal – Staphylococcal

Fifty-three cases of anterior marginal blepharitis were seen by the APCOS IP optometrists. For these, 65 drug prescriptions were issued. The distribution of the types of drugs prescribed is given in table 4.13 with the most common prescriptions shown in table 4.14. Most prescriptions related to lid hygiene and ocular lubrication (eyelid wipes 33% and lubricating drops 27%). Topical antibiotic therapy constituted a large proportion of prescriptions issued (27% antibiotic only, 8% antibiotic/steroid combination). The remaining drugs were steroid drops and systemic tetracycline (1 case).

The therapeutic interventions demonstrated by the IP optometrists reflect the recommendations of the associated CMG (184). Lid hygiene is emphasised with the use of ocular lubrication as first line therapeutic treatment. Antibiotic ointment is recommended as second line treatment with the use of systemic tetracyclines in persistent cases. Steroid/antibiotic ointments can be useful in reducing both eyelid inflammation and bacterial load simultaneously. However, the use of steroids is not mentioned in the CMG. This may be because the evidence of the efficacy of any of the individual treatments (steroid only or antibiotics only) concerning blepharitis is generally scant, however a combination of steroid and antibiotics have shown to give beneficial results (185,186). Given the proportion of drugs containing steroids prescribed for anterior marginal blepharitis (11% for steroid or steroid/antibiotic combination), this CMG may be augmented with the addition of steroid treatment, following further research.

Type of Drug	Percentage prescribed for Anterior marginal blepharitis (staphylococcal)	Agreed/Disagreed with CMG recommendation
Systemic tetracycline	2%	Agreed
Topical steroid	3%	Disagreed
Topical steroid/antibiotic combination	8%	Disagreed
Topical antibiotic (drops and ointment)	26%	Agreed
Ocular lubricating drops	26%	Agreed
Eyelid wipes	35%	Agreed

Table 4.13 The distribution of drugs prescribed for anterior marginal blepharitis

Drug	Drug type	Administration	Duration
Blephaclean eyelid wipes	Eyelid wipes	BDS	On-going
Chloramphenicol 1% ointment	Antibiotic	BDS	1 week
Dexamethasone 0.1%/ Neomycin Sulf. 0.35%/polymyxin B sulf. 6000iu/ml (Maxitrol)	Steroid/antibiotic combination	QDS	2 weeks
Blephaclean eyelid wipes	Eyelid wipes	BDS	2 weeks
Carbomer 0.2% (Lumicare carbomer, Clinitas, Carbomer gel, Viscotears, Gel tears)	Ocular lubrication	QDS	On-going
Carmellose Sodium 0.5% (Celluvisc, Optive)	Ocular lubrication	QDS	On-going

Table 4.14 The most common drug prescriptions for anterior marginal blepharitis (descending order)

4.3.6.2.4 Conjunctivitis – viral – non-herpetic

The CMG for viral conjunctivitis recommends artificial tears/lubrication to relieve the associated symptoms. Topical antihistamines are recommended for cases with severe itching. Low dose topical steroids are listed under possible management by an ophthalmologist. The IP optometrist prescribing data collected however shows that topical steroids (either single or in combination with an antibiotic) constituted 69% of prescriptions issued. Ocular lubrication accounted for the remaining 31% (table 4.15). Low dose steroids are typically prescribed where sub-epithelial opacities affect vision. However, the CMG warns of long-term steroid dependency with the recurrence of infiltrates upon treatment cessation (187). Despite this, the treatment is recommended for severe cases in some authoritative texts with no mention of dependency (177,181,188). The clinician however is warned that although topical steroids can hasten the resolution of symptoms, they can prolong the infectious period. The steroids require tapering over a long period to stop recurrence.

Type of Drug	Percentage prescribed for viral (non-herpetic) conjunctivitis	Agreed/Disagreed with CMG recommendation
Ocular lubrication (drops and ointment)	31%	Agreed
Topical steroid	54%	Disagreed
Topical steroid/antibiotic combination	15%	Disagreed

Table 4.15 The distribution of drugs prescribed for viral (non-herpetic) conjunctivitis

The data collected therefore shows a clear difference in the actual therapeutic management of viral (non-herpetic) conjunctivitis by IP optometrists and the associated CMG. Although the monitoring of a patient prescribed steroids is essential, this can easily be done in a community optometric practice. Table 4.16 shows the most common drug prescriptions given to these patients.

Drug	Drug type	Administration	Duration
Dexamethasone 0.1% minims	Steroid	QDS	Tapered
Dexamethasone 0.1% drops	Steroid	QDS	2 weeks
Sodium hyaluronate P/F (Hylocare)	Ocular lubrication	QDS	2 weeks
Dexamethasone 0.1%/ Neomycin Sulf. 0.35%/polymyxin B sulf. 6000iu/ml (Maxitrol)	Steroid/antibiotic combination	6x day	1 week
Dexamethsone 0.1% minims	Steroid	QDS	2 weeks

Table 4.16 The most common drug prescriptions for viral (non-herpetic) conjunctivitis

4.3.6.2.5 MGD – blepharitis – posterior marginal

Twenty-five patient episodes of MGD resulted in the issuing of 26 drug prescriptions. The distribution of drug-type is shown in table 4.17. Most patients (46%) were prescribed an oral tetracycline, this was closely followed by artificial tears (35%). The remaining 5 cases (20%) composed of eyelid wipes, topical antibiotics, and weak topical steroid treatment.

The prescription of topical steroid therapy for blepharitis and the associated dry eye is not uncommon (as mentioned with anterior blepharitis above). Topical steroid therapy is not mentioned in the associated CMG, however a weak steroid can reduce lid inflammation and palliate symptoms while the underlaying cause of inflammation is addressed (184,185). The associated CMG mentions lid hygiene, artificial tears, topical antibiotic ointment and oral tetracyclines for the treatment of MGD. Consideration should therefore be given to the addition of a short course of weak steroid treatment to alleviate symptoms in severe cases. Table 4.18 shows the most common drug prescriptions given for MGD.

Type of Drug	Percentage prescribed for MGD	Agreed/Disagreed with CMG recommendation
Eyelid wipes	4%	Agreed
Topical antibiotic	4%	Agreed
Topical steroid	12%	Disagreed
Artificial tears	35%	Agreed
Systemic tetracycline	46%	Agreed

Table 4.17 The distribution of drugs prescribed for MGD

Drug	Drug type	Administration	Duration
Doxycycline tablets 50mg	Systemic tetracycline	OD MANE	2 months
Doxycycline tablets 100mg	Systemic tetracycline	OD MANE	1 month
Carbomer 0.2% (Lumicare carbomer, Clinitas, Carbomer gel, Viscotears, Gel tears)	Ocular lubrication	QDS	On-going

Table 4.18 The most common drug prescriptions for MGD (descending order)

4.3.6.2.6 HZO – Herpes Zoster Ophthalmicus

The CMG relating to HZO recommends lubrication and pain relief. Management to resolution is recommended when co-managed with the GP and keratitis is limited to the epithelium. An urgent referral to the HES is indicated if deeper layers of the cornea are involved, anterior uveitis is present

or the IOP raised. The CMG advises emergency referral to the GP for systemic anti-viral therapy in the case of acute skin lesions as this reduces pain and the likelihood of ocular involvement (189,190).

Fifteen cases of HZO were seen over the 12-month period. Nine prescriptions were issued, the majority of which (6 prescriptions) were for artificial tears (as per the CMG recommendation). The remainder (3 prescriptions) however were for antiviral therapy (table 4.19). Each of the 9 prescriptions varied slightly in terms of drug regimen and so lacked a clear hierarchy of drug regimen. Antiviral therapy however involved aciclovir (2 systemic and 1 topical) in all three cases. This activity contrasts the recommendation which leaves the commencing of antiviral therapy to the GP or ophthalmologist if there is skin involvement, or the associated keratitis is severe. Systemic antiviral therapy is a safe and effective treatment for HZO as there is enough antiviral in the tear film following systemic administration for it to be effective on the ocular surface (191). In the author's opinion then, if a patient presents with mild signs of HZO (epithelial) with/without mild skin lesions (the cause for much neurological pain) arguably initiating systemic anti-viral therapy may be the best course of management. This could prevent the development of skin lesions and reduce the risk of encephalitis, a serious complication (192,193). Indeed this is the protocol in well-established eye departments (177). The CMG may need reviewing to reflect this aspect of clinical practice. Recommending systemic antiviral treatment be initiated by IP optometrists in a clinical setting where patients can be monitored appears a reasonable suggestion. Subsequent referral to the GP may also be suggested.

Type of Drug	Percentage prescribed for HZO	Agreed/Disagreed with CMG recommendation
Artificial tears	66%	Agreed
Topical/systemic anti-viral	33%	Disagreed

Table 4.19 The distribution of drugs prescribed for HZO

4.3.6.2.7 HSK – Herpes Simplex Keratitis

The prescribing recommendation of the HSK CMG suggests the initiation of topical antiviral therapy (Ganciclovir 0.15%) in non-contact lens wearing adults exhibiting only epithelial involvement (178). Fourteen cases of HSK were seen over the data-collection period with 16 prescriptions issued. Twelve prescriptions (75%) were for topical antiviral therapy, with the remainder split between systemic antiviral therapy, artificial tears, and topical steroid therapy (table 4.20). Both systemic antivirals and topical steroid treatment fall under the 'possible management by ophthalmologist' section of the CMG. For cases where uveitis presents with the condition, concurrent steroid therapy

is indicated. These patients require close supervision as steroids can promote the progression of corneal lesions. This is perhaps why the use of such drugs is not listed in the 'management by optometrist' section. Although it is difficult to firmly conclude a prescribing practice from the limited number of patient episodes, the concurrent use of steroid and anti-viral therapy is nothing new. The associated CMG could be reviewed for the addition of topical steroid and systemic anti-viral therapies. Table 4.21 shows the most common drug prescriptions for HSK.

Type of Drug	Percentage prescribed for HSK	Agreed/Disagreed with CMG recommendation
Artificial tears	6%	Agreed
Topical steroid	6%	Disagreed
Systemic antiviral	13%	Disagreed
Topical antiviral	75%	Agreed

Table 4.20 The distribution of drugs prescribed for HSK

Drug	Drug type	Administration	Duration
Topical aciclovir 3% ointment	Topical anti-viral	5x day	1 week
Topical ganciclovir 0.15% gel	Topical anti-viral	5x day	2 weeks

Table 4.21 The most common drug prescriptions for HSK (descending order)

4.3.6.2.8 Keratitis – microbial- bacterial

The APCOS team of specialist IP optometrists saw 13 cases of microbial keratitis (bacterial) over the 12-month period. Twelve drug prescriptions were issued. As discussed in section 4.3.3 relating to the management decisions of the IP optometrists there is a significant deviation from the CMG recommendation regarding this condition. With the intervention, once again there is a significant difference. The CMG recommends no treatment and an emergency referral to the HES. However, of the 85% of cases that were kept in the community setting, most (92%), were prescribed a quinolone as mentioned in the CMG under 'possible management by ophthalmologist'. As previously discussed, microbial cultures are not routinely taken at the HES unless the condition is not responsive to initial therapy, the infiltrates are larger than 2mm, or an unusual organism is suspected based on the case history (177). Arguably therefore these patients could be managed and referred only if unresponsive, which appears to be current clinical practice. This contrasts the CMG recommendation.

Table 4.22 shows the proportion of drug types prescribed. Table 4.23 shows the most common prescription.

Type of Drug	Percentage prescribed for microbial keratitis	Agreed/Disagreed with CMG recommendation
Artificial tears	8%	Disagreed
Topical quinolone	92%	Disagreed

Table 4.22 The distribution of drugs prescribed for microbial keratitis (bacterial)

Drug	Drug type	Administration	Duration
Topical Ofloxacin 0.3% (Exocin)	Antibiotic	hourly	3 days

Table 4.23 The most common drug prescription for microbial keratitis (bacterial)

4.3.6.2.9 Cellulitis – Pre-septal

Again, as mentioned in section 4.3.3 the management of pre-septal cellulitis differed significantly to the recommendation in the associated CMG. The CMG recommends these patients are referred as an emergency to the HES, however all the 10 patients were kept within the APCOS service with follow-up. Of the 10 prescriptions issued, 8 (80%) were for systemic broad-spectrum antibiotics with the remainder for topical antibiotics (table 4.24). The most common prescription is given in table 4.25 which is, unsurprisingly, Co-amoxiclav, the first-line treatment for this condition. The CMG relating to pre-septal cellulitis is not representative of the clinical scope of practice and should be reviewed both in terms of the management and pharmacological recommendations (194).

Type of Drug	Percentage prescribed for pre-septal cellulitis	Agreed/Disagreed with CMG recommendation
Systemic antibiotic	80%	Disagreed
Topical antibiotic	20%	Disagreed

Table 4.24 The distribution of drugs prescribed for pre-septal cellulitis

Drug	Drug type	Administration	Duration
Co-amoxiclav tablets 500mg/125mg	Systemic antibiotic	TDS	1 week

Table 4.25 The most common drug prescription for pre-septal cellulitis

4.3.6.2.10 Conjunctivitis medicamentosa

Eight cases of conjunctivitis medicamentosa presented during the data collection period. Five drug prescriptions were issued, 2 of which were for ocular lubricants as recommended in the associated CMG for symptomatic relief (195). The remaining 3 were for topical ocular hypotensive treatment. This implies that the cause of this conjunctivitis was medications prescribed for the treatment of glaucoma or ocular hypertension. The prescription of alternative ocular hypotensive agents would

therefore alleviate the problem. This CMG appears also to mirror practice, however without information on the offending medication a true evaluation could not be undertaken.

4.4 Summary and Reflection on How the CMGs Compared to IP Optometrist Practice

The clinical activity of specialist IP optometrists practising as part of an acute primary care ophthalmology service (APCOS) was compared to the recommendations set out in the CMGs for anterior ocular conditions. This was deemed important as not only do the CMGs serve as a guide to practitioners, but they also represent the scope of IP optometrist practice. The College of Optometrists' website advises that the CMGs may be drawn upon in the event of a fitness to practice or civil case (196). It is therefore extremely important that they are truly representative of the bounds of current clinical practice.

The profession is fast evolving to include the management of ocular disease (12,197,198). Maintaining the appropriateness of clinical guidelines is difficult without a 'live' data-harnessing method that revises recommendations according to evidence but also clinical practice at the 'coalface'. Indeed, as of 01/03/2020, more than half of the CMGs had an expired date of review (chapter 3, appendix 7).

Five aspects of IP optometrists' clinical activity were evaluated against the College CMGs. These were the ocular conditions encountered, patient history and symptoms, clinical signs, and patient management and prescribing decisions. IP optometrist practice was compared to CMG recommendations using the novel metrics of adherence, underperformed and overperformed test rates.

4.4.1 Ocular Conditions Encountered

Considering the first aspect, the CMGs represented, accurately, the majority of anterior ocular conditions encountered in clinical practice (92.7% of activity). It must be noted however that over a third of total episodes (34.6%) related to posterior eye conditions. These conditions are largely bereft of CMG guidance. This may be because most do not indicate intervention by IP optometrists. However, one condition, Post-operative CMO (following cataract extraction) can and indeed is treated by IP optometrists. The extension of the CMGs to cover such conditions may be beneficial to prescribing clinicians.

4.4.2 History and Symptoms

The adherence to CMG guidelines across aspects of history and symptoms was high and the overperformed and underperformed test rates low (figure 4.9 and 4.10). A hierarchy of symptoms does not appear in the CMGs. However, IP optometrists reported ocular discomfort most frequently, suggesting patients presenting with an anterior CMG condition are likely to suffer from this symptom.

4.4.3 Clinical Tests Undertaken

Adherence across the 66 clinical tests (variables) was also high with low overperformed and underperformed test rates (Chapter 4, appendix 4). Most clinical testing (for ocular signs) undertaken was therefore highly appropriate when compared to CMG recommendations. The following changes were suggested as demonstrated from the practice of IP optometrists:

- Conjunctival/episcleral hyperaemia could be added to the CMGs of conjunctivitis medicamentosa, corneal (or other) superficial foreign body, HZO, pre-septal cellulitis and HSK
- Epiphora could be added to the CMG relating to KCS
- Consideration could be given to adding 'ocular discomfort' to HZO, HSK and pre-septal cellulitis
- Consideration could be given to adding viral conjunctivitis, KCS, uveitis and primary angle closure to the differential diagnoses of SAC (seasonal allergic conjunctivitis) as the associated signs are investigated by IP optometrists

The value of negative test results was highlighted in cases involving diagnoses of exclusion such as SAC. In such cases the CMGs may be enhanced with the addition of a few strong negatively associated clinical signs and symptoms (i.e. those definitely not to be observed). However, creating such a list of variables from an evidence-base would be difficult as positive clinical signs for a condition are almost always reported and negative neglected.

4.4.4 Management Decisions

For management decisions, the adherence, overperformed and underperformed rates indicated the favouring of discharge as an outcome. There was an underperformance of referral to HES and follow-up. This indicated that specialist IP optometrists were managing most patients within the APCOS service. Indeed, onward referrals accounted for only $1/10^{th}$ of outcomes. Since APCOS is an ophthalmology patient pathway designed to reduce referral to the HES, this result reflected good practice. On further analysis, MK and pre-septal cellulitis showed the greatest deviation from

management recommendations (section 4.3.5). Both these conditions were managed by IP optometrists despite the associated CMGs recommending a prompt referral to the HES. A limitation, however, was that these patients were not followed-up to determine whether the management decision was appropriate (i.e. whether they required a referral to the HES on the subsequent visit).

4.4.5 Prescribing Decisions

It is unsurprising that the above two conditions (MK and pre-septal cellulitis) were also highlighted as deviant in their prescribing decisions. In both cases, antibiotics were prescribed by the specialist IP optometrists contrary to guidance. The drug therapies were correctly selected. However, in contrast to practice, the CMGs had placed this activity within the realms of ophthalmological management. Nine of the 24 conditions included in the prescribing evaluation (38%) differed in the scope of practice (lacked the range of drugs prescribed). Figure 4.13 shows these conditions along with suggested changes that would make the recommendations reflect clinical practice. The conditions for which significant changes should be considered are shaded red and include HZO, microbial keratitis (bacterial) and pre-septal cellulitis. An overall agreement of 62% shows that the 24 evaluated CMGs are moderately reflective of the scope of IP optometrist practice. However, in the face of potential litigation this level of agreement may be deemed as sub-optimal.

A limitation of the results presented herewith is that a straight comparison of the reported clinical findings against the CMGs was not possible. This is because the variables held within the EPR were refined to include those deemed the most diagnostic for the purpose of machine learning (Chapter 2). It is possible then, that clinicians examined aspects of the history, symptoms, and signs above and beyond those for which data fields existed. The data collected was therefore evaluated with regards to the adherence of the tests listed in the EPR against the CMG recommendations.

Since this audit scoping the activity of IP optometrists, the College of Optometrists have formed an IP review group of which the author and one of the APCOS IP optometrists are members. The group is asked to review the recommendations listed within each of the CMGs. The guideline for pre-septal cellulitis was reviewed in November 2021. As a result of feedback, with reference to the information provided by this audit, the College has now updated the CMG to include management in the community by IP optometrists with prescription of systemic antibiotics in the first instance. This CMG now reflects IP practice as found in the audit.

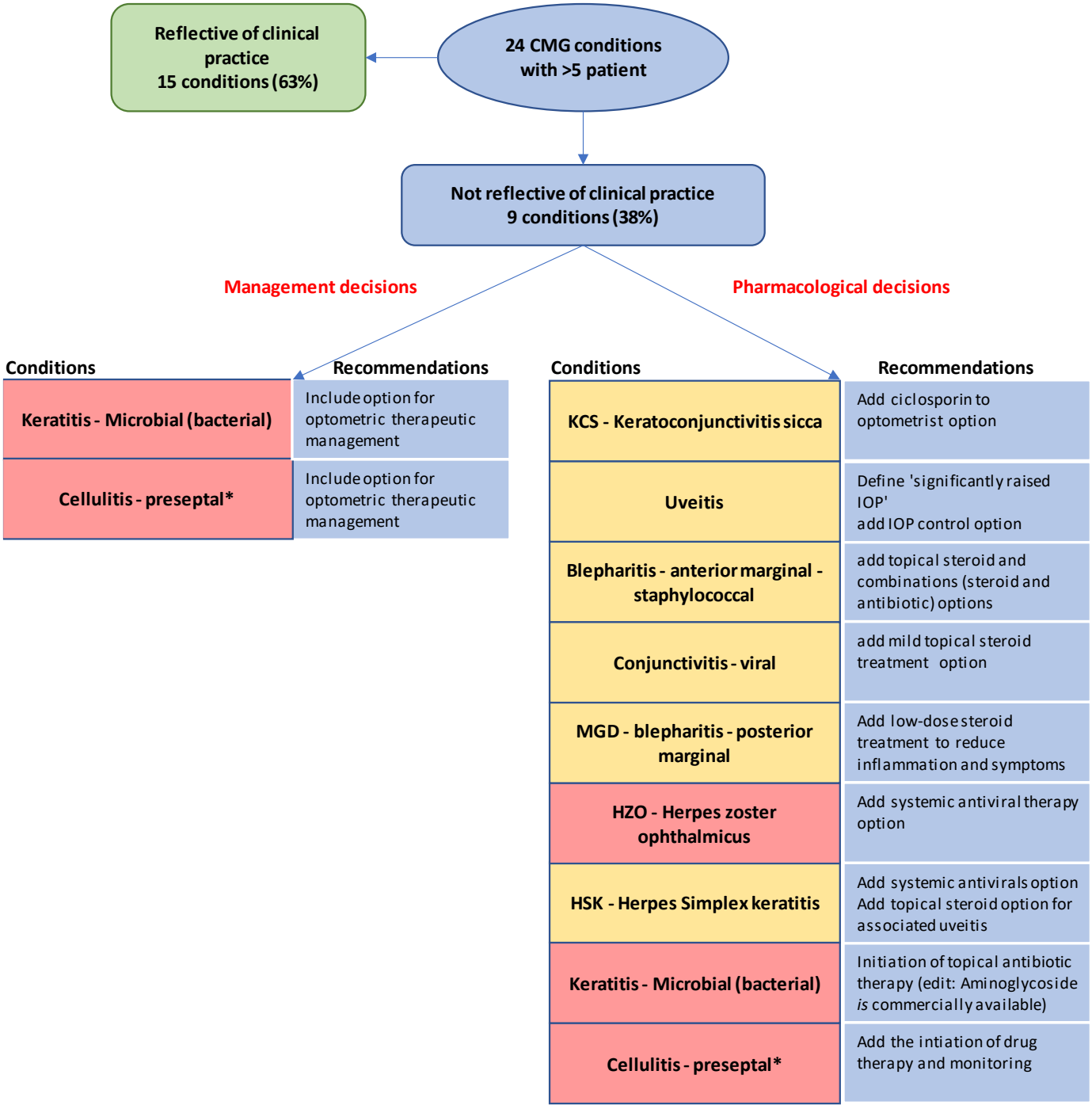


Figure 4.13 Diagrammatic representation of the comparison results. For management decisions, 2 conditions were not reflective of practice. For prescribing decisions, 9 conditions were not reflective of practice. Red denotes major deviation to CMG recommendations and amber denotes small deviations. Each condition is accompanied by the recommendation extracted from the clinical activity demonstrated by APCOS IP optometrists. * The recommendation for pre-septal cellulitis has been reviewed and altered to include the above recommendations since this audit of clinical activity.

Chapter 5 follows on with the aims of the present study. That is, a) to apply machine learning to the clinical activity of IP optometrists, and b) to develop the concept of interactive and evolving evidence-based support systems for IP optometrists and those in training. A digital learning platform is introduced (MyDLP) fulfilling the aims of the study and based on the Bayes' Translational Learning Concept (bTLC) described in chapter 1.

Chapter 5: The Development of MyDLP

5.1 Introduction

The first attempt at applying machine learning to the data collected, resulted in modest model performance (hierarchical Bayes' – chapter 3). Hierarchical Bayes' did not perform well enough to be used as a basis for developing the concept of an interactive and evolving evidence-based support system for IP optometrists. As discussed, this may have been due to a combination of insufficient data, and a high degree of granularity (i.e. too many outcomes and therefore not enough episodes of each outcome to allow optimal machine learning to occur). This resulted in the model performing poorly. To maximise on the data collected by APCOS and to demonstrate a 'proof-of-concept' the crux of the model were re-evaluated.

Chapter 5 describes the development of a digital learning platform (MyDLP). This platform conformed to the concept of the Bayes' translational learning concept (bTLC) as discussed in chapter 1. Figure 5.1 shows the chapter plan.

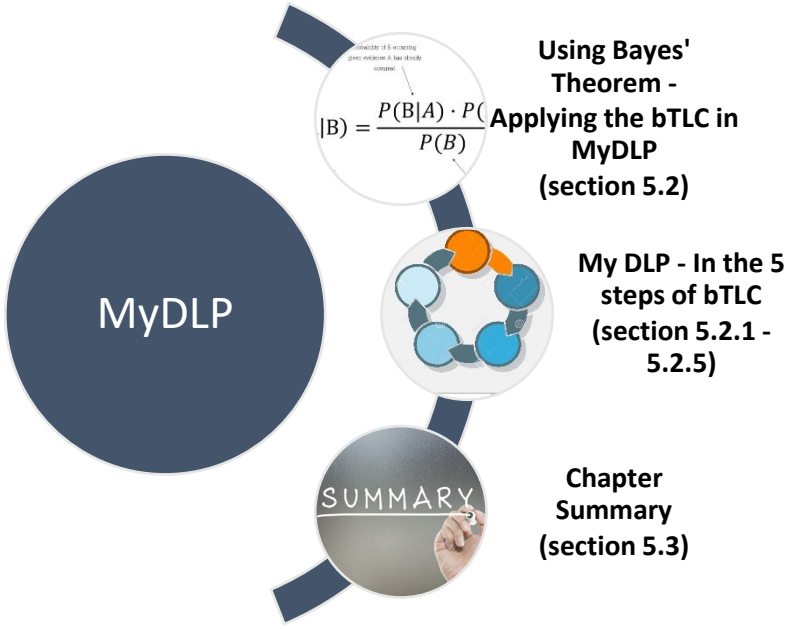


Figure 5.1 The chapter plan for MyDLP development

5.2 Using Bayes' Theorem – Applying bTLC in MyDLP

The digital learning platform developed (MyDLP) demonstrates the potential of machine learning, in the form of Bayes' theorem, for the linking of any specified dependent variables (i.e. outcomes, such as diagnoses and prescribing decisions) with independent variables (such as clinical tests and demographic information). MyDLP is completely housed within the ubiquitous Microsoft Excel® worksheets. When applied to the data collected from the specialist IP optometrists constituting APCOS, MyDLP has the potential to provide 'live' digital clinical support via an "intelligent" electronic patient record (iEPR). It also provides a learning platform allowing training clinicians to explore data in order to understand the basis of machine learning and clinical decision-making ("intelligent" virtual patient tool – iVPT).

Figure 5.2 shows the original five steps of the bTLC. The development of MyDLP in relation to these five steps is summarised in table 5.1. The five steps of the bTLC, and the development and functioning of each of the associated worksheets within MyDLP, are discussed in more detail below. Appendix 1 shows a blank MyDLP (without any data). Appendix 2 shows MyDLP filled with hypothetical data for the detection of cystoid macula oedema. The latter (appendix 2) will be referred to when discussing the functionality of MyDLP. Chapters 6 and 7 will then demonstrate the use of MyDLP for the diagnosis and therapeutic decisions relating to KCS and uveitis using the collected data.

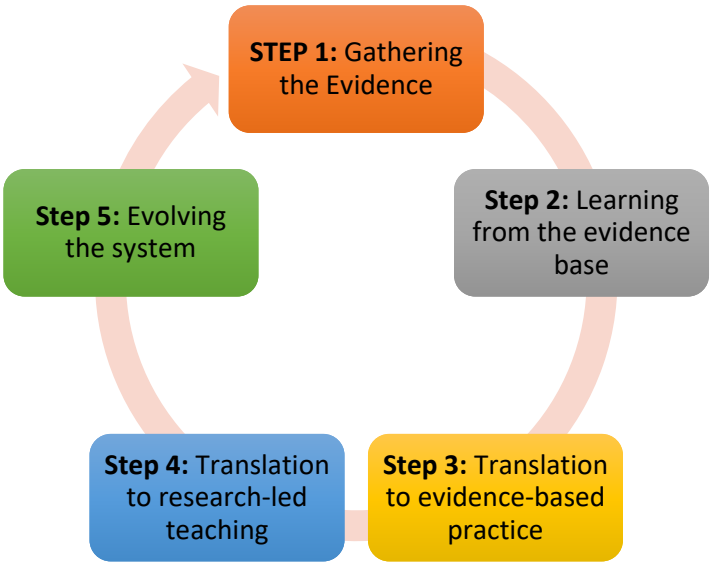


Figure 5.2 The five steps forming the Bayes' translational learning concept (bTLC)

bTLC Step	MyDLP worksheet no. & title (appendix 1)	What it shows
	1 – "Notes"	Summary page including an explanation of each of MyDLP worksheets and quick view of performance, accuracy, and learning
Step 1: Gathering the Evidence	**2** – "Variables"	Specifies the parameters of interest – i.e. the predictors (independent variables such as clinical tests) and predicted (dependent variables such as diagnoses and drug type) variables.
	3 – "Record (gathering the evidence)"	A simplified electronic patient record (EPR) system (containing only the variables selected) producing a code for each patient episode.
	4 – "Database (store evidence)"	The gathered evidence is stored in the form of numerical codes, each representing a patient episode.
	5 – "Multi-converter"	The numerical codes are converted to binary attributes for each variable – making them conducive to machine learning.
Step 2: Learning from the evidence base	**6** – Bayes' (learn from the evidence)"	The calculations for pre-test odds, likelihood ratios (LRs) and their 95% confidence limits for each predictor against each predicted variable.
Step 3: Translation to evidence-based practice	**7** – "iEPR & iVPT (apply the evidence)"	Intelligent electronic patient record (iEPR) uses the Bayes' calculations for decision support – this tool is a combined with the iVPT (below)
Step 4: Translation to research-led teaching	**7** - "iEPR & iVPT (apply the evidence)"	Intelligent virtual patient tool (iVPT) uses Bayes' to grade clinical decisions – this tool is combined with the iEPR (above)
Step 5: Evolving the system	**8** – "Bayes' accuracy (evaluation)"	First stage of evaluating the system – 10-fold cross validation with the calculation of accuracy, informedness and markedness.
	9 – "Bayes' learning"	Second stage of system evaluation – learning efficiency curves to show whether maximum system learning has been achieved

Table 5.1 The five steps forming the bTLC

5.2.1 **Step 1:** Gathering the Evidence

As described in chapter 2, highly structured data is needed to optimise machine learning. As such the data-collection EPR contained clinical parameters (predictor or independent variables) and clinical outcomes (diagnosis, management, and prescribing decisions, that is, predicted or dependent variables). MyDLP uses this highly structured data to power the digital platform. Step 1 of the bTLC, gathering the evidence, is represented in worksheets 2-5 in MyDLP (table 5.1).

5.2.1.1 Worksheet 2: “Variables”

In its current form, up to 25 predictor variables can be added to MyDLP (table 5.1 and appendix 1, worksheet 2 “variables”). MyDLP can predict up to a maximum of 5 dependent variables (outcomes, such as diagnoses and prescribing decisions). It was felt that up to 5 differential diagnoses or prescribing suggestions with probabilities and confidence limits for each was sufficient for proof of concept.

Figure 5.3 shows a screenshot of a “variables” worksheet filled with hypothetical data pertaining to cystoid macula oedema (CMO). Dependent (predicted) variables must only have two levels. That is, an outcome is either present or not. As such, level 1 must always be specified as ‘no’ and level 2 ‘yes’. The independent (predictor) variables can have up to 4 levels (i.e. results of tests). As can be seen in figure 5.3 these can be binary (‘no’ and ‘yes’) or reflect the grade or level of a clinical sign (‘no’, ‘mild’ and ‘severe’ or ‘unilateral’, ‘bilateral’ and ‘asymmetric’). The fourth level is reserved for cases of incomplete data, that is, when an outcome is ‘unknown’, or a test not carried out.

Variables

Enter variable names (in columns 1-30)
Enter variable assignment 1 = entered blank = not entered
Enter variable levels (in rows 1-4) 1 = no 2 = yes or level 1 name e.g. mild 3 = blank or level 2 name e.g. severe 4 = unknown (i.e. N)

NB - for laterality (unilateral, bilateral & asymmetric), code as bilateral (no

Laplace smoothing: 1

	Enter up to 5 dependent (predicted) variables					Enter up to 25 independent (predictor) variables						
	1	2	3	4	5	6	7	8	9	10	11	12
Names >>	CMO					DVA<6/9	NVA<N8	Central bl	cataract s	laterality	hyperaemi	a/c cells
Assignment >>	1					1	1	1	1	1	1	1
1	no					no	no	no	no	unilateral	no	no
2	yes					yes	yes	partial	yes	bilateral	mild	mild
3								yes	asymmetri		severe	severe
4						unknown	unknown	unknown	unknown	unknown	unknown	unknown

Figure 5.3 A screenshot of worksheet 2: “variables” of MyDLP containing hypothetical data pertaining to cystoid macula oedema (CMO).

Note the “assignment” row allows the convenient and rapid inclusion or exclusion of variables (both predictor and predicted) in the analysis. The clinician/trainee can then add or remove variables to explore the effect on decision support. MyDLP automatically updates the calculations upon adjustment of the “assignment”.

Laplace smoothing can be added at the "variables" worksheet simply by placing a "1" in the appropriate box (figure 5.3). As discussed in chapter 3 (section 3.2.1.6), a Laplace correction of 1 added to all counts, mitigates the effect of poor data representation. This would otherwise result in a LR of zero for some tests. A LR of zero, due to the poor representation of one predictor (i.e. a test rarely undertaken), can result in a post-test probability of zero for an otherwise strong outcome determined by a battery of diagnostic tests. The clinician has the choice to add or remove the Laplace correction. MyDLP then automatically updates the calculations and predictions for the clinician to view.

5.2.1.2 Worksheet 3: "Record (gathering the evidence)"

MyDLP includes a simplified EPR (table 5.1, step 1 (worksheet 3); and appendix 1, worksheet 3 "Record (gathering the evidence)"). This is useful should a clinician want to add more of their own patient episodes to the data. It generates a code representing clinical findings. This code, like the data-collection EPR designed in chapter 2, can be transferred between sites without compromising patient confidentiality. The code generated is in the form required for processing by MyDLP. The code can be copied from worksheet 3 and pasted into the "database (store evidence)" (worksheet 4) where MyDLP incorporates it into the analysis.

The variables in this simplified EPR are automatically loaded from those inputted on the "variables" worksheet. Figure 5.4 shows a screenshot of worksheet 3: "Record (gathering the evidence)" from appendix 2 containing hypothetical data for a patient with CMO. It shows this hypothetical patient has reduced distance and near visual acuities (codes = 2 for each of the predictor variables), a partial central blur (code = 2), which is unilateral (code = 1), has had cataract surgery (code = 2) with no hyperaemia or cells in the anterior chamber (codes = 1 for each of the predictor variables).

Record Sheet

Enter data as a number. When finished, copy data row to database

	2						2	2	2	2	1	1	1
number	CMO						DVA<6/9	NVA<N8	Central blur	cataract surgery	laterality	hyperaemia	a/c cells
1	no						no	no	no	no	unilateral	no	no
2	yes						yes	yes	partial	yes	bilateral	mild	mild
3									yes		asymmetric	severe	severe
4							N	N	N	N	N	N	N

Figure 5.4 A screenshot of worksheet 3: "Record (gather evidence)" of MyDLP containing hypothetical data pertaining to a case of cystoid macula oedema (CMO).

5.2.1.3 Worksheet 4: “Database (store evidence)”

The data used to drive MyDLP is stored in a database constituting worksheet 4 (table 5.1 and appendix 1). Any new episodes can simply be added to the end of this list provided they match the numerical form as shown and generated in worksheet 3. Figure 5.5 shows a screenshot of worksheet 4: “Database (store evidence)” with the first 6 episodes of 1024 hypothetical cases pertaining to the detection of CMO. This has been extracted from appendix 2. MyDLP is designed for up to 5000 episodes (or cases). The headings (the variables) are automatically imported from the “variables” worksheet once selected.

	Database Paste data from Record Sheet or from other Excel database											
Record	CMO					DVA<6/9	NVA<N8	Central blu	cataract su	laterality	hyperaemi	a/c cells
1	1					1	1	1	2	4	1	1
2	2					2	2	2	2	1	2	2
3	2					2	2	2	2	1	1	1
4	2					2	2	2	2	1	1	2
5	1					2	1	1	1	1	2	1
6	1					1	1	1	2	4	1	1

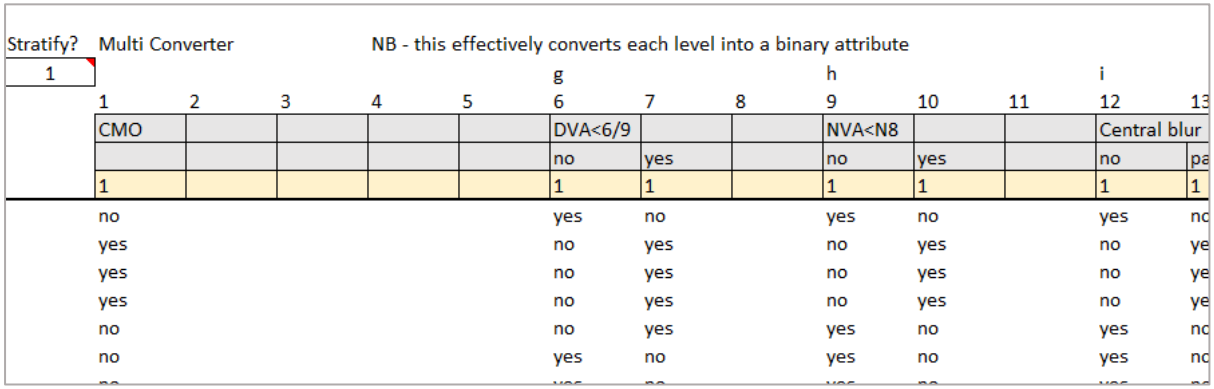
Figure 5.5 A screenshot of worksheet 4: “Database (store evidence)” of MyDLP containing hypothetical data pertaining to 6 patient episodes.

5.2.1.4 Worksheet 5: “Multi-converter”

This is the final worksheet representing step 1 of bTLC (figure 5.2 and table 5.1). In order to apply Bayes’ to the data in the second step of bTLC (learning from the evidence base), the result for each variable needs to be converted into a binary attribute. Bayes’ can then be used to generate the pre-test odds and the LRs for each individual test outcome (multi-level). Worksheet 5 “multi converter” of MyDLP does this (appendices 1 and 2). The variables listed in worksheet 2 (”variables”) are automatically transferred to this worksheet, and the data from worksheet 4: “Database (store evidence)”, are automatically populated and converted into a binary form. This assigns a separate column for each outcome of a test as it is considered separately (so that multi-level LRs can be computed). The counts required for Bayes’ can then be extracted from the data. This completes the second step of bTLC, gathering the evidence in a form conducive to machine learning.

In worksheet 5: “multi converter”, the data is also grouped for 10-fold cross validation (chapter 3, section 3.2.3.1). The clinician has the option to carry out cross validation using stratified or unstratified folds. Here stratification ensures equal numbers of each outcome in each fold used for

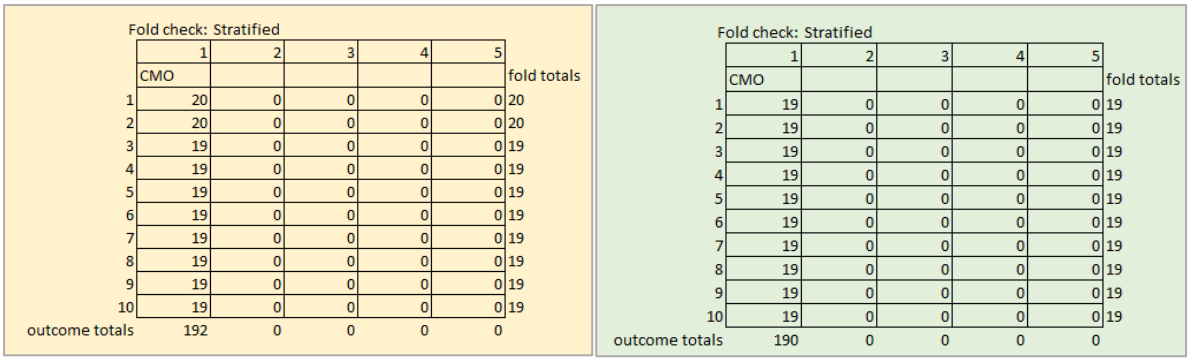
training and testing. Lack of stratification means that each fold has a random allocation of outcomes. Figure 5.6 shows a screenshot of worksheet 5: “multi converter” extracted from appendix 2 containing the hypothetical CMO data. Stratification is selected by simply adding “1” to the “Stratify?” box in the top left of the worksheet.



Stratify?	Multi Converter			NB - this effectively converts each level into a binary attribute												
1						g			h			i				
	1	2	3	4	5	6	7	8	9	10	11	12	13			
	CMO					DVA<6/9			NVA<N8			Central blur				
						no	yes		no	yes		no	pa			
	1					1	1		1	1		1	1			
	no					yes	no		yes	no		yes	nc			
	yes					no	yes		no	yes		no	ye			
	yes					no	yes		no	yes		no	ye			
	yes					no	yes		no	yes		no	ye			
	no					no	yes		yes	no		yes	nc			
	no					yes	no		yes	no		yes	nc			

Figure 5.6 A screenshot of worksheet 5: “Multi converter” of MyDLP for CMO (appendix 2) containing hypothetical data pertaining to 6 patient episodes. Only part of the data is shown in the figure, the entire worksheet can be accessed in appendix 2.

Scrolling to the right of the “multi converter” worksheet leads to areas handling the auto-stratification for Bayes’ learning (beige) and Bayes’ accuracy (green) in readiness for model performance evaluation (step 5 of bTLC, figure 2, table 5.1). The composition of the 10 stratified folds can be examined in the ‘fold check’ tables. Screenshots for both learning and accuracy (extracted from appendix 2) are given in figure 5.7. Note that the fold sizes (see in the column labelled ‘1’ in figure 5.7) are fairly equal (vary from 19 to 20) as stratification has been selected on this worksheet (as described above). Unstratified folds would show a much more unequal distribution of episodes in each fold.



Fold check: Stratified

	1	2	3	4	5	
	CMO					fold totals
1	20	0	0	0	0	20
2	20	0	0	0	0	20
3	19	0	0	0	0	19
4	19	0	0	0	0	19
5	19	0	0	0	0	19
6	19	0	0	0	0	19
7	19	0	0	0	0	19
8	19	0	0	0	0	19
9	19	0	0	0	0	19
10	19	0	0	0	0	19
outcome totals	192	0	0	0	0	

Fold check: Stratified

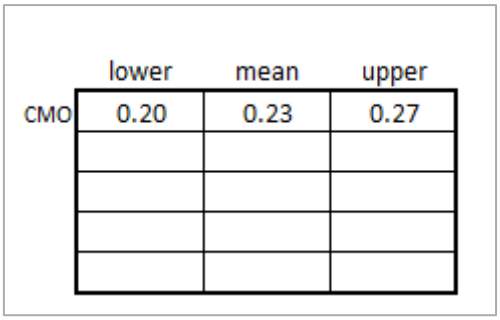
	1	2	3	4	5	
	CMO					fold totals
1	19	0	0	0	0	19
2	19	0	0	0	0	19
3	19	0	0	0	0	19
4	19	0	0	0	0	19
5	19	0	0	0	0	19
6	19	0	0	0	0	19
7	19	0	0	0	0	19
8	19	0	0	0	0	19
9	19	0	0	0	0	19
10	19	0	0	0	0	19
outcome totals	190	0	0	0	0	

Figure 5.7 A screenshot of worksheet 5: "Multi converter" of MyDLP for CMO (appendix 2) showing the fold check tables in readiness for Bayes' accuracy (beige) and Bayes' learning (green) evaluation. Each column represents each dependent variable. As there is only one dependent variable (CMO) in this example, only column '1' is filled. Stratification involving multiple dependent variables would ensure that an equal number were assigned to each fold for later Bayes' learning and accuracy evaluations.

5.2.2 **Step 2:** Learning from the Evidence Base – Worksheet 6: "Bayes'(learn from evidence)"

The second step of bTLC (figure 5.2), learning from the evidence base, is realised with the extraction of the components required for supervised Bayes' machine learning (as described in chapter 3, section 3.2.1). Bayes' involves taking the product of the pre-test odds for a predicted outcome and the multilevel LRs for each of the outcomes of the tests undertaken. The clinical outcome of each test undertaken determines which of the multi-level LRs are used. That is, a test may find that a particular sign or symptom is absent, mild, or severe. Each of these outcomes are assigned individual LRs and therefore the outcome of the test determines which of the multi-level LRs is used in the calculation.

Worksheet 6 of appendix 2: "Bayes' (learn from the evidence)" shows how the pre-test odds, multi-level LRs, 95% confidence limits and fractional uncertainties required for Bayes' predictions are calculated (as described in chapter 3). Here, counts are extracted from the preceding worksheet 5 (containing all individual binary outcomes for the predictor variables as discussed in 5.2.1.4). The pre-test odds (i.e., the rareness of things) are calculated for the outcomes or predicted variables and the multi-level likelihood ratios (i.e., the importance of test outcomes) for each predictor (test undertaken). These are calculated with their 95% confidence limits. Summary tables of the values are given at the top of the worksheet. The first, smaller table, presents the pre-test odds of the possible predicted variables (outcomes) with the lower and upper 95% confidence limits. Figure 5.8 shows a screenshot of this table from appendix 2. This table is reproduced in the summary "Notes" worksheet (Appendix 2, worksheet 1: "Notes") for ease of access.



	lower	mean	upper
CMO	0.20	0.23	0.27

Figure 5.8 A screenshot of the summary table showing the pre-test odds (with 95% confidence limits) of CMO. The pre-test odds for other predicted variables would be listed in the rows below CMO. This has been extracted from appendix 2, worksheet 6: "Bayes' (learn from evidence)" of MyDLP containing hypothetical data.

The second, larger table, shows the multi-level LRs for each individual outcome of the predictor variables. These are also presented with their 95% confidence limits. This table is reproduced in a similar fashion on the summary "Notes" worksheet (Appendix 2, worksheet 1: "Notes").

5.2.2.1 "Safe" predictors

The table of multi-level LRs reproduced in the summary "Notes" worksheet (Appendix 2) includes a column to show the reliability of the LRs for each predictor (i.e. whether they are "safe" or not). Figure 5.9 shows a screenshot of this table. As discussed in chapter 3 (section 3.2.4, analysis 2), a likelihood ratio of 1 has no predictive clinical value as it does not alter the probability of an outcome. Therefore, multilevel likelihood ratios with confidence limits that straddled 1 were deemed spurious or "unsafe" due to the reduced certainty of true clinical value.

Multi-level likelihood ratios (95% confidence limits)

		CMO			
		lower	mean	upper	reliable?
DVA<6/9	no	0.00	0.01	0.04	YES
	yes	10.10	12.76	16.12	YES
NVA<N8	no	0.00	0.01	0.04	YES
	yes	117.00	829.70	5883.55	YES
Central blur	no	0.00	0.01	0.04	YES
	partial	117.00	829.70	5883.55	YES
	yes	0.27	4.30	68.43	no
cataract surgery	no	0.00	0.02	0.16	YES
	yes	1.25	1.29	1.35	YES
laterality	unilateral	0.98	1.01	1.04	no
	bilateral	0.02	0.34	5.36	no
	asymmetric	0.02	0.34	5.36	no
hyperaemia	no	0.65	0.72	0.80	YES
	mild	3.16	4.30	5.84	YES
	severe	0.27	4.30	68.43	no
a/c cells	no	0.36	0.44	0.53	YES
	mild	2.45	2.87	3.37	YES
	severe	0.27	4.30	68.43	no

Figure 5.9 A screenshot of the summary table showing the multi-level LRs (with 95% confidence limits) of the predictor variables of CMO. This has been extracted from appendix 2, worksheet 1: "Notes" of MyDLP and contains hypothetical data. Note, if the confidence limits of a LR straddle 1, it is not deemed reliable and a red 'no' is shown in the associated 'reliable?' column. If, however, the confidence limits do not straddle 1, it is deemed reliable and a green 'yes' is shown in the associated 'reliable?' column.

By highlighting the spurious predictors, the clinician has the choice to either include or exclude these (or indeed any other) predictors in the analysis. In worksheet 6 of MyDLP ("Bayes' (learn from the evidence)"), the assignment (i.e. '1' or blank) is automatically copied from worksheet 2 ("variables") to the "assignment" row under for each variable. In effect variables can be included or excluded from the analysis in either of these places (worksheet 2 or worksheet 6). It is advisable that variables are adjusted in either worksheet 2 or 6 for accurate reference an to minimise confusion. Figure 5.10 shows a screenshot of the "assignment" row in worksheet 6. MyDLP instantly updates the analysis and support tool (iEPR and iVPT) to include only the assigned predictors.

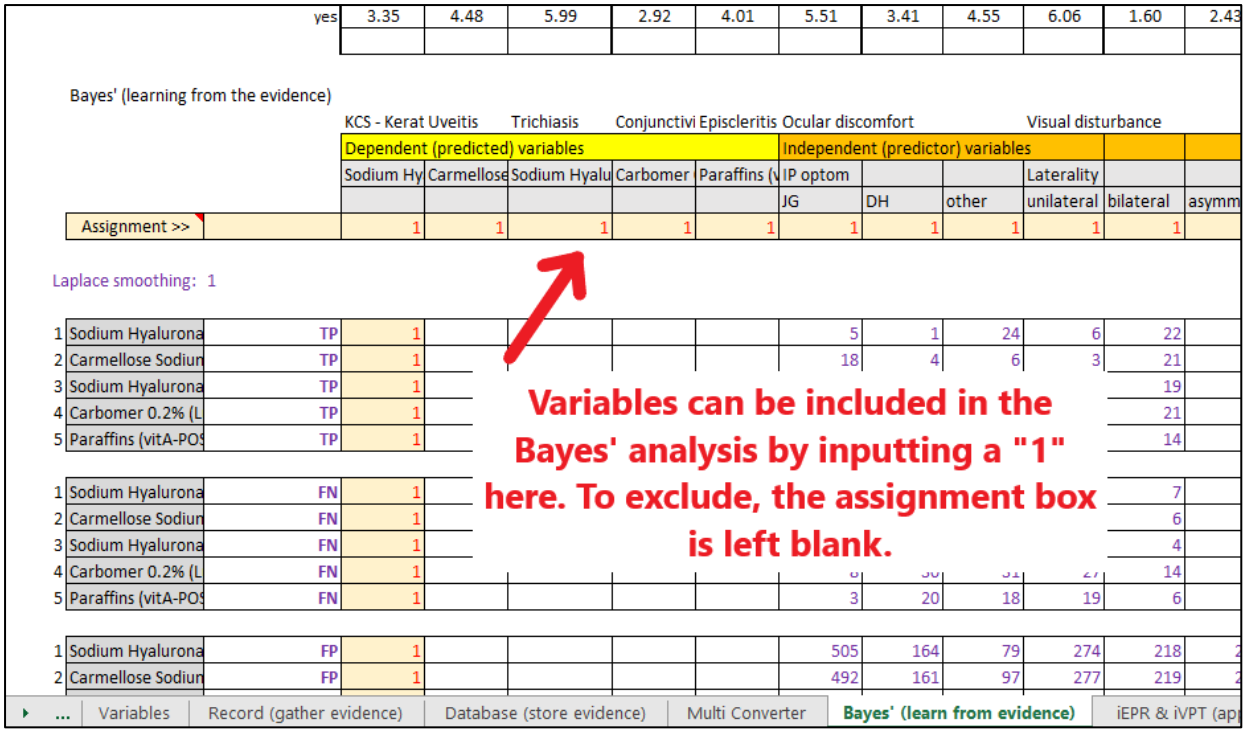


Figure 5.10 A screenshot from MyDLP, Worksheet 6: Bayes' (learn from evidence) showing the "assignment" row used to include or exclude variables in the Bayes' analysis.

5.2.3 **Steps 3 and 4:** Translation to Evidence-based Practice and Research-led Teaching - Worksheet 7: iEPR & iVPT (apply evidence)

Through the combined "intelligent" electronic patient record (iEPR) and the "intelligent" virtual patient tool (iVPT), MyDLP translates the evidence gathered from clinical data into evidence-based practice support and research-led teaching. This pivotal tool combines both steps 3 and 4 of the bTLC (figure 5.2).

A blank combined iEPR/iVPT worksheet can be found in appendix 1 (MyDLP – blank). An iEPR/iVPT worksheet containing hypothetical data for CMO can be found in appendix 2. Figure 5.11 shows a screenshot of the blank combined iEPR/iVPT worksheet within MyDLP. Each area of the interactive worksheet is numbered and a brief explanation relating to these areas follows.

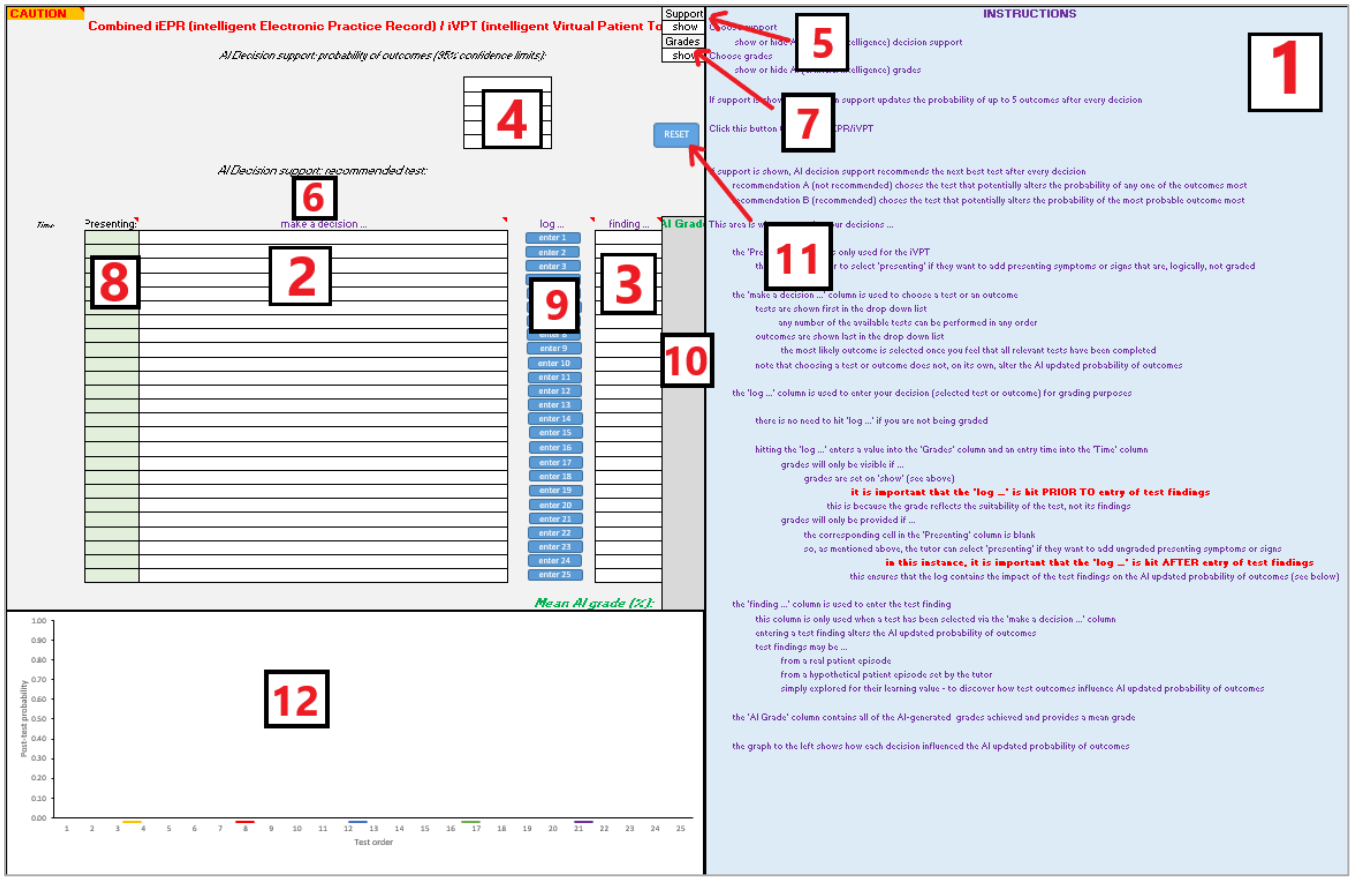


Figure 5.11 A screenshot of the combined iEPR/iVPT tool ("iEPR & iVPT (apply evidence)"). The numbers relate to the interactive areas: (1) Instructions (blue area), (2) Predictor/predicted variable input, (3) Findings of predictor variable, (4) Probability and 95% confidence limits of all predicted variables (outcomes), (5) Drop-down menu to 'show' or 'hide' decision support, (6) Area showing next-test decision support, (7) Drop-down menu to 'show' or 'hide' grading of decisions (iVPT element only), (8) Area to select presenting predictors (iVPT element only), (9) 'Log' buttons to record each clinical decision selected (iVPT element only), (10) Grading area (iVPT element only), (11) Whole system reset button, (12) A graph showing probability change for each predicted variable (outcome) given the result of the predictor selected.

5.2.3.1 Area 1: Instructions

The first interactive area is the blue box of instructions. This is a quick reference guide constantly available to the clinician or trainee. It not only presents practical instructions on use, but also provides a brief explanation of each step undertaken.

5.2.3.2 Area 2: Predictor/predicted variable input

Labelled "make a decision…" this column forms the main body of the combined support tool. Decisions about the tests undertaken (predictor variables, i.e. anterior chamber cells) or outcomes decided (predicted variables, i.e. CMO) are entered here. Upon selecting a row, a drop-down list becomes available which offers a choice of up to 25 possible predictor variables (tests) first, followed by the 5 possible predicted variables (outcomes). A clinician using the iEPR for decision support

would begin use by entering the test undertaken in area 2. Figure 5.12 shows a screenshot of the available predictors in appendix 2 containing hypothetical data.

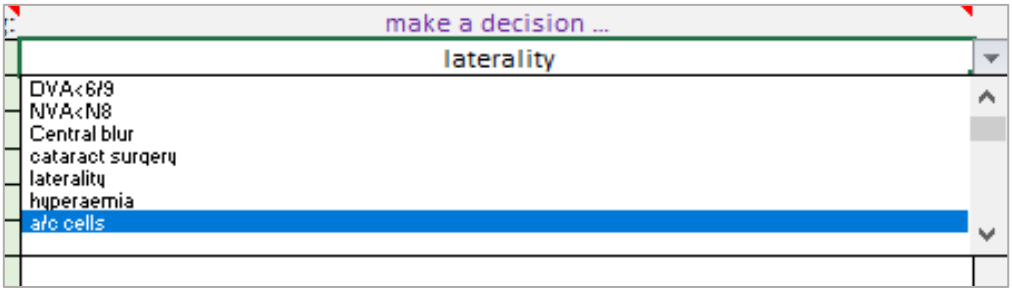


Figure 5.12 A screenshot from the "iEPR & iVPT" worksheet, area 2, showing the drop-down list used to make decisions about test or outcome selection. Here a/c cells has been selected. Predictor variables are listed in the first 25 rows (if there are fewer than 25 predictors, as in this example, remaining rows are left blank). This is followed by the 5 predicted outcomes (again, if there are fewer than 5 predicted variables, the remaining rows are left blank).

5.2.3.3 Area 3: Findings of predictor variable

Entering the test or predictor variable alone does not alter the probability of an overall outcome (i.e. a diagnosis or prescribing decision). The test finding identifies the LR to be used in the Bayes' analysis, and thus, is pivotal to ascertaining final probabilities (much like in clinical practice where the finding of a test dictates further testing or a diagnosis). Area 3 is where this outcome is entered. Labelled "finding…" the corresponding row to the test selected automatically loads the possible findings of that test from the "variables" worksheet. Here, the clinician is to enter the test result. Figure 5.13 shows a screenshot of the corresponding drop-down list of findings relating to a/c cells.

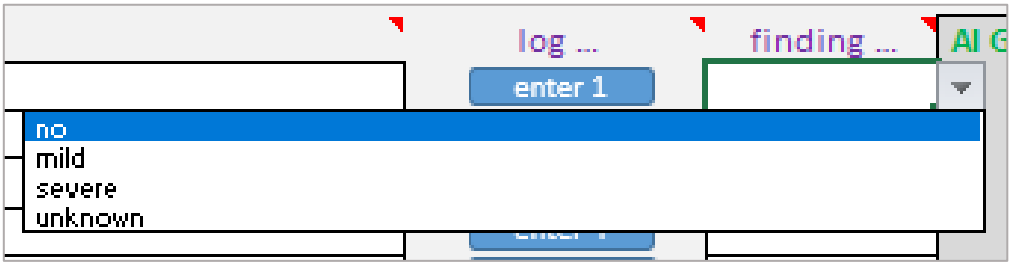


Figure 5.13 A screenshot from the "iEPR & iVPT" worksheet, area 3, showing the drop-down list used to enter the findings of a test selected. Here the possible findings for a/c cells are listed as 'no', 'mild', 'severe' or 'unknown'.

5.2.3.4 Area 4: Probability and 95% confidence limits of all predicted variables (outcomes)

The probability of a predicted variable (i.e. an overall outcome such as a diagnosis or a prescribing decision) is displayed as a table in area 4. This probability adjusts with the results of the tests undertaken (as described above). Figure 5.14 shows how the probability of CMO changes with the selection of two tests with differing results. The differing LRs associated with each clinical finding are calculated from the data collected. Bayes' is then applied, and the outcome probability changes depending on the finding selected. The table in area 4 allows the clinician to see how the probability of five predicted variables (outcomes) adjust with the gathering of more evidence in the form of clinical testing. It also allows the clinician to gauge the reliability of such predictions (with 95% confidence limits – CL, shown in brackets), and consider the other competing outcomes (such as differential diagnoses or different prescribing options).

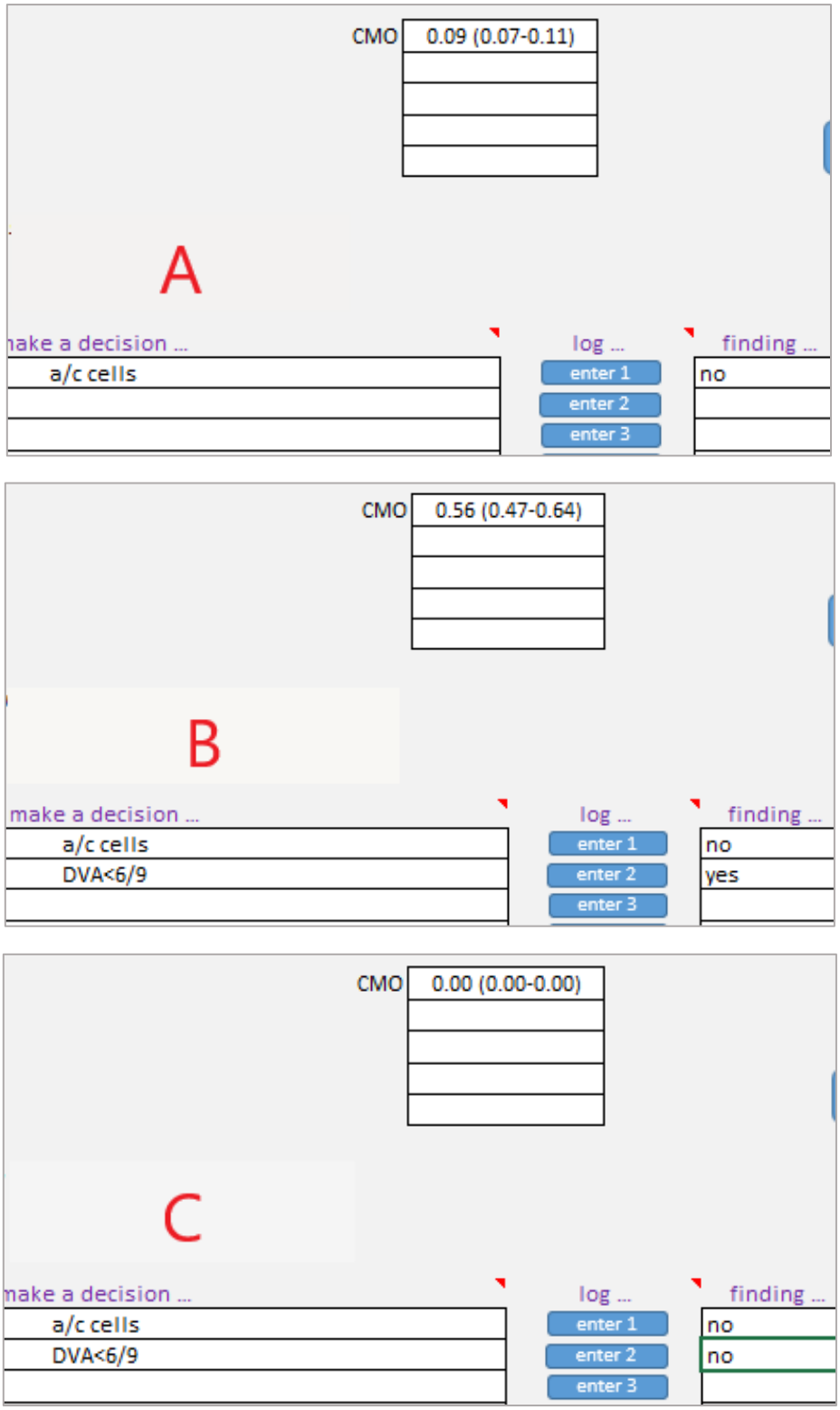


Figure 5.14 Three screenshots from the "iEPR & iVPT" showing the change in probability of a diagnosis of CMO following the results of testing. Screenshot A shows that with no a/c cells the probability of CMO is 9% (CL as percentages: 9-11%). With the additional finding of a reduced distance visual acuity of <6/9 (screenshot B), the probability of CMO rises to 56% (CL as percentages 47-64%). However, if the distance visual acuity is not reduced (screenshot C), the probability of CMO is reduced to 0%. All data used in this example is hypothetical.

5.2.3.5 Area 5: Drop-down menu to 'show' or 'hide' decision support

This drop-down control allows the clinician to show or hide the decision support given by MyDLP. The forms of decision support affected are both the next test suggestion (area 6, discussed below) and the outcomes probability table in area 4.

5.2.3.6 Area 6: Area showing next-test decision support

Aside from showing the changing probabilities of each predicted variable (outcome) with confidence limits, the iEPR/iVPT also provides support selecting the most appropriate next test. This is shown in area 6 (figure 5.11). It presents this in two ways (recommendation A, in red, and recommendation B, in green), both using AI based on Bayes' theorem. The iEPR/iVPT calculates how much each clinical finding (predictor) alters the probability of an outcome of interest. This is based on the pre-test odds of the outcome and the LRs of all previously selected test findings. MyDLP ranks each potential predictor before recommending the highest-ranking test (i.e. the one that would make the most difference to the outcome of interest). Recommendation A (shown in red, as in figure 5.15) suggests the test that would alter the probability of any of the 5 outcomes the most. Recommendation B (shown in green, as in figure 5.15) suggests the test that would alter the probability of the most probable outcome the most. Since there is only one outcome (CMO) in the example used thus far, the next test recommended by both method A and B would be the same. To demonstrate differing recommendations from each method, figure 5.15 shows a screenshot of the iEPR/iVPT containing two diagnoses (uveitis and KCS). Here, two different recommendations have been given.

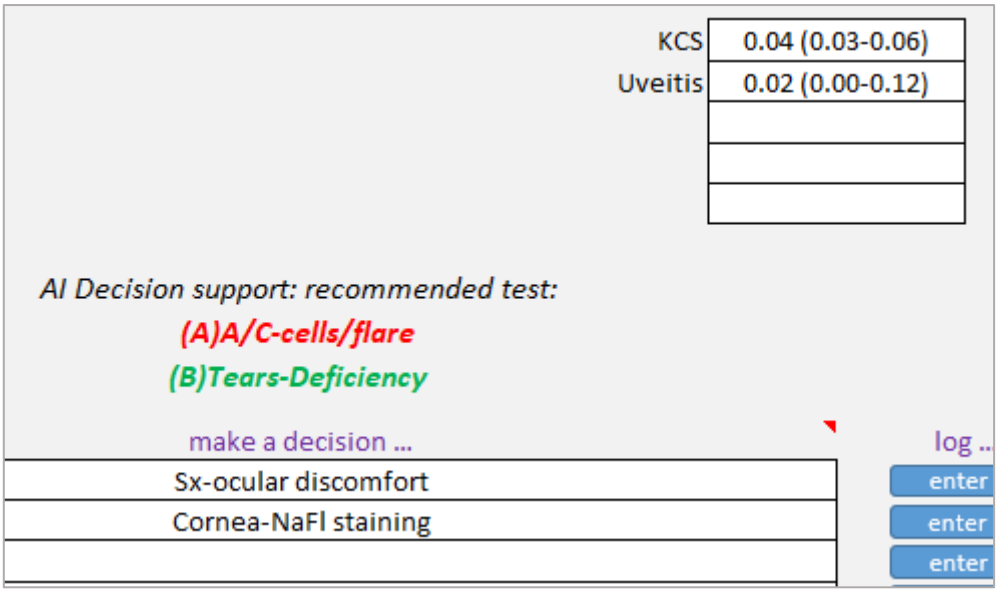


Figure 5.15 A screenshot from a "iEPR & iVPT" showing differing next-test decision support provided by method A (red) and method B (green). The differential diagnoses depicted here are KCS and uveitis.

5.2.3.7 Area 7: Drop-down menu to 'show' or 'hide' grading of decisions (iVPT element only)

This drop-down menu allows the trainee IP optometrist to choose to hide or show the automated AI grading provided by the virtual patient tool (iVPT) aspect of the iEPR/iVPT. It can be turned on or off by simply selecting either 'hide' or 'show' from the appropriate drop-down menu.

5.2.3.8 Area 8: Area to select presenting predictors (iVPT element only)

When the grading is selected to be shown, each variable (test) selected by the trainee IP optometrist (in area 2: predictor/predicted variable input) results in an automatic grade. The tutor may choose to invite trainee interaction by giving a presenting sign or symptom (in the form of a test and its result). These presenting variables must be excluded from grading. This is done by selecting 'presenting' in the corresponding box from the neighbouring column also labelled presenting. Figure 5.16 shows a screenshot of this drop-down.

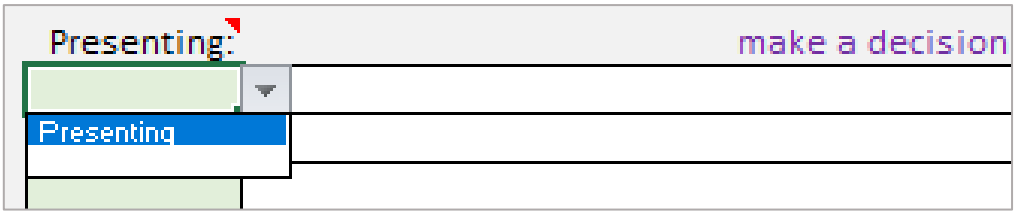


Figure 5.16 A screenshot from a "iEPR & iVPT" showing the presenting drop down menu used to exclude presenting variables from grading.

5.2.3.9 Area 9: 'Log' buttons to record each clinical decision selected & Area 10: Grading area (iVPT element only)

Each 'log' button (to the right of the area 2: predictor/predicted variable input) registers the corresponding test decision made by the trainee. These decisions can then be graded by the iVPT. The button is only pressed when grading is required. The appropriateness of the next test selected demonstrates understanding of the examination process (regardless of the outcome), making it the basis of grading. It is due to this that the 'log' button is hit *prior* to the entry of test findings.

When a test selection decision is logged, the date and time is recorded. Figure 5.17 shows 2 logged tests in the "time" column to the left of the predictor/predicted variable area (where the test is selected). It also shows the grades allocated for each decision to the far right ("AI Grade" column). A mean AI grade is given at the bottom of the "AI Grade", this being the overall mark.

Where the selected variable is a presenting one (provided by the tutor), the 'log' button is hit *after* the outcome of the test is entered in the appropriate column. This updates the background AI (Bayes') without allocating a grade to the presenting variable.

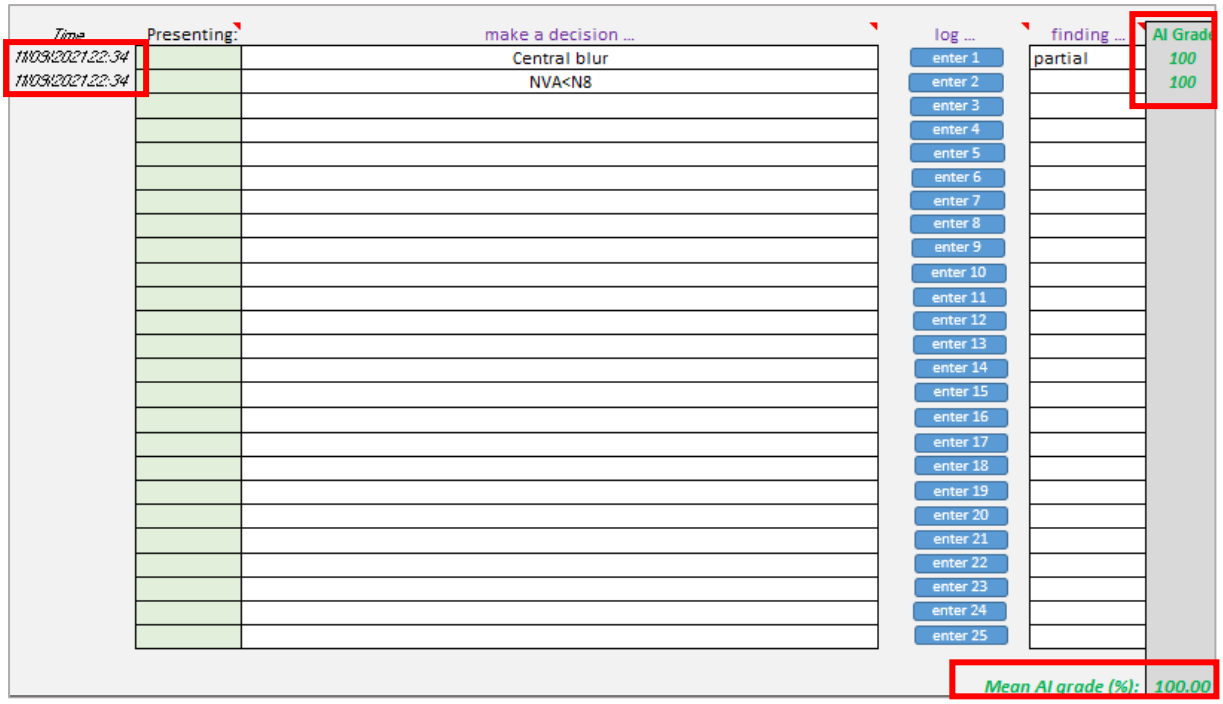


Figure 5.17 A screenshot from a "iEPR & iVPT" showing the logging and automatic AI grading of clinical decisions. The date and time (far left) and the AI grade for each clinical test undertaken (far right) are recorded. The overall AI grade is shown in the bottom right.

5.2.3.10 Area 11: Whole system reset button

Area 11 (figure 5.11) shows the system reset button. As implied, this clears the iEPR/iVPT of any selected variables and their findings.

5.2.3.11 Area 12: A graph showing probability change for each predicted variable (outcome) given the result of the predictor selected

The graph shown in area 12 (that appears empty in figure 5.11) tracks changes in the final probabilities of the predicted (outcome) variables with the tests selected. It highlights the effect of test findings upon diagnoses and prescribing decisions and thus is a very useful learning tool. Figure 5.18 presents a graph showing the effects of random testing upon the diagnostic probabilities of KCS versus uveitis. Here, the effect of each test upon the final probability of with KCS or uveitis can be tracked. After the first 3 tests KCS seems the likely diagnosis. However, unsurprisingly, with the presence of cells/flare in the anterior chamber (test 6), the probability of uveitis rises dramatically before finishing at 92%. At the same time the probability of KCS plummets to zero. The iEPR/iVPT can be interrogated by the clinician/trainee to ascertain which tests are the most diagnostic. The effect of the degree of findings (i.e. mild or severe) on outcome probability can also be explored.

Recall that this whole system is built upon real patient data collected by practising specialist IP optometrists, and thus presents a truly evidence-based way of learning.

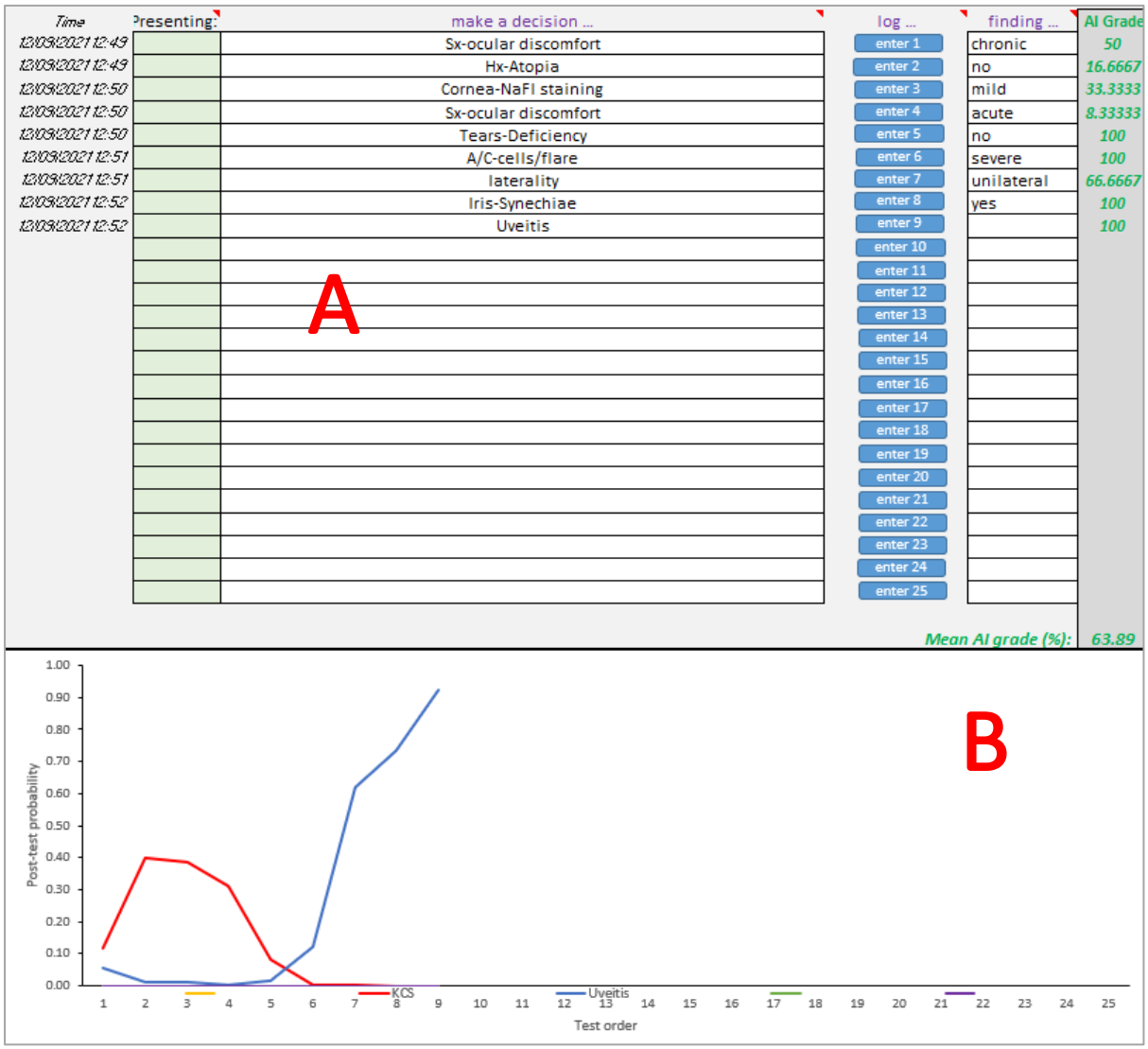


Figure 5.18 A screenshot from a "iEPR & iVPT" showing (A) the tests undertaken, their findings and the grading for each decision (grey area) and (B) the corresponding graph (bottom of screenshot) showing how the probability of each outcome (KCS versus uveitis) is affected by the findings of each test undertaken.

5.2.3.12 Support for the clinician within the iEPR

For the practicing clinician looking to make an accurate diagnosis or prescribing decision, the iEPR presents evidence-based decision support in the following ways:

(a) By presenting an easy reference to the likely differential diagnoses or prescribing decisions, complete with 95% confidence limits for each

(b) By providing a support system presenting the most appropriate next test based upon the pre-test odds, LRs of tests already undertaken (regardless of the order in which the tests were taken), and the probabilities of likely outcomes
(c) Allows decision support to be switched on and off, i.e. used on an 'as required' basis

In the context of this thesis, this support is based upon the clinical activity of experienced specialist IP optometrists. There is an assumption that the practice of experienced practitioners reflects evidence-based practice. This can be viewed as a limitation of using this kind of data. However, as discussed in chapter 4, the clinical activity of APCOS, from whom the data was collected, largely reflects the CMGs and therefore evidence based practice. The advantage is that this activity can be readily updated by adding more data to MyDLP (worksheet 4: Database) and thus has the potential to offer 'live' evidence-based peer support. It could be considered as a mechanism of putting 'old heads' on young shoulders.

5.2.3.13 Support for the trainee within the iVPT

For the training clinician looking to learn about the effects of signs and symptoms on diagnoses and prescribing decisions, the iVPT presents, through evidence-based teaching, a means to:

(a) Interrogate an evidence-based system by making hypothetical clinical decisions based upon presenting signs and symptoms (allowing the practice of clinical investigation and decision-making)
(b) Allow automatic and consistent evidence-based grading of decisions that lack well known inconsistencies of human grading (199)
(c) Track, graphically, the effect of each decision made upon the probability of final outcomes (to understand how some tests may increase or decrease the probability of multiple final outcomes, whilst others are diagnostic)
(d) Easily assess the likely differential diagnoses or prescribing decisions, complete with 95% confidence limits
(e) Switch decision-support and grading on and off, so that the student can get the benefit of instant formative feedback while training which can be hidden during assessments

5.2.4 **Step 5**: Evolving the system

Evolving the system concludes the bTLC as its final step (figure 5.2). In order to evolve or improve a model, a performance evaluation must take place. The performance evaluation answers the pivotal questions:

(a) Are the model predictions accurate enough to provide clinical support?

(b) How do we measure and quantify model performance?

(c) Has the model reached optimal performance?

(d) Can performance be improved? And if so, how can performance be improved?

MyDLP answers these questions by presenting an evaluation of the iEPR/iVPT performance. Within MyDLP there are two stages of evaluating the tool with view to evolving the system:

Stage 1 - Bayes accuracy (evaluation): worksheet 8

Stage 2 - Bayes' learning (evaluation): worksheet 9

These interactive worksheets can be found in both appendix 1 (blank) and appendix 2 (CMO example). These two stages are now discussed in further detail.

5.2.4.1 Stage 1: Bayes' accuracy (evaluation) – worksheet 8

The first stage evaluates the performance of the iEPR/iVPT model. This allows the clinician or trainee to assess how accurate Bayes' predictions and suggestions are. As discussed in chapter 3 (section 3.2.3.2 Measuring performance, - accuracy, informedness and markedness), accuracy alone is insufficient as a predictor of overall model performance due to the effects of chance. Informedness and markedness have therefore been added to accuracy to give an evaluation of performance that is not prone to chance. Worksheet 8 automatically and rapidly generates the accuracy, informedness and markedness of the model. This is done through automated 10-fold cross validation (chapter 3, section 3.2.3.1) which generates the means and standard deviations of accuracy, informedness and markedness for each predicted outcome. Weighted values for all outcomes combined are also generated. The estimates generated are affected by both the addition of the Laplace correction in the "variables" worksheet (section 5.2.1.1) and the stratification status selected ("Multi converter" worksheet, section 5.2.1.4). Figure 5.19 shows a screenshot extracted from appendix 2: MyDLP (CMO example). It shows worksheet 8: "Bayes' accuracy (evaluation)" containing the hypothetical data. Each of the interactive areas of the worksheet are explained below.

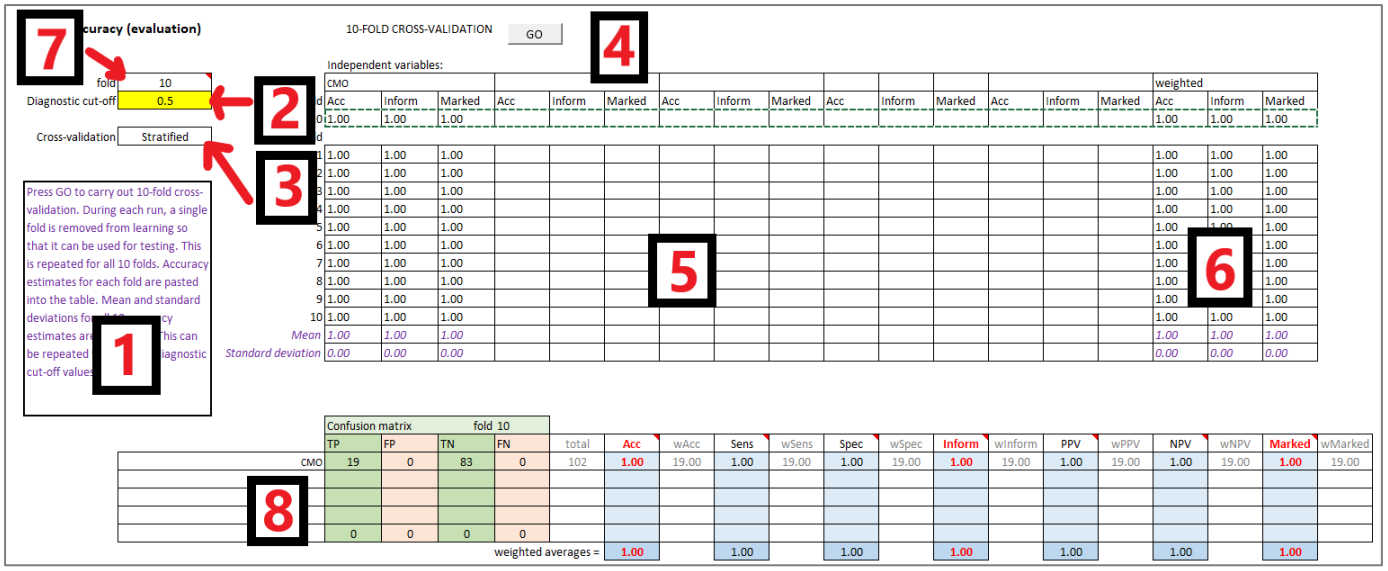


Figure 5.19 A screenshot of worksheet 8: "Bayes' accuracy (evaluation)" taken from appendix 2. The numbers relate to interactive parts of the worksheet, namely: (1) Instructions, (2) diagnostic cut-off value (yellow box), (3) stratification window, (4) 'GO' button to start automated 10-fold cross-validation, (5) The generated means and standard deviations for accuracy (acc), informedness (Inform) and markedness (Marked) of each fold, (6) The weighted values of accuracy, informedness and markedness for all outcomes, (7) 'Fold' box – entering a number here (1-10) allows the clinician to examine the confusion matrix (presented at (8)) associated with that fold along with its corresponding sensitivity (sens), specificity (spec), positive predictive (PPV) and negative predictive values (NPV). The hypothetical data presented here only involves one predicted (outcome) variable (CMO) and thus the weighted performance values (6) are the same as the performance values given in area (5).

5.2.4.1.1 Area 1: Instructions

The instructions contain a brief explanation of the purpose of the worksheet, with 10-fold cross-validation and diagnostic cut-off values highlighted.

5.2.4.1.2 Area 2: Diagnostic cut-off box

The default diagnostic cut-off value is set at 0.5 (50%). Altering the cut-off value to maximise the accuracy of a model or test is an established practice (200). Indeed, receiver-operator curves (ROC) plot the true positives versus the false positives to identify the optimum cut-off value. The clinician is able to alter this cut-off value and inspect the data at multiple values very rapidly due to the automated calculation process within worksheet 8 (which makes use of an embedded MACRO - Excel®). By this means confusion matrices, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for every diagnostic cut-off value selected, and for every fold of data, can be readily inspected. The effect of altering the diagnostic cut-off on the performance of the model can thus be very quickly explored (i.e. can model performance be improved? Which diagnostic cut-off makes the model perform optimally?).

5.2.4.1.3 Area 3: Stratification window

This states whether ‘stratified’ or ‘unstratified’ folds were selected in the “Multi converter” worksheet.

5.2.4.1.4 Area 4: ‘GO’ button

This button starts automated 10-fold cross validation by triggering the embedded MACRO which generates performance measures are generated for all folds of the data.

5.2.4.1.5 Area 5: Performance measures for each fold

This area shows the mean accuracy, informedness and markedness for each fold of the data. Each of the 5 predicted (outcome) variables has a section on the table displaying the results of each fold. The overall mean and standard deviations of each outcome are presented at the bottom of the table. Since the example used in figure 5.19 concerns one predicted (outcome) variable, only one section of the table is filled.

5.2.4.1.6 Area 6: Weighted performance measures

This area of the table shows the weighted measures of accuracy informedness and markedness. The term ‘weighted’ means that the calculated performance measures account for the prevalence of each of the five specified predicted outcomes. That is, it presents the competency of the model as a whole covering all outcomes. As there is only one specified predicted outcome here (i.e. CMO), any values in this area will be identical to those found in the CMO part of area 5.

5.2.4.1.7 Area 7: ‘Fold’ box

Entering a number here (from 1-10) allows the clinician to examine the confusion matrix (presented at (8)) associated with that fold. These values and the corresponding sensitivity (sens), specificity (spec), positive predictive (PPV) and negative predictive values (NPV) can be viewed.

5.2.4.1.8 Area 8: Confusion matrix

This area presents a dynamic confusion matrix for all predicted (outcome) variables. It shows the breakdown of the data for the fold selected (area 7). To the right, it also presents the individual measures of performance (accuracy, sensitivity, specificity, informedness, markedness, PPV and NPV) for each outcome in that fold.

The contents of worksheet 8: “Bayes’ accuracy” therefore allow the clinician to assess the performance of the model underlying the iEPR/iVPT tool. This can be done both for individual outcomes (each predicted variable), but also for the whole model. It also allows the clinician to investigate whether altering the diagnostic cut-off value can improve model performance (i.e. provide better clinical support).

5.2.4.2 Stage 2: Bayes’ learning (evaluation) – worksheet 9

Recall from chapter 1 (section 1.5.2) that power calculations were not undertaken to establish a required sample size. This is because the data collected was composed of the entire population of patients examined by APCOS over a 1-year period. This study, therefore, took the form of big data collection and analysis. Also, Bayesian statistical theory (i.e. conditional probability), as opposed to frequentist probability, was the basis of the model used to analyse the data. This meant that the frequentist alpha and beta levels, required to determine whether a sample had provided enough power to determine statistical significance in frequentist probability methods, no longer applied (201). Instead, the competence of the model was evaluated by examining learning efficiency in the second stage of evolving the system.

The second stage of ‘Evolving the system’ (the final step of bTLC) determines the extent to which Bayes’ has achieved maximum learning. This addresses the questions given in section 5.2.4 namely, (c) has the model reached optimal performance? – i.e. has Bayes’ learnt from enough episodes? and (d) can performance be improved? And if so, how can performance be improved? – i.e. will the addition of more patient episodes improve model performance?

Worksheet 9 “Bayes’ learning (evaluation)” rapidly automates the process of determining the extent to which the model has achieved maximum learning. Once again, this is achieved by using an embedded MACRO® (Excel) which plots a learning efficiency curve showing the change in weighted accuracy, informedness and markedness as increasing numbers of patient episodes (folds) are used for learning. The data is split into 10 folds, the size of each fold can be altered. An increasing number of folds 1-8 are used for learning, while folds 9 and 10 are reserved for testing the model (20% of the total data is used for testing). This sectioning of the data is based on a known approach, though others have used 25% for testing as opposed to 20% (99). Figure 5.20 shows a screenshot of the final worksheet of MyDLP, that is worksheet 9 “Bayes’ learning (evaluation)”. Each of the interactive parts are explained in more detail below.

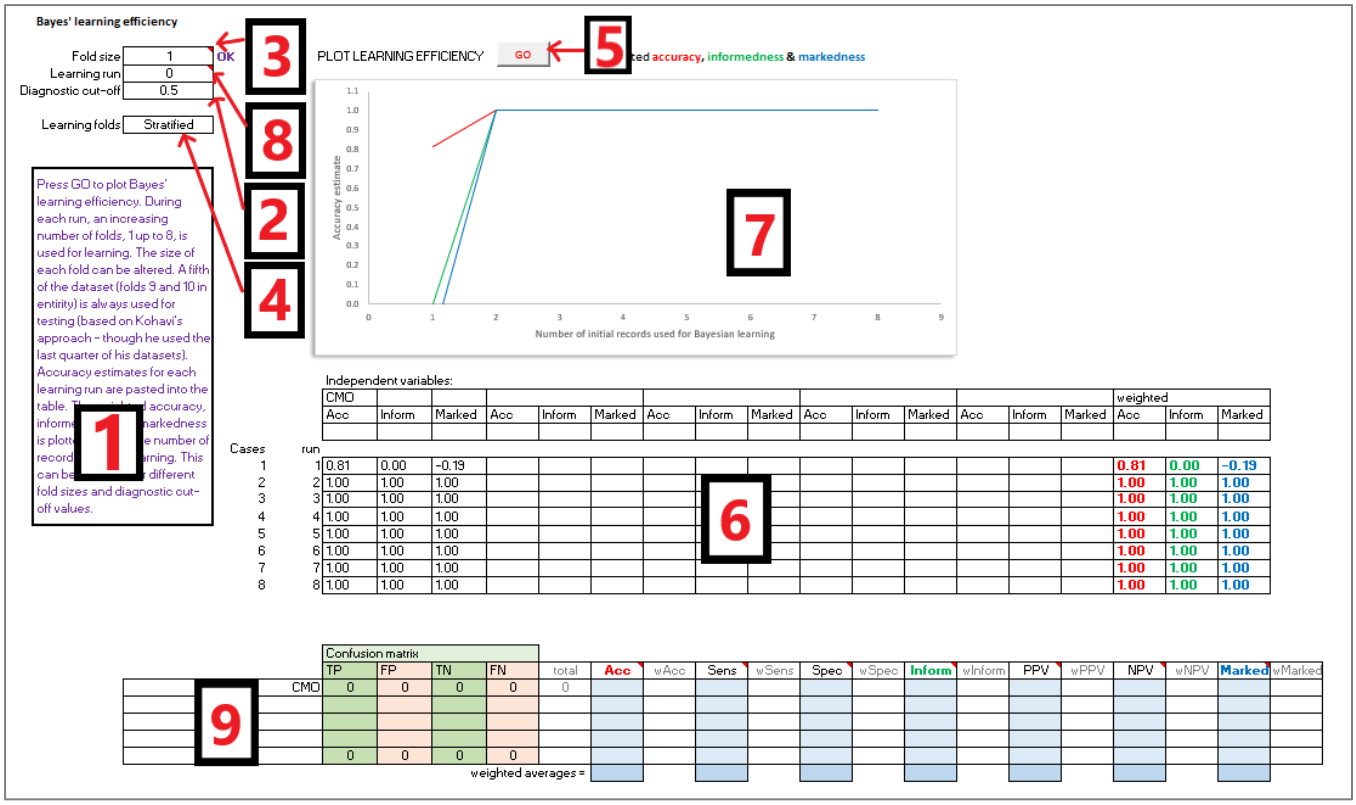


Figure 5.20 A screenshot of worksheet 9: "Bayes' learning (evaluation)" taken from appendix 2. The numbers relate to interactive parts of the worksheet, namely: (1) Instructions, (2) diagnostic cut-off value, (3) learning fold size, (4) stratification window, (5) 'GO' button starts automated calculation of learning efficiency, (6) the generated accuracy (acc), informedness (Inform) and markedness (Marked) of each fold are displayed. The mean weighted performance measures for all predicted (outcome) variables per fold is given in bold to the right of the table, (7) A plot showing the weighted values of accuracy (red), informedness (green) and markedness (blue) against number of patient episodes used for learning – a learning efficiency plot, (8) 'Fold' box – entering a number here (1-8) allows the clinician to examine the confusion matrix (presented at (9)) associated with that fold along with its corresponding sensitivity (sens), specificity (spec), positive predictive (PPV) and negative predictive values (NPV). The hypothetical data presented here only involves one predicted (outcome) variable (CMO) and thus the weighted performance values are the same as the performance values.

5.2.4.2.1 Area 1: Instructions

The instructions contain a brief explanation of the purpose of the worksheet and how it works. It highlights that fold size and diagnostic cut-off values can be altered and the calculations re-run.

5.2.4.2.2 Area 2: Diagnostic cut-off value

Much like the previous worksheet 8, the default diagnostic cut-off value is set at 0.5 (50%). This value can be altered. The effect of altering the diagnostic cut-off on the performance of the model can therefore be very quickly explored.

5.2.4.2.3 Area 3: Learning fold size

Here the size of the learning folds (number of episodes a fold contains) can be altered. For a model that learns very quickly the fold size needs to be smaller to see exactly how many episodes are needed to reach maximum learning. In figure 5.20 the fold size is at 1, this is because highly correlated hypothetical data was used to demonstrate MyDLP (i.e. the predictor variables were a perfect match to the predicted outcome). Therefore the model effectively needed 2 episodes to discover the rules and reach maximum learning. When altering the fold size, the system indicates with "OK" if the fold size selected is possible (i.e. there are enough episodes to fill 10 folds of a specified size) and "TOO HIGH" if not.

5.2.4.2.4 Area 4: Stratification window

Again, as with the previous worksheet 8, this box states whether 'stratified' or 'unstratified' folds were selected in the "Multi converter" worksheet.

5.2.4.2.5 Area 5: 'GO' button

This button starts automated calculation of learning efficiency via the embedded MACRO which generates performance measures for all 8 folds of the data used for learning.

5.2.4.2.6 Area 6: Performance measures for each fold and weighted performance measures

The generated performance measures (accuracy, informedness and markedness) for all predicted outcomes are presented on this table. The weighted mean of overall system performance measures (including all outcomes) in given on the right (accuracy in red, informedness in green and markedness in blue – the colours corresponding to the colours of each performance measure in the learning efficiency plot shown in area 7).

5.2.4.2.7 Area 7: Learning efficiency plot

The learning efficiency curves show the extent to which Bayes' has achieved maximum learning. The weighted values for accuracy (red), informedness (green) and markedness (blue) are plotted against the number of records (patient episodes) used for Bayes' learning. As explained in chapter 3 (section 3.2.3.2.2), markedness is considered the most relevant measure as it is free of the effects of chance and is related to the predictive ability of a model. It is common to see (as in figure 5.20) that there is an initial phase of rapid learning followed by an asymptote. Once this asymptote is achieved Bayes' has reached its maximum learning potential and no further improvement can be made by adding more data (99). In the hypothetical example given, Bayes' reached a maximum learning efficiency of 100% after 2 episodes, however, such a high learning efficiency is not expected in real-life data,

which is more complex. Gurney (2017) found that a maximum accuracy of 94-99% required 69 cases of primary open angle glaucoma (98). His study concerned far fewer predictor variables and so it is expected that the present study will show that many more learning episodes are needed to reach maximum learning efficiency.

5.2.4.2.8 Area 8: 'Fold' box and area 9: confusion matrix

As with worksheet 8 (figure 5.19), entering a number here (from 1-8) allows the clinician to examine the confusion matrix (presented at (9)) associated with that fold. The corresponding sensitivity (sens), specificity (spec), positive predictive (PPV) and negative predictive values (NPV) for that fold can then be viewed.

5.2.5 Worksheet 1: "Notes"

The "Notes" worksheet (appendices 1 and 2) presents an explanation of MyDLP (in relation to the five-steps of the Bayes' translational learning cycle) alongside a summary of all the key findings. This includes the variables included in the analysis, and as mentioned above (section 5.2.2) the pre-test odds and LRs. It also includes summary tables showing the model performance for each of the predicted (outcome) variables (extracted from worksheet 8) and the learning efficiency plots extracted from worksheet 9. This worksheet in not interactive but merely presents a quick summary.

5.3 Chapter Summary

The unexpected poor performance of hierarchical Bayes' (chapter 3) forced the conclusion that an interactive and evolving evidence-based support system for IP optometrists could not be based on that model. Although the aim of the study remained to create a model which would follow through on the 5 steps of the Bayes' translational learning concept (bTLC), the application of machine learning was re-evaluated. This led to the development of a digital learning platform (MyDLP) which would demonstrate a 'proof-of-concept'.

The present chapter introduced MyDLP, its components, and functionality. MyDLP can provide decision-support for practising clinicians and those in training across clinical professions, whilst enabling an understanding of machine learning concepts. Considering IP optometrists, this not only provides clinical support for an evolving profession whose workload increasingly involves the community management of ocular conditions (12,202) but also answers the calls of the Topol review to "prepare the healthcare workforce for a digital future" (28) (See chapter 1, section 1.1.3.3).

MyDLP includes detailed instructions and explanations of each stage of machine learning, using the bTLC (chapter 1, section 1.6). It incorporates a combined interactive 'intelligent' electronic patient record (iEPR) and 'intelligent' virtual patient tool (iVPT). With interactive areas for inputting more data (in-built data collection EPR) and manipulating data analysis (a choice to remove spurious predictors and the addition or otherwise of Laplacian smoothing and stratification), basic machine learning concepts can be explored by the user. Moreover, model performance can be evaluated using performance measures and learning efficiency curves. This has the further potential of opening up research opportunities to practising clinicians or creating "citizen data scientists" (64) bridging the gap between AI and clinicians.

The following chapters (6 and 7) will present MyDLP used for providing clinical support for the diagnosis and therapeutic decisions relating to KCS and uveitis using the data collected (Chapters 2 and 4).

Chapter 6: Using MyDLP for diagnoses

6.1 Introduction

Chapter 5 described the components and functionality of the digital learning platform (MyDLP). Using up to 25 predictor variables, MyDLP (the Bayes' model) can predict up to 5 outcome variables (for example diagnoses or drug treatment options). MyDLP has the potential to provide decision support to clinicians and trainees across disciplines. Moreover, it can be used as an evidence-based teaching tool, automatically grading trainee performance both for each clinical decision made, and overall (chapter 5, section 5.2.3). MyDLP also houses an evaluation system which allows the clinician to explore and evaluate the performance and learning efficiency of the Bayes' machine learning the model (chapter 5, section 5.2.4).

The current chapter presents the application of MyDLP to predict diagnoses of KCS and uveitis (the chosen predicted or outcome variables). Figure 6 shows the chapter plan.

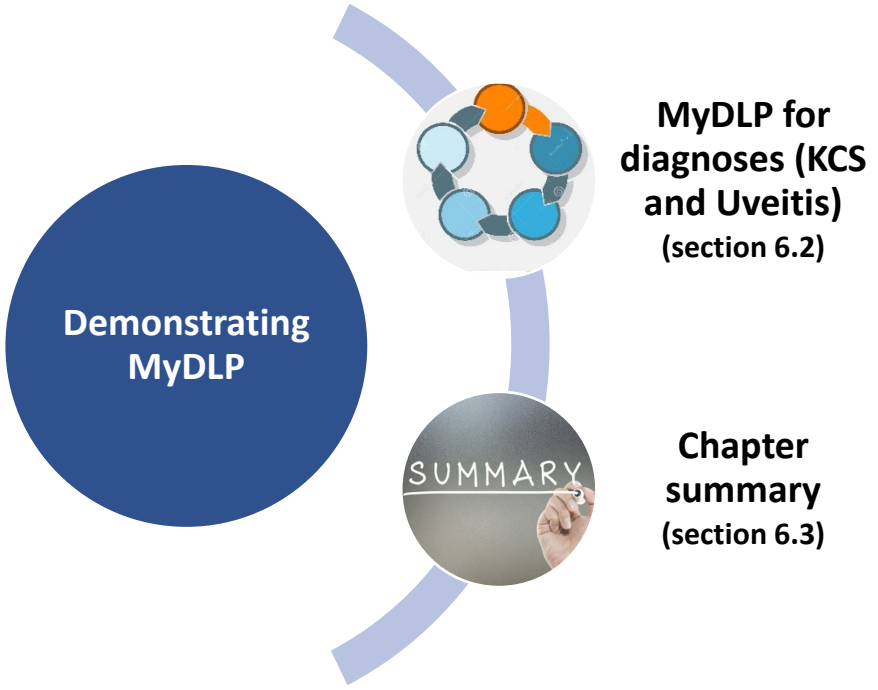


Figure 6.1 The chapter plan for Using MyDLP for diagnoses

6.2 MyDLP for diagnoses (KCS and uveitis)

Following on from chapter 5, in the present chapter MyDLP (for diagnoses) is demonstrated systematically using the 5-step bTLC. Figure 6.2 shows the 5-step bTLC as a reference.

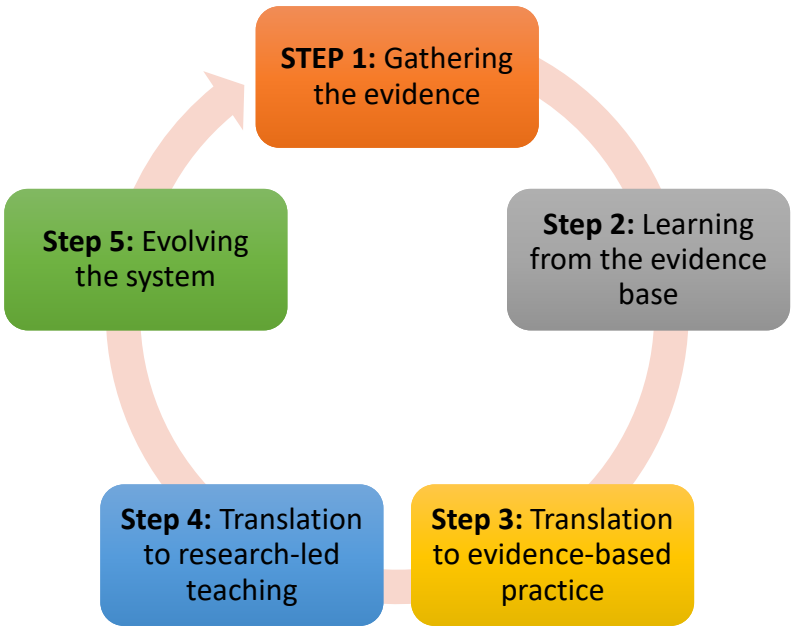


Figure 6.2 The five steps forming the Bayes' translational learning concept (bTLC). The demonstration of MyDLP in this chapter follows these 5 steps.

6.2.1 **Step 1:** Gathering the Evidence

The 1351 patient episodes collected by the specialist IP optometrists over the course of a year, were used to drive MyDLP. This provided the evidence-base. The insufficiency of data for many of the 48 presenting ocular conditions contributed to poor model performance when hierarchical Bayes' was applied (chapter 3). Keratoconjunctivitis Sicca (KCS) and uveitis were therefore selected as the only outcomes of choice (predicted variables) to best demonstrate MyDLP. These ocular conditions were those most frequently encountered by APCOS and thus provided the highest number of patient episodes from which the model could best learn and be evaluated (157 and 75 episodes respectively – chapter 4, figure 4.3). Restricting the diagnoses to KCS and uveitis, provided a 'proof-of-concept' rather than the comprehensive diagnostic support system covering all 48 ocular conditions that was originally planned using hierarchical Bayes'. Appendix 1 shows MyDLP for diagnoses (KCS and uveitis).

The total number of predictor variables required to diagnose both KCS and uveitis (by the specialist IP optometrists) was 12. These were added to the 'variables' worksheet along with the two predicted variables, KCS and uveitis (appendix 1, worksheet "variables"). Pooling the predictor variables onto the 'variables' worksheet demonstrated how MyDLP learned to tease out the relevant variables for each diagnosis.

Laplace smoothing was kept at 1 to mitigate the effects of a zero count (chapter 5, section 5.2.1.1). MyDLP could reveal unreliable predictors (those with confidence limits that straddle 1). However, all predictors were kept in the model. The diagnostic cut-off was kept at the default value of 0.5 (chapter 5, section 5.2.4.1.2).

6.2.2 **Step 2:** Learning from the Evidence Base

The patient episodes were used to generate the pre-test odds, i.e. the rareness of things, and the multi-level likelihood ratios (LRs), i.e. the importance of test results. These two components form the basis of Bayes' (as discussed in Chapter 3, section 3.2), and thus fulfil the second step of bTLC, namely learning from the evidence base. The pre-test odds and LRs for KCS and uveitis are discussed below.

6.2.2.1 Pre-test odds

Table 6.1 shows the mean pre-test odds of both KCS and uveitis along with their lower and upper 95% confidence limits (CIs). Keratoconjunctivitis Sicca (KCS) was encountered more commonly (pre-test odds of 0.11-0.15) than uveitis (pre-test odds of 0.05 to 0.07). It is useful to remember that these values do not represent population prevalence, but only prevalence within the referral refinement pathway from which the data was extracted (i.e. APCOS).

	lower	mean	upper
KCS	0.11	0.13	0.15
Uveitis	0.05	0.06	0.07

Table 6.1 The mean pre-test odds of KCS and uveitis with their lower and upper 95% confidence limits (taken from appendix 1 "Notes" worksheet). MyDLP can predict up to 5 outcome variables. Since two outcome variables were being analysed, the remaining 3 rows of the table remain unfilled.

6.2.2.2 Multi-level likelihood ratios (LRs)

Multi-level LRs were calculated for all 12 predictor variables in relation to the prediction of KCS and uveitis (the predicted variables). These 32 multi-level LRs, along with their confidence limits can be viewed in the summary table of appendix 1 (worksheet: “Notes”). A screenshot of this is shown in figure 6.3. For KCS, 18 of the possible 32 LRs were deemed reliable. Only 12 of the 32 LRs were reliable for uveitis.

		KCS				Uveitis			
		lower	mean	upper	reliable?	lower	mean	upper	reliable?
laterality	unilateral	0.11	0.19	0.32	YES	1.39	1.67	2.01	YES
	bilateral	2.91	3.27	3.67	YES	0.25	0.43	0.74	YES
	asymmetric	0.13	0.23	0.41	YES	0.54	0.83	1.27	no
Sx-ocular discomfort	no	0.39	0.77	1.52	no	0.08	0.33	1.30	no
	acute	0.40	0.48	0.59	YES	1.24	1.33	1.42	YES
	chronic	3.87	5.00	6.47	YES	0.06	0.19	0.58	YES
Sx-Visual disturbance	no	2.96	3.97	5.32	YES	0.59	1.39	3.27	no
	acute	0.37	0.52	0.74	YES	0.80	1.09	1.49	no
	chronic	0.12	0.36	1.10	no	0.07	0.48	3.21	no
Hx-Atopia	no	0.74	0.96	1.24	no	0.40	0.90	2.02	no
	yes	0.56	1.13	2.28	no	0.25	1.29	6.64	no
Hx-CL wear	no	1.18	1.53	1.99	YES	0.64	1.15	2.07	no
	yes	0.10	0.29	0.87	YES	0.13	0.72	4.01	no
Hx-ocular meds	no	0.90	1.22	1.67	no	0.06	0.36	2.11	no
	yes	0.53	0.78	1.16	no	1.10	1.78	2.88	YES
Tears-Epiphora	no	0.40	1.10	3.04	no	0.63	2.77	12.25	no
	mild	0.88	1.39	2.21	no	0.23	0.92	3.77	no
	severe	0.22	0.51	1.17	no	0.40	1.69	7.15	no
Tears-Deficiency	no	0.01	0.04	0.34	YES	1.46	8.67	51.51	YES
	mild	0.68	0.88	1.13	no	0.46	1.03	2.31	no
	severe	1.15	2.96	7.61	YES	0.21	1.04	5.24	no
Conj/epi-hyperaemia	no	1.25	2.45	4.82	YES	0.02	0.15	1.08	no
	mild	0.95	1.08	1.22	no	1.08	1.19	1.31	YES
	severe	0.05	0.19	0.74	YES	0.33	0.66	1.33	no
Cornea-NaFl staining	no	0.70	0.97	1.35	no	2.37	3.49	5.14	YES
	mild	0.59	0.72	0.88	YES	0.05	0.30	1.78	no
	severe	1.60	2.21	3.05	YES	0.14	0.83	5.01	no
A/C-cells/flare	no	2.32	2.99	3.85	YES	0.00	0.02	0.13	YES
	mild	0.04	0.14	0.56	YES	1.42	2.07	3.02	YES
	severe	0.02	0.13	0.92	YES	4.45	12.01	32.43	YES
Iris-Synechiae	no	0.54	1.33	3.29	no	0.24	0.45	0.87	YES
	yes	0.13	0.67	3.49	no	0.67	4.28	27.25	no

Figure 6.3 The 32 multi-level likelihood ratios (with 95% confidence limits) pertaining to the individual outcomes of the 12 predictor variables were generated for KCS and uveitis (predicted variables). The values demonstrating confidence limits straddling 1 were deemed unreliable and labelled as such. Abbreviations: Sx: symptom; Hx: history; CL wear: contact-lens wear; conj/epi-hyperaemia: conjunctival or episcleral hyperaemia; NaFl: Sodium fluorescein; A/C: anterior chamber.

6.2.2.2.1 LRs: MyDLP versus hierarchical Bayes'

Likelihood ratios represent the value of each test (i.e. the importance of a particular result) relative to a diagnosis. Tables 6.2 and 6.3 show the top 5 predictor variables (according to the LR values) for the diagnosis of KCS and uveitis, respectively. Both tables present the predictor variables (outcomes of tests) in descending order generated using both hierarchical Bayes' and MyDLP. The predictors

highlighted green represent those mentioned in the clinical management guideline (CMG) associated with each diagnosis.

Top 5 clinical variables for KCS	
Hierarchical Bayes	**MyDLP**
No visual disturbance	Ocular discomfort
Severe tear deficiency	No visual disturbance
Bilateral	Bilateral
No A/C cells/flare	No A/C cells/flare
No Conj/episcleral hyperaemia	Severe tear deficiency

Table 6.2 The top 5 predictor variables for the diagnosis of KCS using hierarchical Bayes' and MyDLP (descending order). Those highlighted green are mentioned in the KCS clinical management guideline (CMG). Abbreviations: A/C: anterior chamber; conj/epi-hyperaemia: conjunctival or episcleral hyperaemia.

For KCS (table 6.2), 4 of the 5 highest-ranking predictor variables were common to both hierarchical Bayes and MyDLP. This represents good agreement in predictor variable importance (LRs) between both methods. The two differing predictors were 'No conjunctival/episcleral hyperaemia' for hierarchical Bayes' and 'Ocular discomfort' for MyDLP.

Three of the ranked predictors generated by MyDLP correlated with those mentioned in the KCS clinical management guideline (CMG). Whilst for hierarchical Bayes', this was 2 (shaded green). Interestingly, those not mentioned in the CMGs (not shaded green) related to negative findings (i.e. the absence of the predictor examined for). As mentioned in chapter 2, the CMGs lack negative findings and yet these appear to be important for the establishment of a diagnosis.

Top 5 clinical variables for Uveitis	
Hierarchical Bayes	**MyDLP**
Severe A/C cells/flare	Severe A/C cells/flare
No corneal NaFl staining	No tear deficiency
Acute visual disturbance	No corneal NaFl staining
Hx ocular medication	Mild A/C cells/flare
Mild A/C cells/flare	Hx ocular medication

Table 6.3 The top 5 predictor variables for the diagnosis of uveitis using hierarchical Bayes' and MyDLP (descending order). Those highlighted green are mentioned in the uveitis clinical management guideline (CMG). Abbreviations: A/C: anterior chamber; conj/epi-hyperaemia: conjunctival or episcleral hyperaemia; NaFl: Sodium fluorescein.

For uveitis (table 6.3), once again a good agreement was found, with 4 of the 5 highest-ranking predictor variables common to both hierarchical Bayes' and MyDLP. Here, the two differing predictors were ' Acute visual disturbance' for hierarchical Bayes' and 'No tear deficiency' for MyDLP.

Three of the highest-ranking predictors generated using hierarchical Bayes' correlated with those mentioned in the associated CMG. Whilst two of those generated by MyDLP did the same. Once again, most of the highest-ranking predictors that did not feature in the associated uveitis CMG (not shaded green) pertained to important negative test findings.

LRs are not connected to the prevalence of an ocular condition nor the complexity of a prediction model (such as hierarchical Bayes' and MyDLP). LRs should therefore remain consistent between models and populations. The full results of the comparison of LRs for KCS and uveitis are shown in appendix 2. It can be clearly seen that the LRs generated by both hierarchical Bayes' and MyDLP do not differ greatly. This suggests that hierarchical Bayes' generated meaningful LRs despite the overall model underperformance.

6.2.3 **Step 3:** Translation to evidence-based practice & **Step 4:** Translation to research-led teaching

The diagnostic decision support provided by the combined "intelligent" electronic patient record and "intelligent" virtual patient teaching tool (iEPR/iVPT) is now presented. This forms the third and fourth steps of the bTLC, namely translation to evidence-based practice, and research-led teaching (as discussed in chapter 5, section 5.2.3). The iEPR/iVPT was used to establish a diagnosis for both KCS and uveitis (see appendix 1, worksheet "iEPR/iVPT (apply evidence)" for the interactive iEPR/iVPT of MyDLP set for KCS and uveitis). The iEPR/iVPT was applied in three analyses for each diagnosis to demonstrate both the support-tool and teaching-tool capability:

Analysis 1: Random selection of tests

First, the 12 predictor variables (tests) were selected in the order in which they appeared in the "make a decision" drop-down menu of the iEPR/iVPT. This is effectively a random order. The efficiency of the clinical examination was judged by the number of tests undertaken before a firm diagnosis was reached. That is, the lower the number of tests needing to be undertaken to establish a diagnosis, the more efficient the support-tool. The effect of each test-decision undertaken on the probability of a diagnosis (either KCS or uveitis) was presented graphically (as described in chapter 5,

section 5.2.3). To demonstrate the teaching tool element, test decisions were graded both individually and with a mean AI grade.

Analysis 2: Tests selected using method A

The sequence of predictor variables (tests) was then selected following the recommendations of MyDLP decision support method A (given in red, as shown in chapter 5, figure 5.15). This method recommended the next most appropriate test that would alter the probability of either of the diagnoses the most. Once again, the number of tests undertaken to reach the diagnosis (KCS or uveitis) dictated the efficiency of the decision-support tool. These test selections were also automatically graded, demonstrating the teaching tool.

Analysis 3: Tests selected using method B

The final way in which the iEPR/iVPT was demonstrated involved selecting the predictor variables (tests) in the order recommended by MyDLP decision support method B (given in green, as shown in chapter 5, figure 5.15). This method recommended the next most appropriate test, that would alter the probability of the already most probable diagnosis the most. The efficiency of decision support was again judged by the number of tests selected before a diagnosis was reached.

The overall AI grade generated for analysis 3 was expected to be the highest since the automated grading was based on method B for test selection. The automated grading system is able to grade decisions in a hierarchical fashion depending on their proximity to method B. The mean grades of each of the three analyses were used to demonstrate the teaching element of the iVPT.

6.2.3.1 **Diagnosis: KCS** - Analysis 1: Random selection of tests

Figure 6.4 shows a screenshot of the combined iEPR/iVPT with a random selection of tests (as they appear in the "make a decision" drop down menu). The final probability of KCS reached 98% but had a large confidence interval (0-99%). This degree of uncertainty may be related to the fact that with a small dataset the accumulation of broad confidence limits for the LRs of each test undertaken result in large confidence limits for the probability of the final outcome. Having a larger dataset would then reduce the confidence limits but would not improve overall model performance. Figure 6.5(a) shows the change in probability of each diagnosis (KCS versus uveitis) with every test performed (with the outcome entered). KCS became the clear diagnosis after 12 predictor variables were selected in a random order.

The automatically generated AI grades for each test selected are shown in green to the right of figure 6.4. The mean AI grade, also in green, is given in the bottom right. The random selection of tests resulted in a mean AI grade of 77%.

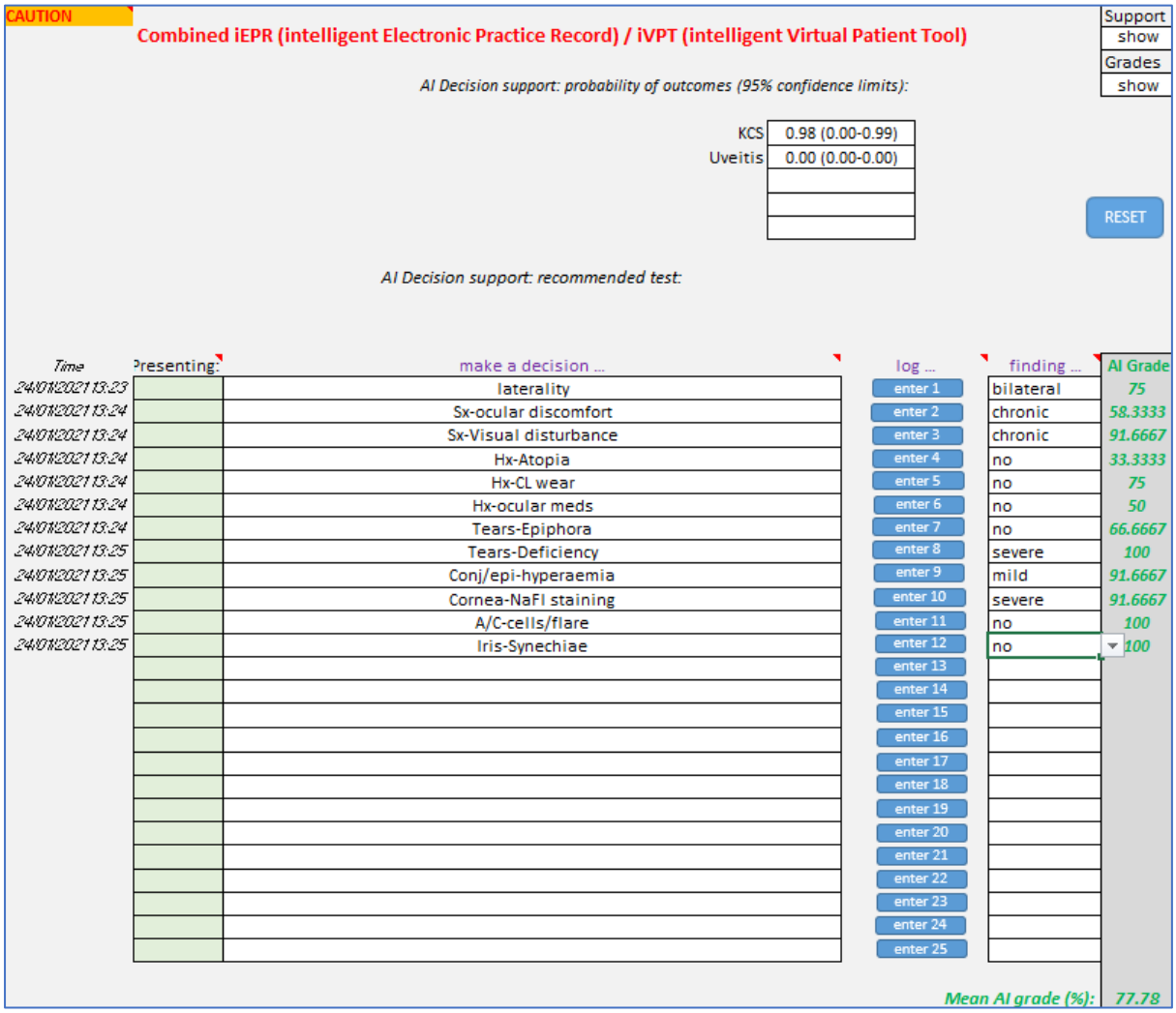


Figure 6.4 A screenshot of the iEPR/iVPT for analysis 1 (random selection of tests for KCS diagnosis)

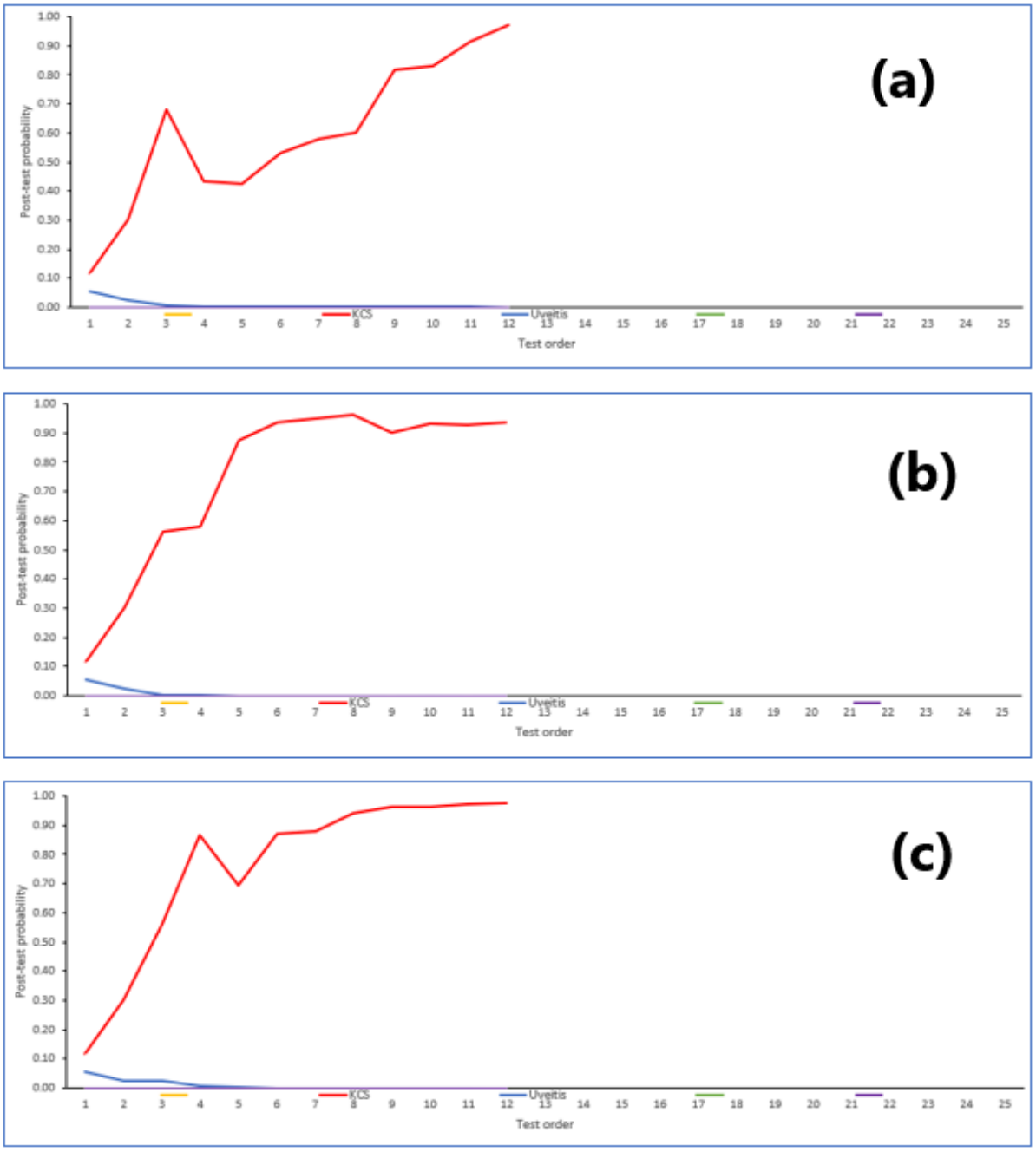


Figure 6.5 Graphs showing the change in post-test probability against the number of tests selected for (a) analysis 1 (random selection of tests for KCS diagnosis), (b) analysis 2 (tests selected using method A for KCS diagnosis), and (c) analysis 3 (tests selected using method B for KCS diagnosis)

6.2.3.2 **Diagnosis: KCS** – Analysis 2: Tests selected using method A

Figure 6.6 is a screenshot of the combined iEPR/iVPT with test selection as recommended by MyDLP decision-support method A. Again, the final probability of KCS reached 98%. Figure 6.5(b) shows the associated post-test probability graph. This graph shows a much steeper gradient in the probability curve than that in analysis 1 (figure 6.5(a)). A steeper gradient indicates a more efficient order of

testing to reach the diagnosis. Indeed a firm diagnosis was reached with the selection of 8 predictor variables (tests with results).

The iVPT aspect and AI grading is given to the right of figure 6.6. This shows that, as expected, following method A (a more efficient method of testing with AI support) gave a mean AI grade of 84%, higher than that for random test selection.

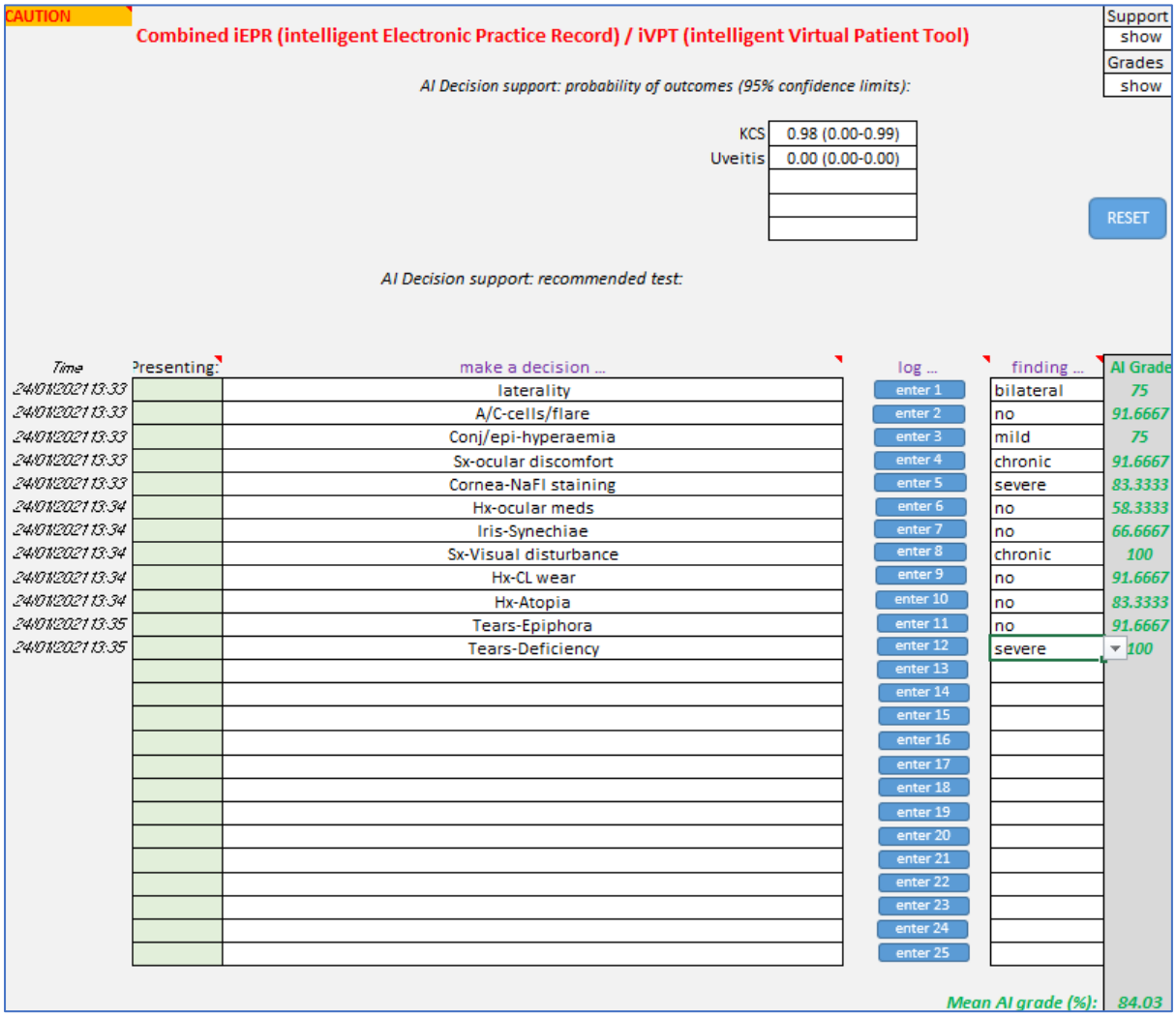


Figure 6.6 A screenshot of the iEPR/iVPT for analysis 2 (tests selected using method A for KCS diagnosis)

6.2.3.3 **Diagnosis: KCS** – Analysis 3: Tests selected using method B

Figures 6.7 is a screenshot of the combined iEPR/iVPT with test selection as recommended by MyDLP decision-support method B. Figure 6.5(c) shows the associated post-test probability graph. The graph displays a steeper increase in the probability of KCS when compared to both random test selection and method A (figure 6.5 (a) and (b) respectively). Although the final (98%) post-test

probability is reached with the selection of 8 predictor variables (tests), the same as with method A, the steeper gradient of the curve indicates a more efficient order of tests.

The AI generated grading showed, as expected, the highest score of the 3 analyses (97%) for following method B.

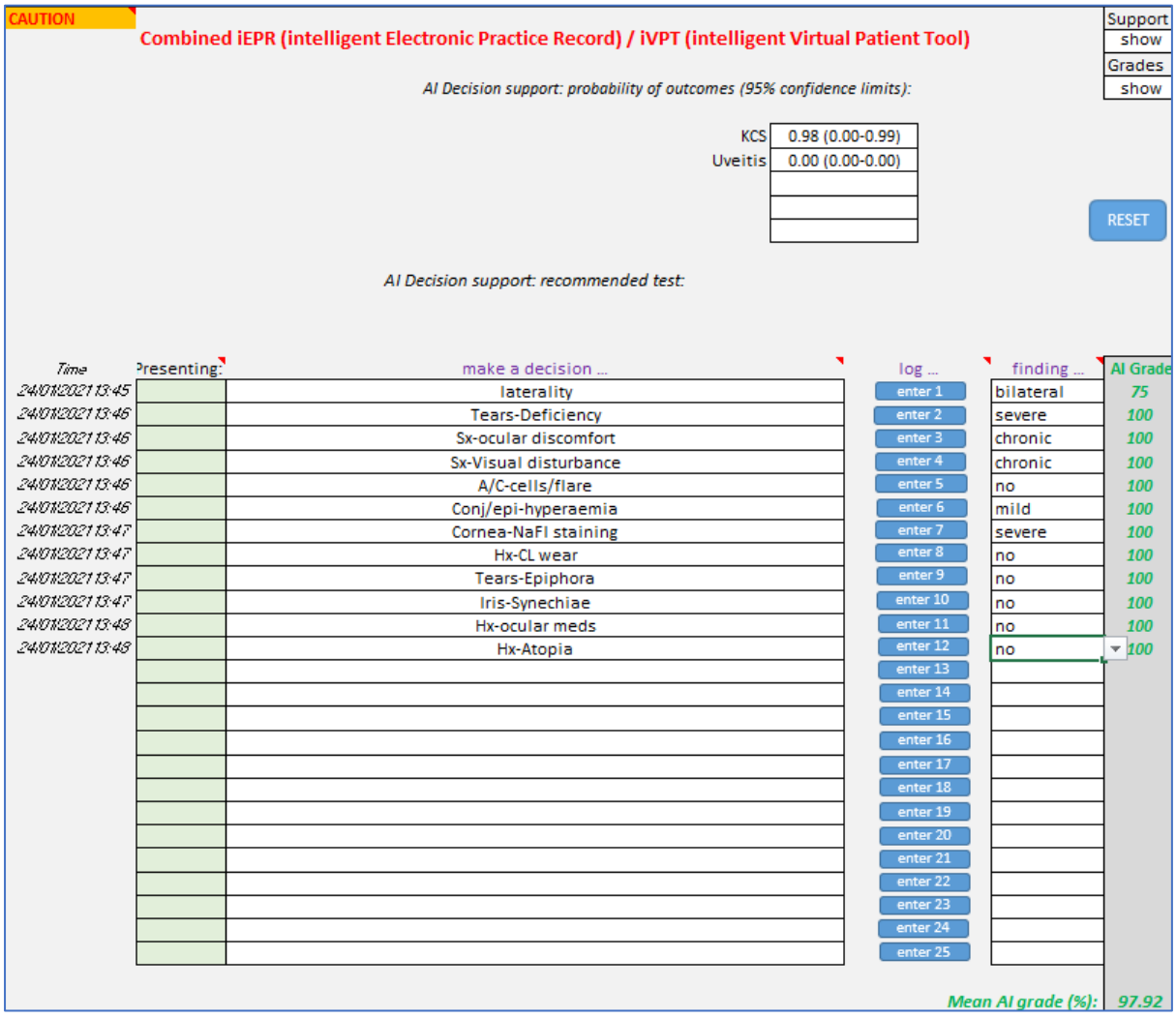


Figure 6.7 A screenshot of the iEPR/iVPT for analysis 3 (tests selected using method B for KCS diagnosis)

6.2.3.4 **Diagnosis: Uveitis**- Analysis 1: Random selection of tests

Figure 6.8 shows a screenshot of the combined iEPR/iVPT with a random selection of tests (selected as described in analysis 1, in the order they appear in the drop-down list). The final probability of uveitis reached 98%. As with KCS, this is accompanied by a large confidence interval (0-100%). Again, this degree of uncertainty may be related to the cumulative effect of broad confidence limits for the LRs of each test undertaken (section 6.2.3.1). Figure 6.9(a) shows the change in probability of each

diagnosis (uveitis versus KCS) with every test performed. Uveitis became the clear diagnosis after all 12 predictor variables (tests) are selected.

The automatically generated AI grades in analysis 1 for uveitis are shown in green to the right of each test selected (figure 6.8). The mean AI grade, also in green, is given in the bottom right. The random selection of tests for the diagnosis of uveitis resulted in a mean AI grade of 76%.

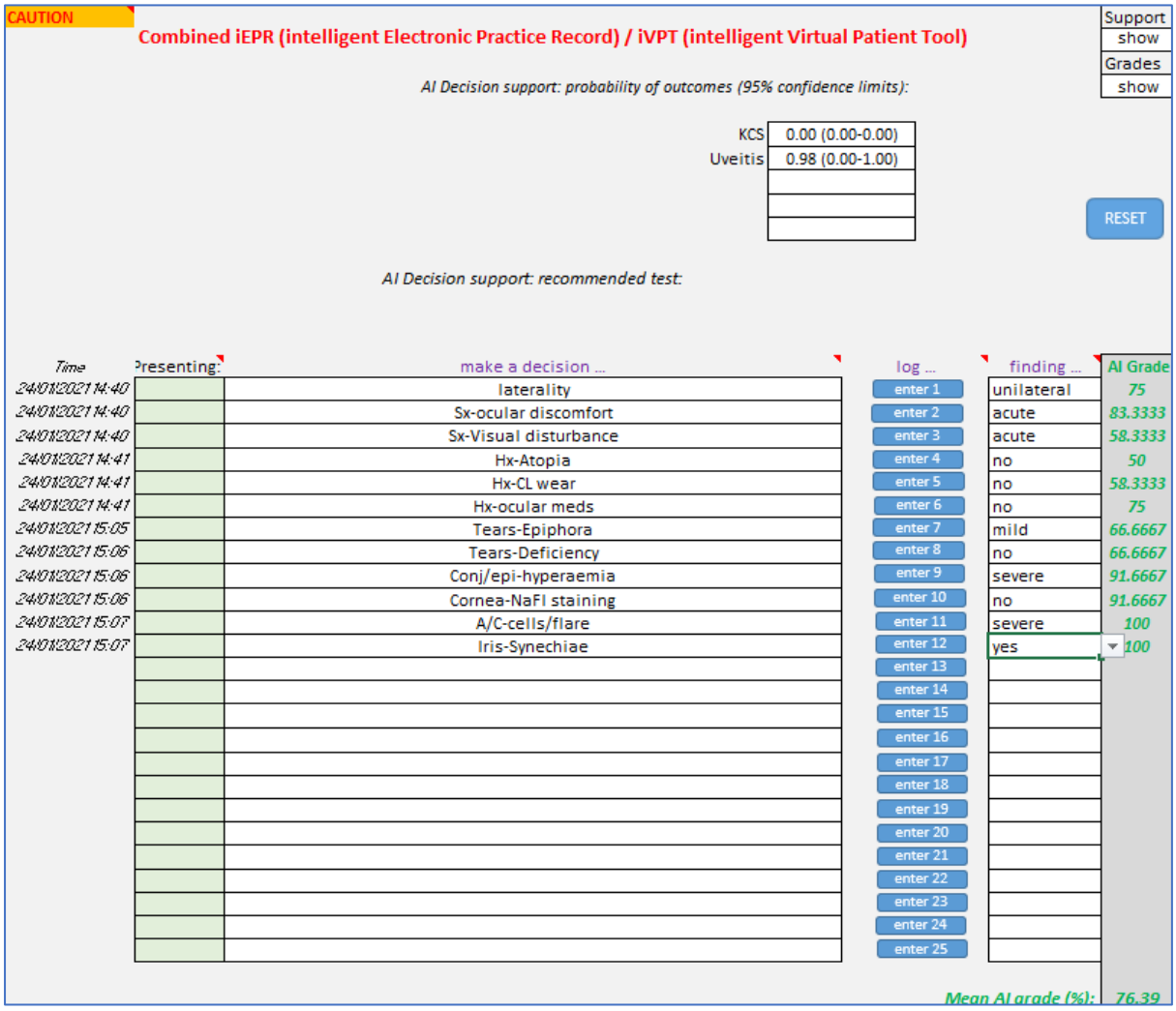


Figure 6.8 A screenshot of the iEPR/iVPT for analysis 1 (random selection of tests for uveitis diagnosis)

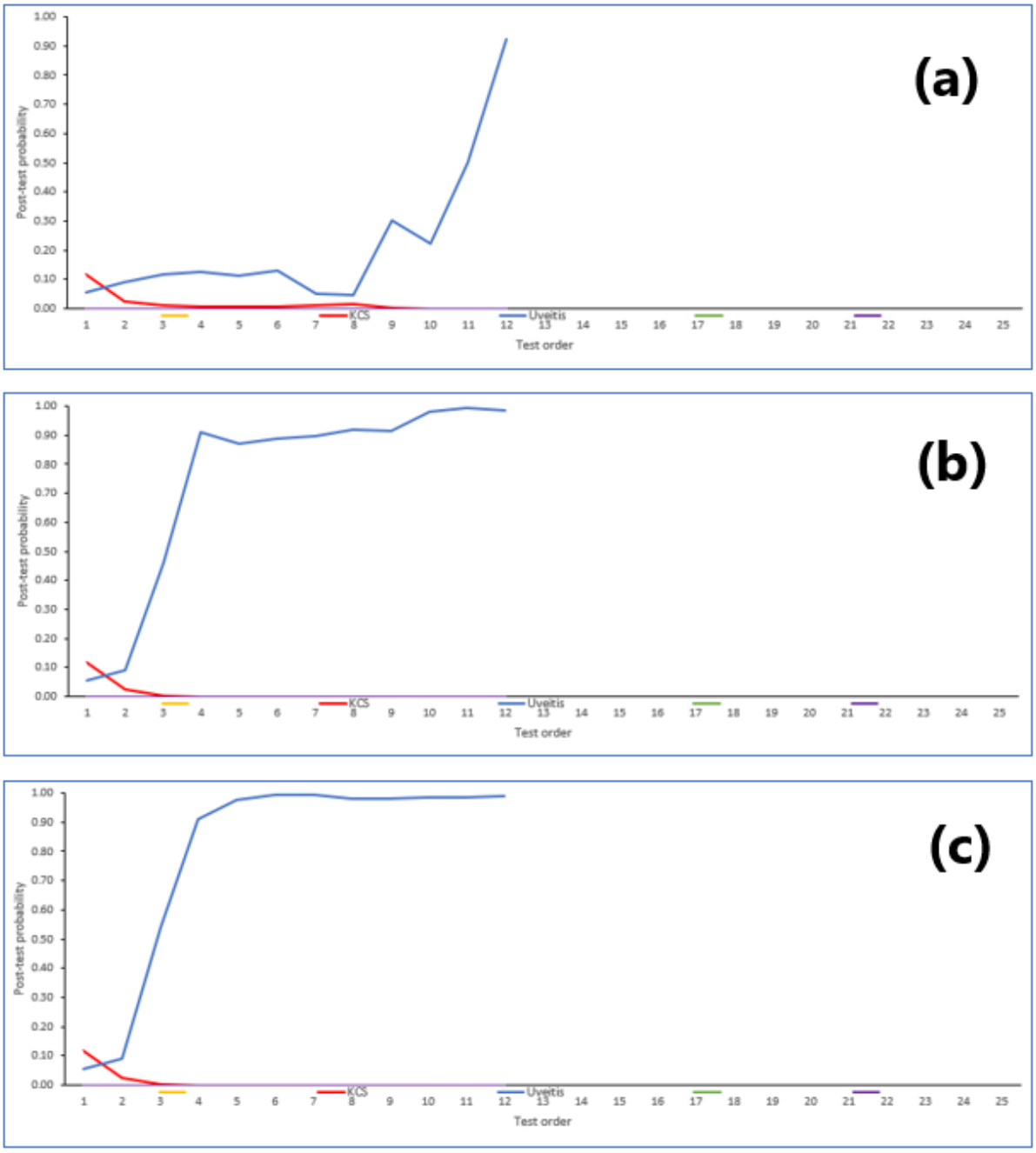


Figure 6.9 Graphs showing the post-test probability against the number of tests selected for (a) analysis 1 (random selection of tests for uveitis diagnosis), (b) analysis 2 (tests selected using method A for uveitis diagnosis), and (c) analysis 3 (tests selected using method B for uveitis diagnosis)

6.2.3.5 **Diagnosis: Uveitis** – Analysis 2: Tests selected using method A

Figure 6.10 shows a screenshot of the combined iEPR/iVPT with test selection as recommended by method A. The final probability of uveitis reaches 98%. Figure 6.9(b) shows the associated post-test probability graph. Clear certainty of the correct diagnosis of uveitis is reached after 4 tests (90% post-test probability) increasing to 98% after 11 tests. This graph shows a much steeper gradient in

the curve than that of analysis 1 (the random selection of tests) shown in figure 6.9(a). As mentioned with KCS, a steeper gradient indicates a more efficient order of testing to reach the diagnosis.

The iVPT aspect and AI grading for analysis 2 is given to the right of figure 6.10. This shows that following the MyDLP AI decision-support (method A) resulted in similar results to random test selection. The mean AI grade of 74% was achieved for following this method.

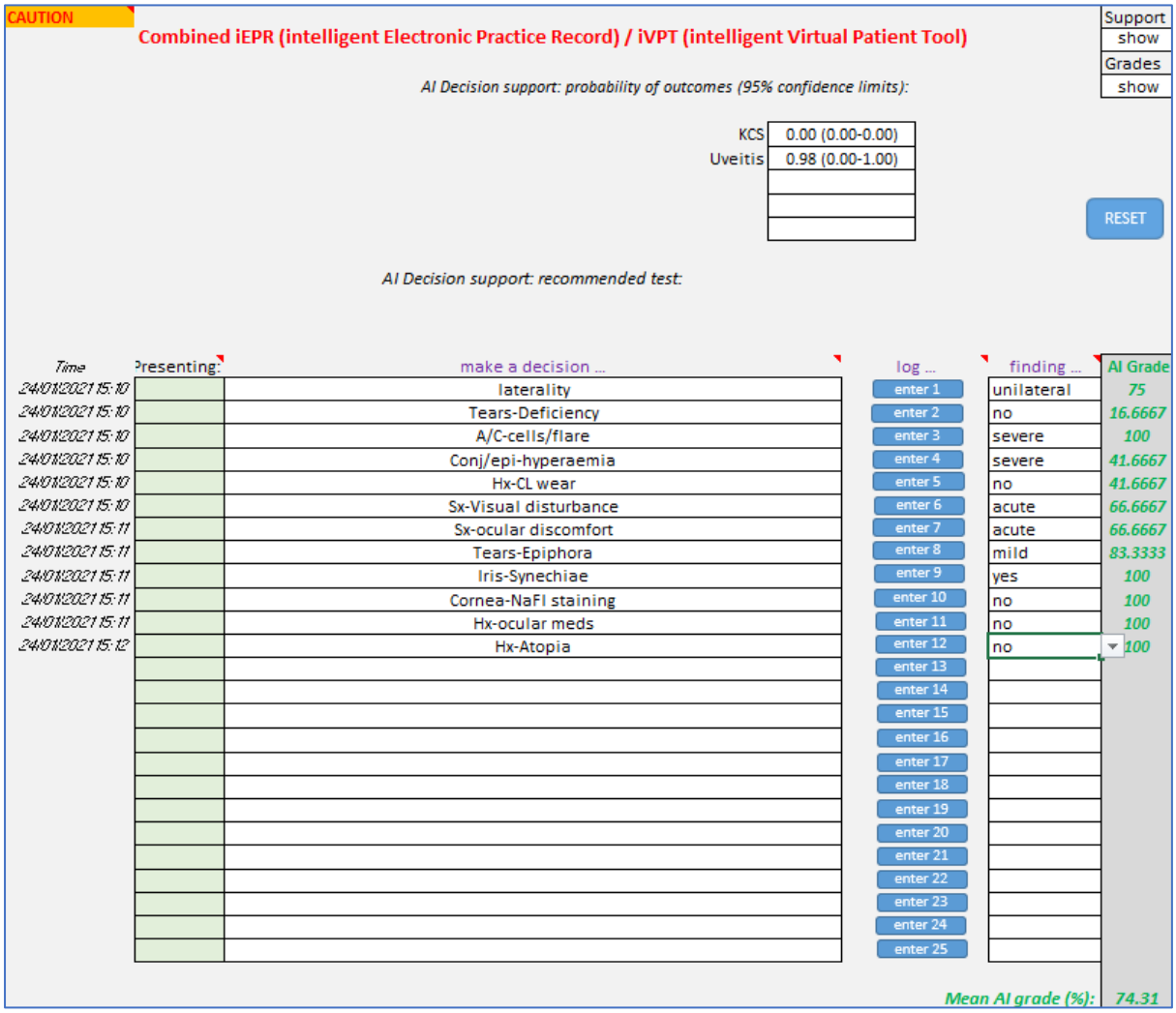


Figure 6.10 A screenshot of the iEPR/iVPT for analysis 2 (tests selected using method A for KCS diagnosis)

6.2.3.6 **Diagnosis: Uveitis** – Analysis 3: Tests selected using method B

Figures 6.11 shows a screenshot of the combined iEPR/iVPT with test selection as recommended by MyDLP decision-support method B. Figure 6.9(c) shows the associated post-test probability graph. The uveitis curve displays a steep increase in the probability of uveitis when compared to random test selection (figure 6.9(a)), but a similar gradient to method A (figure 6.9(b)). That is, uveitis, again

becomes the clear diagnosis by test 4. A maximum post-test probability (98%) is reached by test 5. This shows that following MyDLP decision-support B provides the most definitive trajectory towards a firm diagnosis of uveitis.

The AI generated grading showed, as expected, the highest score of the 3 analyses (97%) for following method B.

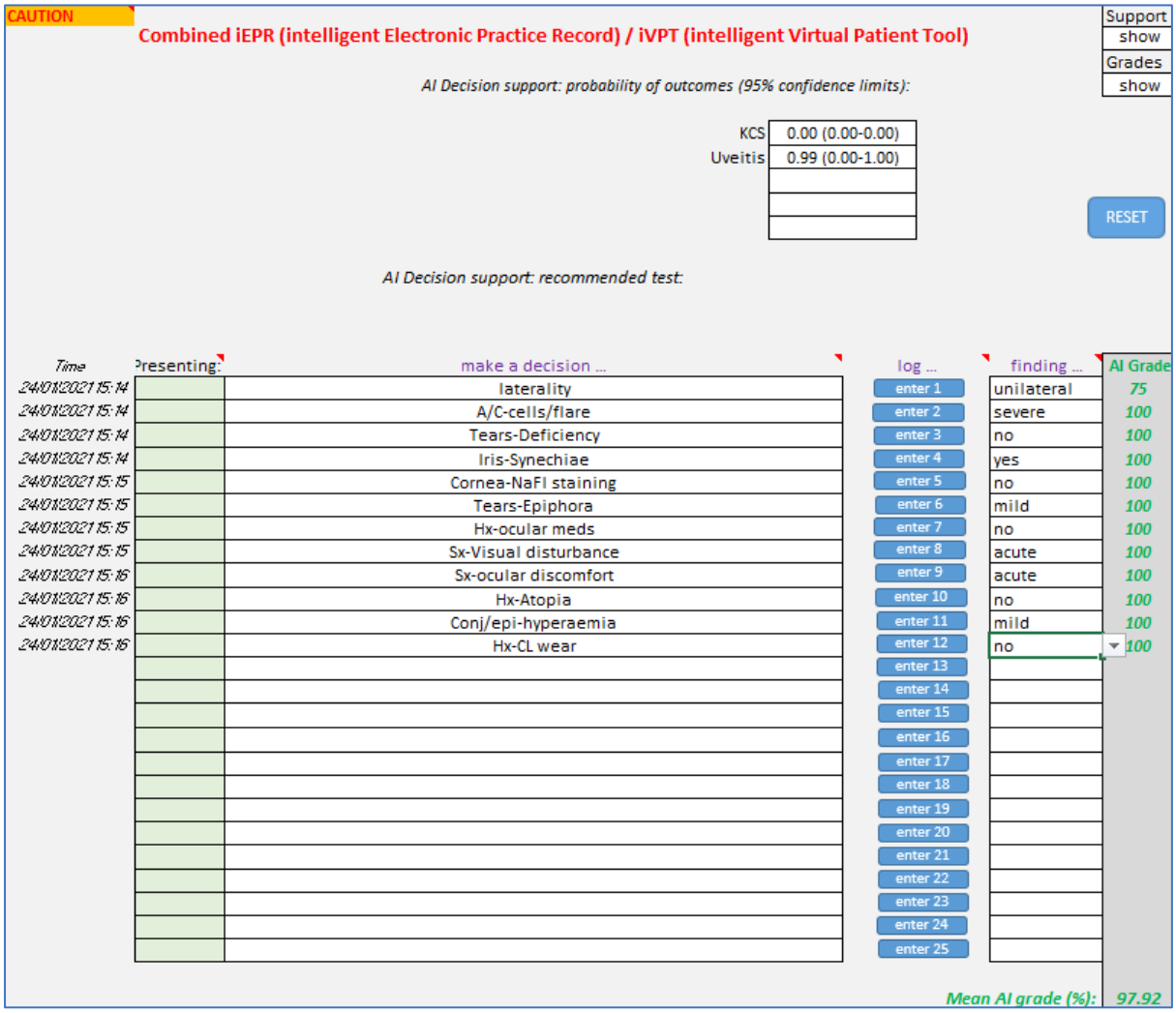


Figure 6.11 A screenshot of the iEPR/iVPT for analysis 3 (tests selected using method B for KCS diagnosis)

6.2.4 **Step 5:** Evolving the system

As described in chapter 5 (section 5.2.4) MyDLP can be used to perform 10-fold cross validation on the underlying model (Bayes'). This is used to interrogate model performance, generating measures for accuracy, informedness and markedness for the clinical support provided (appendix 1, worksheet

"Bayes' accuracy (evaluation)"). The performance of MyDLP (for the diagnosis of KCS and uveitis) is now presented.

Firstly the weighted accuracy, informedness and markedness are presented as overall model performance indicators. This is followed by learning efficiency curves that evaluate the extent to which Bayes' has reached maximum learning. Recall that learning efficiency curves display the change in weighted accuracy, informedness and markedness with an increasing number of patient episodes (grouped into folds of different sizes) (chapter 5, section 5.2.4.2). The fold sizes can be altered to best display the asymptote of the curves. An asymptote translates as maximum learning, that is no further improvement in performance is expected with the addition of more data.

The model performance and learning efficiency were explored in three analyses (for each diagnosis):

Analysis 1 – using all 12 predictor variables

Analysis 2 – using only CMG recommended predictor variables

Analysis 3 – using the CMG recommended predictor variables after removing unreliable predictors (i.e. those for which the confidence limits of the LRs straddled 1).

These three analyses explore whether the model performs and learns most efficiently with all 12 predictor variables as dictated by the activity of the specialist IP optometrists, only those predictor variables (of the 12) that are also recommended by the CMGs, or only those predictor variables (of the 12) that are recommended by the CMGs but also are reliable. In other words, it explores whether (1) fewer but more robust predictor variables affect the models' performance and learning efficiency, and (2) whether maximum learning efficiency has been reached.

6.2.4.1 **KCS** – Model performance

Stratified 10-fold cross validation, (with a diagnostic cut-off of 0.5) was performed to generate the performance measures for KCS diagnosis. Figure 6.12 shows the weighted model accuracy, informedness and markedness for analyses 1-3 (described above) along with the standard deviations.

Weighted accuracy remained high and unaltered across analyses (91%) despite the reduction in predictor variables. Weighted informedness was good for analysis 1 (using all 12 predictor variables) at 59%. It reduced slightly to 52% for analysis 2 (CMG recommended predictor variables only) and analysis 3 (only reliable CMG recommended predictor variables) at 55%. Weighted markedness, the most important measure, remained good and fairly constant across analyses 1-3 (58%, 57% and 57%

respectively). Altering the predictor variables used did not therefore have a bearing on the overall performance of the model, which was deemed good.

Performance for each diagnosis using all 12 predictor variables

	KCS		
	Acc	Inform	Marked
Mean	0.91	0.59	0.58
Standard deviation	0.02	0.09	0.11

Performance for each diagnosis using only CMG recommended predictor variables

	KCS		
	Acc	Inform	Marked
Mean	0.91	0.52	0.57
Standard deviation	0.02	0.08	0.11

Performance for each diagnosis using CMG recommended predictor variables after removing unreliable predictors

	KCS		
	Acc	Inform	Marked
Mean	0.91	0.55	0.57
Standard deviation	0.02	0.09	0.12

Figure 6.12 Weighted performance measures (Acc = accuracy, Inform = informedness, Marked = markedness) of MyDLP (the model) for the diagnosis of KCS. Analysis 1 (green), analysis 2 (peach) and analysis 3 (blue).

6.2.4.2 **KCS** – Learning efficiency

Appendix 3 shows 10 learning efficiency curves (for all 3 analyses mentioned above) generated to demonstrate the learning efficiency for KCS diagnosis. Of these, figure 6.13 shows the learning efficiency curves with clear asymptotes for the diagnosis of KCS using (a) all 12 predictor variables (b) only CMG recommended predictor variables and (c) only reliable CMG recommended predictor variables.

The number of patient episodes needed to reach an asymptote in learning efficiency for analysis 1 – using all predictor variables, was 80 (figure 6.13 (a)). For analysis 2 – using only CMG predictor variables, this was 60 (figure 6.13 (b)). For analysis 3 – using only reliable CMG recommended predictor variables, this reduced to 40 (figure 6.13 (c)). As can be seen from figure 6.13, the overall model performance was unchanged for the analyses 1 to 3 (i.e. the curves asymptote at the same performance values). However, the number of records needed to reach maximum learning efficiency reduced with each analysis. This may suggest that either the reliable CMG recommended predictor variables are the most diagnostic for KCS, or that reducing the model complexity (i.e. reducing the number of variables from which the model has to learn, or noise reduction) increases its performance. Either way, for KCS, the learning efficiency asymptotes for all three analyses was reached and no additional patient data would therefore optimise the model performance further. This served as a sample check, confirming that the conclusions reached in this chapter were the best possible for the included predictor variables.

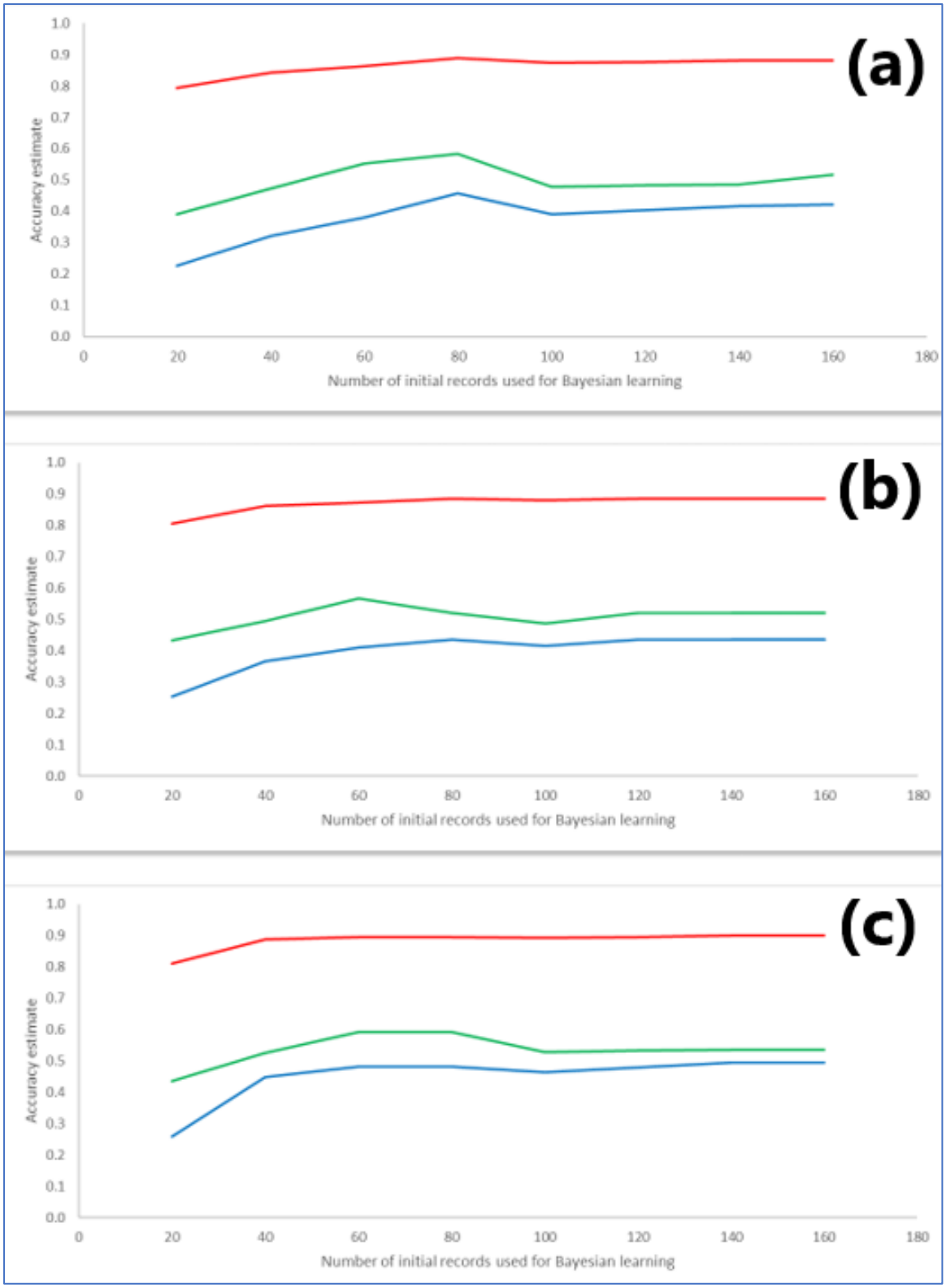


Figure 6.13 Learning efficiency curves showing the model performance for the diagnosis of KCS. (a) using all 12 predictor variables (b) using only CMG recommended predictor variables and (c) using only reliable CMG recommended predictor variables. Curve colours: red = accuracy, green = informedness and blue = markedness.

6.2.4.3 **Uveitis** – Model performance

Again stratified 10-fold cross validation, with a diagnostic cut-off of 0.5 was performed to generate the performance measures for the diagnosis of uveitis. Figure 6.14 shows the weighted model accuracy, informedness and markedness for analyses 1-3 (6.2.4) with their standard deviations.

Weighted accuracy was excellent and remained high and unaltered across analyses at 96%. This is despite the reduction in predictor variables in analyses 2 and 3. Weighted informedness was low and unaltered across analyses 1-3 at 32%. Weighted markedness was very good at 79% for analysis 1 (all 12 predictors) and rose to 86% for analyses 2 and 3 (CMG recommended predictors only, and reliable CMG predictors only respectively). This indicates that the predictors included in the latter 2 analyses were more consistent for a diagnosis of uveitis. The overall performance of the model was better for uveitis than for KCS.

Performance for each diagnosis using all 12 predictor variables

	Uveitis		
	Acc	Inform	Marked
Mean	0.96	0.32	0.79
Standard deviation	0.01	0.12	0.21

Performance for each diagnosis using only CMG recommended predictor variables

	Uveitis		
	Acc	Inform	Marked
Mean	0.96	0.32	0.86
Standard deviation	0.01	0.12	0.18

Performance for each diagnosis using CMG recommended predictor variables after removing unreliable predictors

	Uveitis		
	Acc	Inform	Marked
Mean	0.96	0.32	0.86
Standard deviation	0.01	0.12	0.18

Figure 6.14 Weighted performance measures (Acc = accuracy, Inform = informedness, Marked = markedness) of MyDLP (the model) for the diagnosis of uveitis. Analysis 1 (green), analysis 2 (peach) and analysis 3 (blue).

6.2.4.4 **Uveitis** – Learning efficiency

Appendix 4 shows 10 learning efficiency curves (for all 3 analyses) generated to demonstrate the learning efficiency for a diagnosis of uveitis. Of these, figure 6.15 shows the learning efficiency curves with clear asymptotes for the diagnosis of uveitis using (a) all 12 predictor variables (b) only CMG recommended predictor variables and (c) only reliable CMG recommended predictor variables.

The number of patient episodes needed to reach an asymptote in learning efficiency for analysis 1 – using all 12 predictor variables, was 320 (figure 6.15 (a)). For analysis 2 – using only CMG predictor variables, the patient episodes needed was also 320 (figure 6.15 (b)). Note, however, that

markedness showed a steeper (more definitive) trajectory and achieved maximum learning at around 230 patient episodes for this analysis. This may indicate slightly optimised learning when compared to analysis 1. For analysis 3 – using only reliable CMG recommended predictor variables, the number of patient episodes needed to achieve maximum learning efficiency was 650 (figure 6.15 (c)).

As for KCS, the overall model performance remained unchanged for the analyses 1 to 3 (i.e. the curves asymptote at the same values). Although the number of patient episodes needed to reach maximum learning efficiency remained the same for analyses 1 and 2, they doubled for analysis 3. This analysis used the fewest predictor variables. For uveitis then, reducing model complexity did not achieve more efficient learning. Again, and as for KCS, the learning efficiency asymptote for all three analyses was reached and no additional patient data would therefore optimise the model performance any further.

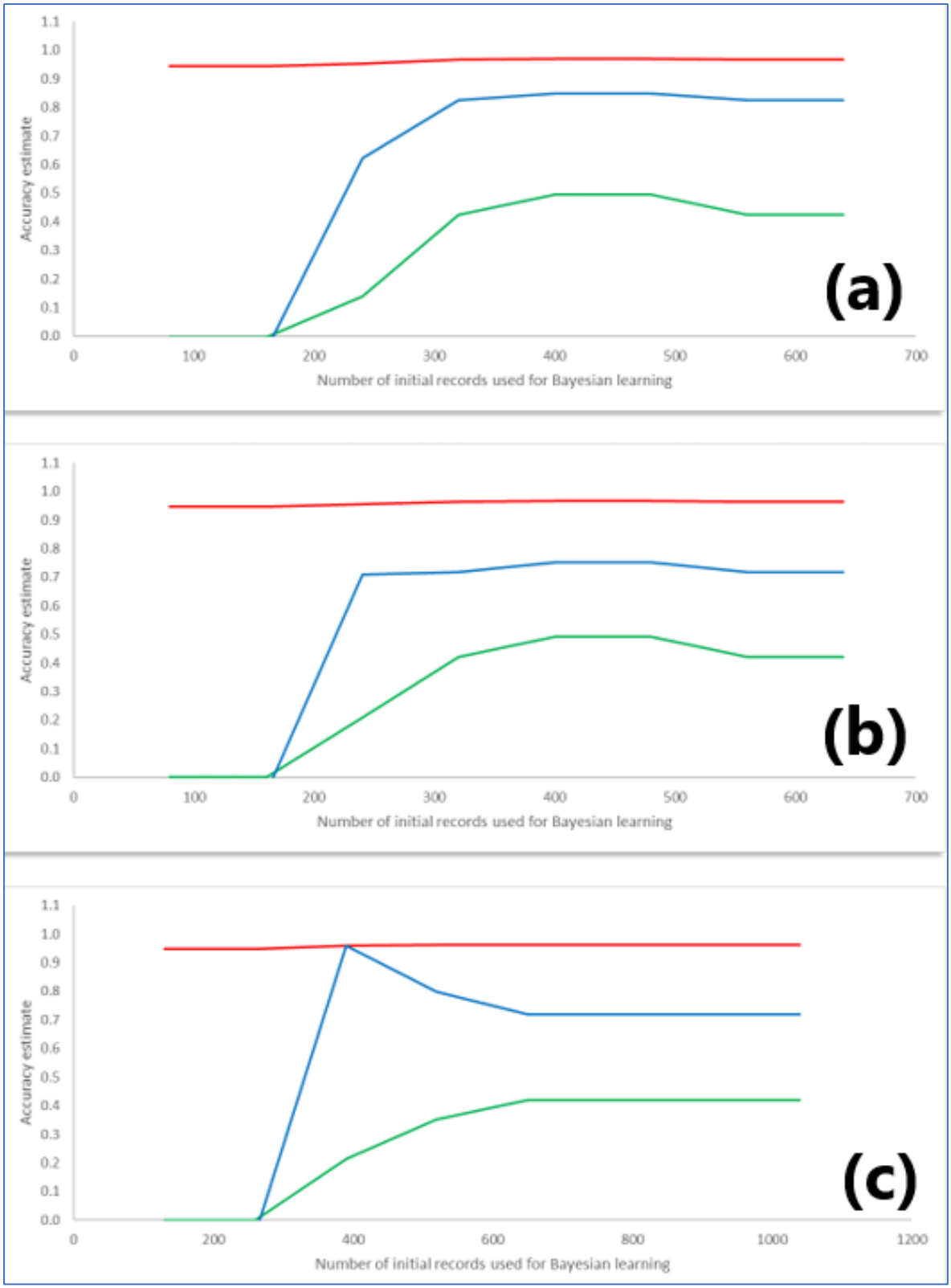


Figure 6.15 Learning efficiency curves showing the model performance for the diagnosis of uveitis. (a) using all 12 predictor variables (b) using only CMG recommended predictor variables and (c) using only reliable CMG recommended predictor variables. Curve colours: red = accuracy, green = informedness and blue = markedness.

6.3 Chapter summary

Chapter 6 presented the application of MyDLP for the diagnosis of KCS and uveitis (predicted variables). The application was presented in the 5 steps of the bTLC. Data collected by the specialist IP optometrists (1351 patient episodes, the evidence) was used to power MyDLP (step 1, gathering the evidence). Twelve predictor variables (covering all tests carried out by the specialist IP optometrists to diagnose the two predicted variables) were included to provide a 'proof-of-concept' for a diagnostic support system.

Step 2 of bTLC consisted of learning from the evidence base. This step generated pre-test odds and likelihood ratios (LRs) which were subsequently used to provide clinical decision support through Bayes' theorem (the model) in steps 3 and 4 of bTLC. The LRs generated by MyDLP were very similar to those generated by hierarchical Bayes'. The ranking of clinical tests according to their importance (as dictated by their LRs) showed an overlap not only for hierarchical Bayes' and MyDLP, but also confirmed that both methods showed good agreement with the clinical testing recommended by the CMGs. Hierarchical Bayes' therefore presented a valuable output concerning clinical tests (predictor variables). This could be applied to future models or in the ranking of clinical tests for guidelines.

The iEPR/iVPT presented how, with the use of AI generated clinical support, the efficiency of clinical testing could be optimised. Healthcare services have always faced a shortage of resources, be that of clinical equipment or time. By encouraging a more focussed and relevant examination, MyDLP could minimise the wastage of resources by encouraging efficient clinical investigation. Indeed section 6.2.3 showed that MyDLP reduced the number of tests required to reach a diagnosis of KCS by a third and uveitis by two-thirds. Aside from the clinical decision support described, the iEPR/iVPT also automatically graded the clinical decisions made by the user to arrive at the diagnoses and therefore could be used as a teaching tool. This was also presented in section 6.2.3.

The final step of bTLC, namely evolving the system, allowed the user to explore the performance and learning efficiency of the model. The performance measures showed a well performing model, both for KCS, but especially for uveitis (highest weighted markedness: 58% for KCS and 86% for uveitis). The manipulation of the predictor variables to include only variables that were both reliable and recommended by the CMGs resulted in the best performing model. The learning efficiency curves showed that maximum learning had been achieved for both KCS and uveitis with the data available. This means that even with the addition of more data, the model is unlikely to perform any better. Kohavi (1996) (99) mentioned that this ceiling effect was to be expected with Bayes' based decision support. It appears, however, that even with the performance level achieved, the AI decision

support provided by MyDLP for the diagnosis of KCS and uveitis can improve the efficiency of clinical testing when compared to random test selection.

In conclusion, having been applied to the diagnosis of KCS and uveitis, MyDLP has the potential to provide an effective decision-support tool for both the clinician and trainee. A full diagnostic model, encompassing all ocular conditions, would require more data (patient episodes) of all presenting ocular conditions on which the model could train. The proof-of-concept for the application of MyDLP to establish an ocular diagnosis was presented herewith. Chapter 7 goes further, presenting the application of MyDLP for therapeutic decisions relating to KCS and uveitis.

Chapter 7: Using MyDLP for drugs prescribed

7.1 Introduction

Chapter 6 presented the application of MyDLP for establishing a diagnosis (using KCS and uveitis as examples). The various aspects of MyDLP (a basis and provision for AI-driven decision support, interaction with machine learning, and the evaluation of the system) were showcased using the 5-step bTLC. Chapter 7 now presents the application of MyDLP for the prediction of drugs prescribed. Following on from the approach taken in chapter 6, KCS and uveitis were used as the two ocular conditions for which the therapeutic intervention was determined. Both these conditions had the most data on which the model could train. Figure 7.1 shows the chapter plan and figure 7.2 shows the 5-step bTLC as a reference.

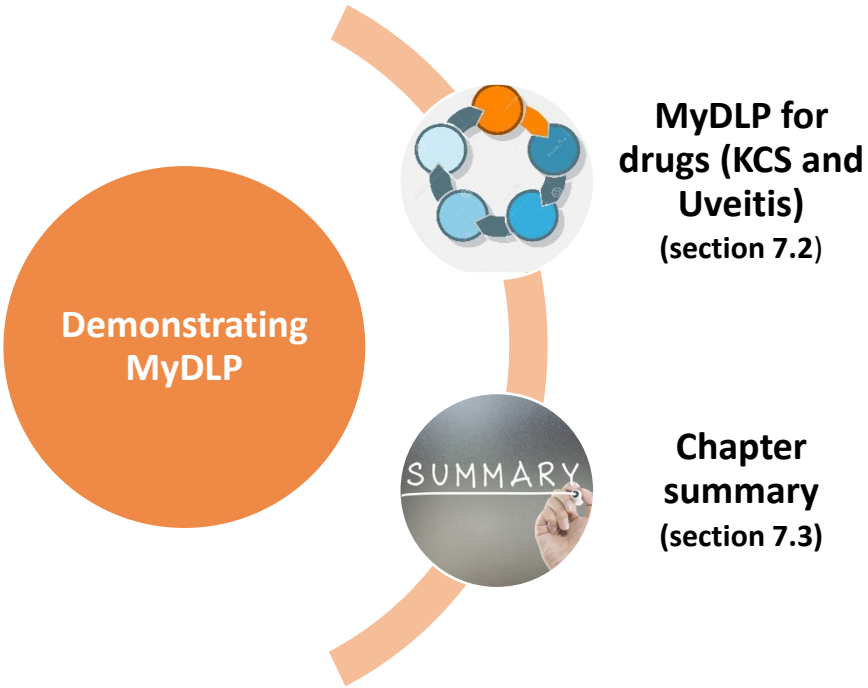


Figure 7.1 The chapter plan for Using MyDLP for drugs prescribed

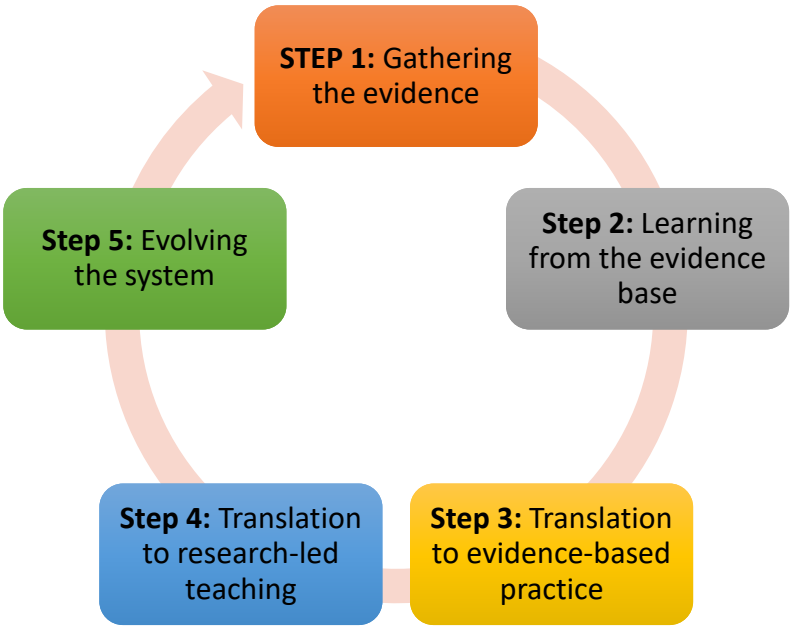


Figure 7.2 The five steps forming the Bayes' translational learning concept (bTLC). The demonstration of MyDLP for drugs prescribed follows these 5 steps.

7.2 MyDLP for drugs prescribed (for KCS and uveitis)

As discussed in Chapter 3 (Hierarchical Bayes'), the first attempt at applying machine learning to the data resulted in a model that did not perform well enough to be used as a basis of decision support. This was especially true for drug prescription where the maximum markedness achieved was 16% over all 5 analyses (chapter 3, section 3.3). The poor performance may have been due to a combination of insufficient data, and a high degree of granularity (i.e. too many outputs and therefore not enough episodes of each output to allow optimal machine learning to occur). In order to mitigate these problems and demonstrate a 'proof-of-concept' for MyDLP, two measures were taken. Firstly, following on from chapter 6, KCS and uveitis were selected, as the anterior ocular conditions most encountered, to demonstrate the model. Secondly, following the audit (chapter 4), consideration was also given to whether a support system for recommending drug therapies needed to be as prescriptive as to name individual drugs (i.e. prednisolone or dexamethasone) as opposed to drug categories based on the clinical action of the drug (e.g. topical steroids). Consider a patient with uveitis. The prescription of guttae prednisolone 1% versus guttae dexamethasone 0.1% to treat this patient are deemed acceptable variations in clinical practice. Such variations are often dictated by external factors such as the availability and cost of the drug and local prescribing recommendations. As such, a support system recommending drug categories may be more useful, particularly for use across borders where specific drug availability can vary. Consider also ocular lubrication options. As discussed in chapter 4 (section 4.2.3), there are 65 ocular lubricant options mentioned in the BNF

(166), of these, only 17 were prescribed by APCOS over the course of the data-collection period. The superiority of one drop over another remains unproven (165), and thus a generic, categorical recommendation for ocular lubrication was considered to be more useful in a support system.

Since the demonstration of MyDLP herewith concerned KCS and uveitis, the drug groupings (predicted variables) followed those described in the audit (chapter 4), that is ocular lubrication, topical steroids and cycloplegics.

7.2.1 **Step 1:** Gathering the Evidence

Once again, the 1351 patient episodes collected by the specialist IP optometrists over the course of a year, provided the evidence-base used to drive MyDLP (the platform housing the Bayesian model, referred to as "the model" in its current form and henceforth). Appendix 1 shows MyDLP for the drugs prescribed for KCS and uveitis.

The predictor variables remained the 12 required for the diagnosis of the two conditions (as in chapter 6) along with the actual diagnoses (totalling 14 predictor variables). The predicted variables were the 3 drug groups used to treat these conditions (as demonstrated by the data collected). That is, ocular lubrication, topical steroids and cycloplegics (appendix 1, worksheet "variables").

Once again, Laplace smoothing was kept at 1 to mitigate the effects of zero data counts (chapter 5, section 5.2.1.1). MyDLP could reveal unreliable predictors (those with confidence limits that straddle 1). However, all predictors, reliable or otherwise, were kept in the model. The diagnostic cut-off (here for the prediction of drugs prescribed) remained at 0.5 (chapter 5, section 5.2.4.1.2).

7.2.2 **Step 2:** Learning from the Evidence Base

Again, learning from the evidence base involved the generating of pre-test odds (the rareness of drug groups prescribed) and LRs (the importance of clinical predictor variables, i.e. clinical test findings). These were needed for Bayes' to be applied to the data and therefore 'learning from the evidence base'.

7.2.2.1 Pre-test odds

The pre-test odds of each predicted variable, along with their 95% confidence limits, is shown in table 7.1 (extracted from appendix 1, "Notes" worksheet). Of the three drug categories, ocular lubrication had the highest pre-test odds and therefore was the most prescribed drug.

	lower	mean	upper
Ocular lubrication	0.20	0.23	0.26
Topical Steroid	0.17	0.19	0.22
Cycloplegic	0.03	0.05	0.06

Table 7.1 The mean pre-test odds of the prescription of ocular lubrication, topical steroids and cycloplegics with their 95% confidence limits on either side (taken from appendix 1 "Notes" worksheet)

7.2.2.2 Multi-level likelihood ratios (LRs)

The multi-level LRs attributed to each of the 14 predictor variables generated By MyDLP, along with their confidence limits, are shown in figure 7.3 (extracted from appendix 1 "Notes" worksheet). The reliability of each LR is indicated in the right-most column for each predicted variable; with "YES", meaning reliable and indicated in green, and "NO", meaning unreliable and indicated in red. As can be seen 20 of the 36 predictor variables were reliable for the prescription of ocular lubrication. For topical steroids, this was 19 out of 36, and for cycloplegics this was 15 out of 36.

		Ocular lubrication				Topical Steroid				Cycloplegic			
		lower	mean	upper	reliable?	lower	mean	upper	reliable?	lower	mean	upper	reliable?
laterality	unilateral	0.24	0.33	0.44	YES	1.19	1.38	1.60	YES	1.54	1.84	2.20	YES
	bilateral	2.41	2.76	3.15	YES	0.46	0.60	0.78	YES	0.13	0.29	0.61	YES
	asymmetric	0.42	0.57	0.76	YES	0.81	1.02	1.29	no	0.47	0.78	1.27	no
Sx-ocular discomfort	no	0.32	0.60	1.10	no	0.05	0.14	0.46	YES	0.10	0.41	1.64	no
	acute	0.59	0.66	0.75	YES	1.28	1.37	1.47	YES	1.23	1.32	1.42	YES
	chronic	3.07	4.07	5.41	YES	0.18	0.30	0.50	YES	0.04	0.16	0.62	YES
Sx-Visual disturbance	no	2.73	3.68	4.96	YES	1.32	1.98	2.97	YES	0.70	1.61	3.70	no
	acute	0.44	0.58	0.76	YES	0.75	0.93	1.15	no	0.71	1.03	1.49	no
	chronic	0.24	0.52	1.16	no	0.06	0.25	0.99	YES	0.08	0.55	3.68	no
Hx-Atopia	no	0.68	0.88	1.12	no	0.76	1.05	1.45	no	0.40	0.90	2.02	no
	yes	0.73	1.46	2.95	no	0.30	0.87	2.50	no	0.25	1.29	6.64	no
Hx-CL wear	no	1.17	1.57	2.12	YES	0.73	1.06	1.54	no	0.64	1.15	2.07	no
	yes	0.16	0.35	0.78	YES	0.40	0.89	1.96	no	0.13	0.72	4.01	no
Hx-ocular meds	no	0.89	1.22	1.68	no	0.61	0.98	1.59	no	0.12	0.61	3.07	no
	yes	0.55	0.79	1.15	no	0.59	1.02	1.75	no	0.64	1.46	3.32	no
Tears-Epiphora	no	0.89	3.75	15.73	no	0.95	2.95	9.14	no	0.63	2.77	12.25	no
	mild	0.67	1.07	1.71	no	0.34	0.92	2.53	no	0.23	0.92	3.77	no
	severe	0.20	0.44	0.96	YES	0.14	0.82	4.67	no	0.40	1.69	7.15	no
Tears-Deficiency	no	0.00	0.02	0.18	YES	0.27	2.04	15.60	no	2.66	13.08	64.33	YES
	mild	0.81	1.26	1.96	no	0.14	0.37	0.99	YES	0.19	0.77	3.09	no
	severe	0.65	1.59	3.93	no	1.73	2.63	3.97	YES	0.39	1.57	6.40	no
Conj/epi-hyperaemia	no	0.83	1.63	3.18	no	0.01	0.04	0.29	YES	0.03	0.19	1.39	no
	mild	0.94	1.05	1.18	no	0.91	1.00	1.10	no	1.05	1.16	1.29	YES
	severe	0.28	0.54	1.05	no	1.44	2.22	3.42	YES	0.36	0.74	1.54	no
Cornea-NaFl staining	no	0.52	0.70	0.96	YES	0.39	0.62	1.00	YES	1.26	2.31	4.22	YES
	mild	0.77	0.90	1.05	no	0.98	1.17	1.39	no	0.03	0.20	1.25	no
	severe	1.44	2.02	2.84	YES	0.71	1.09	1.66	no	1.07	2.26	4.79	YES
A/C-cells/flare	no	2.41	3.26	4.41	YES	0.15	0.23	0.36	YES	0.00	0.03	0.20	YES
	mild	0.13	0.28	0.60	YES	1.43	2.24	3.52	YES	0.77	1.13	1.66	no
	severe	0.01	0.08	0.56	YES	2.49	6.73	18.19	YES	6.71	18.03	48.41	YES
Iris-Synechiae	no	0.77	1.58	3.24	no	0.29	0.56	1.08	no	0.24	0.45	0.87	YES
	yes	0.08	0.48	2.73	no	0.46	2.75	16.59	no	0.67	4.28	27.25	no
KCS	no	0.35	0.41	0.47	YES	1.03	1.07	1.12	YES	1.08	1.12	1.16	YES
	yes	40.92	82.21	165.20	YES	0.29	0.50	0.85	YES	0.02	0.13	0.93	YES
Uveitis	no	1.05	1.07	1.09	YES	0.60	0.66	0.72	YES	0.03	0.07	0.17	YES
	yes	0.01	0.06	0.41	YES	54.81	392.09	2804.71	YES	40.55	63.66	99.95	YES

Figure 7.3 The multi-level likelihood ratios (with 95% confidence limits) pertaining to the individual outcomes of the 14 predictor variables were generated for the 3 predicted variables (taken from appendix 1, "Notes" worksheet). The values demonstrating confidence limits straddling 1 were deemed unreliable and labelled as such.

7.2.2.3 The most important predictors (LRs)

As discussed, the importance of a particular test in relation to its outcome depends on the value of the LR (the higher the value, the more important the predictor) and the reliability of the LR (confidence limits do not straddle 1). Table 7.2 shows the top five reliable predictor variables for

each outcome (predicted variable). These have been extracted from appendix 2, which presents all the results for all 3 predicted variables.

Ocular Lubrication	Topical Steroid	Cycloplegic
Diagnosis: KCS	Diagnosis: Uveitis	Diagnosis: Uveitis
Ocular discomfort – present	Severe A/C cells/flare	Severe A/C cells/flare
No visual disturbance	Severe tear deficiency	No severe tear deficiency
No A/C cells/flare	Mild A/C cells/flare	No corneal NaFl staining
Bilateral	Severe conjunctival/episcleral hyperaemia	Severe corneal NaFl staining

Table 7.2 The top 5 predictor variables (in descending order) for the prescription of ocular lubrication, topical steroids and cycloplegics.

It is immediately obvious for all three predicted outcomes that a diagnosis was the single, and by far, the most important predictor dictating the prescription of the drugs. For the prescription of ocular lubrication, the diagnosis of KCS had a LR of 82.21 compared to the second most important predictor, ocular discomfort, which had a LR of 4.07. Similarly, for the prescription of topical steroids the diagnosis of uveitis was the strongest predictor (LR: 392.09) followed by the presence of severe A/C cells/flare (LR: 6.73). Again, for the prescription of a cycloplegic, a diagnosis of uveitis was the most important predictor variable (LR: 63.66) followed by the presence of severe A/C cells/flare (LR: 18.03). Arguably, this finding mirrors the thought process of a clinician who accumulates evidence either for or against a diagnosis, determines a 'working diagnosis' and formulates a management plan accordingly. These findings indicate therefore that specialist IP optometrists do not simply treat a sign or symptom with therapeutic intervention but rather a diagnosis which is supported by clinical findings.

7.2.3 **Step 3:** Translation to evidence-based practice & **Step 4:** Translation to research-led teaching

The prescribing decision-support provided by the combined iEPR/iVPT tool now forms the third and fourth steps of bTLC. That is, translation to evidence-based practice and research-led teaching. The interactive element of MyDLP (iEPR/iVPT) for prescribing decision-support can be found in appendix 1 (worksheet "iEPR/iVPT (apply evidence)".

The LRs discussed in section 7.2.2.3 suggested that the diagnosis is the most important predictor variable for the drug prescribed. Two analyses were therefore carried out to ascertain the degree of importance of a diagnosis and whether a diagnosis was necessary in the prediction of drug

treatment (i.e. could signs and symptoms alone generate the correct treatment?). The analyses were as follows:

Analysis 1: All 12 predictor variables and diagnoses

The first analysis involved setting up MyDLP (the model) to predict the drugs prescribed (ocular lubrication, topical steroids or cycloplegic – the outcome or predicted variables) using all 12 predictor variables and the diagnoses (14 predictors in total). Two hypothetical patients were presented to the model, one exhibiting the signs and symptoms for KCS and the other for uveitis. As with the analyses described in chapter 6 (section 6.2.3) the efficiency of the clinical examination was determined by the number of tests undertaken before a firm decision for the prescription of the correct drug (the target variable) was reached. The lower the number of tests required to reach an endpoint the more efficient the support-tool. The effect of each test undertaken on the probability of each outcome was also presented graphically.

Analysis 2: Using only the 12 predictor variables

The second analysis used only the 12 predictor variables pertaining to the signs and symptoms to reach a prescribing decision (ocular lubrication, topical steroids or cycloplegic). The analysis therefore excluded the diagnoses as predictor variables. Once again, two hypothetical patients were presented to the model, one exhibiting the signs and symptoms for KCS and the other for uveitis. This analysis was intended to demonstrate whether (a) an accurate prescribing decision could be reached without a confirmed diagnosis, and (b) the effect of excluding a diagnosis on the efficiency of reaching a prescribing decision. As with analysis 1, the efficiency of the clinical examination was determined by the number of tests undertaken to reach a prescribing decision.

The teaching tool element (bTLC, step 4) generated individual and overall AI grades for the decisions in each analysis (as described in chapter 5, section 5.2.3.13). The iEPR/iVPT was used such that all recommended tests (according to method B) were followed exactly. This naturally results an AI grade of 100%.

7.2.3.1 **Predicting drug treatment:** KCS

Figure 7.4 shows a screenshot of the combined iEPR/iVPT used to predict the treatment for a patient exhibiting the signs and symptoms of KCS. This included the diagnoses as predictors (analysis 1). Figure 5 shows that the final probability for the prescription of ocular lubrication was reached after 4

steps (probability 100%) and figure 7.4 shows that this probability had a wide confidence interval (0-100%). As suggested previously (chapter 6, section 6.2.3.1), this degree of uncertainty may be related to the fact that with a small dataset the accumulation of broad confidence limits for the LRs of each test undertaken result in large confidence limits for the probability of the final outcome. Having a larger dataset would then reduce the confidence limits but would not improve overall model performance. A diagnosis of KCS was suggested by the AI decision support at step 3 (figure 7.4). The change in probability with every test decision taken (and result found) is shown graphically in figure 7.5(a). At step three, the diagnosis of KCS dramatically raises the probability of the prescription of ocular lubrication (shown in red) from 50% to 98%. This shows the effect of a diagnosis of KCS on the certainty of drug treatment.

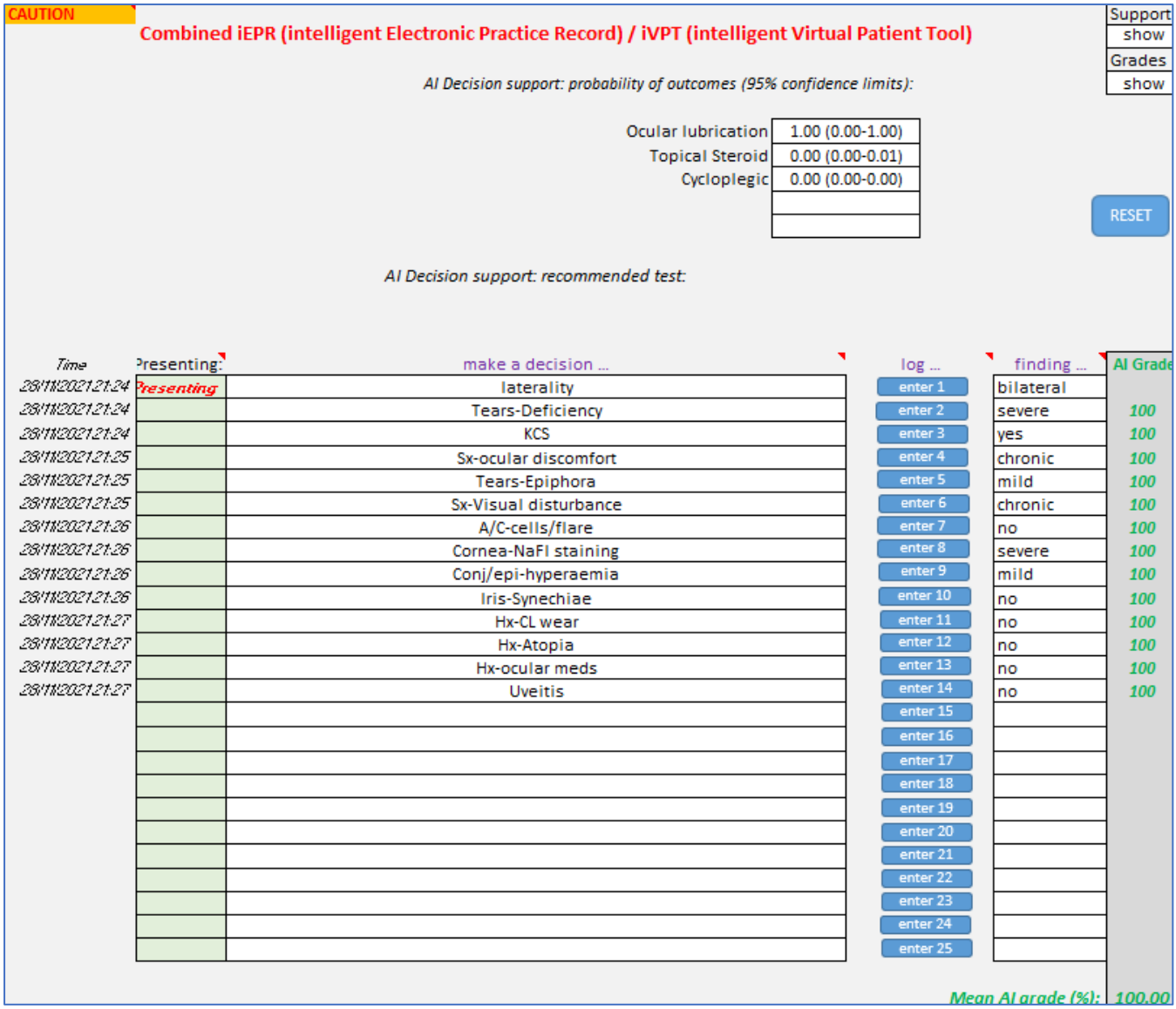


Figure 7.4 A screenshot of the iEPR/iVPT for analysis 1 (the prediction of ocular lubrication using all 12 predictor variables and the diagnoses)

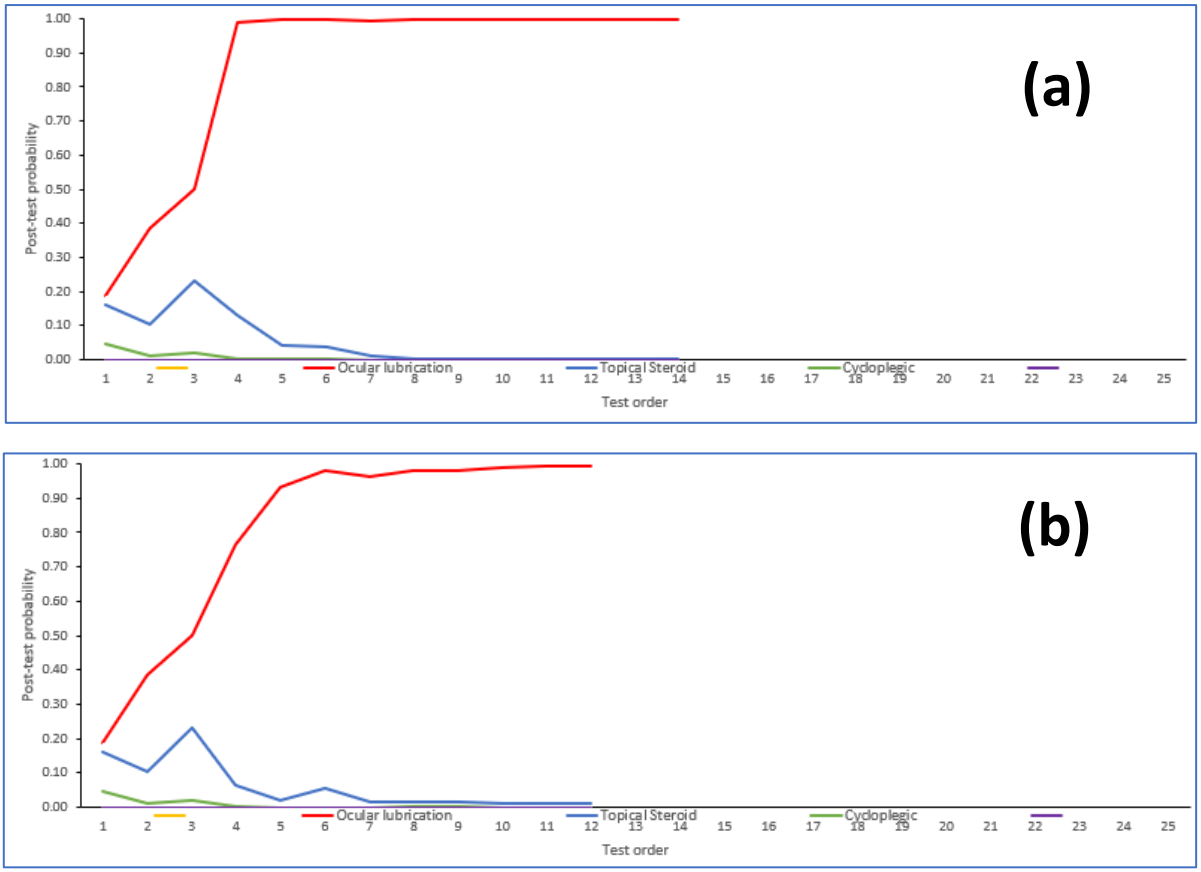


Figure 7.5 Probability graphs showing the change in post-test probability with each test selection for (a) analysis 1 (the prediction of ocular lubrication using all 12 predictor variables and the diagnoses), (b) analysis 2 (the prediction of ocular lubrication using all 12 predictor variables only)

Figure 7.6 shows a screenshot of the iEPR/iVPT, again, used to predict the treatment for a patient exhibiting the signs and symptoms of KCS. This included only the 12 predictor variables (analysis 2, i.e. excluding diagnoses as predictor variables). This time the maximum final probability for ocular lubrication was reached after 6 steps. That is, a final probability of 98%. This was also accompanied with a large confidence interval (0-99%). Figure 7.5(b) shows the associated probability graph. The red curve, indicating the prescription of ocular lubrication, shows a flatter gradient than in analysis 1 (figure 7.5(a)). It does however reach a final clear maximum probability. This suggests that a diagnosis of KCS although not necessary to predict the prescription of ocular lubrication, does make the examination more efficient (4 tests undertaken in analysis 1 to reach maximum probability versus 6 in analysis 2).

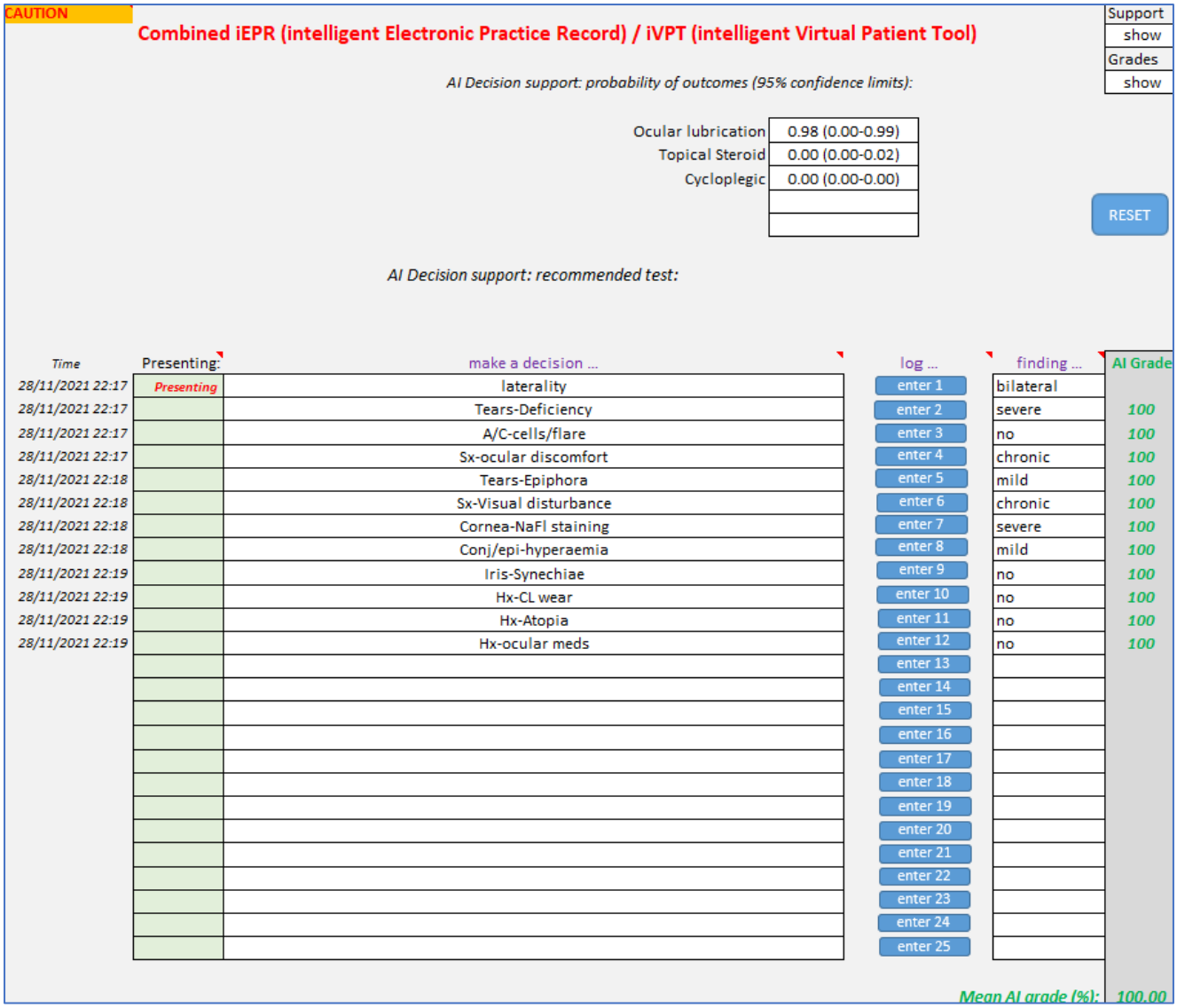


Figure 7.6 A screenshot of the iEPR/iVPT for analysis 2 (the prediction of ocular lubrication using all 12 predictor variables)

7.2.3.2 **Predicting drug treatment:** Uveitis

Figure 7.7 shows a screenshot of the combined iEPR/iVPT used to predict the treatment for a patient exhibiting the signs and symptoms of uveitis. As described for analysis 1, this included diagnoses as predictors. The drug treatment for uveitis included the prescription of two drug groups, namely topical steroids and cycloplegics. The final probability for the prescription of topical steroids was reached after 3 steps (100%) and for cycloplegia 4 steps (100%). A diagnosis of uveitis was suggested by the AI decision support at step 2 (figure 7.7). The change in probability with every test decision taken (and result found) is shown graphically in figure 7.8(a). At step three, the probability for the prescription of topical steroids (shown in blue) rises dramatically from 20% to 100%. For cycloplegia

this rise is from 5% to 85% by step 3. The presence of a diagnosis as a predictor has a dramatic effect on the probability of drug treatment, similar to the results found for KCS.

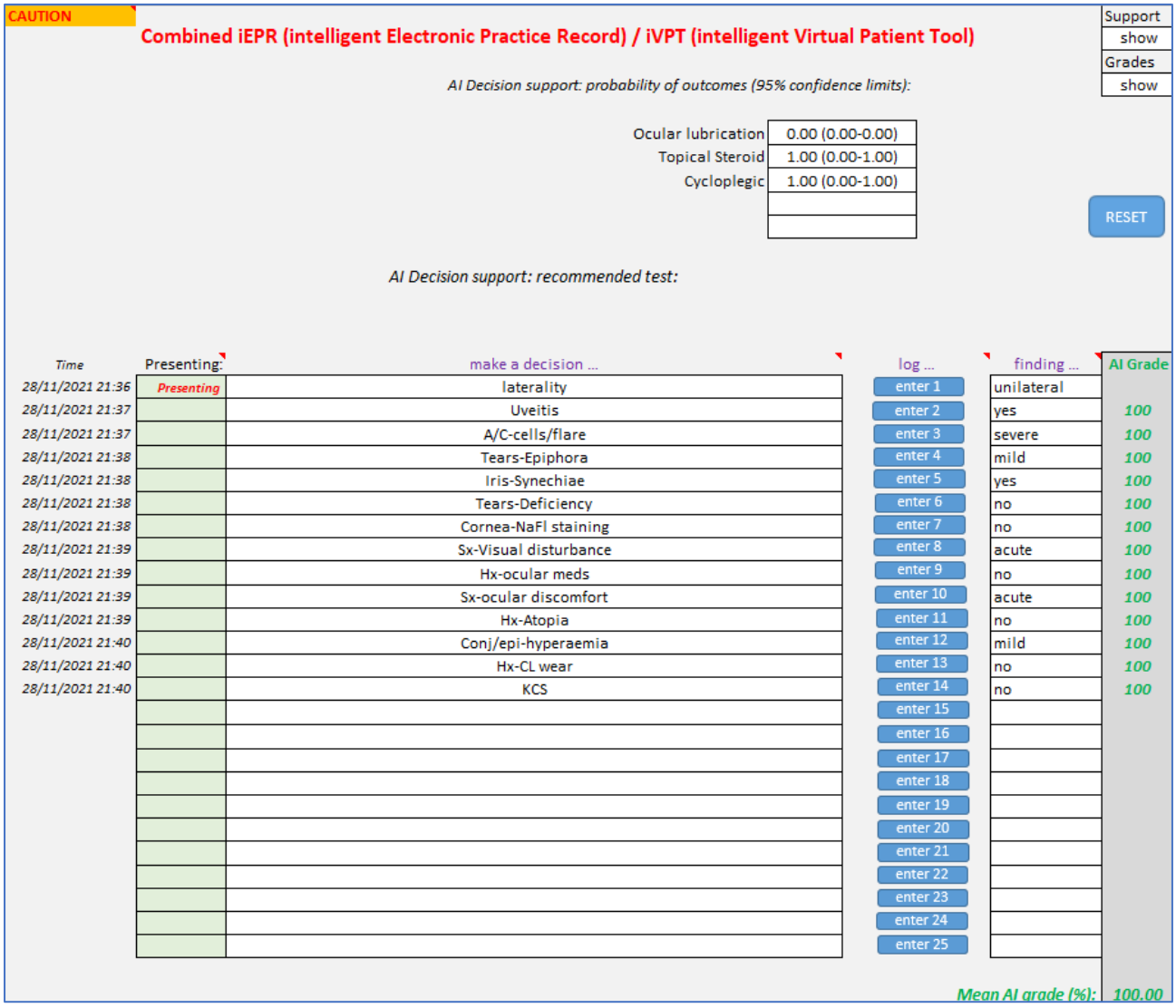


Figure 7.7 A screenshot of the iEPR/iVPT for analysis 1 (the prediction of topical steroids and cycloplegia using all 12 predictor variables and the diagnoses)

Figure 7.9 shows a screenshot of the iEPR/iVPT used to predict the treatment for a patient exhibiting the signs and symptoms of uveitis. This included only the 12 predictor variables (analysis 2, i.e. excluding diagnoses as predictor variables). The maximum final probability for both topical steroids and cycloplegia was reached after 8 steps. That is, a final probability of 89% (CI: 0-99%) for topical steroids and 99% (CI: 0-100%) for cycloplegia. Figure 7.8(b) shows the associated probability graph. Both the blue and green curves, indicating the prescription of topical steroids and cycloplegia respectively, show flatter gradients than in analysis 1 (figure 7.8(a)). They do however reach an

asymptote and clear maximum probability. This suggests that once again, although a diagnosis of uveitis is not necessary to predict a prescribing decision, it does make the examination more efficient in reaching the correct outcome (3-4 tests undertaken in analysis 1 to reach maximum probability versus 8 in analysis 2).

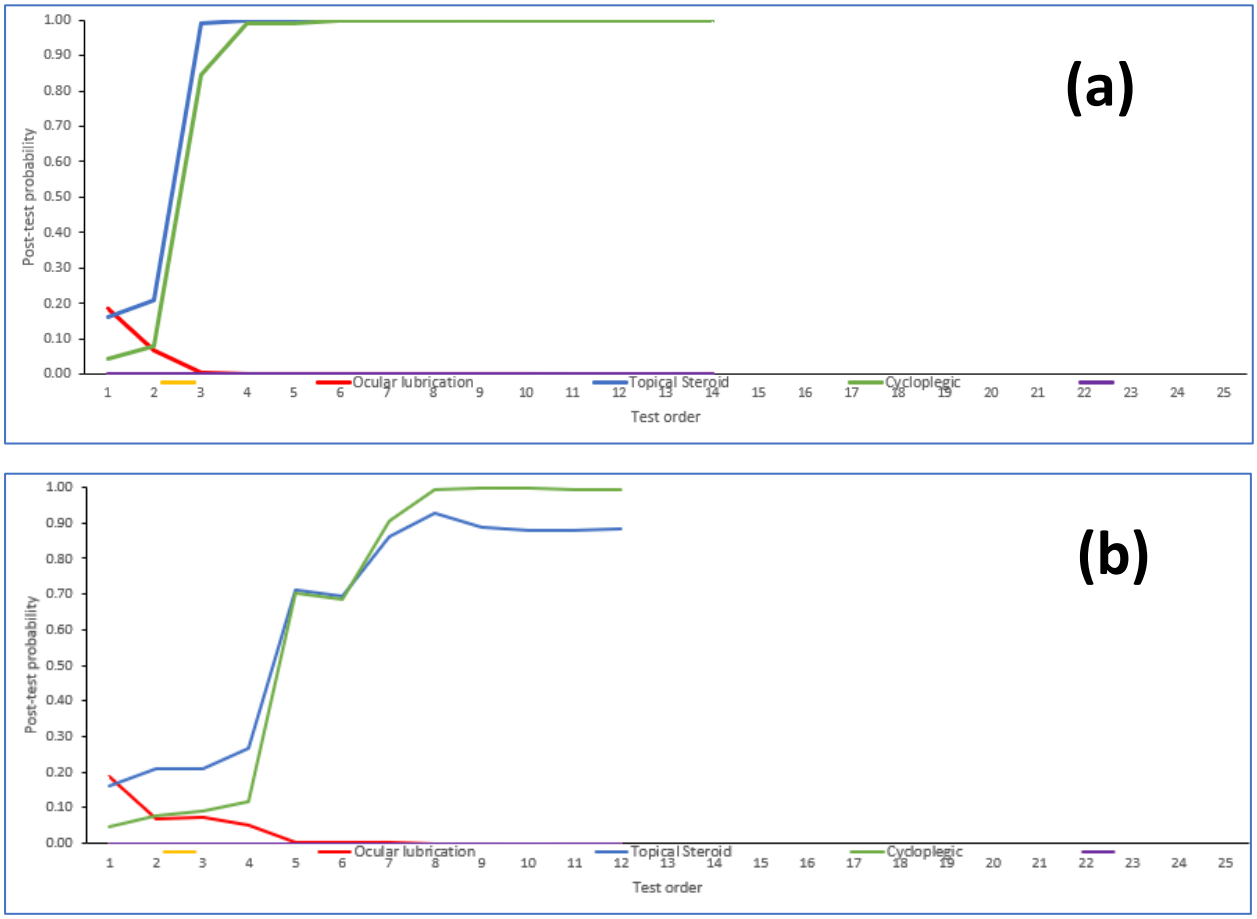


Figure 7.8 Probability graphs showing the change in post-test probability with each test selection for (a) analysis 1 (the prediction of topical steroids and cycloplegia using all 12 predictor variables and the diagnoses), (b) analysis 2 (the prediction of topical steroids and cycloplegia using all 12 predictor variables only)

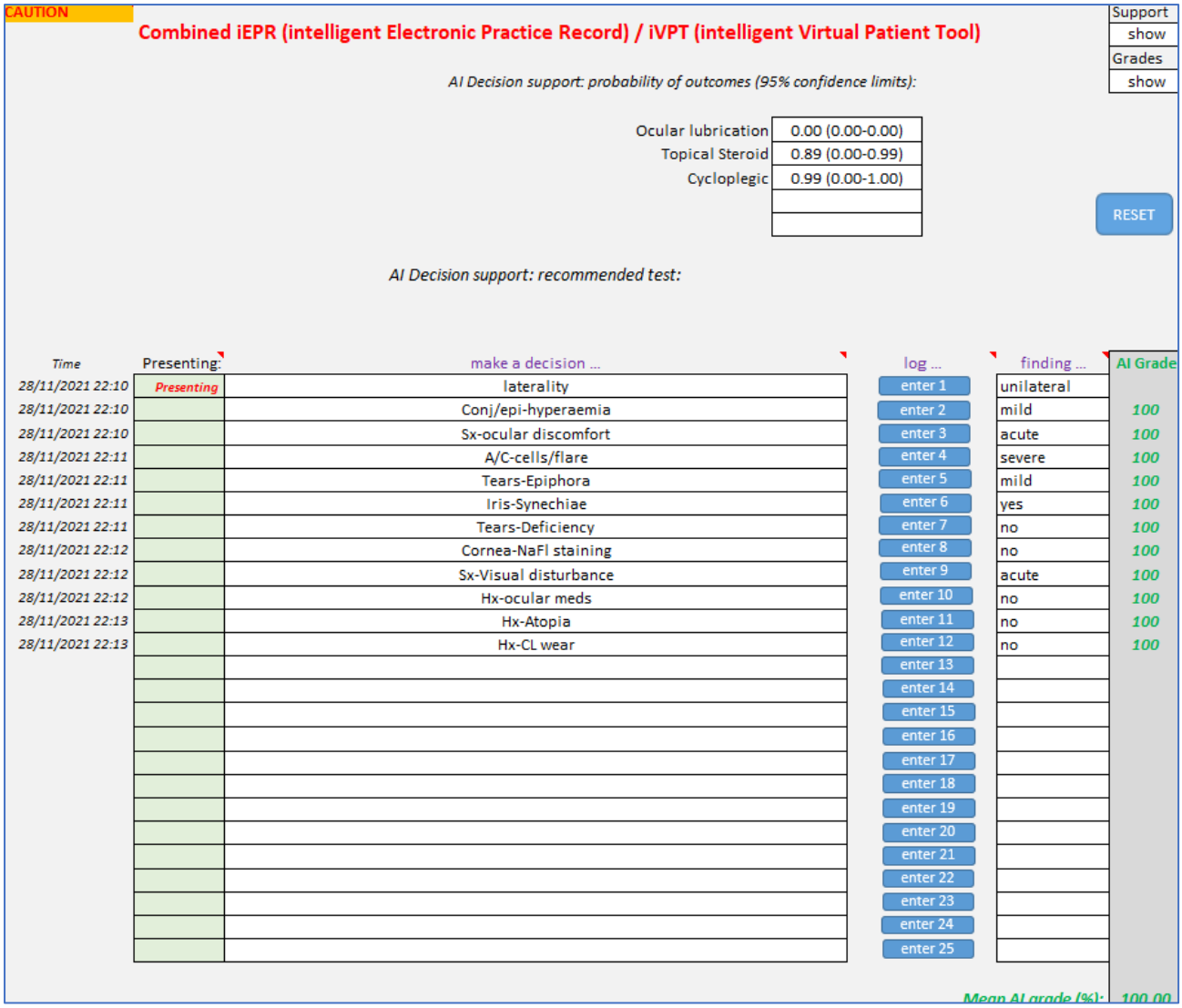


Figure 7.9 A screenshot of the iEPR/iVPT for analysis 2 (the prediction of topical steroids and cycloplegia using all 12 predictor variables)

7.2.4 **Step 5:** Evolving the system

As described in chapter 5 (section 5.2.4) and presented in chapter 6 for diagnoses (section 6.2.4), MyDLP can be used to perform 10-fold cross validation on the underlying model (Bayes'). This is used to interrogate model performance, generating measures for accuracy, informedness and markedness for the decision support provided (appendix 1, worksheet "Bayes' accuracy (evaluation)"). The performance of MyDLP (for the prescribing decisions related to KCS and uveitis) is now presented.

Following on from chapter 6, firstly the weighted accuracy, informedness and markedness are presented as overall model performance indicators. This was done by comparing three analyses (described below). This is followed by learning efficiency curves that evaluate the extent to which

Bayes' (the model) reached maximum learning. The predictors were adjusted to include and exclude diagnoses to explore and determine the best learning model.

7.2.4.1 Model performance

The model performance was explored in three analyses (for each drug type prescribed):

Analysis 1 – using all 12 predictor variables (excluding diagnoses)

Analysis 2 – using only reliable predictor variables and excluding diagnoses (i.e. those for which the confidence limits of the LRs did not straddle 1).

Analysis 3 – using only reliable predictor variables and including diagnoses

Section 7.2.3 has shown that the diagnoses are the most powerful predictors of therapeutic intervention. However, it has also shown that the correct drug treatment can be predicted with signs and symptoms alone, albeit not as efficiently. The three analyses described above explored whether the performance and learning efficiency of the model was affected by the reliability of the predictors (i.e. did only using reliable predictors optimise model performance?). Analysis 1 (using all 12 predictors) acted as the baseline analysis to which analysis 2 (using only reliable predictors) was compared. The diagnoses were then included in analysis 3, alongside the reliable predictors, to determine the difference in model performance with and without a diagnosis.

7.2.4.1.1 Model performance: Ocular lubrication

Stratified 10-fold cross validation, (with a diagnostic cut-off of 0.5) was performed to generate the performance measures for the prescription of ocular lubrication. Figure 7.10 shows the weighted model accuracy, informedness and markedness for analyses 1-3 (described above) along with the standard deviations.

Weighted accuracy remained high and consistent across the first two analyses (analysis 1: 87%, analysis 2: 86%). This is despite the reduction in predictor variables. With the inclusion of diagnoses (analysis 3), the weighted accuracy rose slightly to 90%. Weighted informedness was fair for analysis 1 (using all 12 predictor variables) at 48%. It reduced slightly to 44% for analysis 2 (reliable predictor variables only) and rose sharply for analysis 3 (reliable predictors and diagnoses) at 60%. Weighted markedness, considered the most important measure by the author (as it is based upon the predictive ability of the model), was good for analysis 1 (61%) and dropped slightly for analysis 2 (57%). It did however rise to 72% for analysis 3 (including the diagnoses). Including only the reliable predictor variables did not therefore improve the overall performance of the model. However,

including the diagnosis did improve all measures of accuracy. The overall performance of the model to predict the prescription of ocular lubrication was deemed good.

Performance for each drug type using all 12 predictor variables (no diagnosis)

	Ocular Lubrication		
	Acc	Inform	Marked
Mean	0.87	0.48	0.61
Standard deviation	0.03	0.12	0.12

Performance for each drug type using only reliable predictor variables (no diagnosis)

	Ocular Lubrication		
	Acc	Inform	Marked
Mean	0.86	0.44	0.57
Standard deviation	0.04	0.13	0.15

Performance for each drug type using reliable predictor variables plus diagnosis

	Ocular Lubrication		
	Acc	Inform	Marked
Mean	0.90	0.60	0.72
Standard deviation	0.03	0.14	0.10

Figure 7.10 Weighted performance measures (Acc = accuracy, Inform = informedness, Marked = markedness) of MyDLP (the model) for the prescription of ocular lubrication. Analysis 1 (green), analysis 2 (peach) and analysis 3 (blue).

7.2.4.1.2 Model performance: Topical steroids

Stratified 10-fold cross validation, was performed to generate the performance measures for the prescription of topical steroids. Figure 7.11 shows the weighted model accuracy, informedness and markedness for analyses 1-3 (described above) along with their standard deviations.

Weighted accuracy remained high and consistent across all three analyses (analysis 1: 85%, analysis 2: 86% and analysis 3: 89%). Weighted informedness was poor for analyses 1 and 2 (using all 12 predictor variables and using only reliable predictor variables respectively) at 15% and 16%. It rose

to 35% for analysis 3 (reliable predictors and diagnoses). Weighted markedness was good for analysis 1 (64%) and analysis 2 (67%). For analysis 3 however, markedness rose to 82%. Once again, including only the reliable predictor variables did not therefore improve the overall performance of the model. However, including the diagnosis did improve all measures of accuracy although informedness remained low. The overall performance of the model to predict the prescription of topical steroids was deemed good.

Performance for each drug type using all 12 predictor variables (no diagnosis)

	Topical steroid		
	Acc	Inform	Marked
Mean	0.85	0.15	0.64
Standard deviation	0.01	0.07	0.16

Performance for each drug type using only reliable predictor variables (no diagnosis)

	Topical Steroid		
	Acc	Inform	Marked
Mean	0.86	0.16	0.67
Standard deviation	0.02	0.07	0.21

Performance for each drug type using reliable predictor variables plus diagnosis

	Topical Steroid		
	Acc	Inform	Marked
Mean	0.89	0.35	0.82
Standard deviation	0.01	0.06	0.09

Figure 7.11 Weighted performance measures (Acc = accuracy, Inform = informedness, Marked = markedness) of MyDLP (the model) for the prescription of topical steroids. Analysis 1 (green), analysis 2 (peach) and analysis 3 (blue).

7.2.4.1.3 Model performance: Cycloplegics

Figure 7.12 shows the mean accuracy measures generated by stratified 10-fold cross validation for the prescription of cycloplegics. The weighted accuracy, informedness and markedness are accompanied by their standard deviations for analyses 1-3.

The weighted accuracy, once again, remained high and consistent at 97%, 97% and 98% for analyses 1-3 respectively. Weighted informedness was fair for analyses 1 and 2 (43% and 42%), both excluding the diagnoses as predictor variables. There was an improvement in the standard deviation value for analysis 2 compared to that in analysis 1 (9% versus 26%) indicating that using the unreliable predictors gave less consistent associations between the outcomes and predictors. However, there was a dramatic improvement in weighted informedness with the inclusion of diagnoses as predictors in analysis 3 (98%). Weighted markedness was good and fairly consistent across analyses 1-3 (76%, 80% and 77% respectively). The overall performance of the model in predicting the prescription of cycloplegics was deemed good. The best performing model included the diagnoses, mirroring the findings for ocular lubrication and topical steroids.

Performance for each drug type using all 12 predictor variables (no diagnosis)

	Cycloplegic		
	Acc	Inform	Marked
Mean	0.97	0.43	0.76
Standard deviation	0.01	0.26	0.34

Performance for each drug type using only reliable predictor variables (no diagnosis)

	Cycloplegic		
	Acc	Inform	Marked
Mean	0.97	0.42	0.80
Standard deviation	0.01	0.09	0.20

Performance for each drug type using reliable predictor variables plus diagnosis

	Cycloplegic		
	Acc	Inform	Marked
Mean	0.98	0.93	0.77
Standard deviation	0.01	0.11	0.18

Figure 7.12 Weighted performance measures (Acc = accuracy, Inform = informedness, Marked = markedness) of MyDLP (the model) for the prescription of cycloplegics. Analysis 1 (green), analysis 2 (peach) and analysis 3 (blue).

7.2.4.2 Learning efficiency

The analyses discussed above showed that the presence of a diagnosis as a predictor variable for treatment improved the overall testing efficiency and model performance for all three outcome variables (drug categories).

Evaluating learning efficiency explores whether the model has room for improvement in predictive ability (i.e., will the addition of more data improve its performance). That is, has maximum learning been achieved? This was done by generating learning efficiency curves (as discussed in chapter 5, section 5.2.4.2). Mirroring the analyses carried out for the evaluation of model performance (7.2.4.1), learning efficiency was explored in 3 ways:

Analysis 1 – using all 12 predictor variables (excluding diagnoses)

Analysis 2 – using only reliable predictor variables and excluding diagnoses (i.e. those for which the confidence limits of the LRs did not straddle 1).

Analysis 3 – using only reliable predictor variables and including diagnoses

Analysis 1 (using all 12 predictors) acted as the baseline analysis to which analysis 2 (using only reliable predictors) was compared. The diagnoses were then included in analysis 3, alongside the reliable predictors, to determine the difference in learning efficiency with and without a diagnosis.

As previously discussed (chapter 5, section 5.2.4.2), maximum learning efficiency was shown with an asymptote in the learning efficiency curves representing system accuracy, informedness and markedness.

7.2.4.2.1 Learning efficiency: Ocular lubrication

Appendix 3 shows the 9 learning efficiency curves (for all 3 analyses mentioned above) generated to explore and determine the learning efficiency for the prescription of ocular lubrication with varying fold sizes. The 9 curves were used to determine whether the asymptote for each analysis was reached between 5 – 40 cases (fold size = 5), 10 – 80 cases (fold size = 10) or 20 – 160 cases (fold size = 20). Figure 7.13 only shows the learning efficiency curves with a clear asymptote for (a) Analysis 1: A predictive model using all 12 predictor variables and no diagnoses (b) Analysis 2: A predictive model using only the reliable predictor variables and (c) Analysis 3: A predictive model using the reliable predictor variables and both diagnoses (KCS and uveitis).

The number of cases required to reach an asymptote was 120, 30 and 40 for analyses 1 to 3, respectively (figures 7.13(a), 7.13(b), and 7.13 (c)).

The number of records needed to reach maximum learning efficiency reduced with the exclusion of unreliable predictor variables and increased slightly with the addition of both diagnoses (KCS and uveitis). Maximum learning was achieved for all 3 models with relatively little patient data. As such, no additional patient data would further optimise the performance of any of the 3 models for the prediction of ocular lubrication. The best model for learning efficiency therefore was analysis 2 (shown in figure 7.13(b)). It was expected that increasing the complexity of the model (by adding more predictor variables, including diagnoses weakly related to the outcome such as uveitis), would increase the amount of data required to reach maximum learning. However, the effect was relatively small (the data required being a third of that required for all 12 predictor variables). This indicated that a more complex decision-support model involving many more ocular conditions could indeed negatively affect the learning efficiency (i.e. more data would be required to achieve maximum learning).

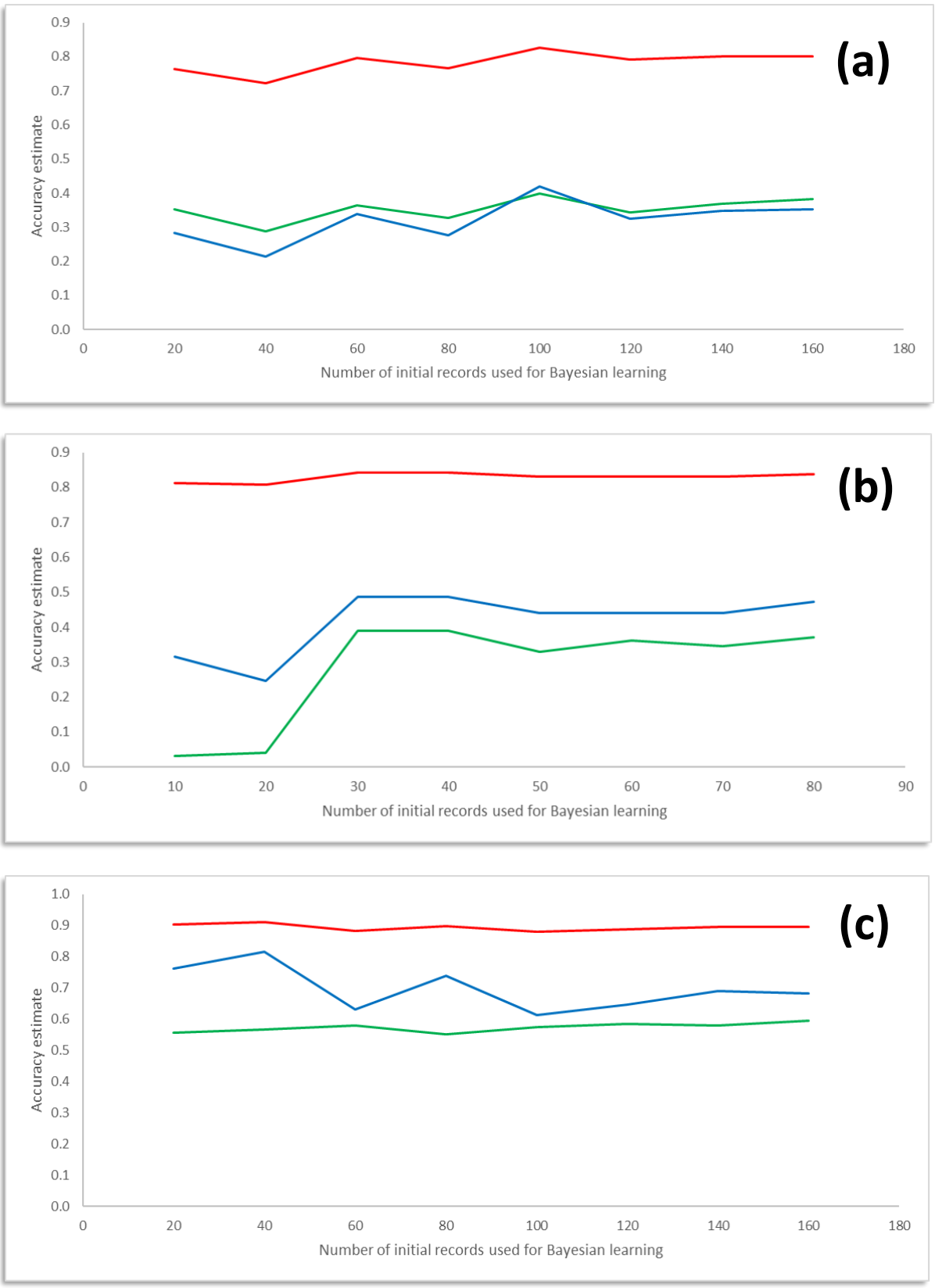


Figure 7.13 Learning efficiency curves showing the model performance for the prescription of ocular lubrication. (a) Analysis 1: Using all 12 predictor variables (b) Analysis 2: Using only reliable predictor variables and (c) Analysis 3: Using the reliable predictor variables and both diagnoses (KCS and uveitis). Curve colours: red = accuracy, green = informedness and blue = markedness.

7.2.4.2.2 Learning efficiency: Topical steroids

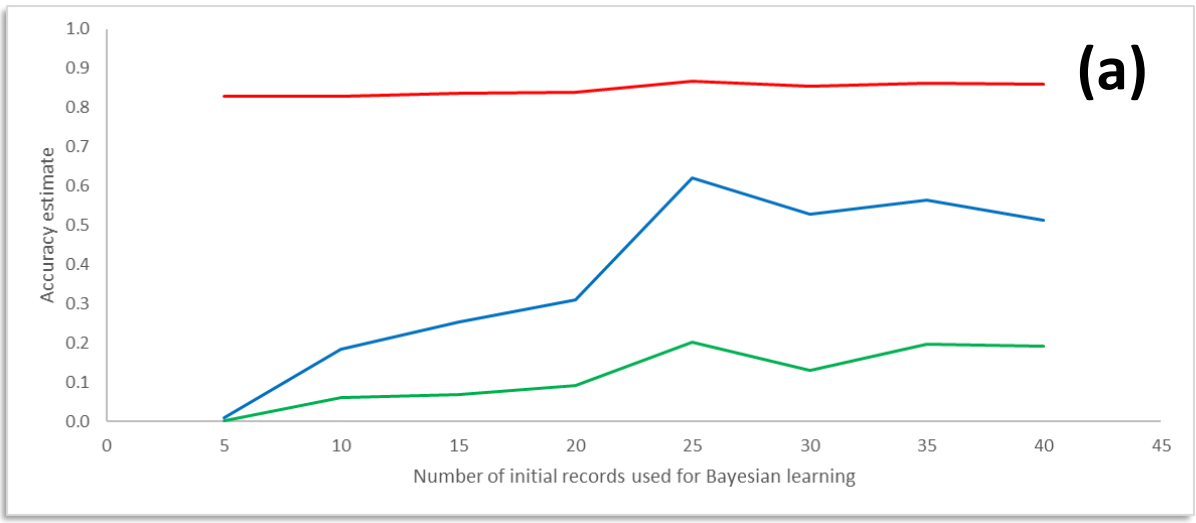
Appendix 4 shows the 9 learning efficiency curves (for all 3 analyses mentioned above) generated to explore and determine the learning efficiency for the prescription of topical steroids. Figure 7.14 only shows the learning efficiency curve with a clear asymptote for (a) Analysis 1: A predictive model using all 12 predictor variables and no diagnoses (b) Analysis 2: A predictive model using only the reliable predictor variables and (c) Analysis 3: A predictive model using the reliable predictor variables and both diagnoses (KCS and uveitis).

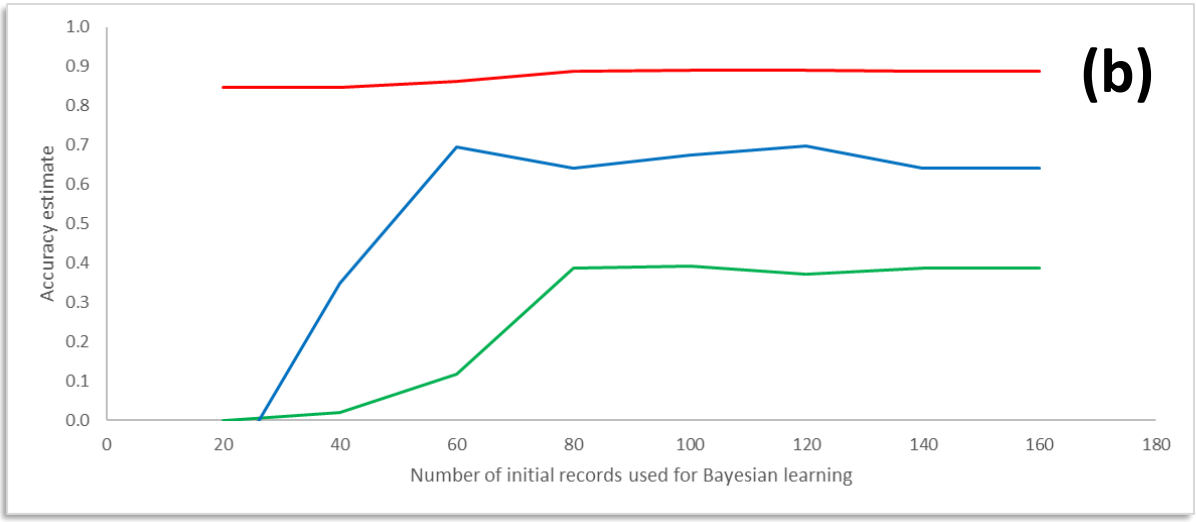
The number of cases required to reach an asymptote was 25, 60 and 15 for analyses 1 to 3, respectively (figures 7.14(a), 7.14(b), and 7.14 (c)).

The number of records required to reach maximum learning efficiency increased with the exclusion of unreliable predictor variables. This was somewhat unexpected and indicated that (1) analysis 2 resulted in the exclusion of strong predictor variables for the outcome, and (2) the importance of the individual excluded variables was underestimated due to their broad confidence intervals. As discussed, (chapter 6, section 6.2.3.1), a small dataset results in large confidence intervals for final outcomes. A larger dataset would reduce individual confidence limits and thus allow the inclusion of more predictor variables for analysis 2. However, despite this improving learning efficiency, it would not have improved overall system performance as an asymptote, indicating maximum learning efficiency, was reached.

The number of records required to reach maximum learning for analysis 3 however, reduced when compared to both the other analyses. This supported the findings in section 7.2.3.2 that a diagnosis of uveitis is the strongest predictor of the prescription of topical steroids.

Maximum learning efficiency was achieved for all three analyses, indicating that the models would not benefit from further data availability.





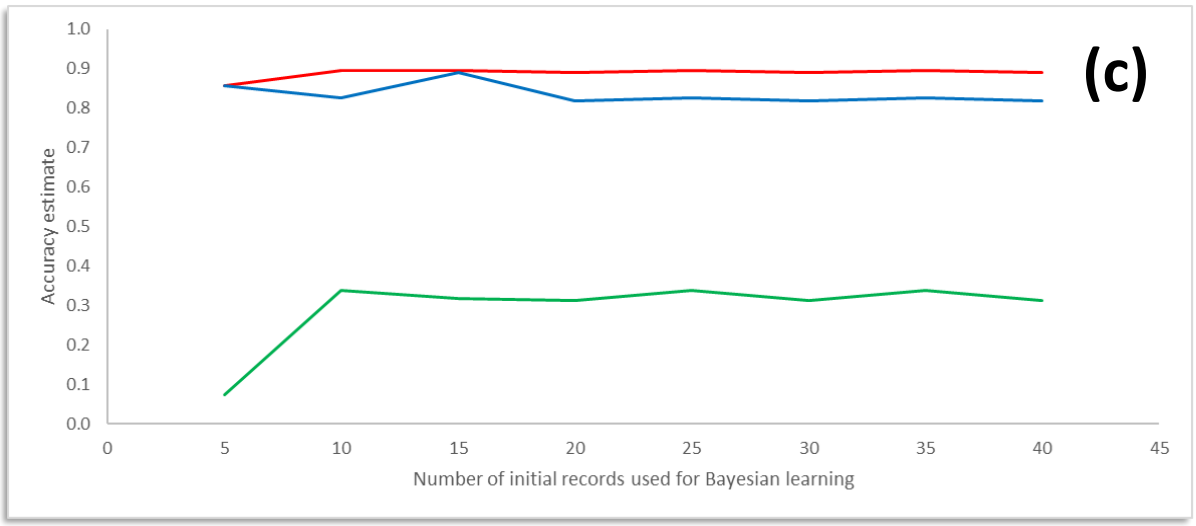


Figure 7.14 Learning efficiency curves showing the model performance for the prescription of topical steroids. (a) Analysis 1: Using all 12 predictor variables (b) Analysis 2: Using only reliable predictor variables and (c) Analysis 3: Using the reliable predictor variables and both diagnoses (KCS and uveitis). Curve colours: red = accuracy, green = informedness and blue = markedness.

7.2.4.2.3 Learning efficiency: Cycloplegics

Appendix 5 shows the 8 learning efficiency curves (for all 3 analyses mentioned above) generated to explore and determine the learning efficiency for the prescription of cycloplegics. The asymptotes indicated that maximum learning was achieved for all 3 models with relatively little patient data.

Figure 7.15 (graphs (a), (b), and (c)) shows only the learning efficiency curves with clear asymptotes, that is, at 300, 100 and 10 for analyses 1, 2 and 3 respectively. As such, no additional patient data would further optimise the performance of any of the 3 models for predicting the prescription of a cycloplegic. The best model for learning efficiency was analysis 3 (shown in figure 7.15(c)) as it required the least number of episodes. A diagnosis, therefore, is a very important predictor in dictating treatment with cycloplegics and enables the most efficient machine learning.

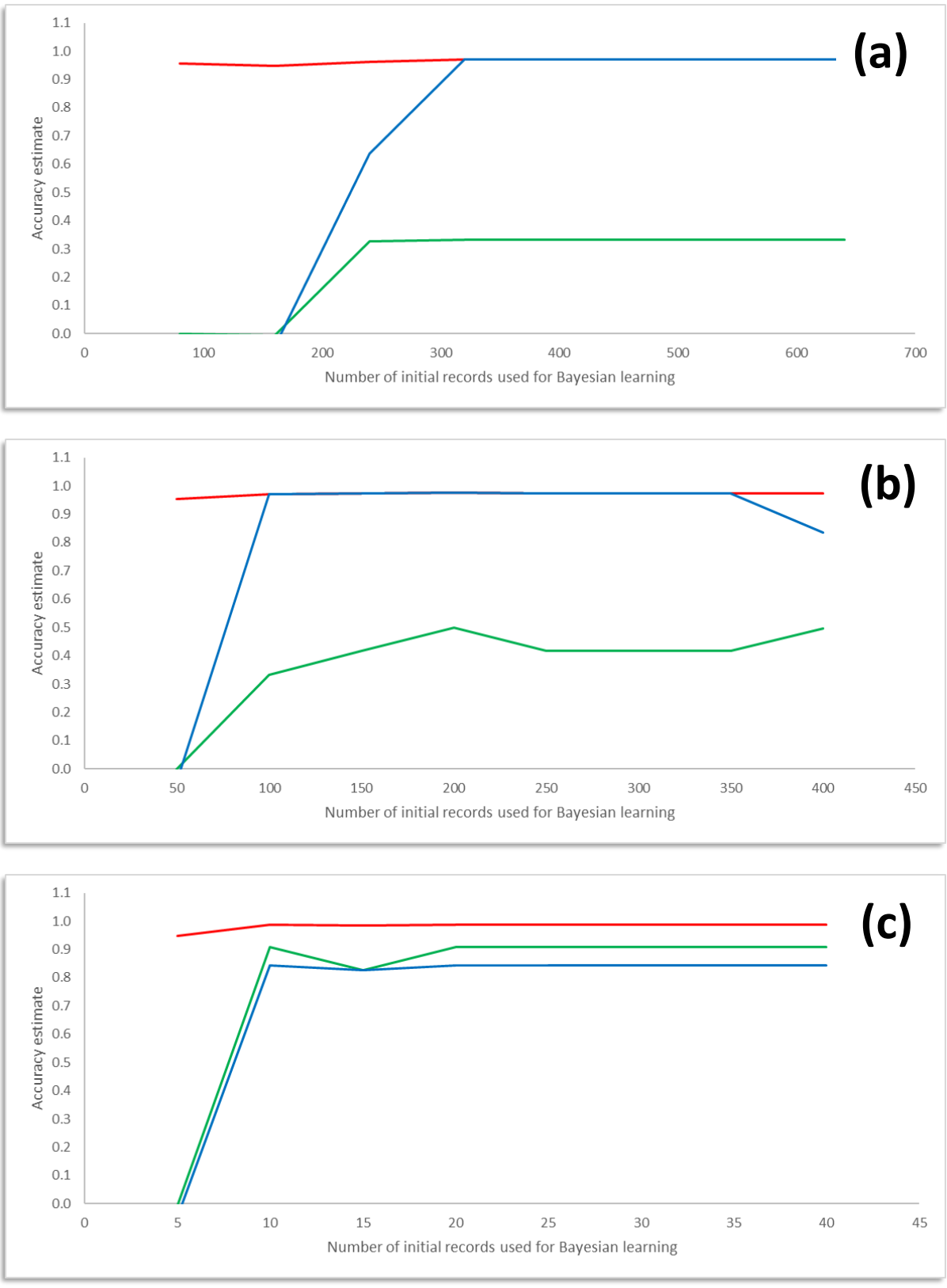


Figure 7.15 Learning efficiency curves showing the model performance for the prescription of cycloplegia. (a) Analysis 1: Using all 12 predictor variables (b) Analysis 2: Using only reliable predictor variables and (c) Analysis 3: Using the reliable predictor variables and both diagnoses (KCS and uveitis). Curve colours: red = accuracy, green = informedness and blue = markedness.

7.3 Chapter summary

Chapter 7 presented the application of MyDLP for prescribing decisions involving ocular lubrication, topical steroids and cycloplegics. Following on from chapter 6, the application was presented in terms of the 5 steps of the bTLC.

Data collected by the specialist IP optometrists (1351 patient episodes, the evidence) was used to power MyDLP (step 1, gathering the evidence). Twelve predictor variables, covering all the tests carried out by the specialist IP optometrists to prescribe the three predicted drug types, were included.

Step 2 of bTLC consisted of learning from the evidence base. This step generated pre-test odds (the rareness of things) and likelihood ratios (LRs – the importance of tests) which were subsequently used to provide prescribing decision support through Bayes' theorem (the model) in steps 3 and 4 of bTLC. An evaluation of the LRs generated revealed that the diagnoses were the strongest predictors of drug treatment. It appeared then that clinicians formulate a working diagnosis and treat accordingly. This supports the presentation of clinical guidelines (such as for the CMGs) which present investigation and treatment guidelines for named conditions and not for signs and symptoms.

The iEPR/iVPT demonstrated how, with the use of AI generated clinical decision support, the efficiency of clinical testing could be optimised. As discussed in chapter 6, healthcare services, clinical equipment and clinician time are finite resources. The current chapter showcased how MyDLP can recommend the order of clinical tests. These leading to the correct prescribing decision (mirroring the decision-making of experienced specialist clinicians). Indeed section 7.2.3 showed that MyDLP reduced the number of tests required for all three drug categories.

The final step of bTLC, namely evolving the system, allowed the user to explore the performance and learning efficiency of the model. The performance measures showed a well performing model for all three predicted variables with accuracy ranging from 85%-98% and markedness ranging from 57% - 92%. Including the diagnoses as predictor variables resulted in the best performing models for all 3 of the drug categories. The learning efficiency curves showed that maximum learning had been achieved for ocular lubrication, topical steroids and cycloplegics. This means that even with the addition of more data, the model is unlikely to perform any better. The assumption of naïve Bayes' (independence of predictor variables) inevitably results in rapid learning to an asymptote that does not improve with further data (99). Despite this, the AI decision support provided by MyDLP for the

prescription of ocular lubrication, topical steroids and cycloplegia can improve the efficiency of clinical investigation and provide robust prescribing support.

In conclusion, chapters 6 and 7 have provided a proof-of-concept for the application of MyDLP as a clinical decision-support and learning tool. In the present study MyDLP was used to (1) provide decision-support to establish an ocular diagnosis, and (2) to provide prescribing support for anterior ocular conditions. A complete diagnostic and teaching support tool would require more data covering many ocular conditions from which the model could learn, and an up-scaling of MyDLP to include all the possible predictor and predicted variables.

Chapter 8 will now summarise the thesis, discuss its limitations and future areas of research.

Chapter 8: Study Findings, Limitations, and Further Work

8.1 Introduction

A summary of the thesis and the key findings are presented in this chapter. This is followed by a discussion about the limitations of the study. Finally, directions for further research are recommended. Figure 8.1 shows the chapter plan.

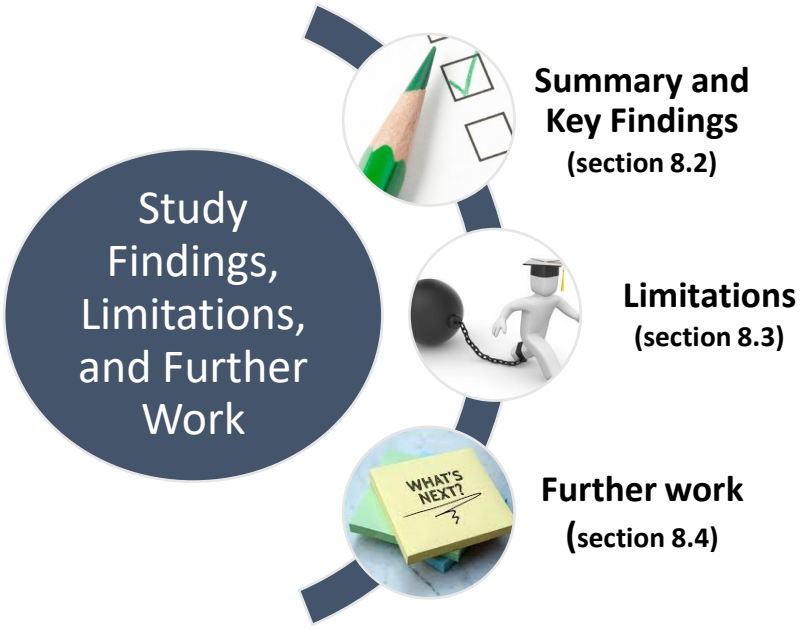


Figure 8.1 The chapter plan for study findings, limitations, and further work

8.2 Summary and Key Findings

The primary aims of this study were to a) apply machine learning to the clinical activity of independent prescribing (IP) optometrists and b) to develop the concept of an interactive and evolving evidence-based support system for IP optometrists and those in training. As far as the author is aware this was the first time machine learning was applied to the clinical activity of IP optometrists. A 5-step Bayesian translational learning concept was developed (bTLC) (Chapter 1, section 1.5). Given the highly specialised data-collection, an opportunity also arose to assess the scope of IP optometrist practice against that dictated by the CMGs (chapter 4). The evidence base was composed of the results of fully anonymised patient consultations (episodes). There was no alteration of routine clinical procedures and no invasion of clinical autonomy.

Previous research undertaken at Aston University using AI in the form of Bayes' theorem showed good potential in replicating optometrist decision-making for the diagnosis of ocular conditions, but especially for the diagnosis and management of chronic open angle glaucoma (chapter 1, section 1.3.2). Unstructured clinical data as well as the lack of negative clinical findings limited the accuracy of naïve Bayes' when considering all ocular conditions (97).

8.2.1 Summary of Thesis

Step 1 of the cycle, gathering the evidence (described in chapter 2), involved the development of a highly structured data collection electronic patient record (EPR). This allowed the input of both positive and negative clinical findings for a refined set of predictor variables (clinical signs, symptoms, and demographical information). This highly structured EPR made the data collected more conducive to machine learning. The ocular conditions included were those listed in the College of Optometrists clinical management guidelines (CMGs) and those concerning the anterior eye (i.e. those most likely to be treated by IP optometrists). Data was collected by four specialist IP optometrists forming part of an Acute Primary Care Ophthalmology Service (APCOS) in West Kent. Over a period of a year, 1351 first patient episodes were collected.

Step 2, learning from the evidence base, described the process of applying machine learning in the form of naïve Bayes' theorem to the data. As described in chapter 3 (section 3.2.1) this involved, extracting from the data collected, the counts necessary to generate the pre-test odds (the rareness of things) and the multilevel likelihood ratios (the importance of tests). These were then used to generate post-test odds and thus post-test probabilities for any outcome of choice (diagnoses, management decisions and prescribing decisions).

Hierarchical Bayes (the model) was the first attempt to apply machine learning to the data collected (chapter 3, section 3.2.2). This model used the data with its complete granularity (all diagnoses, management, and individual drug combinations with their regimes) to train on and to drive a potential clinical decision-support system. This model involved 4 hierarchical stages, each using predictor variables to ascertain diagnoses, management options, drugs prescribed, and the drug regimes (administration and duration) for each of the drugs prescribed (outcome variables). A total of 125,405 multilevel LRs were generated. Five analyses were undertaken on the model to determine its performance (accuracy, informedness and markedness) and thus potential as a basis for clinical decision support. Unexpectedly hierarchical Bayes' performed modestly at best and was thus deemed insufficient for the translation of evidence to clinical practice. The poor performance of the model was attributed to a lack of data for each outcome. That is, too many outcome variables (or high granularity), and thus not enough training data for each outcome to allow adequate

machine learning. There was however a gradual increase in 'safe' LRs with an increase in the data processed (chapter 3, table 3.6). This meant that whilst hierarchical Bayes' was insufficient in its current form, there remained potential should more data be added to the model. Only 3,963 (3.15%) of the LRs were deemed 'safe' after use of all 1351 patient episodes collected and thus it was deduced that many times more data would be required for a full decision support system using the model. Collecting the amount of data required was not possible given the time restrictions of the present study. It was therefore concluded that hierarchical Bayes' could not be developed any further. Hierarchical Bayes performed well (in terms of markedness) for some highly prevalent individual ocular conditions (i.e. those with the most data). This led to an alteration of research objectives, and it was decided that a 'proof-of concept' approach would be adopted. In order to demonstrate the potential of machine learning, two of the most commonly presenting ocular conditions (KCS and uveitis) were used to develop and demonstrate an evidence-based clinical-decision support and teaching tool.

Steps 3 and 4 of the bTLC, namely translation to evidence-based practice and research-led teaching, culminated in the development of a digital learning platform (MyDLP) described in chapter 5. This was developed using the ubiquitous Microsoft Excel® and is composed of 9 worksheets (chapter 5, section 5.2). MyDLP can provide decision-support for practising clinicians and those in training across clinical disciplines, whilst enabling an understanding of machine learning concepts. It houses a combined "intelligent" electronic patient record and virtual patient tool (iEPR/iVPT) capable of providing decision support in the form of next-best-test recommendations alongside automated grading of the decisions made, both of which can be hidden (chapter 5, section 5.2.3). It also generates a probability plot showing how the probability of each outcome alters with each test undertaken. The teaching element (iVPT) was designed to allow students to appreciate the relationship between clinical testing and diagnostic outcomes. This would aid the transition from 'database' style examinations to 'problem-orientated' examinations. The iVPT represents a move from 'first generation' virtual patient tools (199) currently used at Aston University (based upon textbook presentations of ocular conditions) to 'second generation' virtual patient tools that translate clinical evidence to the classroom by way of Bayesian AI. This is particularly important when the exposure to 'real life' patients has been limited due to Covid-19 restrictions. MyDLP can predict up to 5 outcomes (dependent) variables, with confidence limits, using up to 25 predictor (independent) variables.

To fulfil step 5 of the bTLC (evolving the system), MyDLP also allows the user to explore the performance of the machine learning model. Interactive areas allow the addition of more data and the manipulation of data analysis. This includes exploring the effects of spurious predictors,

Laplacian smoothing, diagnostic-cut off points, and altering the stratification status in cross validation to interrogate the performance of the model. Model performance is presented to the user in the form of measures for accuracy, informedness and markedness (chapter 5, section 5.2.4.1). An evaluation as to whether the model has reached maximum learning can also be explored on worksheet 9 (chapter 5, section 5.2.4.2). This allows the user to determine whether more data would improve the performance of the model. Learning efficiency curves are automatically generated for the model (for model accuracy, informedness and markedness). An asymptote in the curves indicates that maximum learning has been achieved and no further data would improve model performance.

Naïve Bayes' was the model used to drive MyDLP using the data collected from specialist IP optometrists. Its use and performance were demonstrated for a 'proof-of-concept' in chapters 6 (to predict the diagnoses of KCS and uveitis) and 7 (to predict the prescribing decisions for KCS and uveitis). These chapters showcased how MyDLP directed the user to more efficient clinical testing (a reduced number of tests to achieve the correct outcome) for both the establishment of a diagnosis and the determination of a prescribing decision.

On evaluating model performance for the prediction of the diagnoses (KCS and uveitis), three analyses were undertaken (chapter 6, section 6.2.3). The manipulation of the predictor variables to include only variables that were both reliable and recommended by the CMGs resulted in the best performing model (highest weighted markedness 58% for KCS and 86% for uveitis). The associated learning efficiency curves showed that maximum learning had been achieved and thus the model would not improve with the addition of more data. In its current form, the model could provide valuable diagnostic-decision support.

Similarly model performance was evaluated for the prediction of drugs prescribed for KCS and uveitis (ocular lubrication, topical steroids and cycloplegics) (chapter 7, section 7.2.4). Again three analyses were undertaken manipulating predictor variables. The best performing model was that composed of only reliable predictor variables with the inclusion of the diagnoses as predictor variables (highest weighted markedness 72% for ocular lubrication, 82% for topical steroids, and 77% for cycloplegics). The diagnoses were identified as the strongest predictors of treatment choice. Learning efficiency curves generated for all 3 analyses showed asymptotes, indicating that model performance could not be improved further with the addition of more data. Once again, in its current form, the model provided valuable decision-support.

8.2.2 Key Findings

- There exists a hierarchy of clinical testing determined by patient symptoms and suspected diagnoses. Ranking of clinical tests according to their importance to a diagnosis (using LRs) and including important negative findings in guidelines, can prove useful in allowing efficacious patient consultations
- Diagnoses are the strongest predictor variables for drug treatment (have the highest LRs)
- MyDLP presents an answer to the Topol review (chapter 1, section 1.1.3.3), The NHS Long Term Plan (chapter 1, section 1.1.3.2) and the science of knowledge translation (KT) (chapter 1, section 1.1.5.3) by introducing the concept of a 'white box' machine learning platform from which clinicians can potentially:

    1) Develop a good working knowledge of the concepts of machine learning, including its evaluation
    2) Incorporate machine learning into their everyday practice without the alteration of clinical activities, bridging the evidence-to-practice gap
    3) Utilise "live" clinical-decision support (including best-next-test suggestions, and diagnostic and prescribing support) for clinicians practising in isolation or those with little experience, complete with confidence intervals
    4) Utilise a virtual patient tool for those in training, presenting the clinical data from real patients and including diagnoses and prescribing decisions from experienced clinicians

8.2.3 The review of IP Optometrist Practice

There remains scarcity in the literature concerning the activity of IP optometrists. As such the opportunity to review the clinical activity of the APCOS specialist IP optometrists was deemed of value. This was done, firstly, to establish whether the ocular conditions presenting to the IP optometrists mirrored those listed in the CMGs and, secondly, to compare the investigation, management, and prescribing decisions of the IP optometrists against the recommendations given in the relevant CMG (chapter 4). The key findings of the audit are given below:

- The CMGs accurately represent the majority of anterior ocular conditions encountered in APCOS IP practice (92.7% of anterior eye activity).
- Posterior eye pathology frequently seen included glaucoma and cystoid macula oedema (CMO) following cataract extraction. Whilst glaucoma has a CMG listing, CMO does not.

- The history and investigations listed in the CMGs reflected the tests undertaken by IP optometrists. However, negative clinical findings were shown to be valuable in the establishment of a diagnosis (from the data). The CMGs do not include strong negative associations.
- Most patients were discharged after the first visit (the most favoured management decision).
- Onward referral accounted for only 10% of management decisions (90% of patients were managed by IP optometrists or discharged).
- Microbial keratitis (MK), pre-septal cellulitis and herpes zoster ophthalmicus (HZO) showed the greatest variation in management and prescribing decisions when compared with the CMGs. These cases were managed therapeutically by the IP optometrists in the first instance despite the CMG recommendation stating a prompt referral.
- Prescribing practice (the drugs prescribed) of IP optometrists showed an overall agreement of 62% against the CMGs.

Since this audit of IP optometrist activity, the College of Optometrists have reformed the way in which clinical guidance is reviewed. This includes seeking feedback from a number of review groups including an IP review group of which the author and one of the data collectors are members. These groups are asked to review the recommendations listed within each of the CMGs. The CMG for pre-septal cellulitis was reviewed in November 2021. As a result the College has now updated the CMG to include management in the community by IP optometrists with the prescription of systemic antibiotics. This now reflects IP practice as reported in chapter 4.

8.3 Limitations

The important limitations to this study are now discussed.

There are some limitations to the generalisability of the IP optometrist activity discussed. Firstly, data was collected from specialist IP optometrists who form part of APCOS, a community ophthalmology service commissioned by the West Kent CCG. The APCOS service was designed to reduce the burden of patient demand in secondary care by acting as an intermediary between referring clinicians (other optometrists and GPs) and the HES. This service includes the provision of NHS prescription pads. Such community services are not commonplace in England. As a result, the patients presenting to APCOS may not be reflective of those who would be expected to present to IP optometrists outside community ophthalmology services or indeed in hospital practice. Whilst this

would affect the pre-test odds generated (the prevalence of the ocular condition or prescribing decision), it would not affect the LRs (the importance of each clinical test undertaken) which would remain transferrable. The author is not aware of any evidence suggesting a regional variation of ocular conditions and therefore so long as the data used to drive the model is from the UK, MyDLP would still provide good decision-support. It may be possible then if this model were to be used across borders where the prevalence of disease may differ, that the pre-test odds generated in this study would be a non-transferrable element. This limitation, however, would be overcome if world location was added as one of the predictor variables, and data was collected from clinics world-wide. Secondly, the IP optometrists from whom the data was collected hold further higher qualifications beyond independent prescribing and were some of the first independent prescribing optometrists in the UK. Unlike more recently qualified IP optometrists (the majority), they may be more confident in treating and discharging patients in the first instance. For the purpose of the audit, this may not be representative of wider IP optometrist activity. Thirdly, local protocols (prescribing practice recommendations) can dictate the drugs chosen to treat patients with ocular conditions. This can be reflective of drug costs as opposed to optimal clinical efficacy. The drugs prescribed therefore may not reflect the prescribing choices in other areas. Fourthly, this study only included first patient episodes. As such, patients were not followed up to ascertain whether the prescribed treatment was effective, or whether they required further support from other clinicians.

Despite the above limitations, the data collected was deemed appropriate for the aims of this study. This is because the aim of the study was to apply machine learning to the clinical activity of IP optometrists, that is, decision replication. To this end, local protocols, confidence in prescribing, NHS prescription pads, higher qualifications, and the effectiveness of prescribing decisions are not important. However, the reader should recall that a) the prescribing-decision support offered by MyDLP was based on drug groups (depending on the action of the drug) as opposed to individual formulations, and therefore bypasses the specific drug recommendations as usually listed in prescribing practice guidelines, and b) the clinical activity of the APCOS IP optometrists was very much aligned to the recommendations of the CMGs. As such it is reasonable to assume that there would be little regional variation and that the data used in this study would reflect best practice.

Another limitation of the study was that only one machine learning method (Bayes') was used to extract the components required for decision replication by MyDLP. This meant that due to the assumptions of Bayes theorem (specifically the independence of each predictor variable) the learning efficiency of MyDLP, although good, reached an asymptote after which the model could not improve further. There are two possible solutions to this limitation presented by Bayes'. One is to refine the predictor variables (known as "reducing dimensionality") by way of selecting or extracting

the strongest predictors in a dataset containing all the raw clinical information from patient consultations (138). The refined predictor variables can be individual or combinations of variables. This works to reduce 'noise' in the data by creating truly independent predictors. Statistical methods to undertake such feature selection include principle components analysis, factor analysis, independent components analysis, wavelet transformation and singular value decomposition (138,203). Bayes' can then be applied to the components by assigning a LR to each component (as was done in this study). The second is to consider a different method of machine learning. There are many other supervised and unsupervised machine learning methods which may have performed better given the data collected. Such methods include logistic regression, neural networks, decision trees and support vector machines (136,137,204). No comparison of machine learning methods was undertaken to determine the superiority of one method over another. Bayes' was decided upon given its relative simplicity and familiarity amongst clinicians. As such it provides a "white box" method of machine learning, something clinicians would have more confidence in (205).

The final limitation is that of record abstraction. Machine learning depends upon the quality of the data on which learning occurs. That is, the clinical episodes on which support systems are built must include all the investigations undertaken in order to create accurate associations between tests and outcomes. It has been shown however that "optometrists ask more questions relating to history and symptoms, perform more clinical tests, and offer more advice than they report in their clinical records" (206). To that end it whilst it cannot be guaranteed, it is anticipated that the data-collectors in this study, having been made aware of the aims of the study and included in the design of the data-collection EPR, understood the importance of accurate data collection. Similarly, the use of a concurrent (data-collection) EPR alongside the usual patient recording system may have produced some bias in the form of the Hawthorne effect (207). That is, when clinicians are aware of being observed, they may alter their behaviour to mirror perceived researcher expectations. Whilst this is a possibility, there is little understanding of the overall effect this bias creates, its mechanism and therefore how to mitigate it completely.

8.4 Recommendations for Further Work

Following on from the limitations mentioned in 8.3, further work could be centred around three areas:

- Further enhancing model performance
- Further development of the decision support tool (MyDLP)

- Preparing clinicians for the digital future

8.4.1 Further enhancing model performance

Naïve Bayes was the machine learning method of choice in this study due to its simplicity and ease of understanding. Other, more sophisticated, machine learning methods (including unsupervised methods) could have achieved greater levels of system accuracy, however, these were considered outside the scope of this study due to their complexity, especially with using Excel® as the delivering software. There are, however, several data mining tools which provide a platform to facilitate machine learning. The techniques available in these tools cover a multitude of supervised and unsupervised machine learning methods. The Waikato Environment for Knowledge Analysis (WEKA), Rapid Miner, KNIME, Python, Oracle data mining and Orange are to name a few (208,209). These platforms would allow a comparison of machine learning methods to optimise decision support. Further work using such platforms to drive and compare the performance of decision support systems could be investigated in the future.

Following on from the limitations of Bayes' theorem and the assumed independence of predictor variables (leading to a plateau in accuracy measures), a further study could include a method by which to isolate principle component predictor variables before analysis. Such methods include principle component analysis, factor analysis, independent components, wavelet transformation and singular value decomposition (section 8.3). This would likely result in the enhanced accuracy of the model as the predictor variables, being individual variables or combined variables (for example blepharitis and tear deficiency as a single predictor), would be truly independent.

The effectiveness of the prescribing decisions on which the model was built were not evaluated in the present study. This is because only first patient consultations were included in data collection. Further studies could include this information by following-up patients post treatment. Only episodes where the treatment proved to be effective could be included as the "evidence-base" on which the model could train. It is anticipated that this would in-turn lead to more effective treatment recommendations.

8.4.2 Further development of the decision support tool (MyDLP)

A limitation of the present study was that data was collected from only one area and service of the UK (section 8.3). In order for a decision support tool to be truly useful, it needs to be reflective of practice in all geographical areas and settings. It follows then, that the data needed to drive such a decision support tool must be harnessed from multiple areas and settings (including community and

hospital settings). The addition of “clinic location” (geographical area) and “practice setting” (community, HES etc.) to the predictor variables on the iEPR, could then be used to generate area and setting-specific decision support. This would be especially useful for portfolio and locum optometrists who work over large areas and settings with differing clinical pathways. Further study involving multicentre data collection is therefore pivotal to the development of a universal decision support tool.

8.4.3 Preparing Clinicians for the Digital Future

As discussed in chapter 1 (section 1.1.3.2) the NHS long-term plan sets out a vision for the transformation of healthcare to include the integration of AI and decision support (23). In response, MyDLP has showcased how this could be integrated into the clinical practice of IP optometrists in the form of an “intelligent” EPR (iEPR). MyDLP has the potential to be used across clinical disciplines, simply by choosing speciality-based predictor and outcome variables.

Such seamless integration of AI into clinical practice is important for the effective implementation of decision support. It is, however, equally important to address the barriers to implementation as mentioned in the Topol review, that is, a resistance to change and scepticism about technology amongst the healthcare workforce (28). It follows then, that an understanding of the mathematical concepts which underpin AI systems along with their strengths and weaknesses is vital for digital progress by alleviating such barriers. It has been suggested that a digital transition of the healthcare industry would mean “our training institutions have a responsibility to prepare current and future physicians and healthcare professionals to become AI competent” (210). Indeed, the Topol review contained a recommendation that AI be “prominent in the undergraduate curricula for healthcare professionals” and that “students gain an appropriate level of digital literacy at the outset of their study” (28). A staged approach to introducing mathematical concepts, AI fundamentals, data science, and the corresponding ethical and legal issues has been suggested for integration into medical curricula (211). However, the author is not aware of any similar published suggestion for AI teaching in undergraduate optometry courses. This is an area that requires further study and implementation by means of bringing together computer scientists and clinicians to deliver an undergraduate programme of study. In the author’s opinion the core learning outcomes of such a programme should centre around basic “white box” machine learning methods such as Bayes’ and extended to include an awareness of “black box” methods. This will develop a basic but transferrable understanding of machine learning concepts (transferrable to any decision augmentation tool that the clinician may encounter in future practice), much like a step-along-ray-trace forms the basis of visual optics.

8.5 Summary

The findings of this study indicate that using a 5-step process (bTLC), Bayesian machine learning can be successfully applied to the clinical activity of IP optometrists. This has shown to result in good decision replication for diagnoses and prescribing decisions. Moreover, the present study provided a 'proof-of-concept', in the form of MyDLP, demonstrating how such machine learning can be used to power "live", "white box" decision support tools, useful to both qualified IP optometrists and those in training.

References

1. Kaplan A, Haenlein M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Bus Horiz. 2019 Jan;62(1):15–25.
2. Susskind R, Susskind D. The Future of the Professions: How Technology Will Transform the Work of Human Experts. paperback. New York: Oxford University Press; 2017.
3. Ansari E, Patel M, Harle D. Acute Community Ophthalmology Services Provided by Independent Prescribing Optometrists Supporting Hospital Eye Services during the COVID-19 Outbreak. J Optom. 2021 Jan 31;
4. Kanabar R, Craven W, Wilson H, Rietdyke R, Dhawahir-Scala F, Jinkinson M, et al. Evaluation of the Manchester COVID-19 Urgent Eyecare Service (CUES). Eye. 2021 Apr 30;1–9.
5. Bengtsson B, Heijl A. A Visual Field Index for Calculation of Glaucoma Rate of Progression. Am J Ophthalmol. 2008 Feb;145(2):343–53.
6. Wesselink C, Heeg GP, Jansonius NM. Glaucoma Monitoring in a Clinical Setting Glaucoma Progression Analysis vs Nonparametric Progression Analysis in the Groningen Longitudinal Glaucoma Study. Arch Ophthalmol. 2009;127(3):270–4.
7. Harvey B. Instruments: Distance Refraction. Optician. 2021;(15 January):28–9.
8. 2020Health. Foresight Project Report: A discussion of the potential impact of technology on the UK optical sector to 2030. 2016.
9. Dabasia PL, Edgar DF, Garway-Heath DF, Lawrenson JG. A survey of current and anticipated use of standard and specialist equipment by UK optometrists. Ophthalmic Physiol Opt. 2014 Sep 1;34(5):592–613.
10. Aravind M, Chung KC. Evidence-Based Medicine and Hospital Reform: Tracing Origins Back to Florence Nightingale. Plast Reconstr Surg. 2010 Jan;125(1):403–9.
11. Slade S, Shickle D. Ophthalmic Public Health Research Project: Final report. 2017.
12. Baker H, Ratnarajan G, Harper RA, Edgar DF, Lawrenson JG. Effectiveness of UK optometric enhanced eye care services: a realist review of the literature. Ophthalmic Physiol Opt. 2016 Sep 1;36(5):545–57.
13. Konstantakopoulou E, Edgar DF, Harper RA, Baker H, Sutton M, Janikoun S, et al. Evaluation of a minor eye conditions scheme delivered by community optometrists. BMJ Open. 2016 Aug 10;6(8):e011832.
14. Swystun AG, Davey CJ. A prospective evaluation of the clinical safety and effectiveness of a COVID-19 Urgent Eyecare Service across five areas in England. Ophthalmic Physiol Opt. 2022 Jan 1;42(1):94–109.
15. Mitchell M. History of the British Optical Association, 1895-1978. London: The Association and the British Optical Association Foundation; 1982. 304 p.
16. Stanton JR. The Ophthalmic Optical Profession. London: Keith V. Aughan Publisher; 1963.
17. GOC. GOC rules and regulations: Rules Relating to Injury or Disease of the Eye 1999 [Internet]. [cited 2018 Aug 26]. Available from: https://www.optical.org/en/about_us/legislation/rules_and_regulations.cfm

18. Kotecha A, Brookes J, Foster P, Baldwin A. Experiences with developing and implementing a virtual clinic for glaucoma care in an NHS setting. Clin Ophthalmol. 2015 Oct 15;9:1915.

19. Bowes OMB, Shah P, Rana M, Farrell S, Rajan MS. Quality indicators in a community optometrist led cataract shared care scheme. Ophthalmic Physiol Opt. 2018 Mar 1;38(2):183–92.

20. NHS England. Transforming elective care services ophthalmology. 2019.

21. NHS Digital. Hospital Outpatient Activity 2019-20 [Internet]. 2020 [cited 2021 May 17]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/hospital-outpatient-activity/2019-20

22. Healthcare Safety Investigation Branch. Lack of Timely Monitoring of Patients with Glaucoma. 2020.

23. Department of Health. The NHS Long Term Plan. 2019 Jan.

24. NHSX. Digitise, connect, transform - NHS Transformation Directorate [Internet]. 2022 [cited 2022 May 23]. Available from: https://www.nhsx.nhs.uk/digitise-connect-transform/

25. NHSX. What Good Looks Like framework - NHS Transformation Directorate [Internet]. 2021 [cited 2022 May 23]. Available from: https://www.nhsx.nhs.uk/digitise-connect-transform/what-good-looks-like/what-good-looks-like-publication/

26. NHSX. Who Pays for What proposals - NHS Transformation Directorate [Internet]. 2021 [cited 2022 May 23]. Available from: https://www.nhsx.nhs.uk/digitise-connect-transform/who-pays-for-what/who-pays-for-what-proposals/

27. Gould M. NHSX moves on - NHS Transformation Directorate [Internet]. 2022 [cited 2022 May 23]. Available from: https://www.nhsx.nhs.uk/blogs/nhsx-moves-on/

28. Department of Health. The Topol Review: Preparing the halthcare workforce to deliver the digital future. 2019.

29. Moorfields Education. An Introduction to Data Science for Clinicians [Internet]. 2021 [cited 2021 May 25]. Available from: https://checkout.moorfields.nhs.uk/product?catalog=2201DSOP

30. Moorfields education. Data science for clinicians: Intensive course in R [Internet]. 2021 [cited 2021 May 25]. Available from: https://checkout.moorfields.nhs.uk/product?catalog=2202DSOP

31. Foot B, MacEwen C. Surveillance of sight loss due to delay in ophthalmic treatment or review: frequency, cause and outcome. Eye. 2017 May 27;31(5):771–5.

32. Jayaram H, Strouthidis NG, Gazzard G. The COVID-19 pandemic will redefine the future delivery of glaucoma care. Eye 2020 347. 2020 May 13;34(7):1203–5.

33. Banes MJ, Culham LE, Bunce C, Xing W, Viswanathan A, Garway-Heath D. Agreement between optometrists and ophthalmologists on clinical management decisions for patients with glaucoma. Br J Ophthalmol. 2006 May 1;90(5):579–85.

34. Marks JR, Harding AK, Harper RA, Williams E, Haque S, Spencer AF, et al. Agreement between specially trained and accredited optometrists and glaucoma specialist consultant ophthalmologists in their management of glaucoma patients. Eye. 2012 Jun 13;26(6):853–61.

35. Ho S, A Vernon S. Decision making in chronic glaucoma - optometrists vs ophthalmologists in

a shared care service. Ophthalmic Physiol Opt. 2011 Mar;31(2):168–73.

36. Clinical Council for Eye Health Commissioning. Priorities for Delivering the NHS Long Term Plan for Eye Health. 2019.

37. AOP. Electronic Eye Care Referral System [Internet]. 2021 [cited 2021 Dec 20]. Available from: https://www.aop.org.uk/our-voice/media-centre/press-releases/2021/11/08/electronic-eye-care-referral-system

38. Department of Health. The NHS Plan: A plan for investment A plan for reform. 2000.

39. Crown J, Review team for the Prescribing supply and administration of medicines. Review of prescribing, Supply & Administration of Medicines: Final Report. 1999.

40. MHRA. Drug Safety Update. Vol. 1. 2007.

41. Loffler G, Henderson R, Bolland S, Gordon GE. Therapeutic prescribing for optometrists: an initial perspective. Optom Pract. 2011;12(3):87–98.

42. Heng JS, Wormald R, Khaw PT. Geographical variation in glaucoma prescribing trends in England 2008-2012: an observational ecological study. BMJ Open. 2016;6(5):e010429.

43. McAllum PJ, McGhee CN. Prescribing trends in infectious keratitis: a survey of New Zealand ophthalmologists. Clin Exp Ophthalmol. 2003 Dec 1;31(6):496–504.

44. Cartwright D. Personal communication.

45. Bornmann L, Haunschild R, Mutz R. Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases. Humanit Soc Sci Commun 2021 81. 2021 Oct 7;8(1):1–15.

46. Sadeghi-Bazargani H, Tabrizi JS, Azami-Aghdash S. Barriers to evidence-based medicine: a systematic review. J Eval Clin Pract. 2014 Dec 1;20(6):793–802.

47. Alnahedh T, Suttle CM, Alabdelmoneam M, Jalbert I. Optometrists show rudimentary understanding of evidence-based practice but are ready to embrace it: can barriers be overcome? Clin Exp Optom. 2015 May 1;98(3):263–72.

48. Toomey M, Gyawali R, Stapleton F, Ho KC, Keay L, Jalbert I. Facilitators and barriers to the delivery of eye care by optometrists: a systematic review using the theoretical domains framework. Ophthalmic Physiol Opt. 2021 Jul 1;41(4):782–97.

49. Woolf SH. The Meaning of Translational Research and Why It Matters. JAMA. 2008 Jan 9;299(2):211–3.

50. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The Quality of Health Care Delivered to Adults in the United States. http://dx.doi.org/101056/NEJMsa022615. 2009 Oct 7;348(26):2635–45.

51. Runciman WB, Hunt TD, Hannaford NA, Hibbert PD, Westbrook JI, Coiera EW, et al. CareTrack: assessing the appropriateness of health care delivery in Australia. Med J Aust. 2012 Jul 1;197(2):100–5.

52. Ho KC, Stapleton F, Wiles L, Hibbert P, White A, Jalbert I. iCareTrack: measuring the appropriateness of eyecare delivery in Australia. Ophthalmic Physiol Opt. 2020 Jul 1;40(4):433–41.

53. Toomey M, Jalbert I. Knowledge translation for the everyday optometrist. Clin Exp Optom. 2021;104(7):744–55.

54. Michie S, van Stralen MM, West R. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. Implement Sci. 2011 Apr 23;6(1):42.

55. Needle JJ, Petchey R, Lawrenson JG. A survey of the scope of therapeutic practice by UK optometrists and their attitudes to an extended prescribing role. Ophthalmic Physiol Opt. 2008 May 1;28(3):193–203.

56. Spillane D, Courtenay M, Chater A, Family H, Whitaker A, Acton JH. Factors influencing the prescribing behaviour of independent prescriber optometrists: a qualitative study using the Theoretical Domains Framework. Ophthalmic Physiol Opt. 2021 Mar 1;41(2):301–15.

57. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med. 2007 Aug 21;147(4):224–33.

58. Millard T, Synnot A, Elliott J, Green S, McDonald S, Turner T. Feasibility and acceptability of living systematic reviews: results from a mixed-methods evaluation. Syst Rev. 2019 Dec 14;8(1).

59. Haynes RB, Sackett DL, Gray JMA, Cook DJ, Guyatt GH. Transferring Evidence from Research into Practice: The Role of Clinical Care Research Evidence in Clinical Decisions. Evidence-Based Med (EBM Notebook). 1996;1(7):196–7.

60. So you want to be an Ophthalmologist? A short guide to a career in ophthalmology in the UK. 2016.

61. Apple Inc. Apple Siri [Internet]. 2022. Available from: https://www.apple.com/uk/siri/

62. Google LLC. Google Home [Internet]. 2022. Available from: https://apps.apple.com/gb/app/google-home/id680819774

63. Amazon.com Inc. Amazon Alexa [Internet]. 2022. Available from: https://www.amazon.co.uk/b?node=14100223031

64. Polson N, Scott J. AIQ: How Artificial Intelligence Works and How We Can Harness Its Power For a Better World. London: Bantam Press; 2018.

65. Bertsch-McGrayne S. The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines & Emerged Triumphant From Two Centuries of Controversy. paperback. New Haven & London: Yale University Press; 2011.

66. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017 Dec;42:60–88.

67. Catania LJ, Nicolitz E. Artificial Intelligence and Its Applications in Vision and Eye Care. Adv Ophthalmol Optom. 2018 Aug 24;3(1):21–38.

68. Hilbert M, Lopez P. The World's Technological Capacity to Store, Communicate, and Compute Information. Sci . 2011 Apr 1;332(6025):60–5.

69. Health Education England. Topol review [Internet]. 2018 [cited 2018 Dec 2]. Available from: https://www.hee.nhs.uk/our-work/topol-review

70. Big data | Definition of big data in English by Oxford Dictionaries [Internet]. Online Oxford Dicionary. 2019 [cited 2019 Apr 10]. Available from: https://en.oxforddictionaries.com/definition/big_data

71. Wong ZSY, Zhou J, Zhang Q. Artificial Intelligence for infectious disease Big Data Analytics.

Infect Dis Heal. 2019 Feb 1;24(1):44–8.

72. Gordon K. What is Big Data? ITNOW. 2013 Sep 1;55(3):12–3.

73. Snijders C, Matzat U, Reips U-D. "Big Data": Big Gaps of Knowledge in the Field of Internet Science. Int J Internet Sci. 2012;2012(1):1–5.

74. De Mauro A, Greco M, Grimaldi M. A formal definition of Big Data based on its essential features Article information. Libr Rev. 2016;65(3):122–35.

75. Boyd danah, Crawford K. CRITICAL QUESTIONS FOR BIG DATA. Information, Commun Soc. 2012 Jun;15(5):662–79.

76. Sudarshan VK, Koh JEW, Acharya UR, Tan JH, Mookiah MRK, Chua CK, et al. Evaluation of Evaporative Dry Eye Disease Using Thermal Images of Ocular Surface Regions with DWT and Gabor Transform. In: Application of Infrared to Biomedical Sciences. Springer, Singapore; 2017. p. 359–75.

77. Long E, Lin H, Liu Z, Wu X, Wang L, Jiang J, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. Nat Biomed Eng. 2017 Jan 30;1(2):0024.

78. González-López JJ, García-Aparicio ÁM, Sánchez-Ponce D, Muñoz-Sanz N, Fernandez-Ledo N, Beneyto P, et al. Development and validation of a Bayesian network for the differential diagnosis of anterior uveitis. Eye. 2016 Jun 8;30(6):865–72.

79. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA. 2016 Dec 13;316(22):2402.

80. Vichnesh J, Oh SL, Wei JKE, Ciaccio EJ, Chua Chua K, Tong L, et al. Thoughts concerning the application of thermogram images for automated diagnosis of dry eye – A review. Infrared Phys Technol. 2020;106(May).

81. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med. 2018 Aug 13;1.

82. Nuzzi R, Brusasco L. State of the art of robotic surgery related to vision: brain and eye applications of newly available devices. Eye Brain. 2018;10:13–24.

83. Edwards TL, Xue K, Meenink HCM, Beelen MJ, Naus GJL, Simunovic MP, et al. First-in-human study of the safety and viability of intraocular robotic surgery. Nat Biomed Eng. 2018 Sep 18;2(9):649–56.

84. Situmorang BH, Setiawan MP, Tosida ET. Decision support system for determining the contact lens for refractive errors patients with classification ID3. IOP Conf Ser Mater Sci Eng. 2017 Jan;166(1):012021.

85. Raghavendra U, Bhandary S V, Gudigar A, Rajendra Acharya U. Novel expert system for glaucoma identification using non-parametric spatial envelope energy spectrum with fundus images. Biocybern Biomed Eng. 2018;38(1):170–80.

86. Bayes T. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. Philos Trans R Soc London. 1763 Jan;53:370–418.

87. Stingler SM. Laplace's 1774 Memoir on Inverse Probability. Stat Sci. 1986;1(3):359–63.

88. Streiner DL. Clinical Medicine and the Legacy of the Reverend Bayes. Int J Clin Pract. 2019 Feb 15;73(4):e13323.

89. Neyman J. Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. Philos Trans R Soc A Math Phys Eng Sci. 1937 Aug 30;236(767):333–80.

90. Aspinall P, Hill AR. Clinical Inferences and Decisions—I. Diagnosis and Bayes' Theorem. Ophthalmic Physiol Opt. 1983 Jul;3(3):295–304.

91. Thomas R, Mengersen K, Parikh RS, Walland MJ, Muliyil J. Enter the reverend: introduction to and application of Bayes' theorem in clinical ophthalmology. Clin Experiment Ophthalmol. 2011 Dec;39(9):865–70.

92. Zhang H. Exploring Conditions For The Optimality of Naive Bayes. Int J Pattern Recognit Artif Intell. 2005 Mar 21;19(02):183–98.

93. Charniak E. The Bayesian Basis of Common Sense Medical Diagnosis*. AAAI-83 Proc. 1983;70–3.

94. Fritz KJ, Polascikt MA, Pottsi AM. COMPUTER ASSISTED DIAGNOSIS FOR OPHTHALMOLOGY *. Compur Biol Med. 1978;8(3):223–8.

95. Singh AD, De Potter P, Fijal BA, Shields CL, Shields JA, Elston RC. Lifetime Prevalence of Uveal Melanoma in White Patients with Oculo(dermal) Melanocytosis. Ophthalmology. 1998;105(1):195–8.

96. Thomas R, Walland M, Thomas A, Mengersen K. Lowering of Intraocular Pressure After Phacoemulsification in Primary Open-Angle and Angle-Closure Glaucoma: A Bayesian Analysis. Asia-Pacific J Ophthalmol. 2016;5(1):79–84.

97. Sagar R V. Application of Naive Bayesian Sequential Analysis to Primary Care Optometry. Aston University; 2014.

98. Gurney JC. Application of Naive Bayesian Artificial Intelligence to Referral Refinement of Chronic Open Angle Glaucoma. Aston University; 2017.

99. Kohavi R. Scaling Up the Accuracy of Naive-Bayes Classifiers: a Decision-Tree Hybrid. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. 1996. p. 202–7.

100. West Kent CCG. Acute Primary Care Ophthalmology Service. 2014.

101. Silcock E, Hardy K. West Kent Population Health and Wellbeing Profile. 2015.

102. Prajapati B, Dunne MCM, Armstrong R. Sample size estimation and statistical power analyses. Optom Today. 2010;(7).

103. College of Optometrists. Clinical Management Guidelines [Internet]. [cited 2019 Jun 11]. Available from: https://www.college-optometrists.org/guidance/clinical-management-guidelines.html

104. Powers DMW. Evaluation: from Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. J Mach Learn Technol. 2011;2(1):37–63.

105. Kharrazi H, Lasser EC, Yasnoff WA, Loonsk J, Advani A, Lehmann HP, et al. A proposed national research and development agenda for population health informatics: summary recommendations from a national expert workshop. J Am Med Informatics Assoc. 2017 Jan 1;24(1):2–12.

106. Ross MK, Wei W, Ohno-Machado L. “Big data” and the electronic health record. Yearb Med Inform. 2014 Aug 15;9(1):97–104.

107. Wasserman RC. Electronic Medical Records (EMRs), Epidemiology, and Epistemology: Reflections on EMRs and Future Pediatric Clinical Research. Acad Pediatr. 2011;11(4):280–7.

108. Wright M-O, Fisher A, John M, Reynolds K, Peterson LR, Robicsek A. The electronic medical record as a tool for infection surveillance: Successful automation of device-days. Am J Infect Control. 2009 Jun;37(5):364–70.

109. Yasnoff WA, Overhage JM, Humphreys BL, LaVenture M. A National Agenda for Public Health Informatics: Summarized Recommendations from the 2001 AMIA Spring Congress. J Am Med Informatics Assoc. 2001 Nov 1;8(6):535–45.

110. Gardner W, Morton S, Byron SC, Tinoco A, Canan BD, Leonhart K, et al. Using Computer-Extracted Data from Electronic Health Records to Measure the Quality of Adolescent Well-Care. Health Serv Res. 2014 Aug 1;49(4):1226–48.

111. Sanders DS, Lattin DJ, Read-Brown S, Tu DC, Wilson DJ, Hwang TS, et al. Electronic health record systems in ophthalmology: impact on clinical documentation. Ophthalmology. 2013 Sep 1;120(9):1745–55.

112. Chan P, Thyparampil PJ, Chiang MF. Accuracy and Speed of Electronic Health Record Versus Paper-Based Ophthalmic Documentation Strategies. Am J Ophthalmol. 2013;156(1):165–72.

113. The Royal College of Ophthalmologists. Electronic Medical Records-Standards for UK Ophthalmology Services.

114. College of Optometrists. Clinical Management Guidelines [Internet]. 2018 [cited 2018 Oct 31]. Available from: https://www.college-optometrists.org/guidance/clinical-management-guidelines.html

115. Gurney JC, Ansari E, Harle D, O’Kane N, Sagar R V., Dunne MCM. Application of Bayes’ to the prediction of referral decisions made by specialist optometrists in relation to chronic open angle glaucoma. Eye. 2018 Jun 9;32(6):1074–8.

116. Cimino JJ. High-quality, standard, controlled healthcare terminologies come of age. Methods Inf Med. 2011;50(2):101–4.

117. World Health Organization. Classifications | ICD-10 online versions [Internet]. World Health Organization; 2018 [cited 2018 Nov 9]. Available from: http://www.who.int/classifications/icd/icdonlineversions/en/

118. College of Optometrists. Clinical Management Guidelines: Ectropion [Internet]. 2018 [cited 2018 Oct 31]. Available from: https://www.college-optometrists.org/guidance/clinical-management-guidelines/ectropion.html

119. College of Optometrists. Clinical Management Guidelines: Entropion [Internet]. 2018 [cited 2018 Oct 31]. Available from: https://www.college-optometrists.org/guidance/clinical-management-guidelines/entropion.html

120. College of Optometrists. Clinical Management Guidelines: Scleritis [Internet]. 2018 [cited 2018 Oct 31]. Available from: https://www.college-optometrists.org/guidance/clinical-management-guidelines/scleritis.html

121. College of Optometrists. College of Optometrists’ Formulary. 2016;

122. Microsoft Corporation. Excel specifications and limits - Excel [Internet]. 2018 [cited 2018 Nov

11]. Available from: https://support.office.com/en-us/article/excel-specifications-and-limits-1672b34d-7043-467e-8e27-269d656771c3

123. Korda N. How humanised machine learning is augmenting human intelligence [Internet]. 2019. Available from: https://www.openaccessgovernment.org/humanised-machine-learning/64178/

124. Firth N. Create a searchable drop down list in Excel - YouTube [Internet]. 2012 [cited 2018 Nov 13]. Available from: https://www.youtube.com/watch?v=vkPoViUhkxU

125. Kellermann AL, Jones SS. What It Will Take To Achieve The As-Yet-Unfulfilled Promises Of Health Information Technology. Health Aff. 2013 Jan;32(1):63–8.

126. Quinlan JR. Induction of Decision Trees. Mach Learn. 1986;1(1):81–106.

127. Skrøvseth SO, Augestad KM, Ebadollahi S. Data-driven approach for assessing utility of medical tests using electronic medical records. J Biomed Inform. 2015 Feb 1;53:270–6.

128. BNF online. MedicinesComplete — Search for hypromellose [Internet]. 2018 [cited 2018 Nov 23]. Available from: https://www.medicinescomplete.com/#/search/bnf/hypromellose?offset=0

129. emc. Electronic Medicines Compendium (eMC) [Internet]. 2018 [cited 2018 Nov 23]. Available from: https://www.medicines.org.uk/emc/

130. West Kent CCG. West Kent Interface Formulary [Internet]. 2018 [cited 2018 Nov 23]. Available from: http://www.formularywkccgmtw.co.uk/

131. College of Optometrists'. Optometrists Formulary. 2011.

132. Meskell P, Murphy K, Shaw DG, Casey D. Insights into the use and complexities of the Policy Delphi technique. Nurse Res. 2014 Jan;21(3):32–9.

133. Karamitri I, Bellali T, Galanis P, Kaitelidou D. The accessibility of vulnerable groups to health services in Greece: a Delphi study on the perceptions of health professionals. Int J Health Plann Manage. 2013 Jan 1;28(1):35–47.

134. Landeta J. Current validity of the Delphi method in social sciences. Technol Forecast Soc Change. 2006 Jun 1;73(5):467–82.

135. Davey CJ, Slade S V., Shickle D. A proposed minimum data set for international primary care optometry: a modified Delphi study. Ophthalmic Physiol Opt. 2017 Jul 1;37(4):428–39.

136. Lu W, Tong Y, Yu Y, Xing Y, Chen C, Shen Y. Applications of Artificial Intelligence in Ophthalmology: General Overview. J Ophthalmol. 2018;2018.

137. Han J, Kamber M, Pei J. Data Mining. Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems). 3rd ed. London: Elsevier Inc.; 2011.

138. Caixinha M, Nunes S. Machine Learning Techniques in Clinical Vision Sciences. Curr Eye Res. 2017 Jan 2;42(1):1–15.

139. Elliott DB. Clinical Procedures in Primary Eyecare. fourth. Elsevier Limited; 2014. 4–7 p.

140. Fenner F, Henderson DA, Arita I, Jezek Z, Ladnye ID. Smallpox and its Eradication. 1988.

141. Villejoubert G, Mandel DR. The inverse fallacy: An account of deviations from Bayes's theorem and the additivity principle. Mem Cogn. 2002;30(2):171–8.

142. Bramwell R, West H, Salmon P. Health professionals' and service users' interpretation of

screening test results: Experimental study. Br Med J. 2006 Aug 5;333(284).

143. Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. BMJ. 2002 Apr 6;324(7341):824–6.

144. Moreira J, Bisoffi Z, Narváez A, Van Den Ende J. Bayesian clinical reasoning: Does intuitive estimation of likelihood ratios on an ordinal scale outperform estimation of sensitivities and specificities? J Eval Clin Pract. 2008 Oct;14(5):934–40.

145. Lyman GH, Balducci L. Overestimation of test effects in clinical judgment. J Cancer Educ. 1993;8(4):297–307.

146. Chernushkin K, Loewen P, De Lemos J, Aulakh A, Jung J, Dahri K. Diagnostic Reasoning by Hospital Pharmacists: assessments of attitudes, knowledge and skills. Can J Hosp Pharm. 2012;(65):258–64.

147. Bergus G, Vogelgesang S, Tansey J, Franklin E, Feld R. Appraising and applying evidence about a diagnostic test during a performance-based assessment. BMC Med Educ. 2004;4(20).

148. Whiting PF, Davenport C, Jameson C, Burke M, Sterne JAC, Hyde C, et al. How well do health professionals interpret diagnostic information? A systematic review. BMJ Open. 2015;5(e008155).

149. Agoritsas T, Courvoisier DS, Combescure C, Deom M, Perneger T V. Does prevalence matter to physicians in estimating post-test probability of disease? A randomized trial. J Gen Intern Med. 2011 Apr;26(4):373–8.

150. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. Lancet. 2005 Apr 23;365(April):1500–5.

151. Gill CJ, Sabin L, Schmid CH. Why clinicians are natural bayesians. Br Med J. 2005;330(May):1080–3.

152. Parikh R, Parikh S, Arun E, Thomas R. Likelihood ratios: clinical application in day-to-day practice. Indian J Ophthalmol. 2009;57(3):217–21.

153. Hayden SR, Brown MD. Likelihood Ratio: A Powerful Tool for Incorporating the Results of a Diagnostic Test Into Clinical Decisionmaking. 1999;575–80.

154. Jaeschke R, Guyatt G, Sackett DL, Bass E, Brill Edwards P, Browman G, et al. Users' Guides to the Medical Literature: III. How to Use an Article About a Diagnostic Test A. Are the Results of the Study Valid? JAMA J Am Med Assoc. 1994 Feb 2;271(5):389–91.

155. McGee S. Simplifying likelihood ratios. J Gen Intern Med. 2002;17(8):647–50.

156. Taylor JR. An Introduction to Error Analysis. 2nd ed. California: University Science Books; 1997.

157. Barber D. Bayesian Reasoning and Machine Learning. Cambridge: Cambridge University Press; 2012. 311–313 p.

158. Sheen NJL, Fone D, Phillips CJ, Sparrow JM, Pointer JS, Wild JM. Novel optometrist-led all Wales primary eye-care services: evaluation of a prospective case series. Br J Ophthalmol. 2009 Apr 1;93(4):435–8.

159. McCracken M. Auditing an Acute Eye Pathway. Optom Today. 2013;(3):34–7.

160. Greenwood L. Auditing of a Reerral Refinement Service in Hull. Optom Today. 2013;(March):24–6.

161. Cottier K. An audit of the Primary Eye-care Referral Scheme (PEARS) within NHS Bromley Clinical Commisioning Group (CCG). Optom Pract. 2015;16(1):21–32.

162. MacEwen C, Davis A, Chang L. Ophthalmology GIRFT Programme National Specialty Report. 2019 Dec.

163. Parkins DJ, Curran R, Pooley JE, Ryan B. The developing role of optometrists as part of the NHS primary care team. Optom Pract. 2014;15(4):177–84.

164. Jones L, Downie LE, Korb D, Benitez-del-Castillo JM, Dana R, Deng SX, et al. TFOS DEWS II Management and Therapy Report. Vol. 15, Ocular Surface. Elsevier Inc.; 2017. p. 575–628.

165. Pucker AD, Ng SM, Nichols JJ. Over the counter (OTC) artificial tear drops for dry eye syndrome. Cochrane Database Syst Rev. 2016 Feb 23;2016(2).

166. British Medical Association, Royal Pharmaceutical Society. BNF 76. 76th ed. London: BMJ Group and Pharmaceutical Press; 2019. 1123–1158 p.

167. College of Optometrists. How to use the Clinical Management Guidelines [Internet]. 2018 [cited 2018 Aug 16]. Available from: https://www.college-optometrists.org/guidance/clinical-management-guidelines/what-are-clinical-management-guidelines.html

168. General Optical Council. A Handbook for Optometry Specialist Registration in Therapeutic Prescribing. 2008.

169. Chawla A, Patel I, Yuen C, Fenerty C. Patterns of adherence to NICE Glaucoma Guidance in two different service delivery models. Eye. 2012 Aug 31;26(11):1412–7.

170. Ho KC, Stapleton F, Wiles L, Hibbert P, Alkhawajah S, White A, et al. Systematic review of the appropriateness of eye care delivery in eye care practice. BMC Health Serv Res. 2019 Sep 6;19(1).

171. The College of Optometrists. Dry Eye (Keratoconjunctivitis Sicca, KCS) [Internet]. 2015 [cited 2016 Dec 2]. Available from: http://www.college-optometrists.org/en/utilities/document-summary.cfm/docid/4111ACA3-058B-45B5-ADDFC10FF163263D

172. Wang MY, Asanad S, Asanad K, Karanjia R, Sadun AA. Value of medical history in ophthalmology: A study of diagnostic accuracy. J Curr Ophthalmol. 2018 Dec 1;30(4):359–64.

173. Peterson MC, Holbrook JH, Von Hales D, Smith NL, Staker L V. Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. West J Med. 1992;156(2):163–5.

174. Diez D, Cetinkaya-Rundel M, Barr, Christopher D. OpenIntro Statistics. Fourth. openintro.org; 2019. 181–182 p.

175. Janson B. Hyperemia, Conjunctival. In: Schmidt-Erfurth U, Kohnen T, editors. Encyclopedia of Ophthalmology (online). Living Edition. Berlin: Springer Berlin Heidelberg; 2014. p. 1–2.

176. Singh Bhinder G, Singh Bhinder H. Reflex epiphora in patients with dry eye symptoms: role of variable time Schirmer-1 test. Eur J Ophthalmol. 2005 Jul 18;15(4):429–33.

177. Bagheri N, Wajda BN. The Wills Eye Manual: Office and emergency room diagnosis and treatment of eye disease. 7th ed. Friedberg MA, Rapuano CJ, editors. Philadelphia: Wolters Kluwer -- Medknow Publications; 2017. 66–67 p.

178. College of Optometrists. CMG: Herpes Zoster Ophthalmicus (HZO) [Internet]. Clinical Management Guidelines. 2019 [cited 2019 Dec 14]. Available from: https://www.college-

optometrists.org/guidance/clinical-management-guidelines/herpes-zoster-ophthalmicus-hzo-.html

179. Baring DEC, Hamilton OJ. An evidence based review of periorbital cellulitis. Clin Otolaryngol. 2011;36(1):57–64.

180. Kikkawa DO, Heinz GW, Martin RT, Nunery WN, Eiseman AS. Orbital Cellulitis and Abscess Secondary to Dacryocystitis. Arch Ophthalmol. 2002;1096–9.

181. Kanski JJ. Clinical Ophthalmology: A Systematic Approach. fifth edit. Butterworth-Heinemann; 2003.

182. Lee S, Yen MT. Management of preseptal and orbital cellulitis. Vol. 25, Saudi Journal of Ophthalmology. 2011. p. 21–9.

183. College of Optometrists. Clinical Management Guidelines: Uveitis (anterior) [Internet]. 2019 [cited 2020 Jan 1]. Available from: https://www.college-optometrists.org/guidance/clinical-management-guidelines/uveitis-anterior-acute-and-recurrent-.html

184. College of Optometrists. Clinical management guidelines: Blepharitis (Lid Margin Disease) [Internet]. 2018 [cited 2020 Jan 1]. Available from: https://www.college-optometrists.org/guidance/clinical-management-guidelines/blepharitis-lid-margin-disease.html

185. Lindsley K, Matsumura S, Hatef E, Akpek EK. Interventions for chronic blepharitis. In: Lindsley K, editor. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd; 2012.

186. Shulman DG, Sargent JB, Stewart RH, Mester U. Comparative evaluation of the short-term bactericidal potential of a steroid-antibiotic combination versus steroid in the treatment of chronic bacterial blepharitis and conjunctivitis. Eur J Ophthalmol. 1996;6(4):361–7.

187. College of optometrists. Clinical Management Guidelines: Conjunctivitis (viral, non-herpetic) [Internet]. 2019 [cited 2020 Jan 1]. Available from: https://www.college-optometrists.org/guidance/clinical-management-guidelines/conjunctivitis-viral-non-herpetic-.html

188. Wilkins MR, Khan S, Bunce C, Khawaja A, Siriwardena D, Larkin DFP. A randomised placebo-controlled trial of topical steroid in presumed viral conjunctivitis. Br J Ophthalmol. 2011 Sep;95(9):1299–303.

189. McDonald EM, De Kock J, Ram FSF. Antivirals for management of herpes zoster including ophthalmicus: A systematic review of high-quality randomized controlled trials. Antivir Ther. 2012;17(2):255–64.

190. College of Optometrists. Clinical Management Guidelines: Herpes Zoster Ophthalmicus (HZO) [Internet]. 2017 [cited 2020 Jan 2]. Available from: https://www.college-optometrists.org/guidance/clinical-management-guidelines/herpes-zoster-ophthalmicus-hzo-.html

191. Gnann Jr. JW. Chapter 65 Antiviral therapy of varicella-zoster virus infections. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.

192. Nakajima H, Hara M, Morita A, Kamei S. Neurologic Complications of Varicella-Zoster Virus Infection. In: Human Herpesvirus Infection - Biological Features, Transmission, Symptoms,

Diagnosis and Treatment [Working Title]. IntechOpen; 2019.

193. Nandi N, Morris P, Blundell A. Herpes zoster ophthalmicus progressing to encephalitis: Beware pain preceding the rash. BMJ Case Rep. 2013 Jul 12;

194. College of Optometrists. Clinical Management Guidelines: Cellulitis, preseptal and orbital [Internet]. 2017 [cited 2020 Jan 2]. Available from: https://www.college-optometrists.org/guidance/clinical-management-guidelines/cellulitis-preseptal-and-orbital.html

195. College of Optometrists. Clinical management guidelines: Conjunctivitis medicamentosa (also Dermatoconjunctivitis medicamentosa) [Internet]. 2019 [cited 2020 Jan 2]. Available from: https://www.college-optometrists.org/guidance/clinical-management-guidelines/conjunctivitis-medicamentosa.html

196. College of Optometrists'. How to use the Clinical Management Guidelines [Internet]. 2017 [cited 2017 Apr 23]. Available from: https://www.college-optometrists.org/guidance/clinical-management-guidelines/what-are-clinical-management-guidelines.html

197. Todd D, Bartlett H, Thampy R, Dhawahir-Scala F, Wilson H, Tromans C. Agreement in clinical decision-making between independent prescribing optometrists and consultant ophthalmologists in an emergency eye department. Eye. 2020 Mar 17;1–11.

198. Harper RA, Dhawahir-Scala F, Wilson H, Gunn PJG, Jinkinson M, Pretty IA, et al. Development and implementation of a Greater Manchester COVID19 Urgent Eyecare Service. Eye. 2020 Jun 29;1–4.

199. Pancholi BR, Dunne MCM. Virtual Patient Instruction and Self-Assessment Accuracy in Optometry Students | The Journal of Optometric Education. Optom Educ. 2018;43(2):1–10.

200. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology. 2010 Jan;21(1):128.

201. Hickey GL, Grant SW, Dunning J, Siepe M. Statistical primer: sample size and power calculations—why, when and how? Eur J Cardio-Thoracic Surg. 2018 Jul 1;54(1):4–9.

202. The College of Optometrists. Our response to the Royal College of Ophthalmologist's guidance on eye care services after COVID-19 [Internet]. 2020 [cited 2020 Jun 3]. Available from: https://www.college-optometrists.org/the-college/media-hub/news-listing/response-to-guidance-on-reopening-services-covid.html

203. Smith LI. A tutorial on Principal Components Analysis. 2002 Dec.

204. Consejo A, Melcer T, Rozema JJ. Introduction to Machine Learning for Ophthalmologists. Semin Ophthalmol. 2019 Jan 2;34(1):19–41.

205. Hayashi Y. The Right Direction Needed to Develop White-Box Deep Learning in Radiology, Pathology, and Ophthalmology: A Short Review. Front Robot AI. 2019;6(24).

206. Shah R, Edgar DF, Evans BJW. A comparison of standardised patients, record abstraction and clinical vignettes for the purpose of measuring clinical practice. Ophthalmic Physiol Opt. 2010 May 1;30(3):209–24.

207. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. J Clin Epidemiol. 2014;67(3):267.

208. Ahmed K. Analysis of Data Mining Tools for Disease Prediction. J Pharm Sci Res.

2017;9(10):1886–8.

209. Chen X, Ye Y, Williams G, Xu X. A survey of open source data mining systems. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag; 2007. p. 3–14.

210. Singh RP, Hom GL, Abramoff MD, Campbell JP, Chiang MF. Current Challenges and Barriers to Real-World Artificial Intelligence Adoption for the Healthcare System, Provider, and the Patient. Transl Vis Sci Technol. 2020 Jan 28;9(2):45–45.

211. Paranjape K, Schinkel M, Panday RN, Car J, Nanayakkara P. Introducing Artificial Intelligence Training in Medical Education. JMIR Med Educ. 2019 Jul 1;5(2).