Optimising the cost-effectiveness of risk-based screening for diabetic retinopathy

Christopher James Sampson

ORCID: 0000-0001-9470-2369

School of Medicine
University of Nottingham

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

April 2020

Declaration

Except where acknowledged, I declare that this thesis is the result of my own work, which was undertaken during my period of registration for this degree at the University of Nottingham.

Christopher James Sampson

Word count: 75,914

For my dad

Acknowledgements and thanks

Thanks to Professors Marilyn James and Dave Whynes for their guidance and their willingness to entertain my ideas through the years. I am also grateful to every member of the ISDR team, past and present. While the work presented — and the liability for any errors — is very much my own, all of these people are represented in the pluralis modestiae ("we") used throughout this thesis. The study could not have been completed without their hard work on linked studies and their support throughout my own. Thank you to all of my colleagues and office mates — from CLAHRC-NDL, the IMH, the School of Sociology and Social Policy, the Division of Rehabilitation and Ageing, the Department of Eye and Vision Science in Liverpool, and finally at the OHE — for egging me on. Thanks to the wider health economics community in the UK and beyond, which has provided encouragement and inspiration along the way. Thanks especially to HESG members for their creative attempts to get me to finish early, from contracts on napkins to promises of printing this thesis on a scarf. Thanks to Alastair Canaway for his spiritual guidance. Thanks to Marilyn and Dave for their patience. Thanks to Matt Sutton for his impatience. Thanks to my family for their support and their patience. And thanks to Rosie, for being the most patient of all.

Disclaimers

This thesis presents independent research funded in part by the National Institute for Health Research (NIHR) under the Programme Grants for Applied Research programme (RP-PG-1210-12016). At the time of submission, Chris Sampson was an employee of the Office of Health Economics (OHE). The OHE receives funding from various organisations including The Association of the British Pharmaceutical Industry, the EuroQol Research Foundation, the NIHR, and UK Research and Innovation. The views expressed are those of the author and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care, or the OHE.

Dissemination of work

Part of Chapter 1 was published in *Diabetic Medicine* [1], and an earlier version of material from Chapter 8 was published as a working paper [2]. A protocol for the systematic review and meta-analysis reported in Chapter 4 was published in *Systematic Reviews* before commencement of the review [3].

An earlier version of Chapter 4 was discussed at the Health Economists' Study Group (HESG) winter 2017 meeting. Findings from Chapters 3 and 5 were presented at the European Association for the Study of Diabetes Eye Complication Study Group (EASDec) 26th Annual Meeting [4, 5]. Findings from Chapter 5 were presented at the Association for Research in Vision & Ophthalmology (ARVO) 2019 Annual Meeting [6]. Parts of Chapter 6 were presented at ARVO 2016 [7]. An earlier version of Chapter 7 was presented at the Third European Health Economics Association (EuHEA) PhD Student-Supervisor and Early Career Researcher Conference, the EASDec 25th Annual Meeting [8], and the Health Technology Assessment International (HTAi) 2019 Annual Meeting [9]. Parts of Chapters 8 and 9 were discussed at HESG winter 2016 and at the first Personalised Medicine and Resource Allocation conference at the University of Oxford.

Work from this thesis was also presented at seminars at the University of Sheffield, the University of Leeds, and the University of Nottingham.

Abstract

Publicly provided health screening programmes tend to offer standardised screening for a fixed eligible population. Recently, the development of risk calculation engines has introduced the potential for the stratification of screening based on individuals' risks of disease onset. This possibility raises practical, methodological, and ethical challenges. To date, no such programme has been the subject of an economic evaluation. In this thesis we present reason and basis for the allocation of screening based on individual risk.

The research is conducted in the context of screening for diabetic eye disease in the UK. Diabetic retinopathy is a common complication of diabetes that can lead to blindness, substantial detriments to quality of life, and significant health care resource use. Our study is linked to a programme of research that includes a cohort study and randomised controlled trial in the city of Liverpool. We review and further develop the evidence base to inform the evaluation of a risk-based screening programme for diabetic eye disease. Specifically, we generate new evidence on the costs and health outcomes associated with the screening and treatment of diabetic retinopathy.

We report on a cross-sectional study of health-related quality of life for people attending screening for diabetic retinopathy and find that people with pre-symptomatic disease tend to report poorer quality of life than people with no disease, with EQ-5D-5L index values of 0.733 on average compared with 0.787 for people with no disease. A meta-analysis of published health state utility values for diabetic eye disease shows a negative impact on health-related quality of life before progression to blindness. Our meta-regression found a utility index decrement of 0.024 for people with proliferative retinopathy.

The costs of screening are low at the individual level, estimated to be £32.03 in our costing study. But the overall budget impact of changes in the frequency of screening can be significant. We analyse a large data set of hospital and community screening activity to identify key treatment pathways for diabetic eye disease. We find that these have changed in recent years, with the introduction of more expensive interventions.

The evidence generated by our work is used to inform the development of a decision analytic model. The model is designed to estimate the cost-effectiveness of risk-based screening for diabetic eye disease, compared with current practice. We find that risk-based screening is likely to be more cost-effective than standardised screening programmes.

Evaluating a programme that allocates screening according to individuals' levels of risk raises theoretical and ethical challenges. To this end, we develop a simple framework for individualised cost-effectiveness analysis that can be used to inform the design of a risk-based screening programme. We also explore the

ethics of risk-based screening, developing the notion of screening need as distinct from treatment need.

Risk-based screening is likely to be cost-effective in the context of diabetic eye disease. The evidence presented in this thesis can be used to support the evaluation of new programmes, which can be designed in order to optimise cost-effectiveness using the methods that we describe. Such an approach is consistent with equitable policy objectives.

Contents

List of Figures

List of Tables

List of abbreviations

AAO . . . . . American Academy of Ophthalmology

BCVA . . . . best-corrected visual acuity

BR . . . . . background retinopathy

CAL . . . . . central acuity loss

CATE . . . . conditional average treatment effects

CCG . . . . . clinical commissioning group

CEA . . . . . cost-effectiveness analysis

CORE-DM . Center for Outcomes Research Diabetes Model

CRAG . . . . Clinical Resource and Audit Group

CSMO . . . . clinically significant macular oedema

CST . . . . . cohort state transition model

CUA . . . . . cost-utility analysis

DCCT . . . . Diabetes Control and Complication Trial

DES . . . . . discrete event simulation

DMO . . . . . diabetic macular oedema

DR . . . . . diabetic retinopathy

DRCRnet . . Diabetic Retinopathy Clinical Research Network

DRS . . . . . Diabetic Retinopathy Study

DRVS . . . . Diabetic Retinopathy Vitrectomy Study

DT . . . . . decision tree

eCRF . . . . . electronic case report form

EDIC . . . . . Epidemiology of Diabetes Interventions and Complications

EQ VAS . . . EuroQol visual analogue scale

ETDRS . . . Early Treatment Diabetic Retinopathy Study

FLP . . . . . focal laser therapy

GLP . . . . . grid laser therapy

HES . . . . . hospital eye service

HMO . . . . . health maintenance organisation

HRG . . . . . Healthcare Resource Group

HRQoL . . . health-related quality of life

HSUV . . . . health state utility value

HTA . . . . . health technology assessment

HTBS . . . . Health Technology Board for Scotland

HUI3 . . . . . Health Utilities Index Mark 3

iCEA . . . . . individualised cost-effectiveness analysis

ICER . . . . . incremental cost-effectiveness ratio

IMIB TOM . Institute for Medical Informatics and Biostatistics Tools for Outcomes Modeling

IRR . . . . . incidence rate ratio

ISDR . . . . . Individualised Screening for Diabetic Retinopathy study

ISM . . . . . individual sampling model

KPI . . . . . key performance indicator

LDES . . . . . Liverpool Diabetic Eye Study

LDESP . . . . Liverpool Diabetic Eye Screening Programme

NA . . . . . not applicable

NHS . . . . . National Health Service

NDESP . . . NHS Diabetic Eye Screening Programme

NHS . . . . . National Health Service

NICE . . . . . National Institute for Health and Care Excellence

NIHR . . . . National Institute for Health Research

NMB . . . . . net monetary benefit

NPDR . . . . non-proliferative diabetic retinopathy

NR . . . . . not reported

NSC . . . . . National Screening Committee

OCT . . . . . optical coherence tomography

PBM . . . . . preference-based measure

PDR . . . . . proliferative diabetic retinopathy

PeT . . . . . person-centred treatment effects

PHE . . . . . Public Health England

PPI . . . . . patient and public involvement

PRISMA . . Preferred Reporting Items for Systematic reviews and Meta-Analyses

PRP . . . . . panretinal photocoagulation

PSA . . . . . probabilistic sensitivity analysis

PSS . . . . . personal social services

QALY . . . . quality-adjusted life year

QWB . . . . . Quality of Well Being Scale

QWB-SA . . Quality of Well Being Scale Self-Administered

RCE . . . . . risk calculation engine

RCT . . . . . randomised controlled trial

RLBUHT . . Royal Liverpool and Broadgreen University Hospitals NHS Trust

SD . . . . . standard deviation

SE . . . . . standard error

SG . . . . . standard gamble

SLB . . . . . slit lamp biomicroscopy

STDR . . . . sight-threatening diabetic retinopathy

SVI . . . . . severe visual impairment

SVL . . . . . severe vision loss

TTO . . . . . time trade-off

UK . . . . . United Kingdom of Great Britain and Northern Ireland

UKPDS . . . United Kingdom Prospective Diabetes Study

VA . . . . . visual acuity

VAS . . . . . visual analogue scale

VBA . . . . . Visual Basic for Applications

VEGF . . . . vascular endothelial growth factor

WESDR . . . Wisconsin Epidemiologic Study of Diabetic Retinopathy

Chapter 1

The decision problem

Summary

This chapter introduces the decision problem that the thesis seeks to address; namely, how to optimise the cost-effectiveness of risk-based screening for diabetic retinopathy. We briefly describe the pathology of diabetic eye disease and current practice in screening and treatment as it pertains to economic evaluation and resource allocation decisions. Key concepts and definitions are introduced in this chapter as they are to be interpreted and used throughout the thesis. We specify the relevance of patient heterogeneity to the evaluation of screening generally and to diabetic retinopathy specifically. The chapter introduces the notion of individual risk estimation and highlights some of the recent changes that have brought it to the fore. We outline the specific policy context in the UK and the ways in which we seek to inform policy decisions. The work is aligned with a National Institute for Health Research (NIHR) study — the ISDR study — which is described in this chapter.

1.1 Diabetic retinopathy

The number of people in the UK with diabetes is currently estimated to be around 4.5 million, with global prevalence expected to reach one person in every ten by 2040 [10]. Vision loss is one of many health complications associated with diabetes. The majority of people with diabetes will have some degree of eye disease within 20 years of diagnosis [11, 12]. Estimates suggest that around 1.5 million people in the UK [13], and almost 100 million worldwide [14], currently have diabetic retinopathy.

There are two types of diabetic eye disease: retinopathy and maculopathy. Retinopathy occurs when microaneurysms appear in the artery walls at the back of the eye. Subsequent to this, abnormal new blood vessels can form in a process known as neovascularisation. This stage of disease is commonly referred to as proliferative diabetic retinopathy (PDR) and haemorrhaging in the fragile new vessels can lead to blurred vision and dark spots. Retinopathy that affects the macula is referred to as maculopathy. Maculopathy can occur in the early stages of retinopathy, when blood vessels leak into the macular region, and can result in rapid loss of vision. The term 'diabetic retinopathy' (DR) is often used as a catch-all term for diabetic eye disease, and will be used as such throughout this thesis. Figures 1.1 and 1.2 characterise the effect of PDR on vision[1].

Disease progression in diabetic retinopathy is well understood, but is classified in a variety of ways. Often, these classifications relate to whether the disease is proliferative (PDR) or non-proliferative (NPDR). In the United States, the most common classification system has five or six levels (no DR/mild NPDR/moderate NPDR/severe NPDR/PDR/PDR with high-risk characteristics), as proposed

[1]Images courtesy of the National Eye Institute (CC0)

Figure 1.1: Normal vision

Figure 1.2: Vision with PDR

by the American Academy of Ophthalmology [15]. The National Health Service (NHS) Diabetic Eye Screening Programme (NDESP) in the UK uses four levels (no DR/background DR/pre-proliferative DR/PDR), categorised as R0, R1, R2, and R3. Recently, the NDESP recommended the disaggregation of R3 gradings into 'active PDR' (R3a) and 'stable PDR' (R3s), the latter corresponding to those who have received successful treatment. The Early Treatment Diabetic Retinopathy Study (ETDRS) used a more detailed grading system [16]. It is also possible to classify disease based on outcome or management pathways; the Liverpool Diabetic Eye Study (LDES), for example, uses levels relating to screening, referral, and treatment [17]. Other systems include the Scottish Diabetic Retinopathy Grading Scheme, the Royal College of Ophthalmologists grading system, and feature-specific grading.

The different levels defined by these grading systems can — to some extent — be mapped to one another. Table 1.1 shows how different levels from a number of key grading systems correspond. Maculopathy, which manifests as diabetic macular oedema (DMO), has fewer classifications, with the later stages described as clinically significant macular oedema (CSMO). The NDESP classifies gradings as either M0 or M1 to indicate the absence or presence of maculopathy.

The earliest stages of DR are asymptomatic and do not require treatment, though progression can be controlled with better diabetes management [18]. More advanced stages of disease are known as sight-threatening diabetic retinopathy (STDR) and require close monitoring by a hospital eye service. In many cases, treatment should be provided to prevent visual impairment. Common treatments include laser photocoagulation, injections into the eye, and surgery. It is important to identify individuals who need treatment before any vision is lost, as treatment cannot always reverse this process.

1.2 Screening

For some diseases, early treatment before symptoms develop can be beneficial. Screening is a means of identifying individuals in the early stages of disease, or those at increased risk of disease, who might benefit from early treatment or preventive management. Screening services are somewhat unusual in health care insofar as they involve intervention for an individual who is apparently in good health or, at least, free of the disease for which they are receiving health care. All screening interventions have the potential to cause harm and the possibility of no benefit. It is therefore necessary to identify those individuals for whom screening is likely to do more good than harm.

The extent to which a screening intervention can be deemed effective depends on the effectiveness of treatment or care following a positive screen event.

ETDRS	**NDESP**	**LDES**	**Narrative**	
10	R0	10	No retinopathy	
20/35	R1	20	Background DR	Mild NPDR
		30	Mild preproliferative DR	
43/47	R2	40	Moderate preproliferative	Moderate NPDR
53		50	Severe preproliferative DR	Severe NPDR
61	R3s	60	Stable treated DR	Mild PDR
65	R3a		PDR	Moderate PDR
71/75		70	PDR with high risk characteristics	High risk PDR
81/85		71/72	Advanced PDR	Advanced
90	U	90	Ungradable	

Table 1.1: Diabetic retinopathy disease classifications. DR = diabetic retinopathy; ETDRS = Early Treatment Diabetic Retinopathy Study; LDES = Liverpool Diabetic Eye Study; NDESP = National Diabetic Eye Screening Programme; NPDR = non-proliferative diabetic retinopathy; PDR = proliferative diabetic retinopathy.

Whether or not an individual is likely to benefit from screening is therefore dependent on their likelihood of having the disease and screening positive; there is likely to be no substantive health benefit from screening negative.

Screening can be defined in a variety of different ways. The World Health Organization (WHO) defines screening as:

> "The presumptive identification of unrecognized disease or defect by the application of tests, examinations, or other procedures which can be applied rapidly. Screening tests sort out apparently well persons who probably have a disease from those who probably do not. A screening test is not intended to be diagnostic. Persons with positive or suspicious findings must be referred to their physicians for diagnosis and necessary treatment." [19]

The WHO definition has several significant implications. Eligibility for screening, by this definition, can include anybody without a particular diagnosed health problem. Therefore, the entire population is potentially suitable for screening of some sort. The WHO definition asserts that screening must be rapid and that treatment or care must be available to justify screening. Thus, the health benefits of treatment are fundamental to the value of screening; individuals should only be screened and referred if they have an expected capacity to benefit from further health care. This fits well with a perspective of cost-effectiveness optimisation, as will be discussed later in this thesis.

The NHS in the UK defines screening in a similar way:

> "Screening is a process of identifying apparently healthy people who may be at increased risk of a disease or condition. They can then be offered information, further tests and appropriate treatment to reduce their risk and/or any complications arising from the disease or condition." [20]

The WHO definition refers to probability of disease being present, and the NHS definition introduces the word risk. The notion of risk is central to the process of screening and will be a key theme throughout this thesis. The NHS defines screening as a means by which an individual's risk of developing complications associated with a disease might be reduced. The definition of screening adopted in this thesis will fit with those of both the WHO and the NHS, incorporating the following characteristics:

1. Applicable to individuals with:
   (a) No disease,
   (b) Asymptomatic disease, or

(c) Unrecognised symptomatic disease.

2. Must provide information on an individual's likelihood of having (or developing) the screened condition.

3. Treatment or care must be available for those who screen positive.

4. The natural history of the disease must be understood.

1.2.1 Evaluation of screening

A particular screening test can be evaluated on the basis of its ability to predict the presence of disease. This is most commonly assessed using a value known as the receiver operating characteristic (ROC). Usually, screening can result in one of four outcomes; true positive (TP), false positive (FP), false negative (FN), or true negative (TN), as shown in Table 1.2. The ROC, and thus the performance of a screening test, is estimated in terms of the distribution of these four possible outcomes within a population. Of key importance are the true positive rate (TPR), also known as 'sensitivity', and the true negative rate (TNR), also known as 'specificity'. These are calculated as:

$$TPR = \frac{TP}{TP + FN} \tag{1.1}$$

$$TNR = \frac{TN}{TN + FP} \tag{1.2}$$

where TP, FP, FN and TN correspond to the number of people with each screening outcome. In ROC analysis, a curve is constructed by plotting the TPR against the false positive rate (FPR; $1 - TNR$) at a range of possible thresholds for identifying a screen-positive event.

Traditionally, screening programmes that use particular screening tests have been evaluated on the basis of clinical trials. The National Screening Committee (NSC), which advises the government and NHS in the UK, states that "There should be evidence from high quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity" [21]. Much work has been done in the context of screening in terms of both economic evaluation and the development of economic theory [22–24]. Such studies identify screening as broadly amenable to standard approaches to clinical and economic evaluation.

	Disease-positive	**Disease-negative**
Test-positive	True positive (TP)	False positive (FP)
Test-negative	False negative (FN)	True negative (TN)

Table 1.2: Possible screening outcomes

1.2.2 Screening for diabetic retinopathy

Screening for diabetic retinopathy satisfies the definitions of screening outlined above [25]. The principal method of screening for diabetic retinopathy is by digital fundus photography, in which photographs are taken of the back of the eye. Trained graders can then examine these photos to identify the presence of microaneurysms and other signs of disease. When taking the photographs, mydriatic eye drops are usually used to dilate the pupils and allow for a higher quality image. A more accurate examination is by slit lamp biomicroscopy, in which an ophthalmologist can carry out a more detailed examination of the eye.

In the UK, the NSC established the NHS Diabetic Eye Screening Programme (NDESP). Since 2007, all people with diabetes over the age of 12 have been eligible for invitations to attend screening annually [26]. Annual recall was adopted on the basis of expert opinion. The NDESP consists of 92 local programmes that undertake screening in primary care. At screening appointments individuals have their visual acuity assessed using logMAR charts and then high quality digital photographs are taken of the back of the eyes after administering mydriatic eye drops. The images are subsequently independently graded multiple times at a grading centre by accredited technicians. If a patient is screened positive for STDR they are referred to their local hospital eye service for further investigation by slit lamp biomicroscopy. Recently, local programmes have begun to allocate individuals to 'digital surveillance' whereby they are followed-up more regularly than annually [27].

There is growing evidence that many individuals who attend screening annually are at low risk of developing STDR, and that these individuals could be safely screened less frequently [28–37]. Furthermore, it is possible that individuals with a higher level of risk might benefit from more regular screening [31, 35]. These findings highlight the presence of heterogeneity in this context and have given rise to calls for more nuanced approaches to screening for DR.

1.3 Heterogeneity

Screening programmes can improve people's longevity and quality of life, and can potentially be cost-saving for health services in the long run. However, it is unlikely that all of these goals could be achieved in a publicly funded programme without rationing and setting eligibility criteria. As such, screening programmes tend to discriminate based on individuals' characteristics, such as age and sex. By identifying particular subgroups in which screening is expected to be cost-effective, and others in which it is not, it is possible to implement limited use criteria, whereby people must fall into particular subgroups in order to be eligible.

The challenge of capturing heterogeneity of treatment effects is a long standing

issue [38]. In economic evaluations, the challenge is even more pronounced [39]. Accounting for patient heterogeneity is a means of improving outcomes, which was rarely considered in economic evaluations until recently [40]. This may be due to perceived difficulties in methodology or ethics. In recent years, the value of accounting for patient heterogeneity has been more widely recognised. The conduct of sub-group analyses in the context of cost-effectiveness analysis is now routinely recommended [41].

Screening is one field in which efforts are being made to acknowledge heterogeneity and to differentiate care accordingly. It is now possible to use algorithms to estimate an individual's risk of disease onset, and the use of such tools has become known as 'predictive medicine'. There have been calls for individualised screening [42] and assertions that, in the future, a greater emphasis must be placed on risk-based screening [43]. In risk-based screening — as the term has recently been used [44, 45] — individuals are only invited to attend screening if their risk of disease onset is deemed sufficiently high. Such an approach offers potential for improved outcomes and lower costs compared with standardised programmes, but presents new challenges. This thesis focusses on individual risk as a particular manifestation of patient heterogeneity.

1.3.1 Individual risk

If an individual is more likely to screen positive, it is more likely that their attendance at screening will result in them receiving the intended benefits. An individual's likelihood of screening positive is dependent on their risk factors. Advancements in our understanding of the causes and correlates of disease, and improvements in routine data collection, mean that it is now possible, in many cases, to estimate an individual's risk of developing a disease or experiencing specific complications. There are an increasing number of risk calculation engines (RCEs) becoming available that can facilitate this process [46–48]. These developments have implications for all areas and stages of treatment, but are particularly pertinent to screening interventions. If an individual is at high risk of developing a disease it is, *prima facie*, likely to be cost-effective to screen them, while if they are not this is less likely. New developments in the quantification of risk factors are therefore particularly relevant to the estimation of the cost-effectiveness of screening interventions.

Current screening programmes implicitly consider risk by offering screening to a limited section of the population. Age is a risk factor common across most areas of health and disease. Similarly, males and females tend to have different risk levels for disease, and some screening programmes are differentiated by sex, or offered to only men or only women. It is clear that risk is a consideration in the design of NSC screening programmes in the UK. Current programmes tend to

consider only one or two risk factors when differentiating screening for different individuals. These are usually general characteristics such as age or sex.

More individualised screening programmes are currently being proposed by academics and decision-makers alike. Methods of economic evaluation, and particularly of decision analytic modelling, have enabled economists to make predictions about complex interventions and their outcomes. However, individual risk is not routinely considered in the evaluation of screening programmes in an explicit way. Modelling methods have yet to be harnessed for the purpose of evaluating screening programmes that are able to considerably differentiate care for individuals based on their level of risk. This thesis will present some ways in which information about an individual's risk can be used to enhance the cost-effectiveness of screening programmes. We will demonstrate the importance of this approach in the context of screening in terms of the potential for improved health outcomes.

We use the word 'risk' in the sense of an individual's hazard rate or hazard function. That is, an individual with a given set of characteristics, at any moment in time, can be ascribed a hazard rate representing the probability that an event will occur. Mathematically this is derived as the limit of a number of events occurring in a given unit of time, divided by the number of individuals at risk over time. In order to estimate an individual's hazard rate we use information collected from relevant hazard ratios, the effect of which can be estimated by treating the log of the hazard rate as a function of baseline hazard and explanatory variables. Hazard ratios represent the estimated effect on an individual's hazard rate of various characteristics: i.e. risk factors. These are calculated using survival analysis or time-to-event analysis, which often involves the use of proportional hazard models informed by existing data and by expert opinion. Such models can be used to estimate the probability that an individual will develop a given disease within a given period of time within a given margin of error.

1.3.2 Individual risk and diabetic retinopathy

In the case of diabetic retinopathy, the risk of disease progression to a level requiring treatment has been shown to be related to age, gender, duration of diabetes [11, 12], severity of retinopathy [35, 36], HbA1c levels [18, 49–53], blood pressure [53–56], blood lipid levels [57] and proteinuria [58]. A risk calculation engine has been produced to estimate the risk of developing STDR at 1 year, 5 years, or 10 years for people in Iceland [34], based on eight risk factors. This model has been validated in different populations [28, 59, 60], and other RCEs are in development. Recently, a new RCE has been developed using data from Liverpool, UK [61]. This RCE will be described in more detail below, as it plays a crucial role in the research described in this thesis.

The principal means by which DR screening can be differentiated according to individual risk is by setting individualised screening recall periods. An individual at high risk of developing STDR could be invited back to screening (following a negative screening result) at an earlier point than an individual with a lower risk of developing STDR. This thesis explores the practical and ethical implications of using such an approach, and estimates the cost-effectiveness of a risk-based screening programme for DR. The need for such research has been demonstrated by recent policy developments in the UK.

1.4 Policy context

The provision of health care involves the distribution of scarce resources. Government budgets, and health spending specifically, are constrained. The same applies to personal budgets of individuals. There is an imperative to achieve allocative efficiency by only funding screening programmes (and specific screening tests) that can be demonstrated to represent good value for money.

There is also an imperative to achieve productive and technical efficiency across and within existing screening programmes. The development of risk-based screening for a disease such as DR represents a means of improving health outcomes by altering eligibility within the programme, rather than by reallocating funds to alternative services or introducing alternative inputs to health. This is the principal driver for policy change within the NDESP.

1.4.1 Development of the NDESP

After a review of the evidence, the NSC recently recommended changes to the NDESP such that: "For diabetics at low risk of sight loss, the interval between screening tests should change from one year to two years. The current one year interval should remain unchanged for the remaining people at high risk of sight loss." [62]

Similar developments can be seen internationally. Recently, the American Diabetes Association issued a position statement recommending that:

> "If there is no evidence of retinopathy for one or more annual eye exams, then exams every 2 years may be considered. If any level of diabetic retinopathy is present, subsequent dilated retinal examinations for patients with type 1 or type 2 diabetes should be repeated at least annually by an ophthalmologist or optometrist. If retinopathy is progressing or sight-threatening, then examinations will be required more frequently." [63]

The NSC recommendation signals a transition from standardisation to stratification; a step in the development of the NDESP beyond a 'one-size-fits-all' approach. There are two aspects to this development.

The first aspect is the ability to identify heterogeneity in individual risk within the population. Risk estimates may be identified at the level of:

1. the population;
2. subgroups defined by disease status;
3. subgroups defined by individual characteristics; or
4. the individual.

The second aspect is the capacity to differentiate screening recall with greater precision. There are at least four levels to this:

1. a fixed interval for the whole population;
2. multiple fixed intervals, whereby individuals are allocated to one of several fixed intervals;
3. variable intervals, whereby individuals can move between different intervals; and
4. variable recall, whereby there are no pre-defined screening intervals.

In principle, each aspect may develop at a different pace. For example, individual-level risk estimates may be used in a programme with multiple fixed intervals. However, the value of identifying heterogeneity and the value of differentiating screening are each dependent on the other, and so the two aspects are likely to develop in tandem.

Four possible types of programme, relevant to the NDESP, are summarised in Figure 1.3. These are defined in relation to the screening recall schedule. Examples are included for the basis of identifying eligible populations. We describe these four programmes as *standardised*, *stratified*, *individualised*, and *optimised*.

Each transition between the stages shown in Figure 1.3 raises new questions requiring research evidence. The NSC has identified several conditions that should be met before stratified screening is introduced. This thesis seeks to inform the development of screening beyond stratification to individualisation and optimisation, both specifically in the context of diabetic retinopathy and in the development of screening pathways more broadly.

There are substantive differences between programmes that are standardised, stratified, individualised, and optimised in practical, ethical, and analytical terms.

Figure 1.3: Development of the NHS Diabetic Eye Screening Programme

Practical differences, particularly relating to greater differentiation in screen recall, are highlighted briefly in this chapter and in Chapter 5. Analytical differences relating to the conduct of cost-effectiveness analysis are discussed in Chapter 8 and ethical issues are considered in Chapter 9.

In moving beyond a standardised programme, stratification must not be arbitrary. There must be a strong basis on which to offer people differential care depending on their allocation to a subgroup [40]. Therefore, a key challenge for risk-based screening is to appropriately define limited use criteria; the grounds on which people should or should not be invited to screening. Additionally, it might also be possible to differentiate the frequency of screening based on risk. Both individual risk and time are continuous and therefore infinitely divisible. As such, an optimal screening programme could lie anywhere between the extremes of no screening for anybody to constant screening for the entire asymptomatic population. In this thesis we present a means of setting optimal criteria for risk-based screening using cost-effectiveness analysis. We discuss some of the practical and ethical challenges presented by such an approach and argue that risk-based screening can improve the effectiveness, efficiency, and fairness of screening programmes.

1.4.2 Previous research

Several studies have considered the cost-effectiveness of alternative screening intervals in the context of diabetic retinopathy. Most of these studies pre-dated recent advances in risk calculation and were therefore not able to evaluate screening programmes that differed according to an individual's level of risk [64–72]. Many of these studies indicated that screening intervals can be safely extended under certain circumstances, achieving greater cost-effectiveness.

More recent studies have evaluated screening programmes that use estimates of individual risk to differentiate screening recall. In this context, efforts have been made to 'optimise' screening intervals for diabetic retinopathy. Aspelund et al. developed a model based on a fixed number of diagnoses across screening intervals [34]. Their model suggested that optimal screening intervals range from 6 to 60 months, dependent on an individual's level of risk. Van der Heijden et al. recently sought to validate Aspelund et al.'s model using Dutch data [59]. The authors found that the model enabled a 23% reduction in screening frequency compared with biennial screening, and a 61% reduction compared with annual screening, and suggested that the use of such a model could help reduce the costs of diabetes care. Mehlsen et al. adopted a similar method, using a fixed risk margin of 0.5% chance of event [73]. They used multiple logistic regression to find the optimal screening interval for low-risk diabetic retinopathy patients, and found that screening intervals should be extended for most low-risk individuals.

Using a discrete event simulation, Day et al. found that the risk of vision loss increased as the screening interval was extended from 1 to 5 years [74], though the authors assert that extending the interval from 1 to 2 years is safe.

These studies constitute important developments in our understanding of risk in the context of screening for diabetic retinopathy, and they can be used to inform screening programmes. However, while these studies are concerned with optimisation problems, they are not concerned with the optimisation of cost-effectiveness. Scanlon et al. recently sought to 'optimise' the screening interval in terms of cost-effectiveness, but only within a restricted set of pre-specified options, all of which might be suboptimal [30]. Specifically, they evaluated screening intervals that were differed according to an individual's allocation to a high-, medium-, or low-risk group defined according to disease characteristics, rather than according to an explicit estimation of their risk of disease onset.

No existing studies have evaluated the cost-effectiveness of a screening programme that differentiates screening intervals according to estimates of individual risk. We are not aware of any such study in the context of either diabetic retinopathy or any other disease area.

1.4.3 The ISDR study

The research on which this thesis reports was embedded within a wider programme of quantitative and qualitative research: a National Institute for Health Research (NIHR) Programme Grant for Applied Research (PGfAR) study, titled 'Introducing personalised risk based intervals in screening for diabetic retinopathy: development, implementation and assessment of safety, cost-effectiveness and patient experience' and known as the ISDR study [75]. The multidisciplinary collaboration and support that was facilitated by involvement with the wider study represents a key strength of the research reported in this thesis.

The ISDR study was based at The Royal Liverpool and Broadgreen University Hospitals NHS Trust (RLBUHT). Several workstreams make up the programme of research, and each workstream relates to this thesis in a variety of ways. This thesis relates primarily to work carried out to support Workstream D, but necessarily draws on aspects from all other workstreams. As such, it is important to provide a brief outline of the work undertaken in each workstream, and how it relates to this thesis.

Workstream A: systematic review

The ISDR study was initially scheduled to complete a full systematic review of the available evidence comparing variable interval and annual recall for screening for diabetic eye disease. The review sought to elicit rates of visual impairment

and treatment in order to inform the other workstreams. However, shortly after commencement of the study, a number of reviews were published by other research groups. Completion of another full review was therefore deemed unnecessary. Nevertheless, two-way communication was maintained with the review team to aid identification of relevant studies. In particular, disease incidence and progression studies identified by the review team were utilised in the modelling described in Chapter 7.

Workstream B1: data warehouse and data processing

Workstream B was made up of several components relating to data collation and analysis. The success of the programme depended largely on the creation of a data warehouse. This combined databases from primary care, hospital patient management, and the screening programme, with patient records matched by NHS Number. Workstream B1 was responsible for the data warehouse and for the creation of data outputs to support other workstreams. Databases included in the data warehouse and used in this study are as follows.

EMIS Web is an IT system used in general practice to maintain electronic patient records. Data from EMIS are key to Workstream B2 because data from other sources are all linked to EMIS records. However, data from EMIS are not directly used in this thesis.

OptoMize is the system used by the screening programme and is provided commercially by Digital Healthcare. It is used to record data at photographic screening appointments in accordance with the standards set by the NSC. OptoMize was preceded by a system called ORION, from which similar data were available. Screening appointment and outcome data were analysed to estimate resource use and to inform event rates and transitions as used in the modelling work described later in this thesis.

Diabolos is an internal Microsoft Access database used in the hospital eye service. Data from slit lamp biomicroscopy assessments were recorded and stored in Diabolos. These data were key for the analyses reported in this thesis as it is these data that identify patients referred following a screen-positive result.

iPM is used to record patient activity across the RLBUHT hospital sites. It includes information on patient appointments in addition to some demographic data. For the purpose of this thesis, iPM was the primary source of information about resource use associated with treatment for DR.

Workstream B2: observational cohort study

Using data from the data warehouse, Workstream B2 consisted of an observational cohort study. The analysis sought to identify population-based estimates of the frequency of a variety of outcomes including STDR and visual acuity. Furthermore, it sought to provide information about risk factors in order to inform future risk calculation estimates. It was initially hoped that findings from Workstream B2 could be incorporated into a model-based cost-effectiveness analysis described later in this thesis. However, due to delays in the delivery of data, this proved not to be possible.

Workstream B3: feature specific retinal grading

The purpose of Workstream B3 was to provide visual acuity data for the data warehouse. This process involved the development of a graphical user interface to be used in a clinic setting, which would also capture feature specific grading for retinopathy and maculopathy; describing the presence of microaneursyms, retinal thickening, and other characteristics. This work was not completed in time for the data to be included in the analyses reported in this thesis.

Workstream C: risk calculation engine

The development of the risk calculation engine (RCE) is central to the development of the cost-effectiveness analysis that is presented in this thesis. The RCE was developed using a Markov model based on panel data from 2009 to 2014 from 11,808 people with diabetes registered with GP practices in Liverpool. The RCE was designed to estimate risk of developing referable disease (that would be considered a screen-positive event if detected) at 6, 12, and 24 months. The development and testing of the RCE is described in detail elsewhere [61], with the key features relating to the cost-effectiveness analysis described in this thesis.

A number of covariates were considered for inclusion from the available data. Variables for inclusion needed to have less than 20% missing data, and multiple imputation was used to replace data that were missing. Covariates were selected in a two stage process by ranking them using the Wald statistic and then selecting them using the corrected Akaike information criterion. The final model included duration of (known) diabetes, HbA1c, age, systolic blood pressure, and total cholesterol. The Markov model was defined in terms of four states: i) no DR, ii) non-referable DR in one eye, iii) non-referable DR in both eyes, and iv) referable disease. The model allowed for six transitions, with the fourth state being the absorbing state.

The RCE estimates hazard rates that are used to inform transition rates in the decision analytic model described in Chapter 7.

Workstream D: health economics

Workstream D represented the health economics work for the ISDR study. An economic evaluation alongside the ISDR trial (Workstream E) will be conducted on its completion. Furthermore, a model-based analysis is used to evaluate the cost-effectiveness of risk-based screening with a lifetime horizon, as described in this thesis. Materials for the collection of resource use and quality of life data were prepared for this purpose. This thesis describes much of the work conducted for Workstream D.

Workstream E: randomised trial

From years 2 to 5 of the programme (for a duration of 42 months), a randomised clinical trial was carried out. The trial is described in detail elsewhere [76], with pertinent details outlined in this thesis. The study was based at Royal Liverpool University Hospital and six community screening centres in the Liverpool Diabetic Eye Screening Programme. Trial participants (n=4,543) were people with diabetes who were under the care of Liverpool clinical commissioning group (CCG) and eligible for screening for diabetic retinopathy. The purpose of the trial was to evaluate the safety and acceptability of a programme of variable interval screening for diabetic eye disease, with attendance as the primary outcome. Individuals were randomised to either current (annual) screening or risk-based variable-interval screening. For those allocated to the risk-based programme, the probability of them screening positive at 6, 12, and 24 months was estimated following a negative screening outcome using the RCE developed in Workstream C. An acceptable risk threshold of 2.5% was established based on guidance from the study's patient and public involvement (PPI) group. This threshold was used to allocate individuals to the longest recall of either 6, 12, or 24 months on the basis of their risk level not exceeding 2.5%. Data were collected from trial participants for the purpose of cost-effectiveness analysis, as described in Chapters 3 and 5.

Workstream F: perceptions of screening

Workstream F consisted of qualitative work in which interviews were conducted with people with diabetes who do and do not attend screening and with health professionals. This package of qualitative work enabled the collection of quality of life data from hard to reach groups, though the data were not available in time for reporting in this thesis.

1.5 Research questions

The decision problem is multi-faceted. There is the more traditional decision problem of determining whether or not screening should be risk-based or standardised. But there is also a more novel decision problem, which is to determine how the risk-based screening programme should in the first place be defined. These two questions are inextricably linked, both in general terms and in the specific case of screening for diabetic retinopathy. The primary focus of this thesis is on the practical challenge of evaluating risk-based screening for diabetic retinopathy. However, we also seek to establish the foundations for the evaluation of more complex risk-based screening programmes that may be realised in the near future. The thesis seeks to address both theory and practice, which are captured by two key research questions:

1. Is risk-based screening for diabetic retinopathy likely to be cost-effective compared with standardised screening?

2. How can individual risk be incorporated into cost-effectiveness analysis in order to inform the allocation of screening?

The validity of these two questions — and their answers — depend upon one another. The application of risk-based screening in practice depends on a better understanding of the underlying principles, but the need for clear principles is borne out of recent developments that make risk-based screening practically possible. However, the practical possibility of risk-based screening does not imply that it should be endorsed. Technical developments in the capacity to deliver health care services can raise ethical questions. In order to answer our primary questions, it is necessary to address a number of secondary research questions, including:

- How have researchers previously evaluated the cost-effectiveness of interventions for diabetic retinopathy?

- How does diabetic retinopathy affect individuals' health-related quality of life?

- What are the costs of screening for diabetic retinopathy?

- What are the costs of treatment for diabetic retinopathy?

- How does individual risk relate to cost-effectiveness in the context of screening?

- Is there an ethical justification for allocating screening according to individual risk?

1.6 Thesis outline

The thesis is structured around the development of a decision analytic model to evaluate the cost-effectiveness of risk-based screening for diabetic retinopathy. Later in the thesis we introduce some novel explication of the theoretical basis for a more sophisticated approach to the evaluation of risk-based screening, and consider a new ethical framework in which to consider the policy implications.

Chapters 2 through 7 chart the development of the cost-effectiveness model. Chapter 2 reviews previous literature that has used decision modelling to evaluate interventions for DR, for the purpose of guiding our choice of model structure. Chapter 3 presents a cross-sectional study of quality of life values in people attending screening within the Liverpool screening programme and Chapter 4 presents a systematic review and meta-analysis of health state utility values (HSUVs) for DR. The findings of these two chapters are used to inform the selection of quality of life parameters for the model. Chapter 5 describes screening pathways within the NDESP in order to understand the resource use involved, and presents a costing study of screening. Chapter 6 presents an analysis of treatment activity and pathways for people with DR in Liverpool in order to estimate appropriate treatment cost parameters for the model. Building on these preceding chapters, Chapter 7 presents the development and findings of a model-based cost-effectiveness analysis of alternative screening programmes for diabetic retinopathy. Chapter 8 introduces the concept of individualised cost-effectiveness analysis (iCEA) and describes how this simple framework could be used to determine an optimised risk-based screening programme. Chapter 9 outlines an ethical basis for risk-based screening. Finally, Chapter 10 presents a discussion of the findings presented in this thesis and some conclusions.

Chapter 2

A review of model-based economic evaluations in diabetic retinopathy

Summary

A primary component of the thesis is the use of model-based economic evaluation. It is important to consider alternative approaches to modelling disease progression and treatment in diabetic retinopathy. In this chapter we describe a narrative review of model-based economic evaluations in DR. We searched MEDLINE and Embase and present our findings according to alternative model structures. Decision trees, cohort state transition models, and individual sampling models are all identified. The findings of the review are used to inform the structure of the model developed in this thesis, as reported in Chapter 7. We discuss some of the shortcomings in existing modelling work and use these to inform the analyses reported in subsequent chapters.

2.1 Introduction

Decision analysis usually requires the evaluation of costs and outcomes beyond the limited time horizon of a clinical trial. This can be particularly important when key outcomes are observed in the more distant future. In the case of screening for diabetic retinopathy, long-term costs and outcomes are important; most notably in relation to the onset of severe visual impairment. Therefore, while the ISDR study does include a randomised trial within which risk-based screening can be evaluated, the cost-effectiveness analysis reported in this thesis will use decision analytic modelling.

There are a variety of candidate approaches to modelling. Decision trees can be used to evaluate decisions based on aggregate level probabilities and pay-offs. They represent a simple approach to decision analysis based on the summation of costs and outcomes associated with alternative scenarios, multiplied by the probability of each scenario in order to obtain expected values. Decision trees do not incorporate any timing. It is possible to conduct simulated decision tree analyses in order to obtain results as a statistical distribution, though such analyses are uncommon.

State transition models — often described as 'Markov' models — have become one of the most popular approaches to model-based economic evaluation in health care. They can be used to simulate aggregate outcomes for a cohort or for individuals. In a state transition model, each state is associated with costs and outcomes and simulated individuals have a probability in each unit of time (known as a Markov cycle) of transitioning from one state to another. The Markov assumption is that the model is 'memoryless'; an individual's probability of transition depends only on their current state. Most state transition models in the context of economic evaluation in health care do not satisfy this assumption and are therefore not strictly Markov models. For example, many models allow transition probabilities to be a function of time. In a state transition model time is usually discrete, meaning that individuals are not continuously observed and instead transitions are only observed at fixed intervals. State transition models can be based on fixed parameters (deterministic) or on statistical distributions of inputs (probabilistic).

When a state transition model is based on an individual-level rather than cohort-level simulation, it can be better described as an individual sampling model (ISM). In this case, the analyst is freed from some of the restrictions of state transition modelling and the model need not be defined only in terms of mutually exclusive and collectively exhaustive states. ISMs can treat time as continuous and estimate the time to events or transitions, rather than their probability of occurring within a fixed time period.

Discrete event simulation (DES) is perhaps the most flexible but also well-

defined approach to decision modelling, with a long history of development in the field of operational research. DES models are structured around a set of possible events that can occur. Individuals' characteristics can be used to determine the likelihood and time to occurrence of each event. Available resources (e.g. health care staff) can be limited in order to create queues (e.g. of patients) within the model.

The pros and cons of various approaches have been evaluated both broadly and within the context of particular conditions [77–79]. In this chapter we describe a narrative review designed to identify key studies adopting a variety of different modelling methods. A review of modelling studies in the context of diabetic retinopathy has not previously been conducted. As we did not seek to provide a complete picture of this literature, a systematic review of all modelling studies was deemed unnecessary. The narrative review will provide a better understanding of the ways in which different approaches have been used and highlight the advantages and disadvantages in this context. The findings of the review are used to determine an appropriate modelling strategy to be pursued in subsequent chapters.

2.2 Methods

Guidelines such as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) were used to guide the methodology and reporting of the review. PRISMA is not defined specifically for use in the review of economic evaluations, but is largely applicable in order to ensure methodological rigour and high reporting standards. However, for a narrative methodological review, a number of items are not relevant and these will be highlighted throughout the chapter as necessary. All stages of the review were conducted by a single reviewer.

2.2.1 Search strategy

The method of identification of studies was through a systematic search of the literature. The search was designed to identify key studies employing a variety of modelling methods. The search used the Ovid MEDLINE and Embase databases and combined terms for diabetic retinopathy, cost-effectiveness analysis, and decision modelling. No date or language restrictions were applied to the electronic search, which included items indexed up to 14th September 2016. The search was conducted on the basis of Medical Subject Headings (MeSH), as shown in Table 2.1.

Search step	Search terms
1	(Diabetic Retinopathy and Cost-Benefit Analysis).sh.
2	(Decision Support Techniques or Decision Trees or Markov Chains or Computer Simulation or Models, Econometric or Patient-Specific Modeling or Systems Analysis or Models, Theoretical or Models, Economic).sh.
3	1 and 2

Table 2.1: Search strategy for MEDLINE and Embase in Ovid

2.2.2 Study eligibility

Included studies needed to be economic evaluations. Cost-benefit, cost-effectiveness, and cost-utility analyses were all included. Studies must have employed decision modelling methods. Specifically, they must not be limited to evaluation within a trial setting or based solely on observed findings from a single study. This is because an important aspect of the review is the identification of approaches to incorporating data from multiple sources. It was expected that most of these analyses would employ either decision trees, Markov or state transition models, or event-based and agent-based simulations. The language of publication must be English.

All types of publication were included if they reported on an economic evaluation using decision modelling. The population must be people either with or at risk of diabetic retinopathy, of any demographic. No inclusion criteria were employed regarding specific interventions or technologies, though models used to evaluate screening are given particular attention in the discussion. The comparator element of the 'Population, Intervention, Comparison, Outcomes' (PICO) statement does not apply, as the purpose of the review is to identify methods used rather than treatment effects. We excluded any study that did not describe a decision analytic model or that was not available in English.

2.2.3 Data collection

Study selection

Studies were assessed for retrieval based on titles and abstracts. Articles were rejected at this stage if it was clear that the study could not satisfy the inclusion and exclusion criteria. Full texts were retrieved for studies not rejected at abstract screening, and subsequently reviewed for satisfaction of the inclusion and exclusion criteria. Reasons for exclusion were recorded. The number of records identified, retrieved, screened, assessed, included, and excluded in the review are summarised in a PRISMA flow diagram.

Data extraction and management

Data were extracted and recorded in a spreadsheet. As the purpose of this narrative review was to identify methodological practice, we extracted data relating to the key features of the model structure and data, with reference to good practice guidelines [80]. In relation to the structure and the operation of the model, we extracted information on i) comparators, ii) perspective, iii) model type, iv) states / pathways, v) time horizon, and vi) cycle length. We also extracted data regarding the sources of data for the key aspects of the model, namely: i) disease progression, ii) costs and resource use, and iii) health state utility values (HSUVs, if applicable). Information was also extracted regarding any i) uncertainty analysis and ii) validation that was conducted, as these could provide further insight into the appropriateness of different methods in this context. Recently, there has been discussion regarding the significance of different software for modelling [81–84]. Therefore, we also extracted information regarding the software used for the analysis. For additional context we also list the countries in reference to which studies were conducted. Where information was missing, no attempt was made to contact authors. Only information relating to the parts of the model(s) specific to retinopathy was extracted. Where models also simulated non-ophthalmic complications of diabetes, these aspects of the studies were not reviewed.

Quality and relevance assessment

The same data collected for the methodological review — and in particular the tendency to not report key information — can be used as an indicator of reporting quality [80]. Studies that do not report the data intended for extraction, or that provide limited details, can be judged to exhibit lower reporting quality. The purpose of this review was to identify the nature of existing studies rather than to make inferences about their findings based on judgements of their methodological quality. A more pertinent consideration is the relevance of these studies to our context, which is the evaluation of screening for diabetic retinopathy. Therefore, our assessment focuses on the appropriateness of the modelling techniques adopted, rather than the methodological quality of the studies overall. Attempts at model validation — if reported — can demonstrate the appropriateness of the modelling approach.

2.2.4 Data synthesis and presentation

Information extracted from studies is tabulated. The key purpose of the review is to identify and characterise the use of alternative approaches to modelling in diabetic retinopathy. Therefore, studies are presented according to the type of modelling approach that they adopt. The classifications used are those specified

in the taxonomy of model structures presented by Brennan and colleagues [85].

Where a particular piece of information is not relevant to a study — for example, the source of HSUVs in a study that does not use quality-adjusted life years — the item was recorded as not applicable (NA). Where an item is relevant but the study fails to report it with clarity, we recorded this as not reported (NR).

2.3 Results

In this section we outline the studies identified by our literature search and summarise their characteristics.

2.3.1 Study selection

The literature search in Embase and MEDLINE identified 37 citations, including one duplicate [30, 65, 67, 86–118]. Of these 36 studies, four were excluded at title and abstract screening either because they were not available in English [88, 89] or reported on a study that did not include decision modelling [112, 116].

When full texts were assessed for eligibility, six were excluded because they did not report modelling studies [86, 94, 95], focussed on non-diabetic eye disease [102], or evaluated treatment for diabetes generally without specific attention to retinopathy progression within the model [105, 115].

The review therefore included 26 eligible articles. Figure 2.1 shows a PRISMA flow diagram of the citation screening procedure.

2.3.2 Study characteristics

The full data extraction is presented in Table A and Table B in the appendices. Table A outlines information relating to the structure of the models reported in the studies. Table B summarises the data used and any uncertainty analysis and validation. In this chapter, findings are presented separately for studies adopting each type of modelling approach, with key characteristics presented in tables.

Some studies did not self-identify as using a particular approach to modelling and it was therefore necessary to infer from the manuscript the type of modelling that was used. This was not always simple and so we highlight the studies that may have been misidentified and outline the basis for our categorisation.

Decision trees

We identified six studies that described models classified as decision trees, as shown in Table 2.2. Several studies did not provide a standard graphical representation of a decision tree according to guidelines for their presentation [93, 98, 108, 114]. In these cases it was more difficult to clearly identify the pathways

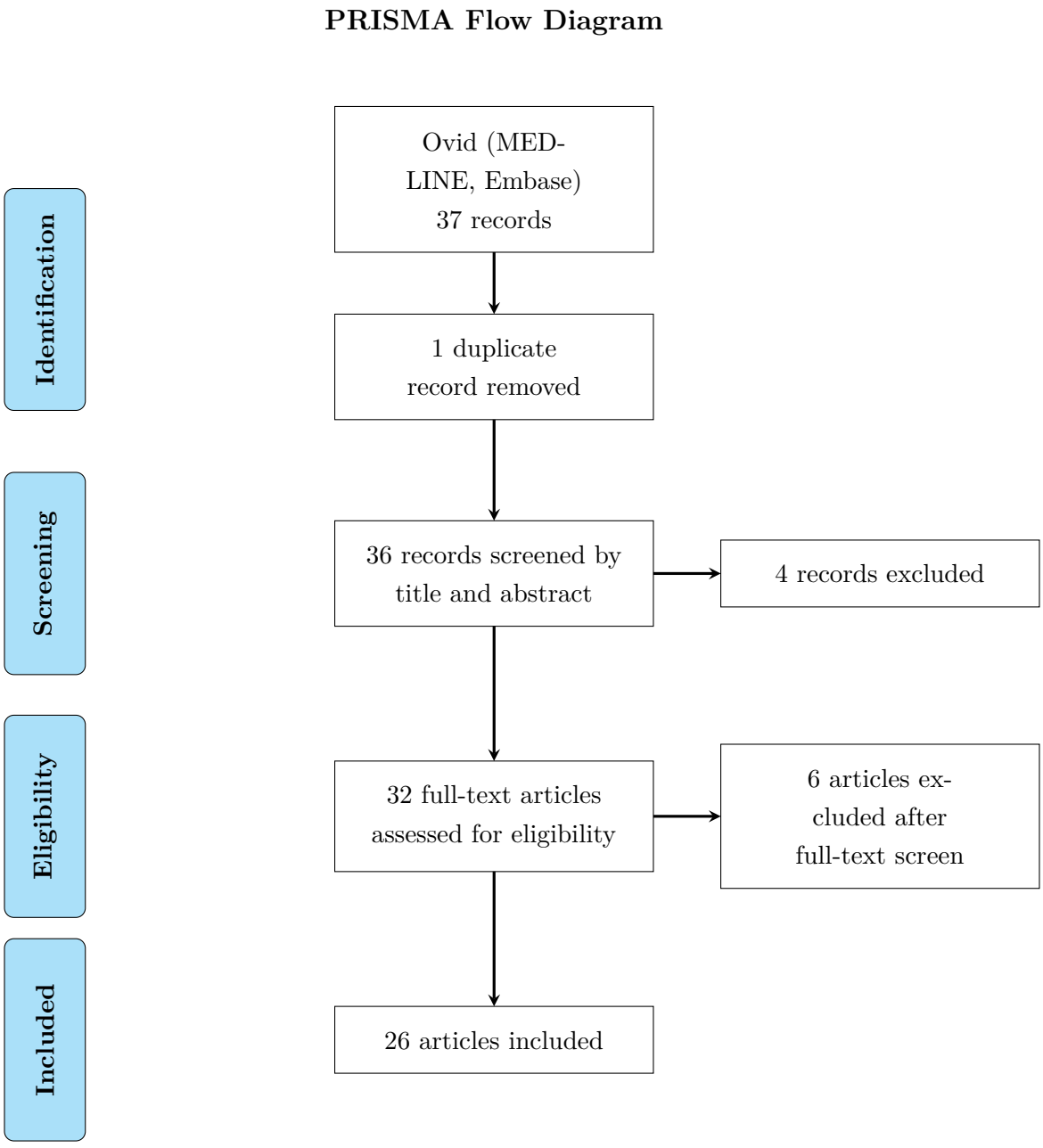

Figure 2.1: PRISMA flow diagram for review of decision models

Study	Events	Software
Sculpher et al. (1992) [114]	Screening outcomes	NR
Sharma et al. (2000) [93]	Treatment; complications	TreeAge
Whited et al. (2005) [97]	PDR; high-risk PDR; PRP; SVL	TreeAge
Scotland et al. (2007) [98]	No/mild DR; observable DR; technical failures	TreeAge
Scotland et al. (2010) [108]	No/mild DR; observable DR; technical failures	TreeAge
Brady et al. (2014) [101]	PDR; no PDR; no treatment; PRP; vitrectomy with laser; vitrectomy with membrane peel	TreeAge

Table 2.2: Decision trees. DR = diabetic retinopathy; NR = not reported; PDR = proliferative diabetic retinopathy; PRP = panretinal photocoagulation; SVL = severe vision loss.

and decisions being evaluated. In their 1992 paper, Sculpher et al. describe their study as "using the data collected on observed single modality screening options to model the results of potential options based on more complex strategies", but do not specify a particular modelling approach. No visual representation is provided, but the description of the model and its use of sample-level probabilities makes it reasonably clear that the analysis is based on a decision tree. Scotland et al. [108] reports on a secondary use of a previously developed model [98] and its characteristics are therefore inferred from the earlier report.

Most of the studies evaluated alternative approaches to screening for DR [97, 98, 101, 108, 114]. Sharma et al. [93] evaluated the use of grid laser photocoagulation for diabetic macular oedema (DMO). The events used to define the pathways in the models varied, but most incorporated both disease progression and treatment. Sharma et al. used disease progression estimates from the ETDRS study, while the other models were based primarily on local data.

All studies that specified software used TreeAge [93, 97, 98, 101, 108]. Three studies conducted both one-way and probabilistic sensitivity analyses [97, 98, 101], while two conducted one-way sensitivity analyses [93, 114] and another conducted only probabilistic sensitivity analysis (PSA) [108]. None of the studies discussed either the internal or external consistency of their findings or reported any formal attempts at validation.

Cohort state transitions

Cohort state transition models were used in 14 studies, making them by far the most common approach identified by our review, as described in Table 2.3. Several studies reported on evaluations of alternative screening strategies [30, 65, 67, 91, 96, 107, 113]. Some studies evaluated alternative treatment strategies [99, 109–111, 117], while others evaluated diabetes management interventions [90, 92, 96]. Dasbach et al. purport to evaluate biannual (twice a year) screening, but the results suggest that they actually evaluated biennial (every two years) screening.

Study	**States / events**	**Time horizon**	**Cycle length**	**Software**
Dasbach et al. (1991) [65]	Low-risk DR; high-risk DR; treated; blind	Lifetime	1 year	NR
Wu et al. (1998) [90]	No DR; any DR	10 years	1 year	Microsoft Excel
Crijns et al. (1999) [91]	no DR; DR excluding both DMO and PDR; PDR; adequate vision; poor central and/or peripheral vision; blindness	Lifetime	3 months	NR
Palmer et al. (2000) [92]	No DR; background DR; PDR; blind	Lifetime	1 year	IMIB TOM
Vijan et al. (2000) [67]	No DR; DR1; DR2; DR3; PDR; DMO; blind	Lifetime	1 year	NR
Sharma et al. (2001) [110]	VA (G1-G5)	Lifetime (55 years)	1 year	TreeAge
Polak et al. (2003) [96]	NR	Lifetime	3 months	NR
Mitchell et al. (2012) [99]	VA (letters)	15 years	3 months	NR

Rachapelle et al. (2012) [113]	No DR; non-STDR; STDR; CSMO; blind from DR	25 years	1 year	TreeAge
Stein et al. (2013) [117]	VA (LogMAR)	25 years	1 year	TreeAge
Pershing et al. (2014) [111]	VA (1-6)	Lifetime	1 month	TreeAge
Kawasaki et al. (2015) [107]	NPDR; severe NPDR; PDR; high-risk PDR; CSMO; blind	Lifetime (50 years)	1 year	TreeAge
Royle et al. (2015) [109]	Moderate NPDR; severe NPDR; early PDR; high-risk PDR; severe PDR; CSMO; SVI; treatment	30 years	6 months	Microsoft Excel
Scanlon et al. (2015) [30]	R0M0 gradings	Lifetime	6 months	Microsoft Excel

Table 2.3: Cohort state transition models. CSMO = clinically-significant macular oedema; DMO = diabetic macular oedema; DR = diabetic retinopathy; HR-PDR = high-risk proliferative diabetic retinopathy; IMIB TOM = Institute for Medical Informatics and Biostatistics Tools for Outcomes Modeling; NPDR = non-proliferative diabetic retinopathy; NR = not reported; PDR = proliferative diabetic retinopathy; STDR = sight-threatening diabetic retinopathy; SVI = severe visual impairment; VA = visual acuity.

Most authors described their model as a "Markov model" [30, 65, 67, 90, 92, 99, 107, 109–111, 113, 117]. Those that didn't were very unclear about the structure of the model [91, 96]. These two studies may not have reported cohort state transitions at all, though we inferred this from the descriptions. Both studies describe models that simulated cohorts rather than individuals and, despite being described as "continuous", appear to use a three-month cycle. Crijns et al. [91] specify disease states as being on an interval (0–1) scale, but thresholds are used that seem to define discrete states.

The majority of studies defined states according to level of DR progression [30,

65, 67, 91, 92, 107, 109, 113]. One study defined DR according to 'risk' level [65] and another only considered a simple no DR / any DR outcome in the context of a wider disease model for diabetes [90]. Some studies also (or alternatively) incorporated visual acuity levels [91, 99, 110, 111, 117]. Polak et al. principally discuss progression to vision loss, but are not clear about what other states or events (if any) are included in the model [96].

Most studies adopted a lifetime horizon [30, 65, 67, 91, 92, 96, 110, 111], though some restricted this to a fixed number of years that may not have seen the model run until the cohort had all died [107, 110]. Other studies used time horizons of 10 years [90], 15 years [99], 25 years [113, 117], or 30 years [109]. One year was the most commonly used cycle length [65, 90, 92, 107, 110, 113]. For some studies the cycle length used was not clearly stated but seemed to be one year [67, 117]. Others used a cycle length of six months [30, 109], three months [91, 96, 99] or one month [111].

The most frequently reported software used was TreeAge [107, 110, 111, 113, 117], followed by Microsoft Excel [30, 90, 109]. One study reported using the Institute for Medical Informatics and Biostatistics (IMIB) Tools for Outcomes Modeling (TOM) software [92]. Several studies did not specify the software used [65, 67, 91, 96, 99].

Most models used disease progression rates taken from landmark studies, such as the Diabetic Retinopathy Study (DRS) [65, 67], the Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR) [65, 91, 96, 99], the Diabetes Control and Complications Trial (DCCT) [90, 92], the Early Treatment Diabetic Retinopathy Study (ETDRS) [67, 109], the Diabetic Retinopathy Vitrectomy Study (DRVS) [110], and the Diabetic Retinopathy Clinical Research Network (DRCRnet) [117]. Some derived their data from other published studies [107, 113], local data [30], or from expert opinion [111]. Most studies did not use quality-adjusted life years as an outcome. Those that did generally used HSUVs from a single published study [30, 107, 110, 111, 117] or from local data [99, 113]. Others derived utility values from multiple sources [109]. Some studies employed mapping to identify HSUVs [90]. Vijan et al. only associated blindness with a utility decrement using a HSUV of 0.69 [67].

Most studies conducted one-way sensitivity analysis [30, 65, 67, 92, 99, 107, 110, 111, 113, 117] or PSA [30, 67, 99, 107, 109, 111, 113, 117]. Some studies did not report on any analysis of uncertainty [90, 91, 96]. Wu et al. [90] state that their model had previously been validated and give some consideration to the external consistency of their results. Pershing et al. [111] do not explicitly investigate internal or external consistency, but calibrate their model to reflect disease progression shown in external epidemiological data. The only study to describe any formal attempt at assessing validity using external data is Kawasaki

et al. [107].

Individual sampling models

Six studies used individual level simulations [87, 100, 103, 104, 106, 118], including one discrete event simulation [104], as outlined in Table 2.4. Three of the studies appeared to report different applications of the same decision model: the Prospective Diabetes Retinopathy Simulation (PROPHET) model [87, 106, 118]. Only one of the studies described their model as having a Markov structure [100].

Most of the models were structured around the progression of DR and DMO [87, 103, 104, 106, 118], while one study simulated patients with DMO and modelled their level of VA [100]. Three studies explicitly adopted a lifetime horizon [104, 106, 118], while two others did so implicitly by implementing 50-year [103] and 60-year [87] time horizons. Dewan et al. [100] appear to have used a 10-year horizon. Three studies used a one-year cycle length [87, 103, 106], while one used two months [118] and another used one month [100]. The DES modelled time continuously [104]. Half of the ISM studies did not specify what software was used to develop the model. TreeAge [100], Microsoft Excel [103], and Turbo Pascal [106] were each specified once.

Sources for disease progression estimates were principally the landmark studies, including WESDR [87, 106, 118], DRS [87], ETDRS [87, 103], and DCCT or the Epidemiology of Diabetes Interventions and Complications (EDIC) trials [103]. One study used local data [100] and the DES operated on the basis of calibration with local prevalence estimates [104].

All studies reported on some form of sensitivity analysis, with most conducting one-way sensitivity analyses [87, 100, 104, 106, 118] and some conducting PSA [103, 104]. Some studies reported on validation attempts in terms of the internal [103, 104] and external [100, 103] consistency of their results.

2.4 Discussion

This review included 26 articles published between 1989 and 2015. Many model-based economic evaluations in diabetic retinopathy — including in the context of screening — employ cohort state transition models. Most of these use a lifetime horizon and cycle lengths of up to one year.

Most of the models identified in this narrative review were structured around progression of DR. The studies classified disease progression in a variety of ways, but most incorporated the distinction between NPDR and PDR. Several studies modelled progression according to visual acuity, but this approach was primarily adopted for studies that modelled specific cohorts that did not span disease progression from onset to sight loss. Due to the well-defined (if inconsistent) na-

Study	States / events	Time horizon	Cycle length	Software
Javitt et al. (1989) [106]	Background DR; DMO; PDR; SVL; CAL	Lifetime	1 year	Turbo Pascal
Fendrick et al. (1992) [87]	Background DR; DMO; PDR; SVL; CAL	60 years	1 year	NR
Javitt & Aiello (1996) [118]	Background DR; DMO; PDR; SVL; CAL/blind	Lifetime	2 months	NR
Dewan et al. (2012) [100]	VA	10 years	1 month	TreeAge
Wolowacz et al. (2015) [103]	PDR; blind	50 years	1 year	Microsoft Excel
Wu et al. (2015) [104]	No DR; NPDR; PDR; DMO; blind	Lifetime (100 years)	Continuous	NR

Table 2.4: Individual simulation models. CAL = central acuity loss; DMO = diabetic macular oedema; DR = diabetic retinopathy; NPDR = non-proliferative diabetic retinopathy; NR = not reported; PDR = proliferative diabetic retinopathy; SVL = severe vision loss; VA = visual acuity.

ture of disease classification in DR, state transition models appear to represent an appropriate basis for the simulation of disease progression. Furthermore, as screening programmes operate on the basis of gradings that explicitly group people according to disease level, state transition modelling can accurately represent the relevant pathways. This is perhaps one reason why state transition modelling was the most common method identified by our review.

Many of the studies used landmark epidemiological studies as a source for disease progression rates. These studies were conducted many years ago and may not be representative of current epidemiology, either because diabetes management has improved generally or because national screening programmes have since been introduced.

None of the studies that used quality-adjusted life years as a health outcome in their model identified health state utility values in a systematic way. Most relied on estimates from single studies. It is not clear that these values are either accurate or representative of the populations being modelled. We discuss the role of health state utility values and quality-adjusted life years in more detail in Chapter 3. Resource use data were also limited in many studies. Those that sought to evaluate alternative screening programmes tended to assume consistent progression to laser treatment. Modern treatment pathways are not represented in many of the studies. We give further consideration to the appropriate modelling of screening and treatment activity in Chapters 5 and 6.

Some studies — principally those published more than 10 years ago — did not clearly describe the structure or operation of their models. Very few of the studies included any validation, making it difficult to judge the appropriateness of the various modelling approaches in this context. However, Kawasaki et al. [107] provide some evidence to support the validity of cohort-based state transition modelling in the evaluation of DR screening programmes. The review revealed that TreeAge, which was used in 11 of the 26 studies identified, was the most popular software package for modelling in this context. Being based on a pre-programmed software package, these models may lack transparency [84, 119].

2.4.1 Strengths and limitations

We conducted the first review of model-based economic evaluations in diabetic retinopathy and extracted many key details regarding the structure of models and sources for parameters. Our findings facilitate a more informed development process for the structure and parameterisation of a decision model.

There are several limitations to this review. We used a restricted search strategy as our goal was not to identify all modelling studies but rather to provide an overview. It is therefore possible that the studies identified in this review are not representative of all modelling studies that have been conducted in DR. Some

of the studies that we describe were conducted more than 20 years ago and use software that is no longer available. As such, some of the studies may not be relevant to the present analysis. The extent to which our review can inform the development of a decision model is limited by the lack of validation attempts reported in the existing literature.

2.4.2 Implications

Most of the models identified in our review adopted a cohort state transition approach. The nature of disease progression and classification in DR lends itself well to this model structure. Furthermore, we seek to evaluate a screening programme and are therefore interested in outcomes at the population level. Many of the studies identified in our review evaluated alternative screening programmes, with several considering alternative screening intervals. Most of these used cohort state transition modelling. One of these studies [30] formed part of the basis for the recent NSC recommendation described in Chapter 1.

However, there may be analytical advantages to the use of an individual sampling model. In this review, studies reporting ISMs were more likely to describe validation attempts, and these tended to demonstrate validity of the models. Furthermore, ISMs are far better able to account for patient heterogeneity, which is key to the value of stratification. ISMs are therefore more likely to give externally consistent results where individual risk is used to determine treatment pathways.

In the following chapters, we use the information collected in this review to help guide the development of a new state transition model. The model will be based around the NSC 'R0M0' classifications described in Chapter 1 and used (either implicitly or explicitly) in a number of studies included in this review. Having identified a variety of shortcomings in the evidence used to populate the models described in this chapter, chapters 3 through 6 describe our efforts to improve the available evidence for this purpose. Chapter 7 will combine all of this information to guide the development and population of our own model-based economic evaluation.

Chapter 3

Quality of life and diabetic retinopathy: a cross-sectional study

Summary

As part of the ISDR randomised controlled trial, health-related quality of life data were collected from people attending screening for diabetic retinopathy. In this chapter we describe the data that were collected at baseline using the EQ-5D-5L and HUI3 questionnaires. We discuss the relationship between health-related quality of life and individuals' level of retinopathy as determined at screening. The study examines whether or not health state utility values differ according to retinopathy level and explores some of the determinants of health-related quality of life in this population. On average, people with background retinopathy reported lower health state utility values than people with no retinopathy. We also find that HUI3 values tend to be lower than EQ-5D-5L values and that this difference is associated with visual function. The chapter presents health state utility values that can be used as parameters in future model-based economic evaluations.

3.1 Background

Health-related quality of life (HRQoL) has become a key outcome measure used in the evaluation of health technologies. It captures a person's quality of life in the domain of health, where health is often defined in terms of the WHO definition as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" [120]. HRQoL is important because it is of interest to patients, while physiological markers may lack significance and not evidently impact on a person's life. It also addresses heterogeneity in individuals' responses to physiological changes, which may be detrimental to some people but irrelevant to others. Thus, HRQoL captures the effects of changes in health to the extent that they actually matter to patients.

It is possible to use either generic or condition-specific HRQoL measures. For example, research could focus on vision-related quality of life. However, in the context of economic evaluation and health technology assessment (HTA), generic measures are usually preferred. This is because they are more informative where decision makers have to make decisions across multiple disease areas. By using a generic measure of health outcome it is meaningful to compare the cost-effectiveness of treatments for diabetic retinopathy with — for example — breast cancer screening.

3.1.1 QALYs

In order to inform resource allocation decisions, HTA agencies such as the National Institute for Health and Care Excellence (NICE) prefer generic measurement of health outcomes [121]. The quality-adjusted life year (QALY) is a well-established measure of health outcome that incorporates both the quality of life associated with health states and the amount of time spent in these health states. The quality-adjustment aspect of QALY calculation generally requires the use of a generic preference-based measure (PBM) of health outcome. NICE and other HTA agencies identify QALYs as a preferred measure of health outcome [122]. An economic evaluation using QALYs is known as a cost-utility analysis because it values health outcomes according to preferences. Preferences can be elicited from the individual using a variety of methods including time trade-off, standard gamble, and choice modelling.

It is possible to directly elicit preferences from individual patients within a study. However, it is more common — and is recommended by NICE — that preferences be elicited indirectly using generic PBMs. This form of estimation involves a two stage process. First, a classification system is used in the form of a questionnaire to elicit health state profiles of individuals. Second, a valuation process is used to attach weighted values to each of these health state profiles. In

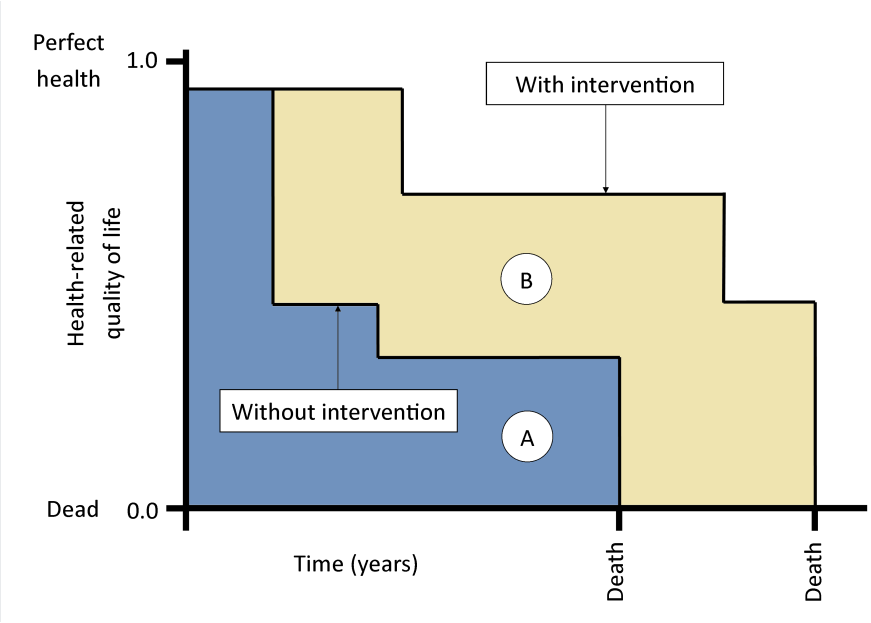

Figure 3.1: Quality-adjusted life years associated with two hypothetical pathways

the UK and elsewhere a consensus has developed that HTA should be conducted on the basis of societal values for these health states, rather than patients' preferences [123]. Societal values for health states are derived from a representative sample of the general public. Values are scaled from 0 to 1, where 1 is 'full health' and 0 is a health state of equivalent value to being dead; negative values indicate a health state worse than being dead. The resulting index value derived from each possible health state defined by the classification system is referred to as the health state utility value (HSUV). Figure 3.1 can be used to demonstrate how QALYs are estimated using an area-under-the-curve approach. Area 'A' represents the number of QALYs associated with a control group, while area 'A' plus area 'B' represents the number of QALYs associated with an intervention group.

Health state classification systems — also known as multi-attribute utility instruments — generally include multiple domains that are selected such that they provide adequate information to describe a full range of possible health states. The domains define those aspects of life that are valuable and relevant to health [124]. There are a variety of health state classification systems available for use as generic PBMs; some of the most popular include the EQ-5D, SF-6D, HUI3, AQoL, and QWB [125]. Some HTA agencies, including NICE, specify a preferred health state classification system for use as a generic PBM. NICE recommend use of the EQ-5D [121]. The EQ-5D, developed by the EuroQol Group [126], is one of the most popular and most researched health state profile elicitation questionnaires used in the UK context and across Europe. Domains used in the EQ-5D were selected on the basis of a review of existing non-preference-based HRQoL measures.

The EQ-5D has five domains: i) mobility, ii) self-care, iii) usual activities, iv) pain/discomfort, and v) anxiety/depression. Each of these domains is associated with alternative levels of response. The original version of the EQ-5D (the EQ-5D-3L) used three response levels. A newer version of the EQ-5D (the EQ-5D-5L) uses five response levels [127]. For both versions, societal values have been elicited from the general population using various techniques [128, 129]. The EQ-5D-5L value set for England is based on a combination of time trade-off and discrete choice tasks [130]. The EQ-5D-5L is still relatively new, but is seen to be preferred to the EQ-5D-3L because of a reduced ceiling effect and a more even distribution of responses [131, 132].

The EQ-5D is less dominant outside Europe, and alternative instruments are available. Some researchers have preferred the Health Utilities Index Mark 3 (HUI3) to the EQ-5D and other classification systems in the context of vision due to it having greater sensitivity to sight problems [133]. This is because the HUI3 includes specific domains relating to sensory perception [134, 135]. Recently, this was identified as a key domain missing from the EQ-5D [136]. The HUI3 was developed on a similar basis to the EQ-5D but with a greater focus on impairment, consisting eight domains: i) vision, ii) hearing, iii) speech, iv) ambulation, v) dexterity, vi) emotion, vii) cognition, and viii) pain. The index scores can be estimated using a multi-attribute utility function based on Canadian values. No UK values are currently available. While UK values are recommended by NICE, exceptions have been made where HUI3 is deemed more appropriate [137].

3.1.2 HRQoL and DR

Diabetic retinopathy is one of the most common causes of blindness and severe vision loss [138, 139], which has a substantial negative impact on HRQoL [140–143]. Effective and cost-effective preventive and remedial interventions are available for diabetic retinopathy [144, 145], which can prevent disease progression and reduce the risk of vision loss and its impact on health-related quality of life [143].

The major risk to health-related quality of life posed by the onset of diabetic retinopathy is vision loss. When considering HRQoL it is important to distinguish between vision loss in one eye or both eyes. People may be able to adapt well to having good sight in only one eye, but less able to adapt to sight impairment in both eyes. Therefore, the impact of vision loss in the better seeing eye may be greater than the impact of vision loss in the worse seeing eye. Severe sight impairment — often termed 'blindness' — can be defined in a variety of ways in terms of visual acuity and reductions in the visual field, usually in relation to the better seeing eye. The most extreme form of sight impairment is to have no light

perception.

To have no light perception (in both eyes) has been associated with utility values as low as 0.26 [146]. The purpose of screening for DR is to avoid vision loss, and the availability of a comprehensive screening programme in the UK means that the number of people progressing to vision loss is expected to reduce. However, diabetic retinopathy can still be associated with reductions in visual acuity that may impact on health-related quality of life before progression to blindness. It is therefore important to consider the utility associated with these health states. In the NDESP, individuals are invited to screening before symptoms develop and most should therefore not have experienced vision loss and the associated decrement in HRQoL. Nevertheless, people attending screening for diabetic retinopathy have diabetes and this may be associated with a reduced quality of life before the onset of retinopathy. Therefore, people with no retinopathy cannot be assumed to be either in full health or to exhibit equivalent health states to the general population.

Some people who attend screening for diabetic retinopathy may screen positive without any vision loss having occurred. These people may differ from those who screen negative in ways that affect their quality of life. This may relate directly or indirectly to their level of retinopathy or to other complications of diabetes. Therefore, it may not be correct to assume that people with asymptomatic retinopathy exhibit equivalent health states to people with no retinopathy. It may be important that model-based economic evaluations are able to differentiate between such people in order to accurately estimate the effects of treatment, though there is a lack of evidence to inform this.

In Chapter 2 we reviewed model-based cost-effectiveness analyses of interventions for diabetic retinopathy. The review highlighted several limitations in the evidence base. In the context of economic evaluation there is a tendency for decision models to rely on single HSUV estimates from a small selection of studies.

Of the 26 articles identified in the review of modelling studies, 15 made some use of HSUVs. Several studies [93, 110, 111, 117] used HSUV estimates from a series of small overlapping studies by a specific group of researchers [147–149]. Scanlon et al. [30] — whose model formed part of the basis for the NSC's recent policy recommendation regarding stratification — used values from a single study ([150]) with no justification. Furthermore, some studies only considered the utility loss associated with blindness [67, 108, 118], which may not adequately represent the effect of DR progression on quality of life.

Recent studies also describe limitations in the available evidence regarding HSUVs associated with DR [109, 151]. In order to address some of the shortcomings in the evidence base, we carried out two new studies of health-related quality

of life in people with diabetic retinopathy. This chapter describes a cross-sectional study conducted alongside the ISDR randomised controlled trial (RCT). Chapter 4 describes a synthesis of previously published HSUVs and the combination of these with the data reported in this chapter. The chapters are presented in this order due to the findings of the former being incorporated into the latter.

In this chapter, we seek to inform future decision analyses using findings from a cross-sectional study in which we associate HSUVs with screening results. Both the EQ-5D and the HUI3 have previously been used in the context of people with diabetic retinopathy, and both have been used to show that people with progressed DR have a lower HRQoL than people without [152]. However, no study has yet compared EQ-5D and HUI3 values for a cohort of people attending screening for DR. Furthermore, no study has yet reported index scores estimated using the new EQ-5D-5L for a large sample of people either with or at risk of developing diabetic retinopathy.

3.2 Methods

3.2.1 Data

Clinic sample

As described in Chapter 1, the ISDR study recruited 4,543 people to a randomised controlled trial of risk-based variable-interval screening compared with annual screening. From the beginning of trial recruitment, individuals were also recruited for additional data collection as part of Workstream D. Target recruitment for a convenience sample was at least 700 patients, as described in the trial protocol [76].

Participants were recruited at screening attendance by a research nurse, and the screening visit at which they enrolled in the study became their baseline visit for the trial. At this baseline visit, usually following administration of mydriatic eye drops, individuals in the health economics sample were asked to self-complete a paper questionnaire. The questionnaire consisted of large print versions of the EQ-5D-5L and the HUI3, as well as a set of questions relating to the costs of attending screening, as described in Chapter 5. If participants were unable to self-complete the questionnaire they were assisted by a research nurse, who could complete the questionnaire on the participant's behalf. All research nurses were provided with a guidance document that contained appropriate script to aid both self- and nurse-completion of the questionnaire. Whether or not the participants self-completed the questionnaire was recorded, and any additional notes made by the research nurse were also available.

Ethical approval was obtained from the Preston Research Ethics Committee

for all data collection, and all participants gave written informed consent.

Screen-positive follow-ups

Individuals who were otherwise eligible, but who experienced a positive screen event either at baseline or follow-up, were excluded from the trial. Most people screen negative, so it was expected that few people with STDR would be included in the sample described above. As such, we anticipated that our convenience sample would provide limited information about quality of life for people who screen positive. This group of patients may be the most likely to exhibit poorer HRQoL due to the presence of DR or other complications of diabetes, so it was deemed important to oversample this group. Therefore, an additional cohort was recruited for HRQoL data collection.

When recruited to the trial, individuals were asked if they would consent to be contacted by the research team after their involvement in the trial had ended. This meant that — for those who consented — it was possible to approach individuals excluded from the trial either at baseline or follow-up due to their screening results. All individuals who were excluded from the trial at baseline due to their screening outcome and who had consented to be contacted were sent a questionnaire consisting of the EQ-5D-5L and HUI3 by post with a return envelope. This group could include people screening positive for DR, exhibiting non-diabetic eye disease, or those for whom images were ungradable.

ISDR data warehouse

The HRQoL data collected from trial participants were combined with data from the ISDR data warehouse. This included screening outcome data from OptoMize. Because questionnaires were collected at baseline screening attendance, HRQoL data could be matched to subsequent grading and referral outcomes based on the date of screening. The OptoMize database is described in more detail in Chapter 1. For the purpose of this chapter, we extracted data for those people from whom HRQoL data were collected and associated their HRQoL questionnaire responses with 'R0M0' gradings as recorded in OptoMize. For individuals who screened positive (and who were therefore excluded from the trial), we replaced photographic grading results with gradings recorded at their subsequent visit to the hospital eye service. These data were recorded as part of the trial and were less likely to be missing. In particular, they enabled us to use outcomes data for people whose images were classified as ungradable at screening. Age is likely to be an important confounding variable associated with both HRQoL and retinopathy level, and so participants' ages in years were also extracted from the data warehouse.

3.2.2 Analysis

We present and analyse our data in line with recent recommendations [153]. For the primary analysis, EQ-5D-5L index scores were estimated using the tariff described by Devlin et al. [130], which was elicited using a combination of time trade-off and discrete choice experiments in a representative sample of 996 adults in England. Additionally, we present index values estimated using the EQ-5D-3L mapping function, as recommended by NICE [154, 155]. HUI3 index scores were estimated using the multiplicative multi-attribute utility function described by Furlong et al. [135].

We report descriptive statistics for dimension-level and index-level responses and distributions for both HRQoL measures. Results from the different cohorts are reported separately to maintain the representativeness of the sample of people attending screening. As outlined in Chapter 2, state transition models are often based on retinopathy level. Therefore, in order to inform studies of this nature, we report EQ-5D-5L and HUI3 index scores according to 'R0M0' gradings.

Health state utility data are invariably skewed, can be negative, and have an upper limit of 1. Therefore, we performed a simple linear transformation to HSUV scores and conducted generalised linear modelling to estimate the association between screening outcome and EQ-5D-5L and HUI3 index scores. Selection of the most appropriate distribution and link function was informed by the modified Park test [156], implemented using the `glmdiag` program in Stata [157, 158], link tests, and visual inspection of plotted predictions. To understand any effects associated with different data collection methods, we included a dummy variable to indicate whether the individual self-completed. This model took the form

$$HSUVd_i = \beta_0 + \beta_1 R_i + \beta_2 M_i + \beta_3 Self_i + \beta_4 Age_i + \epsilon_i, \tag{3.1}$$

where $HSUVd_i = (1 - HSUV_i)$, and $HSUV_i$ is the estimated index value (EQ-5D-5L, EQ VAS, or HUI3) for individual i. R_i and M_i are the individual's retinopathy and maculopathy levels according to the NDESP 'R0M0' grading system, $Self_i$ is a dummy variable equal to 1 if the individual self-completed and 0 if they did not, Age_i is the individual's age in years at the time that the questionnaire was completed and ϵ represents a random error term. Given the two distinct sampling methods (i.e. in clinic or by post), cluster standard errors were estimated to account for correlation within the groups.

We additionally conducted pairwise comparison between the EQ-5D-5L and HUI3 at the individual level in order to identify which groups of people responded differently to the two measures. Given the importance of vision in this context, we give further consideration to the vision domain of the HUI3. All analyses were conducted using Stata 15 [159].

	Clinic sample	**Postal sample**
Sample size	868	73
Self-completion		
Yes	573 (66%)	73 (100%)
No	271 (31%)	0 (0%)
Missing	24 (3%)	0 (0%)
Age in years		
Under 20	2 (0%)	1 (1%)
20–40	42 (5%)	3 (4%)
40–60	320 (37%)	10 (14%)
60–80	453 (52%)	46 (63%)
Over 80	51 (6%)	11 (15%)
Age missing	0 (0%)	2 (3%)
Age (mean [SD])	62 [13]	69 [14]
R0M0 gradings (worse eye)		
R0M0	674 (78%)	51 (64%)
R1M0	157 (18%)	13 (16%)
R1M1	11 (1%)	8 (10%)
R2M0	3 (0%)	3 (4%)
R2M1	1 (0%)	0 (0%)
R3M0	0 (0%)	0 (0%)
R3M1	1 (0%)	0 (0%)
Missing	21 (2%)	3 (4%)

Table 3.1: Sample characteristics

3.3 Results

3.3.1 Data

Recruitment could not be stopped immediately upon reaching our target sample of 700 due to administrative processing, and Workstream D over-recruited with 868 people asked to complete the HRQoL questionnaire at baseline. Two people who were eligible chose not to participate. Characteristics of the sample are shown in Table 3.1. The median (and mean) age in years of the clinic sample was 62 (range 17–90), and 66% self-completed. Due to an administrative error, 23 participants (2.6%) did not receive the first page of the HUI3 questionnaire, which contained the 'vision' and 'hearing' domains. These data can safely be assumed to be missing completely at random, as the printing error was not related to recruitment centre, time, or any individual participant characteristics.

In January 2017, 214 postal questionnaires were dispatched to individuals

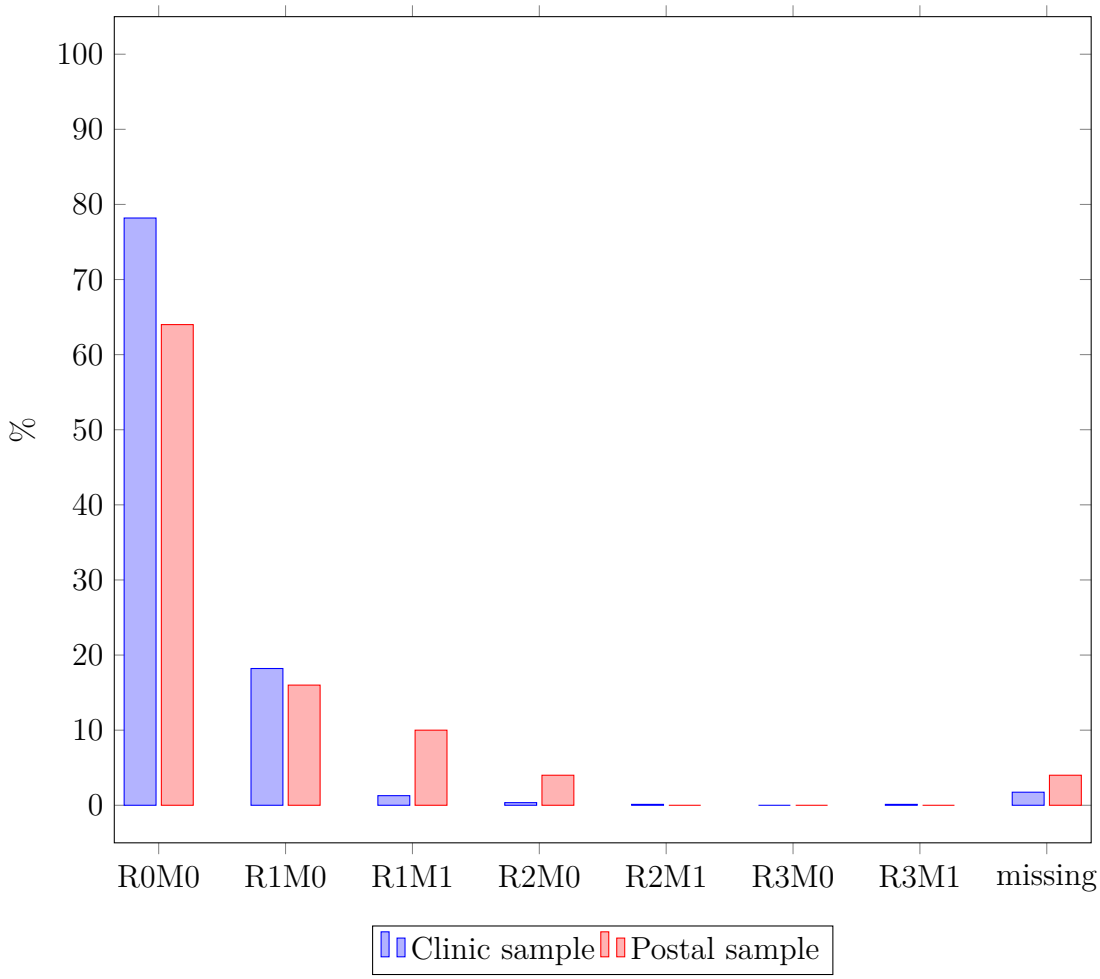

Figure 3.2: Distribution of R0M0 gradings

excluded from the trial at baseline, of which 80 (37%) were returned. For the purpose of the regression analyses, postal responses were assumed to have been self-completed. Seven individuals appeared in both samples. We excluded these individuals' postal data from our analyses. All seven reported different health states at each time point.

Figure 3.2 shows the distribution of R0M0 classifications in the two samples. In the clinic sample, 78% of those screened were graded as having no retinopathy or maculopathy, and 18% were graded as having background retinopathy without maculopathy. Gradings in the postal sample included a higher proportion with early stages of retinopathy, but nobody with PDR.

3.3.2 Domain-level responses

In the clinic sample, 823 (95%) participants fully completed all parts of the EQ-5D-5L and 735 (85%) completed all parts of the HUI3. Even excluding the 23 participants for whom the first page was missed, the 'vision' domain of the HUI3 had the highest rate of non-completion or unclear responses. Table 3.2 shows the

Dimension	Response level 1	2	3	4	5	6	Missing
EQ-5D-5L							
Mobility	51%	15%	17%	15%	0%	NA	2%
Self-care	74%	9%	9%	4%	1%	NA	2%
Usual activities	56%	15%	14%	10%	3%	NA	2%
Pain/discomfort	42%	20%	19%	13%	4%	NA	2%
Anxiety/depression	66%	14%	12%	5%	1%	NA	2%
HUI3							
Vision	28%	56%	3%	2%	1%	0%	10%
Hearing	72%	10%	6%	3%	1%	0%	8%
Speech	91%	4%	2%	0%	0%	NA	2%
Ambulation	60%	17%	11%	5%	2%	0%	5%
Dexterity	79%	14%	2%	2%	1%	0%	2%
Emotion	65%	21%	9%	3%	1%	NA	2%
Cognition	67%	8%	17%	4%	2%	0%	2%
Pain	39%	26%	15%	12%	7%	NA	1%

Table 3.2: Domain-level responses to EQ-5D-5L and HUI3 in the clinic sample. HUI3 = Health Utilities Index Mark 3; NA = not applicable.

distribution of domain-level responses, including missing data, for the 868 trial participants. For most domains, on both questionnaires, the majority of people reported to have no problems (level 1). Exceptions included 'pain/discomfort' on the EQ-5D-5L and 'pain' on the HUI3. Few people reported more severe states (levels 4, 5, or 6). Importantly, most people identified as a level 2 on the 'vision' domain of the HUI3, with only 6% of respondents indicating that their vision was not fully corrected by glasses.

Of complete responses, 251 unique states were observed from the EQ-5D-5L, with 31% of respondents reporting no problems. From the HUI3, 387 unique states were observed, with 11% reporting no problems.

3.3.3 Index values

Across both samples, including incomplete cases, the mean EQ-5D-5L index score was 0.774, compared with 0.697 for the HUI3. Individuals whose subsequent screening outcome was R1 (background retinopathy) in at least one eye had a lower HRQoL on average than individuals with R0 (no retinopathy), for both the EQ-5D-5L and HUI3 index. Median self-assessed health from the EQ-5D visual analogue scale (EQ VAS) was similar across groups, at 80 across all samples.

Table 3.3 shows mean index scores and their standard deviations for the EQ-

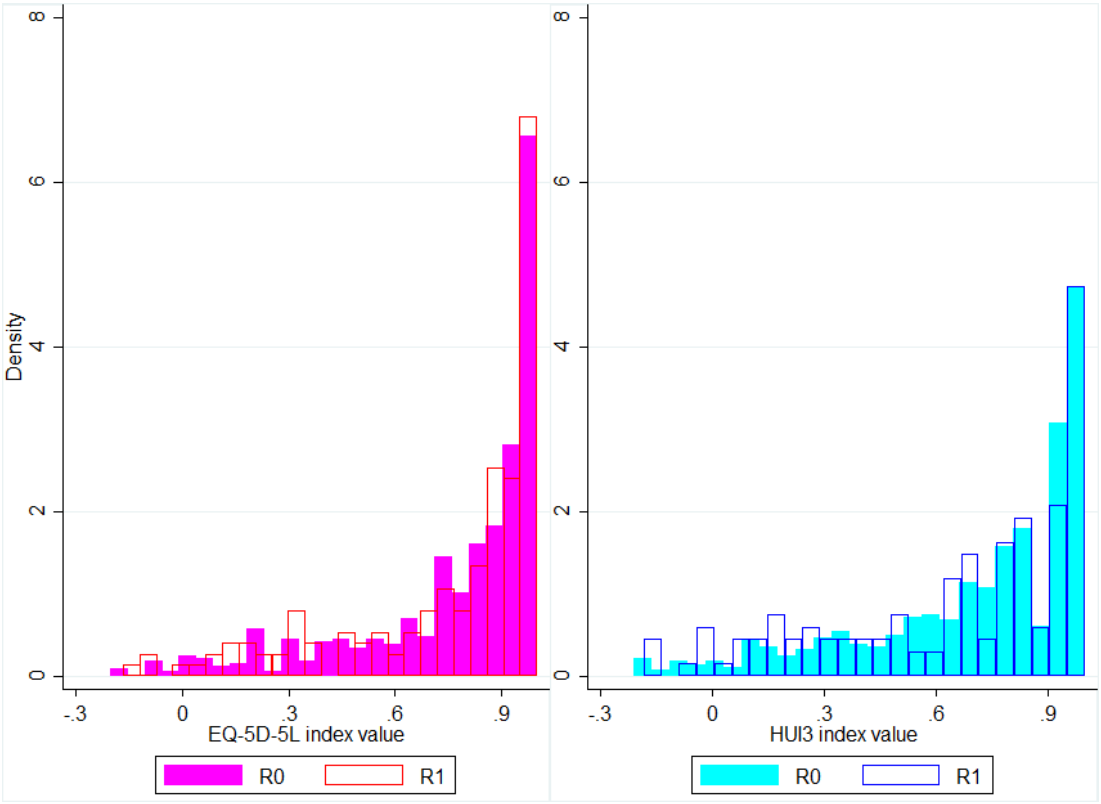

Figure 3.3: Distribution of index values for R0 and R1

5D-5L and HUI3 and for the EQ VAS, grouped by retinopathy grading, for clinic sample cases with complete HRQoL data (n=709). The results in Table 3.3 suggest that people with background retinopathy report poorer HRQoL than people with no retinopathy. For example, the mean EQ-5D-5L score was 0.787 for people with no retinopathy and 0.733 for people with background retinopathy. Due to the small numbers of respondents, we are not able to infer anything about the effect of PDR on HRQoL.

We additionally considered the potential for differences between people with background retinopathy in either one eye or both eyes, because bilateral background disease may be associated with other factors that influence HRQoL. However, index values were approximately equal in each group. For example, across both samples, mean EQ-5D-5L index values were 0.724 for people with background retinopathy in one eye, and 0.721 for people with background retinopathy in both eyes.

Figure 3.3 shows density distributions for EQ-5D-5L and HUI3 index values for people with no retinopathy and people with background retinopathy. Clearly, any difference in index values between people with no retinopathy and people with background retinopathy is subtle. However, the distributions do highlight a tendency for people with background retinopathy to report lower values, especially for HUI3.

R0M0 grading	n	EQ-5D-5L Mean (SD)	EQ-5D-3L Mean (SD)	EQ VAS Mean (SD)	HUI3 Mean (SD)
Any grading	709	0.780 (0.272)	0.723 (0.296)	74 (21)	0.711 (0.298)
R0M0	542	0.787 (0.265)	0.731 (0.286)	75 (20)	0.718 (0.291)
R1M0	134	0.733 (0.308)	0.668 (0.342)	73 (21)	0.658 (0.332)
R1M1	10	0.871 (0.229)	0.817 (0.264)	79 (22)	0.842 (0.226)
R2M0	2	0.873 (0.091)	0.774 (0.090)	70 (28)	0.879 (0.057)
R2M1	1	1 (NA)	1 (NA)	90 (NA)	1 (NA)
R3M0	0	NA	NA	NA	NA
R3M1	1	1 (NA)	1 (NA)	90 (NA)	1 (NA)
Missing	19	0.843 (0.198)	0.798 (0.227)	72 (24)	0.780 (0.238)

Table 3.3: Index scores by 'R0M0' grading outcome, clinic sample. HUI3 = Health Utilities Index Mark 3; SD = standard deviation; NA = not applicable.

The results of the regression analyses — combining both samples — are shown in Table 3.4. For consistency, EQ VAS scores were rescaled from 0–100 to a 0–1 scale. Formal tests indicated that the most appropriate model used a Poisson family distribution with log link, and visual inspection of plotted predictions supported this choice. It was necessary to collapse the small number of R2 and R3 observations in order to achieve model convergence. The inclusion of age-squared improved model fit. Due to the linear transformation, results relate to decrements in index values. Thus, a positive coefficient corresponds to a negative marginal effect on health-related quality of life. The coefficients represent the ratio effect of a unit change in that variable on the log of the mean utility decrement. For example, a coefficient associated with R1 of 0.099 shows that the mean HSUV decrement associated with R1 would be $e^{0.099} = 1.104$ times that of the decrement associated with R0.

Due to the small number of people graded as R2, R3, or M1, we cannot meaningfully interpret the effect of these states on HSUVs. However, we can interpret the effect of R1. Our findings show that both EQ-5D and HUI3 index scores were, on average, lower for people with background retinopathy when compared with people with no retinopathy. The difference was statistically significant within a 99% confidence interval for the HUI3 index value. The marginal effect of background retinopathy on HRQoL (compared with having no retinopathy), derived from the model, was a decrement of 0.024 for the EQ-5D-5L index, 0.035 for the EQ-5D-3L mapped values, and 0.067 for the HUI3. Across all indices, self-completion of the questionnaire was positively associated with HRQoL.

3.3.4 Pairwise comparison

We estimated Pearson's correlation coefficient for the EQ-5D-5L index and HUI3 index to be 0.867, indicating a relatively strong positive relationship. Figure 3.4 shows a Bland-Altman plot, which presents the average of individuals' EQ-5D-5L and HUI3 index values plotted against the difference between them. The graph also plots the average difference (dashed green line) and upper and lower limits of agreement (coloured area). The plot was generated using the `batplot` command in Stata [160]. The figure demonstrates that, on average, the EQ-5D-5L was associated with higher index values than the HUI3, and that there was a lot of variation in values across the scale. The (95%) limits of agreement were differences of -0.224 and +0.373.

Our findings suggest that the HUI3 may be sensitive to differences in retinopathy level where the EQ-5D-5L is not. A Wilcoxon signed-rank test indicated that HUI3 index values were lower than EQ-5D-5L index values across different retinopathy gradings. We conducted a (normal) linear regression analysis, with the same independent variables outlined in equation 3.1, to investigate whether

	EQ-5D-5L n=878			**EQ-5D-3L n=878**			**EQ VAS n=884**			**HUI3 n=774**		
	Coef.	**SE**	**P-value**	**Coef.**	**SE**	**P-value**	**Coef.**	**SE**	**P-value**	**Coef.**	**SE**	**P-value**
Constant	-3.661	0.016	0.000	-3.101	0.049	0.000	-1.585	0.191	0.000	-2.684	0.402	0.000
R0	(base)											
R1	0.099	0.060	0.100	0.115	0.055	0.036	0.024	0.039	0.543	0.207	0.040	0.000
R2/R3	0.187	0.037	0.000	0.149	0.079	0.060	0.204	0.070	0.003	0.309	0.461	0.502
M0	(base)											
M1	-0.206	0.186	0.267	-0.176	0.128	0.167	-0.067	0.165	0.687	-0.268	0.194	0.168
Self-complete	-0.208	0.074	0.005	-0.209	0.071	0.003	-0.166	0.058	0.004	-0.198	0.080	0.014
Age	0.067	0.004	0.000	0.057	0.001	0.000	0.014	0.009	0.128	0.034	0.009	0.000
Age squared	-0.000	0.000	0.000	-0.000	0.000	0.006	-0.000	0.000	0.197	-0.000	0.000	0.009
Deviance	273.982			280.896			159.004			216.924		

Table 3.4: Determinants of HRQoL index scores: generalised linear model results. Coef. = regression coefficient; SE = standard error; HUI3 = Health Utilities Index Mark 3.

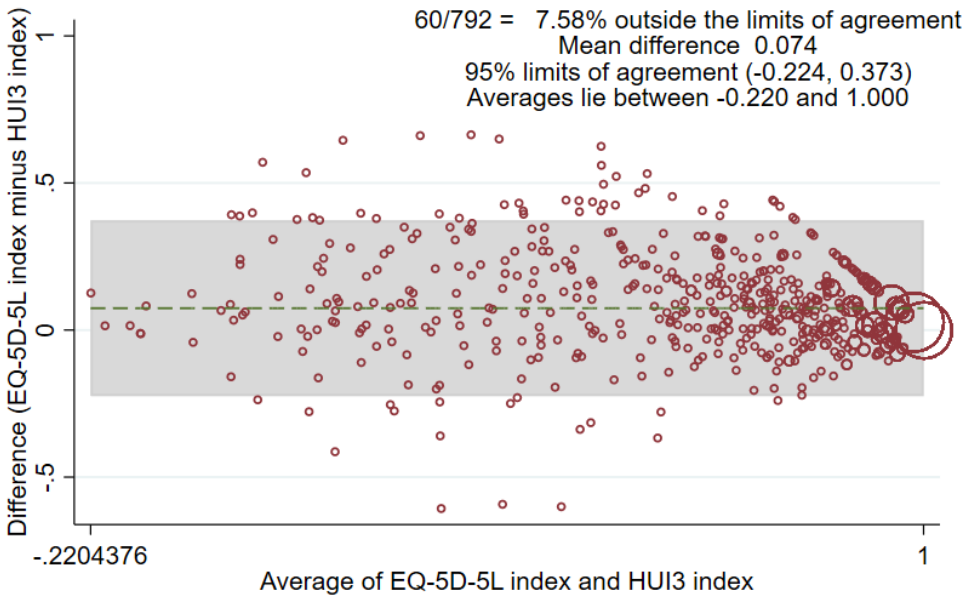

Figure 3.4: Bland-Altman plot for EQ-5D-5L and HUI3 index values. HUI3 = Health Utilities Index Mark 3

retinopathy level was an important driver of the difference between the two index scores (EQ-5D-5L index minus HUI3 index). The mean individual-level difference between EQ-5D-5L and HUI3 values was 0.011 greater for people with background retinopathy. This corresponded to a greater HRQoL decrement on the HUI3 relative to the EQ-5D-5L. However, this finding was not statistically significant.

The distribution of responses for the vision domain of the HUI3 were similar for R0 and R1 groups. Of those people with no retinopathy, 30% reported no problems on the vision domain of the HUI3, while 64% reported a level 2. Of those people with background retinopathy (R1), 32% reported no problems on the vision domain and 59% reported a level 2. Around 8% of people with background retinopathy and no maculopathy reported a level 3 or worse, while for people with no retinopathy around 6% reported a level 3 or worse. Table 3.5 shows mean index scores for individuals with complete data reporting each level of response on the vision domain of the HUI3. These findings show that the difference between EQ-5D-5L scores and HUI3 scores is positively associated with a person's level of visual impairment. This suggests that the HRQoL impacts of more advanced vision limitations, that can be included in HUI3 state descriptions, are not fully represented by EQ-5D index values. Therefore, the slightly greater number of people with background retinopathy reporting more serious vision problems partly explains the difference in HRQoL, despite it being unlikely that retinopathy is the cause of any vision problems.

Vision domain	n	EQ-5D-5L Mean (SD)	HUI3 Mean (SD)	EQ-5D-5L - HUI3 Mean (SD)
1	233	0.831 (0.247)	0.792 (0.276)	0.038 (0.141)
2	485	0.774 (0.268)	0.697 (0.288)	0.077 (0.143)
3	26	0.541 (0.294)	0.411 (0.296)	0.130 (0.171)
4	20	0.614 (0.265)	0.359 (0.263)	0.255 (0.196)
5	10	0.442 (0.374)	0.137 (0.288)	0.305 (0.240)
6	0	NA	NA	NA

Table 3.5: Index scores by HUI3 'vision' domain level. SD = standard deviation; HUI3 = Health Utilities Index Mark 3; NA = not applicable.

3.4 Discussion

This study has provided health state utility values for people attending screening for diabetic retinopathy. We found evidence that individuals with background retinopathy tend to report being in poorer health than people with no retinopathy.

A key finding of our analysis is that the HUI3 and EQ-5D-5L can provide divergent HSUVs, and that the difference between them might systematically relate to individual characteristics. This means that any measure of relative effect could differ depending on the choice of instrument. As such, it is important to consider and justify the choice of HRQoL measure in this setting.

The HSUVs identified for our sample are lower than those reported in previous studies [151, 161]. For example, Lloyd et al. reported mean values of 0.83 and 0.81 for EQ-5D (3L) and HUI3 indices in people with no retinopathy [150], while for our clinic sample the figures are 0.795 and 0.718 respectively.

Our findings support the notion that the HUI3 may be more sensitive to differences in health states in the context of vision-related disorders. However, the completion rate for the HUI3 was substantially lower than that for the EQ-5D-5L and non-completion was particularly common for the vision domain of the HUI3. This may be due to the greater complexity of the HUI3 domain descriptors and implies a trade-off in the use of the HUI3 in terms of data quality. Use of the HUI3 may result in more specific information at the cost of poorer completion. Further research is warranted to understand why respondents refuse or struggle to provide usable responses to HUI3 questions, particularly the vision domain.

As identified in Chapter 2, some previous model-based economic evaluations of screening for DR have treated screen-negative populations as homogeneous in terms of their health-related quality of life. Our findings in this chapter challenge that assumption, as individuals with background retinopathy exhibit a lower HRQoL than people with no retinopathy even when controlling for age. The estimated difference of 0.067 in the HUI3 index may exceed a minimally important

difference (see, for example, for the EQ-5D-5L [162]). Assuming that these groups are homogeneous may lead to inaccurate cost-effectiveness estimates. Despite it being unlikely that people in either of these groups have experienced any vision loss due to retinopathy, our findings suggest that the difference in HRQoL is at least partly explained by differences in visual function. It is also possible that the risk (rather than experience) of sight loss associated with background retinopathy influences people's self-reported HRQoL. The mechanism by which this operates remains unclear.

Our analysis focussed on heterogeneity in terms of disease level, but it may also be important to consider other forms of heterogeneity. For instance, age is an important determinant of HRQoL independent of retinopathy disease state. Self-completion of questionnaires was associated with higher values. This is suggestive of better functional capacity, though may reflect other contextual individual-level factors, including the fact that participants in the clinic sample would have received mydriatic eye drops. There is scope for further work to understand sources of heterogeneity in HSUVs in this population.

3.4.1 Strengths and limitations

This chapter describes the largest UK-based collection of HSUVs from people attending screening for diabetic retinopathy, and the first to use the EQ-5D-5L and to make direct comparison with the HUI3. The study had a high response rate and our approach to recruitment should ensure that the sample studied is representative of people attending screening for diabetic eye disease in Liverpool, UK. As such, our findings will be valuable to future model-based economic evaluations of interventions for DR.

There are some limitations to this study. While we collected health state utility data from a variety of people participating in screening for DR, there are some groups that we did not sample. In particular, we were not able to obtain data from people not attending screening. Furthermore, we did not collect sufficient data from individuals who screened positive to make assertions about the impact of proliferative retinopathy and maculopathy on HRQoL. We also did not collect information from people in receipt of treatment. Some treatments could be detrimental to HRQoL in the short term due to their invasive nature [163].

Our results may have been affected by the context in which data were collected from the clinic sample – that is, at a screening attendance. Respondents completed the questionnaire at the most opportune time, which was usually following receipt of mydriatic eye drops, which affect vision and can be unpleasant. We are not able to identify whether individuals completed the questionnaire before or after receiving eye drops. It is possible that receipt of eye drops may

explain some of the heterogeneity in responses, which we are unable to capture, particularly between the clinic sample and the postal sample.

Our statistical model is based on gradings in the worse eye. Evidence suggests that vision affects quality of life more strongly in accordance with better seeing eye [164]. Therefore, we might expect the level of disease in the worse eye to be indicative of the impact of diabetes and other health problems rather than vision.

3.4.2 Implications

We have identified some important dynamics in the HRQoL of people attending screening for diabetic retinopathy, which haven't previously been discussed in the literature. People with background retinopathy tend to report poorer HRQoL than people with diabetes who exhibit no signs of retinopathy. Therefore, it may be important to distinguish between individuals with no retinopathy and those with background retinopathy, in terms of HRQoL, when developing model-based cost-effectiveness analyses of screening programmes. These findings will inform the development of the cost-effectiveness model described later in this thesis.

Future research should investigate the determinants of HRQoL in people attending screening for diabetic retinopathy, and the interplay between retinopathy level and visual function. We also found evidence that the EQ-5D-5L is likely to provide more complete HRQoL data in this population, compared with the HUI3. Researchers should further explore the reasons for this and consider its implications when formulating data collection strategies.

Informed by the findings of this study, the model described in Chapter 7 distinguishes between people with background retinopathy and people with no retinopathy, in terms of HRQoL. The point estimates identified in this study are used in Chapter 4 and ultimately to guide the identification of model parameters in Chapter 7.

Chapter 4

Quality of life and diabetic retinopathy: a systematic review and meta-analysis

Summary

People with diabetic retinopathy tend to have lower levels of health-related quality of life than individuals with no retinopathy. Strategies for screening and treatment have been shown to be cost-effective. In order to reduce the bias in these cost-effectiveness estimates, systematic reviews of health state utility values (HSUVs) are recommended for health technology assessment and the development of decision analytic models. However, most models adopt a single source and are therefore susceptible to biased estimates. A synthesis of HSUVs for disease states in diabetic retinopathy has not previously been carried out. We conducted a systematic review of studies reporting HSUVs for people with diabetic retinopathy, in correspondence with disease progression. MEDLINE, Embase, EconLit, and other databases were searched to identify relevant articles. Data from included studies were extracted and subsequently synthesised using linear mixed effects modelling meta-regression, incorporating our findings from Chapter 3. Reported disease severity classifications were mapped to a four-level grading scale for diabetic retinopathy. The search identified 1,472 unique citations, from which 41 studies were included for analysis. Far more articles were identified than in previous reviews. In light of this we discuss appropriate methods for searching and screening citations to identify HSUVs. The findings from our meta-regression can be used by analysts to identify relevant HSUVs for model-based evaluations of interventions for diabetic retinopathy. We discuss the role of meta-analysis in estimating parameters for model-based economic evaluation with a view to informing best practice.

4.1 Background

When a new technology is developed, an economic evaluation is often conducted to determine whether it is cost-effective. It is common for such economic evaluations to require decision modelling, where evidence is synthesised to determine long-term costs and health outcomes. In order to minimise bias in these decision analyses, it is crucial that the process of selecting the evidence for the model is robust, transparent, and systematic [165]. This is stipulated as a requirement by the National Institute for Health and Care Excellence (NICE), which publishes guidance for the NHS based on the clinical and cost-effectiveness evidence for health technologies [166].

As outlined in Chapter 3, NICE and other HTA agencies identify quality-adjusted life years (QALYs) as their preferred outcome measure. In order to estimate QALYs in a decision modelling framework, it is usually necessary to specify health state utility values (HSUVs) that can be associated with states or events within the model. It has been identified that many submissions to NICE do not satisfy their requirements with regard to the transparent and systematic selection of HSUVs [137]. A growing number of systematic reviews are being carried out to inform better selection of HSUVs [151, 167–186], with some synthesising the data using meta-analysis [168–171, 173, 174, 176, 177, 181, 183–185]. Such an approach is increasingly being seen as an important step in the process of a model-based economic evaluation [187]. However, meta-analysis is still seldom used to identify HSUV parameters for decision modelling. There is debate about the proper use of meta-analysis in this context [188], and so this thesis seeks to contribute to this debate and to inform the development of best practice.

Chapter 2 identified that many existing modelling studies in the context of diabetic retinopathy do not select HSUVs in a systematic and transparent way. Some researchers choose HSUVs based on estimated visual acuity levels, rather than on the disease state itself [116, 151]. It is unclear whether or not the effect of visual acuity on HSUVs is consistent across different visual disorders [161]. Any given level of visual acuity may be associated with different levels of health-related quality of life because of the diverse impacts of disease on vision. Furthermore, the prevalence of comorbidities and concurrent health problems may differ across visual disorders. Therefore, such an approach is unlikely to be valid in accurately estimating the impact of an intervention on health-related quality of life. Most modelling studies in the context of DR differentiate between health states based on the stages of disease progression, and appropriate HSUVs for these disease states need to be identified.

We carried out a systematic review and meta-analysis in order to identify HSUVs associated with different levels of diabetic retinopathy, as reported in the

literature. The principal aim of this chapter is to identify HSUVs for use in the model described in Chapter 7. However, our findings will be informative to the development of other models. Although our focus is upon HSUVs associated with specific stages of disease progression in DR, we also review HSUVs associated with visual function in people with DR. The meta-analysis is carried out with a view to enabling modellers to estimate the most appropriate HSUVs with which to populate their models. This chapter as such reports on several aims of the thesis:

1. To provide a narrative overview of published studies reporting HSUVs for diabetic retinopathy.
2. To derive pooled estimates for HSUVs that correspond to disease states based on the most commonly used disease classification systems.
3. To quantify the effects on reported HSUVs of differences in study design.
4. To map reported values to a consistent grading scale for use in a modelling study.

4.2 Methods

Guidelines such as the PRISMA statement [189] are not wholly applicable to the review, though we developed our methods in line with published recommendations [190–194] where appropriate. The Patient, Intervention, Comparison, Outcome (PICO) question is not usually applicable to reviews of HSUVs [187]. For example, this review does not focus on a specific intervention or comparator. Furthermore, it is also necessary to define additional requirements; for example, it is important to define which HSUV elicitation methods will be included.

The protocol for this study was registered with a database of prospectively registered systematic reviews (PROSPERO; registration number: CRD42014012891) and published before commencement of the review [3].

4.2.1 Search strategy

Our search was necessarily broad due to inconsistency in the nature and reporting of studies that include HSUVs. Databases for searching included MEDLINE, Embase, Web of Science, Cost-Effectiveness Analysis Registry, Centre for Reviews and Dissemination Database, and EconLit. Specific pre-defined thesaurus terms for HSUVs do not exist, though broader terms may apply. Our search used general, instrument-specific, and method-specific terms, which were combined with terms for diabetic retinopathy. Diabetic retinopathy is a term which is used to describe progressive retinal changes (for example no DR/background DR/pre-proliferative DR/PDR) but is also used broadly to cover diabetic retinopathy and

maculopathy. We include studies that report HSUVs for maculopathy. Given that HSUVs are often reported as secondary outcomes, it was deemed likely that they would not be mentioned in titles or abstracts. As such it was necessary to carry out full text searches. The search terms used for Ovid (including MEDLINE and Embase) are outlined in Table 4.1. No date or language restrictions were applied to the electronic searches.

4.2.2 Study eligibility

Inclusion criteria

Studies of any design were included, and it was expected that all could be categorised as either clinical decision analyses or outcomes studies. Clinical decision analysis studies include randomised controlled trials and economic evaluations, while outcomes studies are those designed specifically to elicit HSUVs.

Studies must have used a recognised method of direct (for example, standard gamble or time trade-off) or indirect (for example, EQ-5D or Health Utilities Index) utility assessment. HSUVs from visual analogue scales, mapping algorithms, and expert opinion were also included. The language of publication must be English.

All types of publication (both full publications and abstracts) were included. The population included is people with diabetic retinopathy or those attending screening for diabetic retinopathy. There are no inclusion criteria relating to specific interventions or technologies. The comparator element of the PICO statement does not apply. The studies must report either mean or median HSUVs and these must be related either to visual function or disease states specific to diabetic retinopathy. Such disease states are likely to be based on the grading systems described in Table 1.1, though studies adopting other retinopathy grading systems were included.

Exclusion criteria

We excluded editorials, reviews, and meta-analyses that do not report original data. Studies that report data from health state classification systems but do not estimate HSUVs were also excluded.

4.2.3 Data collection

Study selection

Studies were initially assessed for retrieval based on titles and abstracts. It was expected that many titles and abstracts would not mention HSUVs, despite their inclusion. Indeed, it has been reported that rejecting studies based on title and

Step	Search term
1	exp diabetic retinopathy/
2	(diabetic retinopathy or stdr or dmo or dme).af.
3	diabet$.af.
4	(maculopathy or macular edema or macular oedema or csmo or csme).af.
5	3 and 4
6	1 or 2 or 5
7	exp quality adjusted life year/
8	quality adjusted life$.af.
9	disability adjusted life$.af.
10	(daly$ or qaly$ or qald$ or qale$ or qtime$).af.
11	hsuv.af.
12	health status/
13	health state.af.
14	(sf6d or sf 6d or short form 6d or shortform 6d or sf six d or sfsixd or shortform six d or short form six d).af.
15	(euroqol or euro qol or euro-qol or eq5d$ or eq 5d$ or eq-5d$ or rosser).af.
16	(hql or hqol or hrqol or hrql).af.
17	(healthy years equivalent or hye$).af.
18	(hui or hui1 or hui2 or hui3).af.
19	(15d or 15 d).af.
20	aqol$.af.
21	addqol$.af.
22	disutilit$.af.
23	(qwb or wellbeing or well-being or well being).af.
24	(standard gamble or sg).af.
25	(time trade off or time trade-off or time tradeoff or tto).af.
26	(person trade off or person trade-off or person tradeoff or pto).af.
27	(visual analogue scale or vas).af.
28	exp quality of life/
29	(quality adj2 life).af.
30	(cost utility or cost-utility or cua).af.
31	7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30
32	6 and 31

Table 4.1: Ovid search terms

abstract can result in lost citations when reviewing HSUVs [190]. Articles were rejected based on the title and abstract only if it was clear that the study could not have recorded the necessary data for inclusion. Full texts were retrieved for studies not rejected at abstract screening. These were assessed for satisfaction of the inclusion and exclusion criteria. Reasons for exclusion were recorded. The number of records identified, retrieved, screened, assessed, included, and excluded in the review, and reasons for exclusions, are summarised in a PRISMA flow diagram [189].

After commencing the review, we found that studies could not easily be excluded based on title or abstract and that the screening process therefore had a very low specificity. Title and abstract screening was therefore abandoned and full text screening was conducted for all records identified in order to streamline the process. All screening of abstracts and full texts was conducted by a single reviewer.

Data extraction and management

Data were extracted using an electronic data extraction form and automatically recorded in a spreadsheet. The data extraction form is included in Appendix C. A separate form was completed for each reported HSUV as studies could include different sub-populations; for example, from different countries. Based on previous reviews of HSUVs (cited above), we recorded the following information for each study: first author, publication year, study title, publication name, study design, interventions/comparators, sample size, and the number of separate HSUVs reported in the paper. For each reported HSUV we recorded: point estimate type (mean/median), reported HSUV, measure of variance type (standard deviation/variance), reported variance statistic, retinopathy state, maculopathy state, grading system used, visual function measurement method, visual acuity, sample size, country, age range, other sample specifics, valuation method, valuation source, value set country, upper anchor, lower anchor, administration method, study arm, and treatment status. Where data were reported for individuals without retinopathy — for example, for those attending screening — these data were also recorded. The data extraction form was successfully piloted on three pre-identified studies. All data extraction was conducted by a single reviewer.

After commencing the data extraction it became clear that it would be valuable to extract some additional information, namely: whether the retinopathy or visual function level was defined by the better eye or worse eye. This information was also extracted.

Rank	Data components
1	Direct utility assessment for the specific study from a sample either: (a) of the general population (b) with knowledge of the disease(s) of interest (c) of patients with the disease(s) of interest Indirect utility assessment from specific study from patient sample with disease(s) of interest, using a tool validated for the patient population
2	Indirect utility assessment from a patient sample with disease(s) of interest, using tool not validated for the patient population
3	Direct utility assessment from a previous study from a sample either (a) of the general population (b) with knowledge of the disease(s) of interest (c) of patients with the disease(s) of interest Indirect utility assessment from previous study from patient sample with disease(s) of interest, using a tool validated for the patient population
4	Unsourced utility data from previous study - method of elicitation unknown
5	Patient preference values obtained from a visual analogue scale
6	Delphi panels, expert opinion

Table 4.2: Cooper quality ranking for HSUV data [195]

Quality and relevance assessment

Standard means of assessing quality in systematic reviews are not appropriate for reviews of HSUVs, as they may be at odds with the quality of the evidence reported. For example, though randomised controlled trial data may be the 'gold standard' for capturing treatment effects, such a study design may be inferior when eliciting HSUVs due to low external validity or lack of relevance. There is limited guidance for assessing the methodological quality of studies reporting HSUVs. As a starting point, we use the hierarchy of data sources presented by Cooper and colleagues [195] to rank studies from 1 to 6, as shown in Table 4.2. In order to enable researchers to judge the quality of the study, our data extraction form additionally recorded: study sample size, inclusion/exclusion criteria, response rates, loss to follow-up, missing data, and the Cooper rank. Reporting quality is not formally assessed, but will be indicated by the completeness of the data extraction for each study.

The relevance of studies to particular research questions may be more important than quality. For example, the extent to which a study's results can be used to satisfy the NICE reference case may be crucial. Furthermore, the determinants of relevance may differ for future users of the review. Our data extraction form recorded information that will enable users of the review to judge the relevance of the reported HSUV, namely information on interventions/comparators, inclusion/exclusion criteria, country, age range, other sample specifics, valuation method, and valuation source. There is currently no accepted generalisable method for assessing the relevance of HSUVs for a particular study. As such, these data are summarised qualitatively.

4.2.4 Data synthesis and presentation

All HSUVs and the characteristics of their associated studies were tabulated. Saramego et al. identify that quantitative synthesis of aggregate preference-based HSUVs is limited by i) between-study heterogeneity in instruments used, ii) the value set used to quantify utilities, and iii) the models used to approximate scores for health states [192]. Furthermore, a previous review, which reviewed HSUVs associated with different visual acuity levels in diabetic retinopathy, found variation in the methods of elicitation [151]. It is important to measure the effect of these methodological differences on HSUV estimates, and meta-regression is a way of achieving this. Our prior knowledge of the literature suggested that our review would provide sufficient data to carry out an analysis of this kind. A meta-regression model can also facilitate the prediction of expected HSUVs for a given set of study characteristics.

The data synthesis included our findings from Chapter 3. For the most commonly used retinopathy grading systems, HSUVs were pooled, with observations weighted by the inverse of the variance of the mean HSUV, such that:

$$Y_i = y_i(\frac{n_i}{\sigma_i^2}) \tag{4.1}$$

where y_i is the observed average HSUV and Y_i the weighted average when n_i is the number of respondents and σ_i^2 the observed variance of y_i. In order to address the limitations of HSUV synthesis previously identified [192], we used linear mixed-effects modelling to account for fixed and random effects associated with between-study heterogeneity. This methodology is consistent with previous studies [168, 183]. The model allowed for random variation on three levels: i) variation between mean HSUVs across studies, ii) variation between mean HSUVs across groups of individuals within studies, and iii) error variation. Studies tend to report multiple HSUVs from overlapping samples, so the meta-regression adopted

a hierarchical approach such that:

$$Y_{ijk} = \beta_0 + \sum_h \beta_h x_{hijk} + v_k + u_{jk} + \epsilon_{ijk} \tag{4.2}$$

where Y_{ijk} is the weighted mean of the ith HSUV of the jth group being estimated for study k, x_{ijk} are the variables used to explain the between study heterogeneity, v_k is the random effects term of study i, u_{jk} is the random effects term for the jth group of study k, and ϵ_{ijk} is the random error term with fixed variance to be estimated.

Predictor variables were generated to include retinopathy state, maculopathy state, publication year, study design, country, valuation method, valuation source, and administration method. We explored the inclusion of other covariates and used a stepwise procedure of model selection in order to reduce the likelihood of errors. We tested for heteroscedasticity associated with the inclusion of particular predictor variables. Covariates were only included where the existing evidence suggested that an association with the HSUV outcome might exist. We estimated variance inflation factors to test for collinearity, and any strongly correlated variables were removed or collapsed where possible. Selection of variables to be included in x_{ijk} in the final model was informed by the Akaike information criterion. The base case was — as far as possible — matched to the NICE reference case (for example, using EQ-5D values) [166]. If a study did not have sufficient data for inclusion in the model, the data were assumed missing completely at random and the study was dropped from the model.

We additionally mapped values to a disease state classification with four levels of retinopathy and two levels of maculopathy. The mapped value for each HSUV was recorded using the data extraction sheet and the mapping of the states was subsequently agreed with a clinician. We used the same regression methods described above to pool values based on these classifications. We estimated the intraclass correlation coefficient associated with studies classified in this way when no moderators were included, in order to quantify the heterogeneity associated with such an approach. Publication bias was not a concern in the review, as HSUVs are usually used as a secondary outcome and therefore should not influence the likelihood of publication.

All analyses were conducted in Stata 15 and the meta-regression used the `gllamm` package [196].

4.3 Results

4.3.1 Study selection

After deduplication, 1,472 unique studies were identified. Figure 4.1 shows a PRISMA flow diagram. Full-text screening was carried out for the full set to review for inclusion and exclusion criteria. Full texts for a small number of studies could not be retrieved via the Internet. However, titles and abstracts for all of these studies were screened and it was determined that they were not likely to contain relevant data. 1,415 studies were excluded on the basis of the criteria outlined above, and data extraction was conducted on 57 articles. At this stage, 16 studies were excluded due to missing data, principally because the studies only reported mean utility decrements or regression coefficients from which absolute estimates could not be derived [197–212]. The meta-analysis therefore included 41 studies plus our findings from Chapter 3.

4.3.2 Study characteristics

Table D in the appendices lists the key characteristics of the 41 studies included in the meta-analysis. Of these, 36 were outcomes studies and five were clinical evaluations. The sample sizes of the studies varied greatly, from 68 to 20,705. A variety of retinopathy grading systems were used, but only 12 studies explicitly differentiated between different levels of retinopathy. A majority of the studies incorporated visual acuity in the definition of states, with nine studies reporting values based only on vision states without any distinct classification of retinopathy levels.

4.3.3 Study findings

The studies identified by the review together reported 317 average HSUVs with a combined number of observations of almost 70,000; a far greater number than anticipated. HSUVs from all studies are therefore not presented in this thesis.

Table 4.3 shows the range of HSUVs reported for each classification of disease reported across studies. VAS values reported on a 0–100 interval are rescaled to 0–1. Tables 4.4 to 4.6 show the range of HSUVs reported for each classification of visual acuity, where values are converted to a decimal for consistency where necessary. Tables 4.3 and 4.4 highlight the range of classifications and groupings that are used in the published literature and — where more than one study reports values — the between-study heterogeneity associated with average HSUVs. For example, average observed values for people with PDR ranged from 0.520–0.931, while HSUVs for people with DMO ranged from 0.547–0.860. Tables 4.5 and 4.6 show results where only better- or worse-eye VA is available, for clarity of

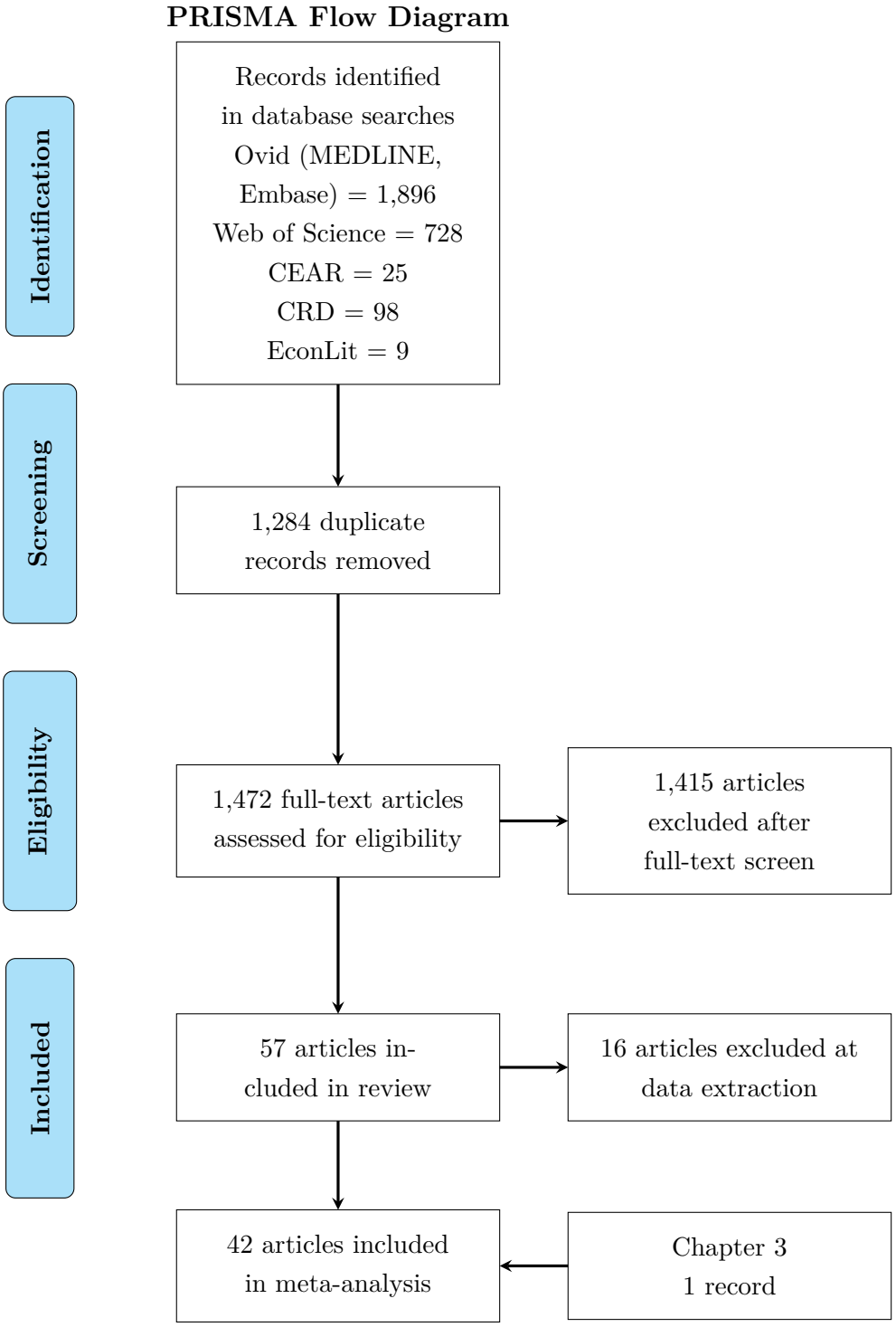

Figure 4.1: PRISMA flow diagram for review of HSUV values

presentation. These results show that while HSUV estimates based on level of disease tend to be associated with worse seeing eye, HSUV estimates based on visual acuity tend to be associated with better seeing eye.

Table 4.7 shows the unweighted mean average HSUV associated with alternative elicitation techniques for mapped R0M0 values. All methods fail to show a gradual decline in HRQoL associated with progression of DR. Only blindness and proliferative retinopathy are shown to consistently have a large negative association with HRQoL. Generally speaking, direct elicitation techniques — such as standard gamble and time trade-off — appear to provide higher values than indirect methods such as the EQ-5D.

4.3.4 Meta-analysis

A large number of studies reported median rather than mean values. Therefore, in order to increase the number of studies included in the regression analysis, a normal distribution was assumed and medians, interquartile ranges, and confidence intervals were used to generate a weighted mean for studies that did not report mean values.

Generally, the R0M0 mapping was not associated with strong intraclass correlation (ICC). For retinopathy, good intraclass correlation was found between samples within studies, but only fair agreement between studies. For maculopathy, agreement was poor at both levels.

The results of our main model, using weighted means with our assumption about median values as described above, are shown in Table 4.8. The model includes 198 HSUVs from 87 subgroups from 25 studies (including our findings from Chapter 3), based on a sum total of 21,570 observations. The final model — the selection of which was informed by the AIC — included mapped retinopathy state, whether or not the state related to blindness, whether the study was an outcomes study or clinical evaluation, whether the state being valued was a patient's own or hypothetical, whether the valuation source was patients or a public sample, the valuation method, and the mode of administration. Other variables were judged not to add enough information to warrant inclusion. Few studies reported both maculopathy and retinopathy level. When included in our main model, maculopathy did not demonstrate a statistically significant effect. It was on the basis of this and of the weak ICC that maculopathy was excluded from the model. The reference case for the model was a state with no retinopathy and no blindness in either eye, collected by face-to-face interview in an outcomes study, whereby individuals assessed their own health state using EQ-5D with index values estimated on the basis of public valuation.

The mean HSUV for the reference case was 0.699. R1 and R2 retinopathy states were associated with effects in the expected (negative) direction, but were

Retinopathy		Maculopathy		Lower estimate		Upper estimate	
Better eye	**Worse eye**	**Better eye**	**Worse eye**	**Mean**	**Source**	**Mean**	**Source**
	Any DR			.300	[150]	.920	[147]
	No DR			.561	[213]	.940	[214]
	No DR / NPDR			.965	[215]		
	BR			.720	[152]	.850	[152]
BR				.747	[152]	.860	[152]
	Mild or moderate NPDR			.740	[216]	.800	[216]
	NPDR			.690†	[217]	.870	[214]
	Severe NPDR / PDR			.660	[216]	.700	[216]
	STDR			.700	[113]		
	PDR			.520†	[217]	.931	[215]
PDR				.767	[152]	.870	[152]
	(either) PDR		(or) CSMO	.668	[218]		
	Symptomatic DR			.530	[219]		
	At least A1			.703	[220]	.860	[220]
			Any DM	.694	[221]	.756	[221]
			DMO	.547	[99]	.860	[99, 152, 222]
		DMO		.769	[152]	.890	[152]
	13-15 / 10/20			.800†	[223]	1.00†	[224]
	20 / 30			.760†	[223]	1.00†	[224]
	31-41/40			.730†	[223]	.990†	[224]
	51, 60-80 / 50			.950†	[224]		
	31,63			.780	[225]	.810	[225]
	41,64			.770	[225]	.780	[225]
	51, 64			.730	[225]	.760	[225]

Table 4.3: DR groups. †median

Better eye	Worse eye	Lower estimate		Upper estimate	
		Mean	Source	Mean	Source
1.00	1.00	.600	[226]	.930	[226]
.80–1.00	<.50	.850	[147]	.900	[147]
.80–1.05	<.50	.840	[227]		
.50–1.00	<.10	.810	[228]	.980	[228]
.40–.67	<.50	.780	[147, 227]	.920	[147]
.25–.40	<.10	.760	[228]	.910	[228]
<.29	<.50	.780	[227]		
.20–.33	<.50	.780	[147]	.840	[147]
.13–.20	<.10	.700	[228]		
.05–.10	<.5	.640	[147]	.710	[147]
<.10	<.10	.550	[113]	.670	[228]
blind	<.50	.590	[147]	.700	[147]
blind	blind	.200	[229]	.760	[229]

Table 4.4: VA groups, both eyes

not statistically significant within a 95% confidence interval. R3 (PDR) was associated with a 0.024 decrement to HRQoL, while blindness in at least one eye was associated with a decrement of 0.172. Clinical evaluations within the model were predicted to report HSUVs 0.157 lower on average, while collection of data by telephone was associated with a slightly higher reported HSUV.

Most alternative HSUV elicitation methods were associated with important differences in mean values. Values derived using the 15D, SF-6D, standard gamble, or VisQoL, were higher than the EQ-5D index, while visual analogue scale and time trade-off values tended to be lower. In general, methodological choices were more important predictors of HSUVs than disease states.

Figure 4.2 plots observed values against those predicted by the model, with observations weighted by the sample size associated with each estimate. The plot highlights the limitations of the predictive model. Several small samples, as well as some medium-sized samples, reported values far from those predicted by the model.

4.4 Discussion

We conducted an extensive review of the literature, identifying 57 articles suitable for data extraction, with 41 ultimately included in a meta-analysis in addition to our study reported in Chapter 3. Studies varied greatly in terms of their reporting of HSUVs and the underlying health states with which they were as-

Better eye	Lower estimate		Upper estimate	
	Mean	Source	Mean	Source
1.00	.860	[230]		
1.05–2.00	.860	[99]		
>.80	.800†	[164]	.881	[231]
.80–1.00	.910	[232]		
.80–2.00	.762	[152]	.830	[152]
.67–1.00	.738	[150]	.860	[99]
>.67	.990†	[224]		
<.67	.860	[232]	.990†	[224]
>.50	.660†	[217]	.780†	[223]
<.50	.760†	[223]		
.50–.67	.800	[230]		
.42–.63	.813	[99]		
.40–.67	.786	[231]		
>.33	.980†	[224]		
.33–.50	.300	[150]	.750	[150]
.32–.63	.758	[152]	.840	[152]
<.32	.730†	[164]		
.26–.40	.802	[99]		
<.25	.610	[150]	.700	[150]
.20–.40	.770	[230]		
.20–.33	.728	[231]		
.17–.33	.950†	[224]		
.17–.25	.770	[99]		
<.17	.880†	[224]		
.13–.25	.530	[152]	.850	[152]
.10–.25	.630	[216]		
.10–.16	.760	[99]		
<.10	.400	[152]	.810	[214]
.07–.10	.681	[99]		
.05–.10	.520	[150]	.730	[231]
0–.06	.547	[99]		

Table 4.5: VA groups, better eye only. †median

Worse eye	Lower estimate		Upper estimate	
	Mean	Source	Mean	Source
.80–2.00	.760	[152]	.820	[152]
.67–2.00	.860	[222]		
<.50	.540	[227]		
.32–.63	.778	[152]	.900	[152]
.25–2.00	.722	[233]	.780	[233]
.13–.25	.675	[152]	.810	[152]
.10–.25	.763	[233]	.800	[233]
<.10	.390	[152]	.770	[233]
0–.06	.550	[222]		
blind	.380	[219]	.740	[226]

Table 4.6: VA groups, worse eye only

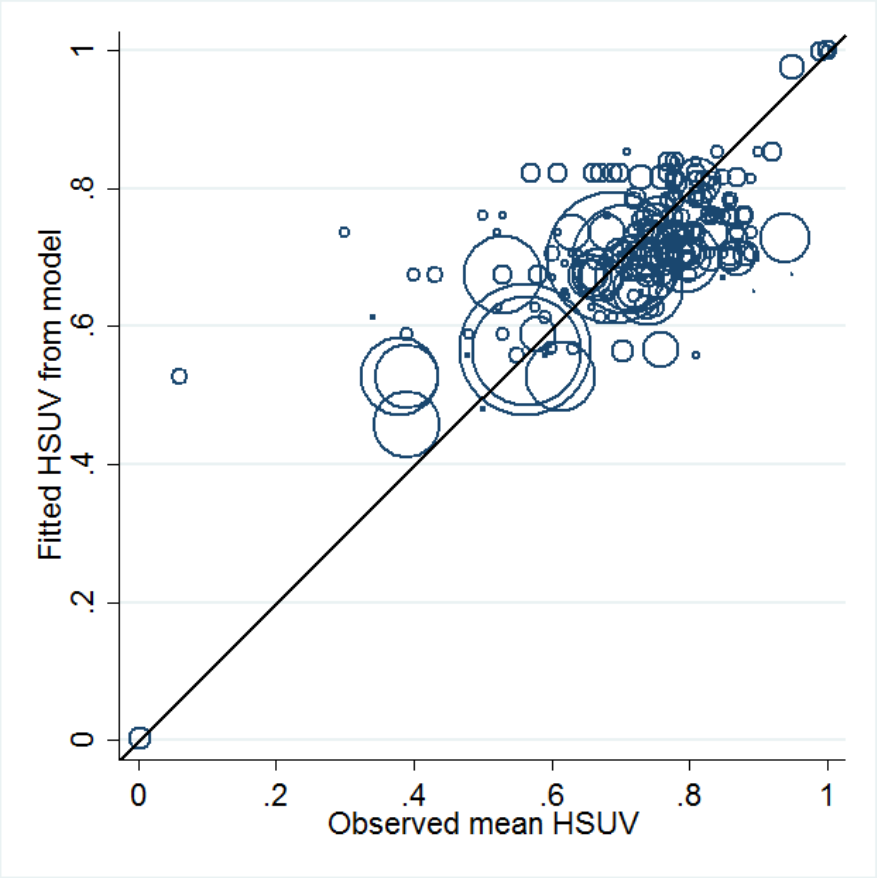

Figure 4.2: Observed versus fitted HSUVs for the model. The diagonal line indicates equivalence. The areas of the circles are weighted by the study sample size.

	R0M0	**R1M0**	**R2M0**	**R3M0**	**R1M1**	**R2M1**	**R3M1**	**Blind**
EQ-5D	.760	.700	.753	.739	.772	-	.626	.501
index	(.014)	(.027)	(.037)	(.036)	(.011)	(-)	(.048)	(.086)
	[31]	[12]	[4]	[9]	[24]	[0]	[5]	[7]
EQ	.709	.639	.703	.762	.720	-	.624	.584
VAS	(.019)	(.063)	(-)	(.006)	(.018)	(-)	(.047)	(.029)
	[20]	[5]	[1]	[4]	[8]	[0]	[4]	[4]
HUI3	.736	.830	-	.845	.827	-	.553	.443
	(.047)	(.025)	(-)	(.021)	(.020)	(-)	(.100)	(.038)
	[15]	[3]	[0]	[4]	[3]	[0]	[4]	[4]
15D	.889	.781	.826	.826	-	-	-	-
	(.077)	(-)	(-)	(.105)	(-)	(-)	(-)	(-)
	[2]	[1]	[1]	[2]	[0]	[0]	[0]	[0]
SF-6D	.757	.774	.775	.745	-	-	-	-
	(-)	(.023)	(.005)	(.015)	(-)	(-)	(-)	(-)
	[1]	[3]	[2]	[2]	[0]	[0]	[0]	[0]
SG	.808	-	-	.843	-	-	.698	.603
	(.076)	(-)	(-)	(.047)	(-)	(-)	(.023)	(.044)
	[4]	[0]	[0]	[4]	[0]	[0]	[12]	[9]
TTO	.817	.807	.787	.763	.847	-	-	.560
	(.016)	(.014)	(.049)	(.018)	(.022)	(-)	(-)	(.043)
	[28]	[6]	[3]	[30]	[3]	[0]	[0]	[11]
Any	.775	.739	.787	.775	.773	-	.648	.525
	(.011)	(.019)	(.025)	(.013)	(.010)	(-)	(.024)	(.026)
	[110]	[34]	[12]	[56]	[39]	[0]	[25]	[41]

Table 4.7: Mean (standard error) [n] unweighted scores associated with R0M0 states

	Coef.	**95% CI**
Intercept	0.699	0.689 – 0.708
R0	Reference	
R1	-0.001	-0.009 – 0.006
R2	-0.002	-0.014 – 0.011
R3	-0.024	-0.047 – -0.001
Not blind	Reference	
Blind in at least one eye	-0.172	-0.201 – -0.142
Outcomes study	Reference	
Clinical evaluation	-0.157	-0.181 – -0.133
Own health state	Reference	
Hypothetical health state	-0.030	-0.063 – 0.003
Public valuation	Reference	
Patient valuation	0.087	0.070 – 0.103
EQ-5D index	Reference	
15D	0.111	0.099 – 0.124
EQ VAS	-0.133	-0.144 – -0.122
HUI3	-0.025	-0.082 – 0.031
QWB-mapped	-0.007	-0.034 – .020
SF-6D	0.056	0.042 – 0.070
SG	0.091	0.051 – 0.131
TTO	-0.056	-0.085 – -0.027
VAS	-0.582	-0.628 – -0.535
VisQoL	0.301	0.292 – 0.311
Face-to-face interview	Reference	
Phone	0.087	0.072 – 0.102
Various methods	0.022	-0.002 – 0.046

Table 4.8: Main model meta-regression results

sociated. Meta-regression identified small and uncertain effects associated with early stages of disease, but a large utility decrement of 0.172 associated with blindness. Methodological features of the studies were very important in explaining variation in HSUVs between studies.

It is common for modelling studies to use utility values from a single study deemed to be most relevant. Guidelines state that the choice of utility values should be transparent and systematic. However, systematic reviews are not common practice and this may result in biased estimates of cost-effectiveness. By reporting all available HSUVs alongside study characteristics, modellers will be able to select the most appropriate values. Furthermore, the results of the meta-regression will enable the estimation of HSUVs based on specific criteria, such as those that match the NICE reference case.

A number of reviews have previously been published that identified HSUVs associated with DR. While Poku et al. identified four studies [151], and Tosh et al. identified two [161], we identified and described 41. Both of these reviews had a broader scope, including other visual disorders. However, for the purpose of informing parameter selection in decision modelling, it is important that all available information is considered. Reviews that exclude the vast majority of relevant studies are not a sufficient basis for the transparent and systematic selection of HSUVs for model parameters. We believe that the far-reaching, comprehensive, and inclusive nature of our review provides an important lesson for future reviews of HSUVs. We recommend that future reviews conduct full-text searches, full-text screening, and a pre-specified review methodology. Several studies were identified in Web of Science, highlighting the importance of searching databases other than MEDLINE and Embase.

We identified a fundamental challenge in the pooling of HSUVs associated with diabetic retinopathy, and that is the incompatibility of estimates based on visual acuity and estimates based on level of disease. The former tends to be reported in relation to the better seeing eye and the latter in terms of the worse affected eye. Sight and disease level in each eye might only be weakly correlated, and it is therefore not possible to pool such values within a single model when studies do not report both.

The published protocol for this study was the first of its kind to be published, and the first to be registered prospectively. By creating a public record of the intended review process it is possible to maintain transparency in the process of selecting parameters to be used in decision analytic models of health technologies. We hope that this approach will become standard practice as part of the modelling process.

4.4.1 Strengths and limitations

The review described in this chapter is by far the largest of its kind conducted in the context of diabetic retinopathy. Furthermore, the far-reaching review and description of studies provides a wealth of information not previously readily available in the literature. The summarised findings and the results of the meta-analysis make our study the best source for HSUV parameters for diabetic retinopathy. We also provide new evidence regarding the importance of methodological choices, both in primary HSUV elicitation studies and subsequent evidence synthesis.

Our study has a number of limitations. For the review reported in this chapter, all screening and data extraction was conducted by a single reviewer, which may be associated with some bias. There are also several limitations relating to our meta-analysis. First, a high proportion of studies were not included in our final meta-regression model, which may have introduced bias. This is principally because many studies did not report all of the required data. Second, we were not able to include maculopathy as a predictor in the model, which limits the applicability of our findings. This is because few studies reported both retinopathy level and maculopathy level. Third, we were not able to construct an alternative model based on visual acuity, which would have been informative. This is because of widespread inconsistency in the definition of subgroups according to visual acuity. Fourth, our attempt to overcome heterogeneity in disease classification by coercing values into R0M0 states is open to criticism. It should be noted, however, that meta-regression would not be feasible without employing such an approach. Fifth, our model specification did not account for the censoring (at 1) associated with HSUVs. However, the model did not predict any values greater than 1 or less than zero. Finally, it could be argued that our meta-regression model had limited predictive capability.

4.4.2 Implications

Our findings with regard to the impact of DR progression on HRQoL are, perhaps, unsurprising. People graded as having retinopathy at R1 or R2 levels, without maculopathy, *shouldn't* be experiencing any symptoms associated with the disease. However, as posited earlier in this thesis, it is possible that people with more progressed disease may have correlated characteristics that do impact HRQoL. We found that this dynamic was not sufficiently strong to indicate statistically significant differences in HRQoL between the pre-symptomatic stages of disease. It may therefore be justifiable to model pre-symptomatic disease as homogeneous in terms of HRQoL. However, in order to more accurately reflect reality, we would nevertheless recommend using differential values for different

levels of retinopathy, within their observed variance.

Future research on HSUVs associated with diabetic retinopathy should consider the need to disaggregate utility values by both disease level and visual acuity and for both better eye and worse eye. At present, there is limited research on the importance of these factors in determining HRQoL for people with diabetic retinopathy.

Our study also has several implications beyond the context of diabetic retinopathy. There is some debate about the use of meta-analysis to identify HSUVs for model-based economic evaluation. A key question regards whether the purpose of such analyses is to identify 'true' values or 'best' values. Our modelling helps to demonstrate that the former simply cannot be done. The model demonstrated that methodological choices were more important in the determination of HSUVs than disease progression. Therefore, unless consensus can be achieved with regard to the correct method of utility elicitation, 'true' HSUVs for health states can never be estimated. Rather, we see the meta-regression approach as a basis for the prediction of the best possible values to include in a decision model. Values derived from a meta-analysis are better in the sense that any bias associated with single study estimates is likely to be diminished and a clearer understanding of the uncertainty associated with mean values can be obtained. Thus, we see meta-regression of HSUVs as a means of reducing bias and of increasing the representativeness of chosen parameters. Furthermore, it can be used to reveal the impact of alternative methodological choices and model assumptions.

Researchers ought to dedicate more time to the consideration of the trade-off in meta-analysis of HSUVs between minimising bias and increasing representativeness or relevance. In this context, whether or not different HSUV measures should be pooled remains an open question. For example, an analyst may have to choose between synthesising a small number of highly relevant values (say, from the EQ-5D index) or a large number of less-relevant values (say, from HUI3, TTO, and other methods), and this choice involves a trade-off. In this study we adopted an exploratory approach in the hope that blanket coverage of all published values can be more informative for future research.

It is important to consider the likely effect of using different HSUV values, either from a single study or from a meta-regression. We recommend that further research be conducted to assess the importance to decision making of using HSUV parameters from meta-analyses compared with single study estimates.

The primary purpose of this chapter was to provide parameter estimates for utilities associated with retinopathy, to be used in our model described in Chapter 7. We can use the coefficients and confidence intervals shown in Table 4.8 as parameters in our model.

Chapter 5

The cost of screening

Summary

Current unit costs for diabetic eye screening in England do not exist. This creates uncertainty in the evaluation of alternative screening programmes. In the research reported in this chapter, we sought to estimate the average cost of screening in the Liverpool Diabetic Eye Screening Programme and to provide data on the resource use associated with different aspects of the programme. Resource use associated with the grading of photographs was obtained from the screening programme and used to estimate staff and capital costs. A time study conducted at screening clinics was used to estimate the duration of photographic screening and associated staff costs. Information about capital and consumables and programme staff costs was obtained through meetings with programme staff. Data on personal expenses were collected from a cohort of trial participants. Top-down and bottom-up costings were calculated using activity levels, ingredient costs, and the total programme budget. The top-down estimate for the mean cost of a screening appointment was £33.61. The bottom-up estimate was £32.03 per screening episode. The additional cost to society associated with productivity losses and travel costs was £8.62. The attendance rate was 63% and we estimated a cost per non-attendance of £15.97. Our study can be used to guide estimates applicable to other settings.

5.1 Background

In England, around 2.5 million people are invited to attend screening for diabetic retinopathy and of these around 2 million attend each year [234]. Considering the size of the population that is invited to attend screening, it is important to obtain precise and accurate estimates of the cost per screening attendance and to understand potential sources of heterogeneity. In evaluating alternative screening strategies, small differences in the estimate of the mean cost per visit could translate to large differences in the estimated incremental cost. Imprecise estimates cannot provide a clear picture of the associated budget impact. However, there has to date been limited research into the cost of screening for diabetic eye disease in the UK.

NHS Reference Costs listed an average unit cost of diabetic retinal screening of £33 in 2009/10 [235] and of £31 in 2010/11 [236], which corresponds to around £37 and £34 at 2017 prices. Reference costs are no longer provided for diabetic eye screening services. The screening programme has changed in a number of ways since reference costs were published, and the activity used to calculate these figures may not even be representative of screening as carried out at the time, due to the limited number of sites reporting activity. At 1997 prices, James et al. estimated the cost per attendance to be £26.25 (£38.10 in 2017 prices) [237]. Scanlon et al. recently cited an unpublished microcosting study that identified a cost of £33 per person screened [30]. Yeo et al. report on the patient-borne costs of attending screening, such as parking, public transport fares, child care costs, and lost earnings [238]. There are no recently published estimates of the cost of screening from the perspective of the NHS, and limited information available regarding the wider societal costs.

In the final quarter of 2015, the national average uptake of screening was 83.6% [239]. It is therefore important to estimate costs associated with non-attendance. This could prove crucial in estimating the cost-effectiveness of alternative screening programmes, which could influence attendance rates.

5.1.1 Costing methods

Cost estimation is an important part of any economic evaluation, as the choice of cost estimates can be a key determining factor in cost-effectiveness results. The goal is to identify average costs associated with alternative strategies in order to estimate expected cost-effectiveness per individual. There are a variety of methods adopted by health economists for the purpose of costing health care programmes. Alternative methods can be broadly characterised as adopting either a 'bottom-up' or a 'top-down' approach.

Top-down costing involves the use of unit costs or block funding to estimate

average costs. For example, in the case of diabetic eye screening, the total cost of running the programme reported by the Department of Health and Social Care could be divided by the number of screening episodes to provide an average. This could be done at either a local or national level. Arguably, top-down costing is the most accurate approach, as it captures costs incurred at all levels of the organisation. However, top-down costs provide little or no information about sources of variation and cannot provide precise estimates of the costs associated with particular activities.

Bottom-up costing involves attaching separate estimates to each distinguishable component of resource use. For the case of diabetic eye screening this could include separate costs for cameras, eye drops, staff time, and clinic overheads. This approach is often referred to as micro-costing and can additionally be used to collect individual-level data in order to provide information about variability and uncertainty. Micro-costing also facilitates generalisability to other settings, as the component costs can be included, excluded, or altered as necessary. However, this approach demands far more of a researcher's time.

The key consideration in deciding what mix of top-down and bottom-up costing to adopt is how accurate and precise the estimates need to be. If an individual cost estimate is not likely to influence the results of a cost-effectiveness analysis then it need not be identified with great accuracy or precision. Chapter 2 highlighted that many model-based studies used fixed estimates for the cost of screening, with little evidence to support their accuracy or precision. In the evaluation of risk-based screening, a key source of divergence between the alternatives being considered is the frequency of screening. Therefore, the magnitude of any difference in costs will depend on the unit cost of screening. While it would be possible to evaluate the marginal cost (or saving) of additional (or fewer) screening attendances, it is important to estimate average costs for the purpose of cost-effectiveness analysis. As such, we judged it important to identify more accurate and precise costs for photographic screening than are currently available, including fixed costs associated with running the screening programme.

In this chapter we seek to address the paucity of information available regarding the cost of screening, in order to inform the parameters for our model described in Chapter 7. We conducted a costing study with collection and analysis of primary and secondary data on resource use and costs. Data were collected using a variety of methods, including i) primary data collection from ISDR trial participants, ii) extracts from the ISDR data warehouse, iii) a time study conducted within screening clinics, and iv) meetings with staff within the screening programme and NHS England. We sought to estimate the cost of screening from a health service and societal perspective.

5.2 Methods

The bottom-up costing exercise involves four steps:

1. Specification of the processes associated with photographic screening and identification of the relevant pathways.
2. Measurement of resource use associated with each pathway.
3. Estimation of the unit cost associated with each item of resource use.
4. Estimation of average total cost associated with screening pathways.

In order to identify all relevant costs associated with screening, we use a number of different methods and sub-studies. Each is described in turn below.

5.2.1 Definition of the process

The full screening pathway is described in a flow diagram produced by Public Health England, and a simplified version is reproduced in Figure 5.1 [27]. The process can be summarised as consisting of 8 key steps:

1. Individual is identified as having diabetes,
2. Individual is sent an invitation to screening,
3. Individual attends screening clinic,
4. Individual receives VA assessment and dilatation,
5. Photographs are taken of the individual's retinas,
6. Photographs are graded a number of times (as per Figure 5.1),
7. Individual is informed of outcome,
8. Individual is referred back to screening or to the hospital, as necessary.

Each rectangular process node shown in Figure 5.1 is associated with resource use. We have numbered these from P1 to P6 to facilitate our description of the results. These screening and follow-up pathways were clarified through discussion with senior clinicians and clerical staff in the Liverpool Diabetic Eye Screening Programme (LDESP).

Individuals whose photographic images are graded as referable to the hospital eye service are invited to attend for an examination using slit lamp biomicroscopy. Broadly speaking, because patients are referred to hospital management, this can be considered outside of the remit of the screening programme. Individuals with

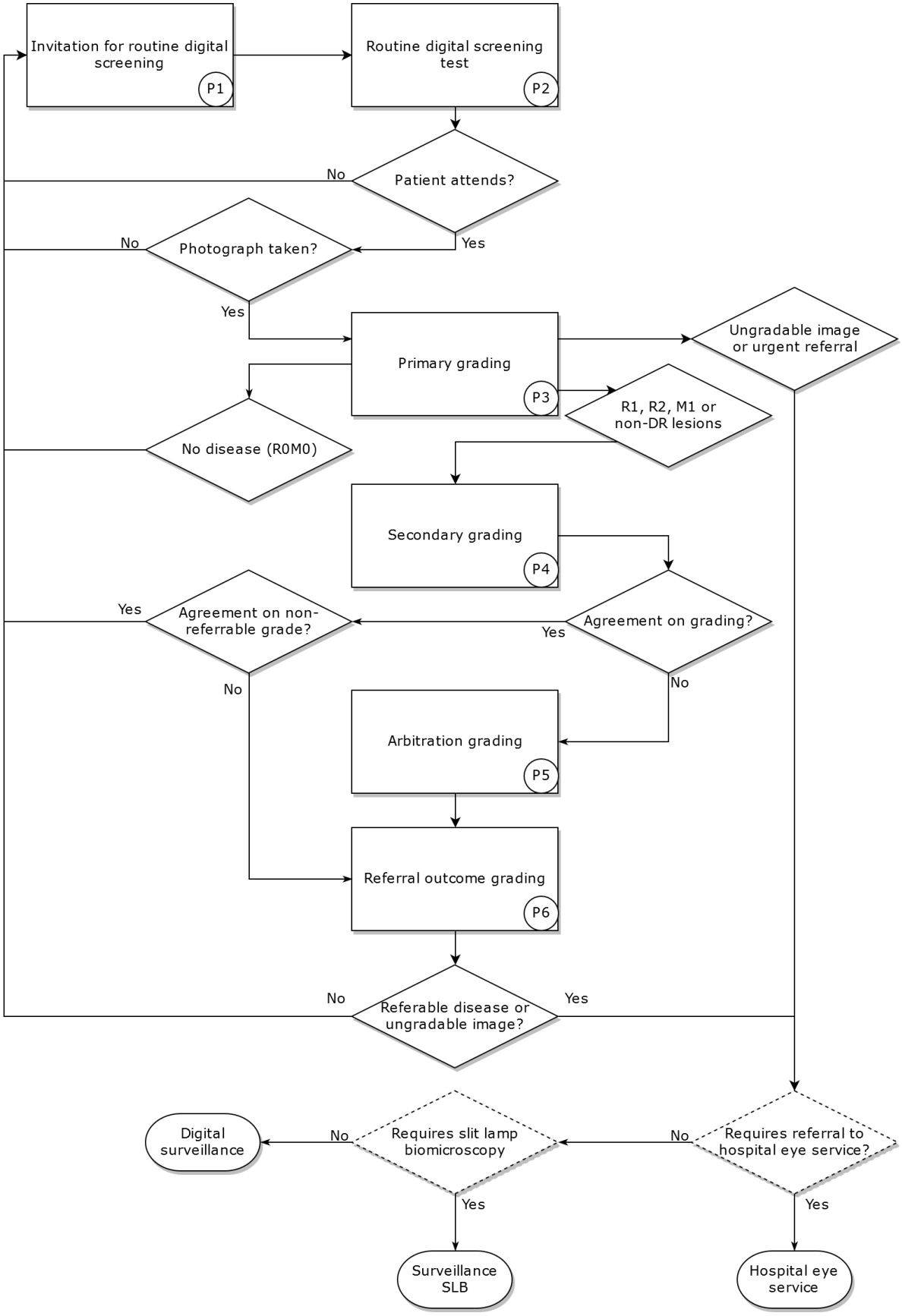

Figure 5.1: NDESP patient flow

ungradable images may be referred to the hospital eye service for assessment within the remit of the screening programme, but this tends to be funded separately by Public Health England. In this analysis of the cost of screening, we focus on photographic screening and assume slit lamp biomicroscopy examinations to lie outside of this scope. Furthermore, we do not consider aspects of the screening programme that are conducted at the national level, which might include training and quality assurance processes.

In the context of this study, it is also necessary to cost an alternative screening programme not represented in current practice. The process of risk stratification using the ISDR RCE — as it would be conducted in practice — also required identification. This was achieved through conversation with the data scientists and clinicians engaged in its development and use. There are a number of stages involved in the process of individualising eye screening for people with diabetes. The process is shown in Figure 5.2. In practice, this process is appended to that shown in Figure 5.1 for all individuals referred back to routine screening.

For individuals receiving risk-based screening intervals, the risk calculation engine is used to generate the required screening interval. This information is fed into the screening software (OptoMize). OptoMize automatically generates a letter to the patient informing them of the screening result and their next planned review date within 3 weeks of the screening appointment. Thus, the invitation process is equivalent to annual screening except in its frequency.

5.2.2 Data

Data were collected from multiple sources. In this section we describe the resource use data that we collected and the sources that were used to estimate unit costs.

ISDR data warehouse

Data for screening appointments since 2006 were available in the data warehouse maintained as part of the ISDR study. These data included information about attendance and non-attendance at screening visits and records of grading activity. For this analysis, we focussed on the most recent complete calendar year for which data were available, which was 2016.

Data associated with all screening appointments scheduled to take place during 2016 were extracted from the data warehouse. These included the date of the screening appointment and, for each grading that was conducted, the grading date, type, and outcome. Individual-level attendance data (including non-attendance) were only available for appointments arranged in earlier years, up to June 2015. Therefore, data from the first quarter of 2013 to the second quarter of 2015 were used to estimate attendance rates and assumed to be applicable to

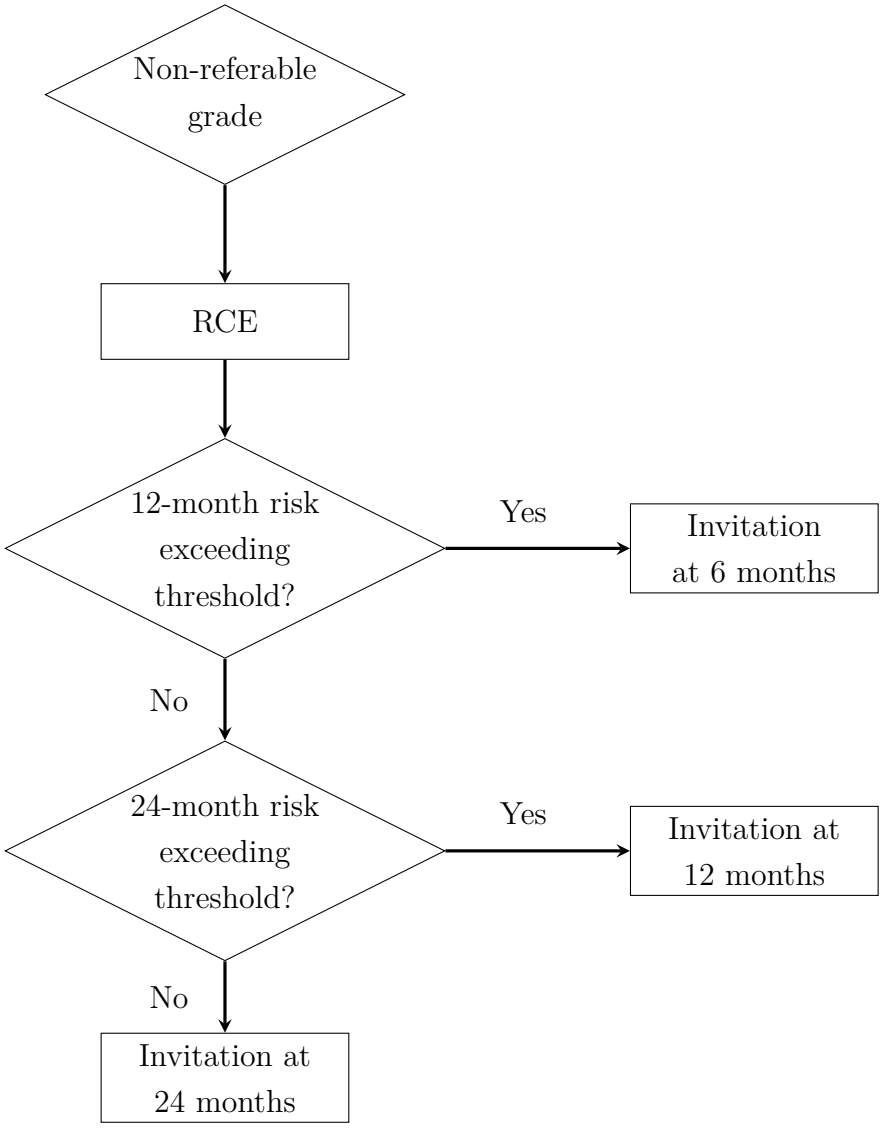

Figure 5.2: RCE patient flow

2016.

Time study of screening

A time study of screening was conducted through a single full-day visit to each of the seven participating screening sites in Liverpool. Data were recorded in a spreadsheet, which included the time (to the nearest minute) at which the following events occurred:

1. Patient enters the clinic,
2. Screener calls patient for VA assessment and dilatation,
3. Patient returns to waiting room,
4. Screener calls patient for photography,
5. Patient returns to waiting room,
6. Patient exits the clinic.

The time study provides information to facilitate the estimation of screeners' time use. Furthermore, the research facilitated a clearer understanding of usual practice in screening clinics with regard to non-attendance and barriers to efficiency, to which we return in the Discussion. The time study did not capture screeners' travel time.

Cross-sectional study of the cost of screening attendance

We developed a bespoke questionnaire designed to elicit information about personal and societal costs associated with individuals' attendance at screening. As part of the ISDR RCT, individuals who were included in the health economics subsample (as introduced in Chapter 3) were asked to complete the questionnaire at their baseline screening attendance.

The 'ISDR Visit Questionnaire' [240] (provided in Appendix E) included 10 questions:

1. Which modes of transport did you use in travelling to and from the centre today?
2. If using public transport or taxi, what is the total cost of your return travel?
3. If travelling by car, how many miles is your return journey?
4. If travelling by car, what is the total cost of parking?
5. How much time did you spend in total on this visit, including preparation, travel time and attending?

6. Are you currently in employment?
7. Did you take time off work to attend today?
8. Did a friend, family member or any other person assist you in attending your appointment today?
9. Did they take time off work to do so?
10. How much time did they spend helping you to attend this visit?

Participants were asked to self-complete these questions as applicable, and asked to provide a best estimate if they were unsure. Participants who were not able to self-complete the questionnaire were assisted by a research nurse, who could complete the questionnaire on the participant's behalf. A script was prepared to assist research nurses in administering the questionnaire. The questionnaire and script were submitted — and accepted — for inclusion in the Database of Instruments for Resource Use Measurement [241]. Whether or not the participant self-completed was recorded, along with the date of the visit. All data were subsequently transferred to the electronic case report form (eCRF) and subjected to consistency and validation checks.

LDESP data

Additional resource use estimates, including capital, consumables, overheads, and staff time, were obtained from screening programme staff and supported by local internal data sources. We obtained the asset register from the RLBUHT finance department and assessed requisitions. These data enabled the estimation of ingredient costs not available from other routinely collected local data or from national sources.

Unit costs

Unit cost estimates for health care resource use, including staff costs, were derived from the Unit Costs of Health and Social Care 2017 [242]. Where specified, prices paid by RLBUHT were used in place of unit costs, including VAT if applicable.

5.2.3 Analysis

We estimated costs accruing through each process in the screening pathway (as per Figure 5.1). Each cost is categorised according to whether it is borne by i) the NHS, ii) the patient, or iii) society. For each category, cost estimates are summed to estimate an average cost per screening attendance in 2017 prices. Where required, the CCEMG-EPPI-Centre Cost Converter was used to inflate prices [243].

Number of observations	Full data set	2016
Individuals	22,253	12,841
Screening attendances	130,264	13,068
Non-attendances	18,867	-
Gradings (all) †	234,418	18,624
Primary gradings	159,444	13,058
Secondary gradings	54,723	4,178
Arbitration gradings	10,834	251
Referral outcome gradings‡	9,376	1,137

Table 5.1: Summary of OptoMize screening and grading data. †Older data include some redundant grading types. ‡Includes ophthalmology gatekeeper.

We used grading data from the data warehouse (2006-2016) to estimate the number of gradings associated with different pathways according to the level of disease identified. We conducted count regressions to predict the number of gradings associated with each R0M0 screening outcome, using negative binomial regression models to predict the number of gradings per screening attendance and variation according to final agreed R0M0 grading and year and month of the screening appointment. Thus, we are able to estimate the average resource use per screening attendance, for the whole population and according to different grading results.

Non-attendance is perceived to be an important source of inefficiency in the NDESP. We used data from the data warehouse to estimate the attendance rate in each month of available data and estimate the proportion of costs that may be attributable to non-attendance.

5.3 Results

5.3.1 Data summaries

ISDR data warehouse

Table 5.1 shows counts from OptoMize for the full data set (dating back to 2006) and for the calendar year 2016. These data show that, in 2016, the total number of screening attendances was more than one per person. This highlights the tendency for a small group of people to be invited back to screening at an interval of less than 12 months, even within the current fixed interval programme.

Duration (in minutes)	Mean	SD	Range
Total for visit	32.70	11.77	13.11–83.01
Dilatation	4.58	2.56	2.18–15.29
Photography	3.30	1.93	0.00–8.74
Screener time	7.50	2.94	2.18–13.11

Table 5.2: Screening centre time study results. SD = standard deviation.

Time study of screening

Table 5.2 shows the findings from the time study conducted through visits to the screening clinics. Information was collected on 104 screening attendances. Eighty one (78%) participants attended screening on a day on which recruitment to the ISDR trial was taking place and three participants were also attending for optical coherence tomography (OCT; another imaging technique). Both of these factors may have extended the duration of visits. Four participants (two couples) attended screening with another individual, which reduced the amount of the screener's time used (and explains the apparent discrepancy in Table 5.2 in the total screener time). A duration of zero could be recorded for photography where the procedure lasted less than one minute.

Cross-sectional study of the cost of screening attendance

Table 5.3 shows the findings from the cross-sectional study of screening attenders. Eight hundred and sixty eight people were asked to complete the visit questionnaire. There were very few missing data. Around half of the sample were able to self-complete. Around half travelled to their appointment by car, on average a 4.7 mile round-trip, and nobody had to pay for parking. Those travelling by public transport (around a third) paid on average £6.30. A third of the sample reported being in employment, and half of these reported taking time off work to attend their appointment. Around half of respondents received assistance from a friend, family member, or other individual, but only 13% of assistants were reported to have taken time off work. In total, people reported on average spending 90 minutes in attending their appointment. Assistants spent on average 65 minutes.

LDESP data

Given the current climate of tendering of NHS-provided services in England, information relating to the cost of screening in Liverpool is now deemed to be commercially sensitive. This meant that it was difficult to obtain precise and complete cost estimates from the LDESP for the purpose of this research. It also means that we are not at liberty to fully report costs that relate to the operation

Variable	Mean/proportion (SD)
Self completion	
Yes	52.30%
No	44.93%
Missing	2.76%
Mode of transport	
Car	49.42%
Taxi	13.48%
Bus/Train	23.85%
Hospital transport	1.27%
Bicycle/on foot	10.71%
Other	0.12%
Missing	1.15%
Public transport cost	
For all (n=856)	£1.36 (3.03)
Where > £0 (n=185)	£6.30 (3.38)
Miles driven	
Where > 0 (n=422)	4.72 (3.94)
In employment	
Yes	33.53%
No	65.21%
Missing	1.27%
Time off work	
Yes (for all)	17.51%
Yes (if in employment, n=291)	52.23%
No	81.45%
Missing	1.04%
Received assistance	
Yes	48.27%
No	50.23%
Missing	1.50%
Assistant took time off work	
Yes (for all)	6.80%
Yes (if receiving assistance, n=419)	13.60%
No	92.17%
Missing	1.04%
Time spent on visit (minutes)	90.40 (116.13)
Assistant's time spent	
For all	31.83 (51.29)
Of those receiving assistance (n=403)	64.74 (57.59)

Table 5.3: Results from the ISDR Visit Questionnaire. SD = standard deviation.

Staff band	Basic pay†	Oncosts‡	Overheads	FTEs	LDESP costs
Administration and management					
2	£16,850	£3,572	£5,718	3	£78,420
3	£18,777	£4,114	£6,409	1	£29,300
4	£21,417	£4,856	£7,356	1	£33,629
5	£25,735	£6,069	£8,905	0.2	£8,142
6	£31,989	£7,826	£11,148	1.2	£61,156
8a	£45.428	£11,603	£15,969	0.3	£21,900
Subtotal					£232,546
Screening					
3	£18,777	£4,114	£6,409	0.6	£17,580
4	£21,417	£4,856	£7,356	2.8	£94,161
5	£25,735	£6,069	£8,905	0.9	£36,638
6	£31,989	£7,826	£11,148	0.05	£2,548
Subtotal					£150,927
Grading					
4	£21,417	£4,856	£7,356	1.2	£40,355
5	£25,735	£6,069	£8,905	0.9	£36,638
6	£31,989	£7,826	£11,148	0.55	£28,030
8a	£45,428	£11,603	£15,969	0.1	£7,300
Subtotal					£112,323
Total					£495,796

Table 5.4: LDESP staff costs. FTEs = full-time equivalents; LDESP = Liverpool Diabetic Eye Screening Programme. †Based on mean average across all NHS staff groups. ‡National Insurance contributions plus 14.3% employer's superannuation contribution.

of the LDESP. As such, some of the data are necessarily presented here in a summarised or otherwise aggregated form, and some figures are approximated.

The LDESP employs staff to handle invitations, call and recall, and follow-up non-attenders. Pre- and post-clinic administration relating to slit lamp biomicroscopy follow-up is also handled by the programme in Liverpool, such as completion of Diabolos data. Table 5.4 lists the staff employed by the programme in 2016-2017, grouped by NHS pay band and according to role relating to administration and management, screening, or grading. Screening and grading were performed mainly by screener-graders with some senior clinical arbitration supported by an administration and management team.

In addition to standard NHS Trust overheads, the screening programme also pays for interpreter services. Non-English-speaking patients are called to the

Item*	Unit purchase cost (2016)	Life in years
Digital camera (x3)	£18,518.02	10
Docking units	£17,301.06	10
Eye screening system	£18,290.32	10
Furniture	£12,127.94	10
IT equipment	£15,259.04	7
Monitors etc	£17,390.43	7
New camera and elevation table	£15,526.80	10
Computer hardware	£1,144.89	7

Table 5.5: LDESP asset register data. *As listed.

hospital and have a particularly high DNA rate. A sign language service is also made available in the community screening clinics.

Key capital items for the screening programme include cameras and a server to store images. These also require maintenance contracts. Table 5.5 shows a selection of capital costs derived from the LDESP asset register. Table 5.5 shows the number of years of life expected, as recorded in the asset register. However, most of the equipment lasts much longer. Anecdotally, cameras are known to last for as long as 20 years if maintained.

5.3.2 Health service costs

For the year 2016-2017, based on the total LDESP budget (£562,466) and the total number of attended diabetic eye screening appointments (16,736) reported in annual Public Health England (PHE) key performance indicator (KPI) data, the mean average cost per screening attendance in Liverpool was £33.61. In this section, we summarise the ingredients to this health service cost.

Our time study showed that the duration of screening attendance, dilatation, and photography was relatively consistent. Dilatation took on average 4.6 minutes and photography took on average 3.3 minutes. Accounting for those couples who received screening simultaneously, the total time required of a screener for dilatation and photography was on average 7.5 minutes, as shown in Table 5.2. Thus, we infer that the average contact time for staff conducting photographic screening is 7.5 minutes.

On average, there were 1.4 gradings for every 1 photographic screening appointment attended. The negative binomial regression model showed that both retinopathy and maculopathy grading result and the year and month in which screening took place influenced the number of gradings conducted. Table 5.6 presents the incidence rate ratios (IRR) compared with an agreed R0M0 grading of no disease. From these results we can assert that higher retinopathy or posi-

Predictor	IRR	95% CI		
R1	1.841	1.825	–	1.857
R2	1.964	1.894	–	2.037
R3	1.330	1.231	–	1.436
M1	1.220	1.196	–	1.245
Year	0.963	0.962	–	0.964
Month	0.998	0.997	–	0.999

Table 5.6: Negative binomial regression of grading count. CI = confidence interval; IRR = incidence rate ratio.

tive maculopathy grades are associated with a greater number of gradings being conducted. For example, a grading of R2 is associated with a rate of grading occurrence 1.964 times greater than that of an R0 grading. This means that for every 1 grading that takes place for R0 outcomes, 1.964 take place for R2 outcomes. These findings are to be expected, as positive screen results are referred for further grading and arbitration. The predictors relating to the year and month in which screening took place demonstrate that year-on-year (and month-on-month) fewer gradings are being conducted per screening attendance.

Information on attendance and non-attendance was available for 51,099 appointments, of which 36.92% were not attended. Attendance data were also associated with the clinic location. Figure 5.3 shows the quarterly attendance rate for the seven current screening locations in Liverpool. On the basis of the whole sample, the Jeffreys 95% confidence interval was estimated as 36.49%–37.33%. No clear time trend was identified, so we used the total rate of 36.92% in our estimates. Here, the attendance rate refers to non-attendance for each invitation, rather than the LDESP operational definition of uptake as an individual who attends following multiple invitations, as reported in the KPI.

Table 5.7 summarises the estimated average cost per screening attendance associated with each part of the screening pathway, as denoted by P1-P6 in Figure 5.1, for the year 2016/17. The average cost to the NHS of an appointment that was attended was £26.14.

We also estimate the cost per non-attendance. This is based on several assumptions: i) that all programme costs are equally divisible across invitations, regardless of whether the appointment is attended; ii) that staff contact time associated with photographic screening can be reallocated for non-attendances, but that non-contact time cannot; and iii) that the cost of cameras and equipment is not reduced by non-attendance. Based on these assumptions, the average cost of an appointment that is not attended is £15.97. Given an attendance rate of 63.08%, this implies a cost per screening episode of £32.03.

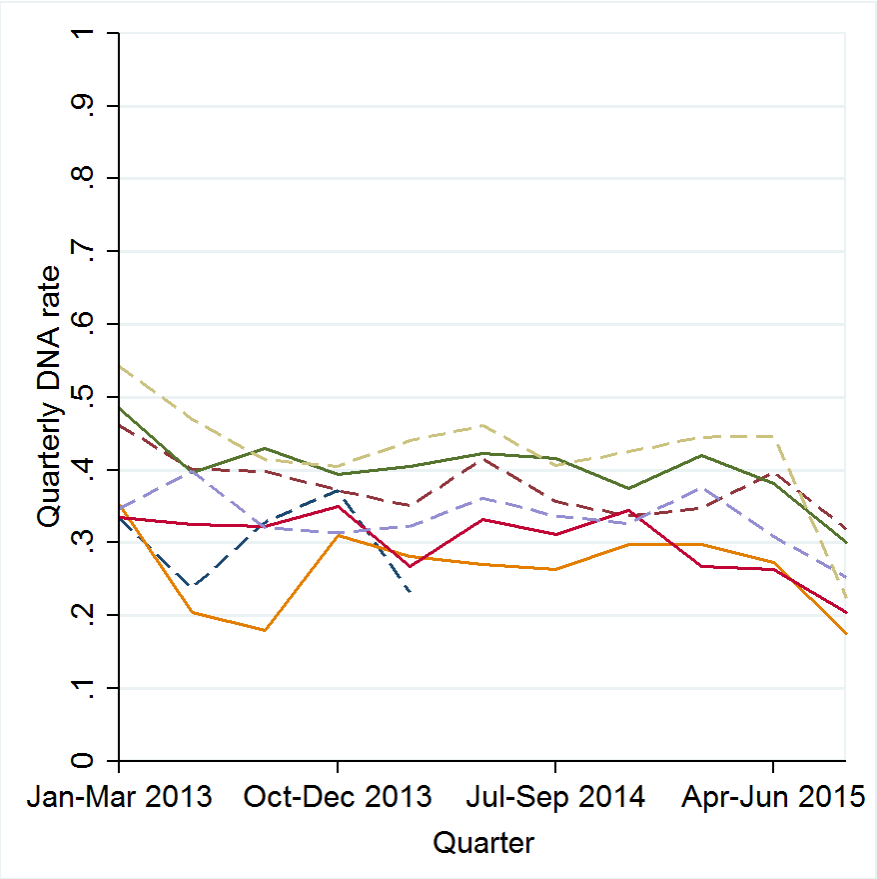

Figure 5.3: Quarterly non-attendance rate for seven screening clinics in Liverpool

Item	Total cost	Cost per attendance (n=16,736)	Cost per non-attendance (n=6,179)
Programme costs (P1)			
Staff (including oncosts)	£232,546	£10.15	£10.15
Stationary	£2,928	£0.13	£0.13
IT	£22,025	£0.96	£0.96
Subtotal	£257,499	£11.24	£11.24
Photography (P2)			
Staff contact time	£53,723	£3.21	£0.00
Staff non-contact time	£97,204	£4.24	£4.24
Cameras and equipment	£11,284	£0.49	£0.49
Medical consumables	£4,052	£0.24	£0.00
Subtotal	£166,263	£8.19	£4.73
Non-NHS costs			
Patient-borne costs	£42,342	£2.53	£0.00
Productivity loss	£101,922	£6.09	£0.00
Grading (P3-P6)			
Staff	£112,323	£6.71	£0.00
Total NHS cost	£536,085	£26.14	£15.97
Total societal cost	£680,349	£34.76	£15.97

Table 5.7: Health care costs for 2016/17. †Programme costs include the costs of invitations.

Risk stratification

Based on our experience in the ISDR research programme, we estimate that a risk-based screening programme will require 0.2 FTE administrator, probably based at the local CCG. We also expect that each screening programme will require a database manager. There may be additional costs associated with software — some of which will be sunk costs associated with setting up the risk-based programme — and additional capital in terms of computers.

For the purpose of our calculations, we estimate staff costs of £40,000 per year. Thus, for the year 2016/17, in which 22,915 invitations were sent, the cost per invitation would be an additional £1.75. Fixed recall or stratified-recall screening programmes can be managed centrally, where individualised programmes may not be. Roll-out of a risk-based screening programme may therefore involve some form of localisation, which may require additional costs.

5.3.3 Societal costs

The greatest proportion of respondents travelled by car and nobody had to pay for parking. For those travelling by car, we estimated the average cost to be £2.36 (assuming a cost of £0.50 per mile). For those travelling by public transport — excluding people with free travel — the average cost of fares was £6.30. This corresponds to an average personal travel cost of £2.53 per attendance.

Assuming a national average wage and hours worked, the average productivity loss for an invitee who took time off work was £26.08. The average productivity loss for an assistant was £23.74. The total productivity loss associated with this time, averaged across the whole sample, was £6.09 per attendance, as shown in Table 5.7.

5.4 Discussion

Our analysis of the ingredient costs summed to £32.03 per screening episode in 2016/17, including the costs incurred from non-attendance. We estimate that each appointment attended incurs an average cost of £26.14, while each appointment not attended incurs an average of £15.97. Costs associated with personal expense and productivity losses are small, but these should be included when evaluating screening programmes with large eligible populations and major budget impacts. An additional cost of £8.62 per visit equates to a societal burden of more than £20 million in the UK.

We have identified important sources of heterogeneity in the average cost of screening. In particular, the costs incurred vary according to the initial grading outcome. Considering wider societal costs, an important determinant is the

amount of time required to attend screening, and whether a friend, family member, or carer attended to assist.

Our estimate of the average cost of screening is comparable to previous estimates from the UK and implies an improvement in productivity over time, at least in the LDESP (n.b. national estimates of around £33 in 2010 and £31 in 2011). Estimates of the average cost of screening from other countries are also similar [244–246]. In future, a budget impact analysis should also be conducted from both a local programme-level perspective as well as from a national perspective. This could inform the most appropriate and cost-effective implementation strategy.

We have provided up-to-date estimates of the costs associated with screening for sight-threatening diabetic retinopathy, based on data from a programme in England. Our study is the first to estimate the cost associated with appointments not attended and further provides data on societal costs. The results provide vital information to guide the development and evaluation of screening programmes for diabetic retinopathy.

5.4.1 Strengths and limitations

There are several strengths to this study, which should be noted. We analysed a large and comprehensive dataset on screening and grading activity, enabling us to produce robust estimates of resource use. Our analysis and results were prepared with input from screening programme staff and members of the ISDR Patient and Public Engagement Group, ensuring an accurate representation of current practice. The key strength of our study is that the disaggregation of costs facilitates the application of our results to other settings.

There are several limitations to our work. Though the estimates were similar, our approach was not able to fully account for the top-down estimate of £33.61, which could be due to missing or inaccurate data. We are not able to provide estimates of uncertainty or variability associated with the average cost of screening because most of our data are not recorded at the level of individual patients or photographs. Furthermore, we relied on 'top-down' costing methodologies for some of the ingredient costs. In particular, we were not able to disaggregate the costs associated with staff time in running the programme. Our estimate of the cost associated with an appointment that is not attended may be an inaccurate estimate as a result of this. More research is required to understand the extent to which clinics are able to efficiently overbook screening sessions and the productivity of screeners in these circumstances. Participants in the cross-sectional study were asked to estimate the total time that they spent on their visit before it was complete. Participants may have tended to over- or under-estimate the total time that they would spend on the visit. It is likely that the participants

will have had many previous visits to the same clinic for the same purpose, yet there may have been recall bias in the estimates they provided.

It is possible that our findings are not representative, due to specific characteristics of the organisation of the LDESP and the population it serves. The ISDR trial commenced on 11th November 2015, and persisted throughout our data collection, which may have inflated resource use. Furthermore, on 2nd February 2015, LDESP adopted a two-image photographic screening protocol having previously taken four images. The resource implications of this are unknown. The current climate of tendering in the NHS, with competition from private providers, meant that certain data were considered commercially sensitive and that we could not report certain aspects of our work in full detail. Our analysis also excluded any expenditure realised at the national level as part of the NDESP.

Our estimated costs for a risk-engine-based approach to individualised screening recall may not apply to other settings where screening programmes are less well-established or require more significant changes to facilitate implementation. There are a variety of alternative implementation models for risk-based screening, which might include i) a fixed, standard risk engine to be used in all settings; ii) a general approach to risk engine development that operates separately within each programme; or iii) the identification of a key 'hub' for risk engine development and administration. The cost implications of each of these alternatives should be addressed by future research.

5.4.2 Implications

In the specific case of screening for diabetic retinopathy, Scotland and Bryan [247] have highlighted the importance of 'technology management' – that is, the continuing evaluation of technologies throughout their life cycle. In this context, top-down costings can become redundant and uninformative. By providing a set of disaggregated costings and resource use estimates, our findings can inform changes in service delivery, including changes in the frequency of screening recall. Our results can be used alongside estimates of the costs of new technologies, such as automated grading [108], to predict the cost of future programmes. Furthermore, revised cost estimates can be produced to predict the economic impact of sociodemographic changes, including increases in the prevalence of diabetes. Our estimates can also be used to inform tendering processes in the UK, to ensure that potential providers quote costs fully with reference to the screening pathway.

A priority for decision-makers is to ensure that screening programmes are run efficiently. Patients who do not attend appointments (around 37%) may represent a substantial opportunity cost to the health service, despite efforts to limit the waste of resources. Previous studies have assumed that non-attendance is associated with the same cost as attendance (e.g. [248]), which is not realistic.

Costs associated with personal travel expense and productivity losses are relatively small. Nevertheless, these could be considered when evaluating screening programmes with large eligible populations and substantial budget impacts. Future research should explore the extent to which personal expenses may act as a barrier to attendance in countries where screening is free at the point of use.

Our findings facilitate the estimation of an average cost – to the NHS and to society – of an invitation to screening that depends on i) whether or not the invitee attends, and ii) the grading result for the photographs. These estimates can be used as parameters in a decision model, meaning that the model can incorporate the cost implications of alternative programmes with respect to their impact on attendance rate and on the casemix of screening attenders in terms of the level of disease present. We use the findings from this chapter in Chapter 7 to model the costs associated with variable-interval screening.

Chapter 6

The cost of treatment

Summary

The purpose of this chapter is to estimate the costs of treatment for diabetic retinopathy in Liverpool, UK. We investigated costs for three key groups of procedures: laser, intravitreal injections, and vitrectomy. In addition to these three key treatment types for diabetic retinopathy, we also identified cost estimates for hospital appointments and follow-up within the screening pathway. Having identified recommended treatment and follow-up pathways from clinical practice guidelines, we used local data from a hospital patient management system to ascertain common procedure codings for the relevant treatments. The data include inpatient and outpatient procedures for more than 20,000 individuals, with 93,086 inpatient admissions and 690,818 outpatient appointments. We estimated the frequency of alternative treatment pathways and of specific procedures for people who screened positive. We find that laser is the most common procedure in our data set, but that intravitreal injections are increasingly included in treatment pathways. Based on national unit costs, we estimated the monthly costs incurred through procedures both related and unrelated to diabetic retinopathy.

6.1 Introduction

In the evaluation of a risk-based screening programme (or indeed any form of screening) it is crucial to take into consideration the costs of follow-up treatment for people who screen positive. As described in Chapter 1, the purpose of risk-based screening is to improve the identification process such that more people are able to benefit from early intervention. As such, a risk-based screening programme — compared with a standardised programme — could exhibit differential resource use and costs associated with treatment in the long run. Furthermore, the incremental cost of screening compared with no screening is often very low, while the incremental cost of treatment may be very high. Therefore, it is possible for treatment costs to be the dominant driver in any long-term cost differences between alternative screening strategies. This chapter sets out work designed to elucidate treatment pathways for people with diabetic retinopathy, and to establish estimates of the cost of treatment for use in a decision model for the evaluation of risk-based screening.

There are a variety of effective treatments available for diabetic retinopathy and maculopathy, as described in management guidelines produced by The Royal College of Ophthalmologists [249]. Panretinal photocoagulation (PRP) laser therapy is a well-established and effective treatment for proliferative diabetic retinopathy [250]. There are different types of lasers and settings that can be used for treatment and different methods including central or peripheral PRP [251]. Side effects of laser can include pain, vitreous haemorrhage, and a reduction in a patient's visual field.

Intravitreal injections are an increasingly popular alternative or complementary treatment to laser, as new delivery systems are developed. The use of anti-vascular endothelial growth factor (anti-VEGF) — also known as antiangiogenic drugs — and intravitreal steroids is supported for maculopathy [252, 253]. There is also evidence that anti-VEGF treatment is effective for retinopathy [254].

In some cases with specific indications, it may be appropriate to carry out vitrectomy surgery, whereby parts of the vitreous can be removed. There are also management strategies that can limit the progression of retinopathy, such as blood pressure control [255].

All treatments may be used in conjunction with others and patients' treatment pathways can vary. In England, people are referred for hospital assessment following a screen-positive result. It is at this stage that a decision about treatment can be made by an ophthalmologist.

There are no up-to-date studies on treatment patterns for people with diabetic eye disease in the UK, or the costs associated with alternative pathways. As risk-based screening programmes are developed, it will become important to understand the treatment patterns associated with different screening pathways.

This chapter reports on research with several specific aims:

1. to identify key treatment pathways associated with diabetic retinopathy,
2. to estimate the frequency of alternative treatments, and
3. to obtain cost estimates for the different treatment pathways.

6.2 Methods

6.2.1 Data

Our analyses were conducted using data from the ISDR data warehouse. Secondary care data from iPM consists of inpatient data and outpatient appointments as detailed below. All people with diabetes registered with general practitioners in Liverpool were eligible for inclusion in the data warehouse.

We were also able to link iPM data with screening programme data from Diabolos, which records slit lamp biomicroscopy appointments and outcomes. These data facilitate the identification of people who have 'true positive' screening outcomes.

Inpatient data

The inpatient data include dates of admission and discharge for each individual who was admitted to hospital over the period. Following each admission, any number of procedures can take place and all of these procedures are recorded. Where an individual's location in the hospital has changed — for example, if they move wards — the duration can be further subdivided into episodes. Procedures are defined as either primary or secondary, with secondary procedures additionally given a sort order. For each procedure event, the date it took place is recorded. Each procedure event is also associated with a procedure code and the corresponding description. Each event is associated with a patient reference number, which can be linked to the other databases within the ISDR data warehouse.

Outpatient data

Outpatient data record four levels of coding — location code, clinic code, session code, and primary procedure code — each with their corresponding description. Each event is associated with a date and a patient identifier. Whether or not the patient attended the appointment is also recorded.

Coding

For the purpose of estimating activity levels relating to treatment, the key variables for this analysis are the various codings. The meaning of these codes is not self-evident and additional research was necessary to determine how they ought to be interpreted.

Location codes identify the place in the hospital that an outpatient appointment is due to take place. Clinic codes are related to the location code but are more specific and can change over time. Session codes relate to specific clinicians' activity. All outpatient data should be associated with a location code, a clinic code, and a session code, but these codes are not applicable for inpatient data. Procedure codes are central to our analysis of treatment activity. Both inpatient and outpatient data record procedures using the OPCS Classification of Interventions and Procedures version 4 (OPCS-4) [256].

In order to determine a list of relevant procedures, we extracted all OPCS-4 codes and also the frequency that each occurred in the data set. The current version of the OPCS-4 consists of more than 9,000 different codes. In collaboration with a clinical expert from the ISDR team, each code that occurred at least once in the dataset was reviewed and determined as being either relevant to diabetic eye disease or not. Additionally, each treatment was categorised based on how likely it was that the treatment was directly related to diabetic eye disease.

In practice, coding staff select codes based on clinical notes. Laterality codes should always be used on any procedure on a paired organ (such as the eyes), sequenced after the main procedure code and never in a primary position. An increasingly important procedure in DR is injections, which often involves the use of high-cost drugs. Codes for drugs such as Lucentis (ranibizumab) should be assigned as a secondary procedure to the injections code.

Data quality

It is important to identify what data might be missing and why. We therefore investigated missing and bad data. In general, we cannot know if data are missing as the events could simply not have been recorded. We have reason to believe (from personal communication with NHS Trust staff) that data recording may have improved over time, and so looking at the changes in the frequency of admissions, episodes, and procedures per person over time could be informative. It is also possible that coding behaviour may change with financial year boundaries as changes are made to the NHS Payment by Results system. In the early years of coding (2008–2010), it was possible to code procedures more liberally and so procedures may appear more frequent.

Label	Description	Included OPCS-4 codes
Laser	Laser procedures	C82.1, C82.5, C82.6, C82.8, C82.9
IVT	Intravitreal treatments	C79.4, C89.1, C89.2, C89.3
VR	Vitrectomy-related procedures	C54.3, C54.6, C54.8, C79.2, C79.3, C79.5, C79.6, C79.7, C79.8, C79.9, C80.1, C80.2, C80.3, C80.4, C80.8, C80.9, C81.1, C81.2, C82.2, C85.1, C85.2, C85.5, C89.8, C89.9
Invest	Diagnostic investigations	C86.5, C87.1, C87.2, C87.3, C87.4, C87.5, C87.8
Cataract	Cataract-related codes	C64.7, C71.1, C71.2, C71.3, C71.8, C72.9, C73.1, C73.2, C73.3, C73.4, C73.9, C74.3, C75.1, C75.2, C75.3, C75.4, C75.8, C77.6, C77.8, C79.1
Glaucoma	Glaucoma-related codes	C01.2, C03.2, C52.2, C54.5, C59.2, C60.1, C60.5, C60.6, C60.8, C61.2, C61.4, C61.5, C62.2, C62.3, C64.8, C65.3, C65.4, C66.2, C66.3, C66.4, C66.5, C69.1

Table 6.1: Procedures related to diabetic retinopathy

6.2.2 Identification of treatment pathways

We are principally interested in laser, injection, and vitrectomy treatments. These are the main recommended treatment strategies for people who screen positive for diabetic eye disease. Through a series of meetings with clinical and clerical staff at RLBUHT, a list of OPCS codes of relevance to diabetic retinopathy was identified. This list was categorised into procedures relating to laser, injections, vitrectomy, and other procedures related to — but not specifically for — the treatment of DR, as shown in Table 6.1. In addition to the procedure codes, a set of clinic codes (specific to RLBUHT) were used to identify those hospital appointments that related to the hospital eye service.

The iPM data were merged with data from Diabolos that indicated whether or not a person had a positive disease state classification following a slit lamp biomicroscopy examination and, if so, the date at which this was recorded. Our analyses of treatment pathways focus on people who have screened positive and

the treatments that they received in the years immediately following the positive screening outcome.

6.2.3 Analysis of treatment frequency

In order to understand the costs associated with treatment for diabetic retinopathy, it is necessary to identify the level of resource use associated with alternative treatment pathways. We recorded the total frequency of all OPCS codes and reviewed common procedures to ensure that no important treatments were missed. The number of times that OPCS procedure codes from the Laser, IVT, and VR categories were used in each month of available data was plotted in order to identify broad trends in treatment behaviour. The same was done for each code within each category.

We identified the number of times that people who had screened positive received each relevant intervention in the years immediately following their slit lamp biomicroscopy examination.

6.2.4 Treatment costing

Unit costs were attached to OPCS codes by cross-referencing with Healthcare Resource Group (HRG) codes listed in NHS Reference Costs. For all injection procedures we further assumed that Lucentis was used in all instances at a cost of £551.00 per vial [257]. HRG costs were estimated for all relevant procedures, and totals and averages were tallied for each procedure. We estimated the average total cost incurred in a week in which a relevant treatment was received.

6.3 Results

6.3.1 Data description

Inpatient data

The inpatient data included information on 14,341 individuals with 93,086 admissions, 74,052 episodes, and 80,822 procedures. All admissions were associated with an admission date and a discharge date. All episode reference numbers were associated with an admission, and each had a start date and end date. There were 21,721 admissions that were not associated with any episodes or procedures. All procedure codes were associated with a date, a sort code, an episode, and an admission. All episodes had one or more procedures associated with them.

The ISDR data warehouse includes 22,091 individuals in total, meaning that 65% were represented in the iPM inpatient data and therefore implying that 35% of individuals were not admitted to an RLBUHT hospital during the period

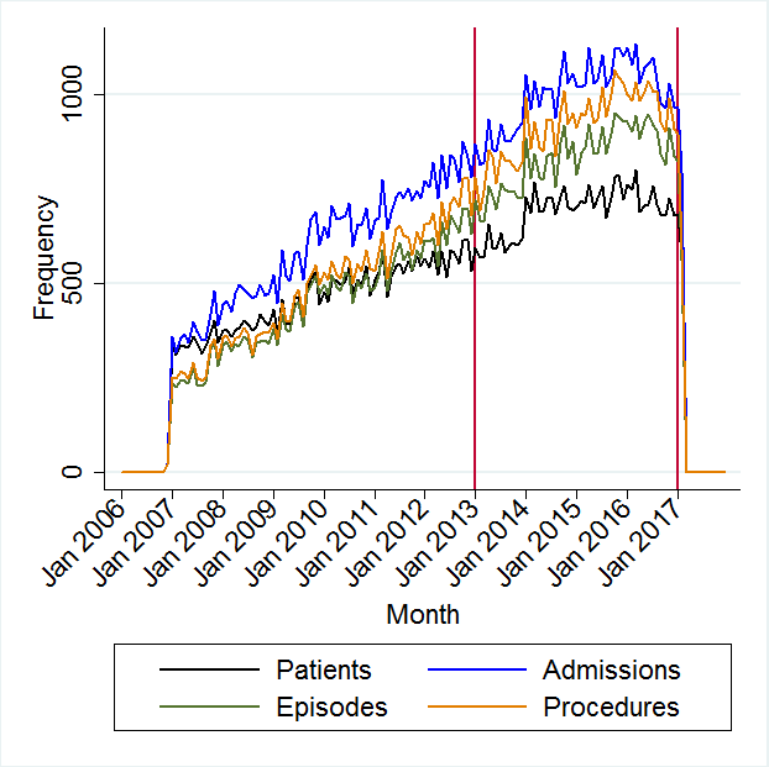

Figure 6.1: iPM inpatient activity

covered by the data. Of all admissions recorded, 23% did not include any procedures. It is likely that the majority of this 'missingness' in the inpatient data is an accurate non-observation, where patients had not been admitted or received a procedure. While there may be some data quality issues at play we are unable to identify these.

Figures 6.1 and 6.2 show monthly admission, episode and procedure activity over time, in absolute numbers and as a rate per patient. Figure 6.2 shows that inpatient activity appears to level-out from January 2013 onwards, which we interpret as indicative of full adoption of current coding practices. Our analysis therefore focusses on data from 2013–2016 inclusive, as identified by red reference lines in Figure 6.1 and Figure 6.2.

Outpatient data

Outpatient data included 20,830 individuals with 690,818 appointments. All appointments were associated with a date and a description of attendance or non-attendance. 88% of appointments were not associated with any procedure codes. Figure 6.3 shows the number of appointments and the number of procedures over time, with the same reference period specified above.

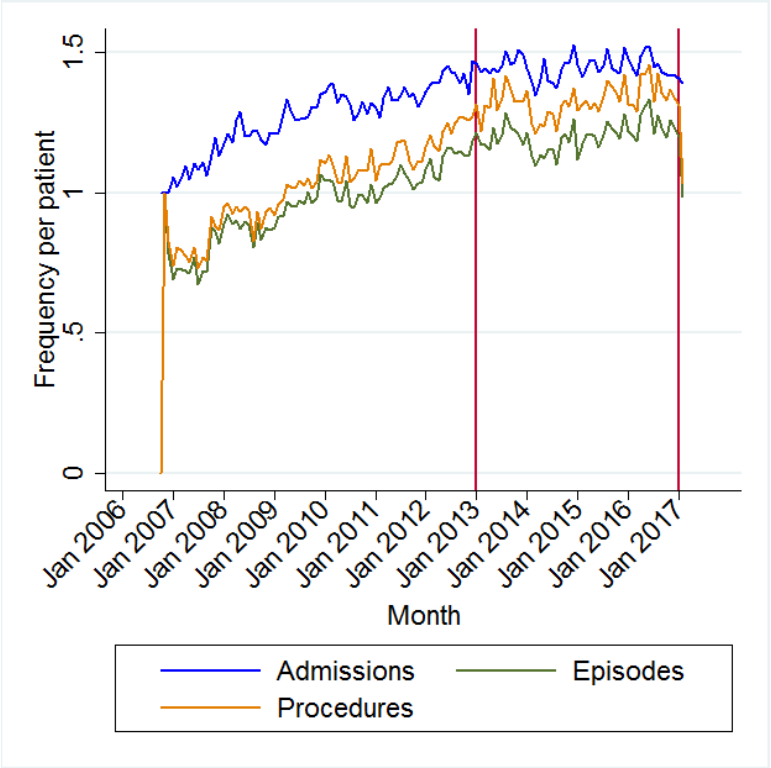

Figure 6.2: iPM inpatient activity per patient

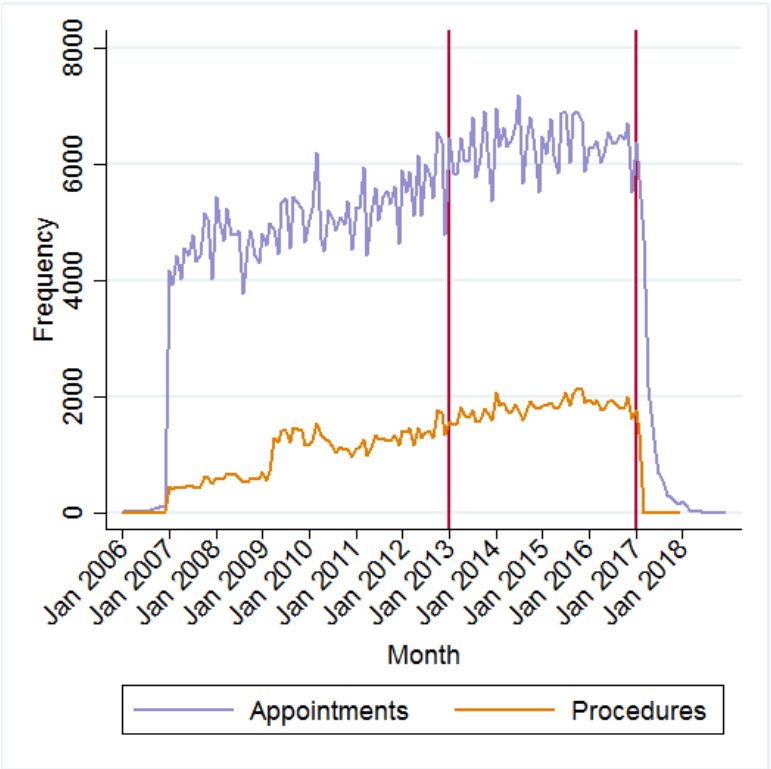

Figure 6.3: iPM outpatient activity

6.3.2 Pathways

The most recent full calendar year of data available from Diabolos was for 2012. The number of people receiving biomicroscopy in the year 2012 was 816, of which 246 screened positive. Table 6.2 shows six treatment pathways observed in the cohort in the four years following a positive slit lamp biomicroscopy exam: i) laser only, ii) injections only, iii) both laser and injections, iv) both laser and vitrectomy, v) laser, injections, and vitrectomy, and vi) hospital follow-up only.

It is clear that laser is still the dominant treatment strategy for people who screen positive for diabetic retinopathy and receive intervention within four years of screening positive. However, a large majority of people did not receive any procedures within our main classifications within the first four years. All people who did not receive a procedure attended the hospital eye service at least once within one year of their positive screen result.

6.3.3 Resource use

Figure 6.4 shows the monthly frequency of laser, IVT, and vitrectomy procedures over the full period of data, with the same four-year period identified as previously. This graph clearly shows that there is high month-on-month variability in the frequency of laser and vitrectomy procedures carried out, though no obvious trend of change year-on-year. The number of injections, on the other hand, shows a steady increase.

Table 6.3 shows the number of procedures observed in the data for people who have screened positive, in the four years following a screen-positive result, for the whole sample (n=246) and per recipient of that treatment. On average, a person who receives laser in the first year following a screen-positive will receive 2.83 laser procedures, while a person who receives injections will receive 1.63 injections procedures. Within four years, a person who receives injections will on average have received 3.79 injections procedures, while a person who receives laser will receive 3.21 laser procedures. No glaucoma-related procedures were observed.

6.3.4 Costs

Table 6.4 shows the average monthly costs incurred by someone who receives a laser, IVT, or VR procedure, or attends the HES (without receiving a procedure) in a given month. The mean 'relevant' cost is the sum of costs associated with the subset of procedures listed in Table 6.1, while the mean 'total' cost is derived from the sum of costs for all possible procedures received in that month. The most notable finding is that intravitreal injections are associated with much higher costs than laser. However, it is also clear that, despite a low unit cost, hospital

Treatment pathway†	Number of years since SLB			
	One	Two	Three	Four
Laser only	20 (8.13%)	22 (8.94%)	24 (9.76%)	26 (10.57%)
IVT only	5 (2.03%)	5 (2.03%)	4 (1.63%)	4 (1.63%)
Laser + IVT	3 (1.22%)	5 (2.03%)	7 (2.85%)	9 (3.66%)
Laser + VR	1 (0.41%)	2 (0.81%)	2 (0.81%)	3 (1.22%)
Laser + IVT + VR	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (0.41%)
HES only	217 (88.21%)	212 (86.18%)	209 (84.96%)	203 (82.52%)

Table 6.2: Number of screen-positive individuals on different treatment pathways. SLB = slit lamp biomicroscopy, IVT = intravitreal treatments, VR = vitrectomy-related procedures, HES = hospital eye service. †All individuals attended follow-up at the hospital eye service.

	One year		
Intervention	Count	Per recipient	Per screen-positive
Laser	68	2.83	0.28
IVT	13	1.63	0.05
VR	1	1	0.00
HES	863	3.98	3.51
Invest	62	1.19	0.25
Cataract	9	1.28	0.04
	Two years		
Intervention	Count	Per recipient	Per screen-positive
Laser	85	2.93	0.35
IVT	30	3	0.12
VR	3	1.5	0.01
HES	1191	5.62	4.84
Invest	73	1.33	0.30
Cataract	9	1.29	0.04
	Three years		
Intervention	Count	Per recipient	Per screen-positive
Laser	102	3.09	0.41
IVT	43	3.91	0.17
VR	3	1.5	0.01
HES	1502	7.19	6.11
Invest	94	1.42	0.38
Cataract	12	1.20	0.05
	Four years		
Intervention	Count	Per recipient	Per screen-positive
Laser	125	3.21	0.51
IVT	53	3.79	0.22
VR	5	1.25	0.02
HES	1794	8.84	7.29
Invest	120	1.64	0.49
Cataract	13	1.18	0.05

Table 6.3: Intervention counts for screen-positives

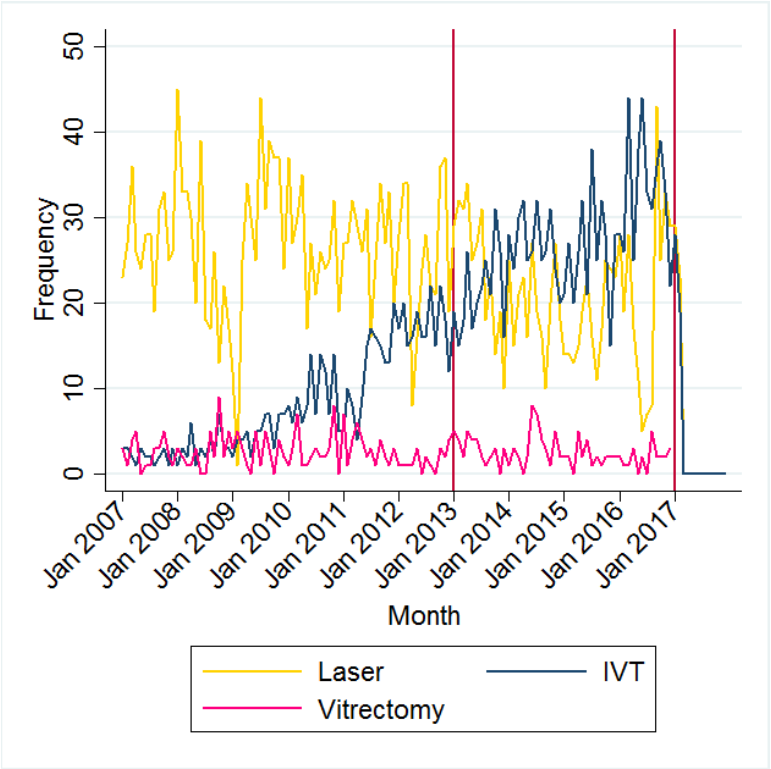

Figure 6.4: Monthly DR treatment frequencies

	n†	Relevant		Total	
Procedure		Mean	SD	Mean	SD
Laser	948	£207.23	134.51	£235.77	296.68
IVT	1,282	£748.39	101.70	£763.36	135.11
VR	112	£283.73	250.12	£299.38	269.78
HES	46,320	£27.09	120.86	£44.84	249.63

Table 6.4: Total costs per treatment month. SD = standard deviation. †n = person-months

attendances without treatment incur a greater cost in aggregate due to their high frequency.

6.4 Discussion

We interrogated a large dataset from a hospital's patient management system, which was linked to screening programme data. Our findings show that the majority of people who receive treatment will receive laser procedures. However, the use of injections is increasing over time and people who receive injections — on average — receive more procedures and incur greater costs than people who receive laser.

The trends in treatment pathways that we have observed are likely to be rep-

resentative of wider developments in the UK and elsewhere [258]. This change has arisen due to the development of new therapies and more robust evidence to support their use [259, 260]. Intravitreal injections are being recommended as a standard treatment choice in many health care settings [63, 261, 262]. There is little quantitative evidence available on the adoption of injections at a national level in recent years, though some studies have identified an increase in the adoption of injections for the treatment of diabetic retinopathy in specific groups (e.g. [263])

6.4.1 Strengths and limitations

A key strength of our analysis is that it includes all people with diabetes who have attended an RLBUHT hospital between 2006 and 2017. This provides real world data without selection effects often involved in observational studies. OPCS codes change over time, so it is not possible to determine a definitive list of codes that do or do not relate to a particular disease. Nevertheless, a strength of our study is that we were able to work with clinical and clerical staff to ensure that our analyses identified key treatments.

It is a significant shortcoming of our work that the screening outcome data and treatment data did not sufficiently overlap chronologically to allow for an analysis of the most up-to-date treatment pathways. The relatively low number of people progressing to laser within four years of screening positive may be an artefact of this.

Our dataset was also limited insofar as outpatient appointments were only associated with a primary procedure code. This means that we were not able to attribute costs to secondary codes. As a result, our cost estimates are likely to be lower than actual hospital costs. For the case of injections, which can involve expensive drugs that would be coded in a secondary position, assuming no drug costs would have resulted in a substantial underestimate of costs. Therefore, we assumed the cost of Lucentis, which is one of several therapeutics available in the NHS. On balance, we expect that this would have resulted in an overestimate of the cost of injections because hospitals are able to negotiate prices for branded medicines such as Lucentis, which are considered commercially sensitive and therefore not available to us.

We found that there is very limited guidance — either nationally or locally — around the coding of ophthalmology procedures, which hampered the interpretation of our data.

6.4.2 Implications

The findings of this study have several important implications for model-based cost-effectiveness analysis. We identified that multiple pathways are significant in the population of people who screen positive for diabetic eye disease. It is not appropriate to assume that everybody within a model receives a standard treatment. Some model-based cost-effectiveness analyses have done this in the past (see, for example, [264] and [265]). Furthermore, it cannot be assumed that patients only receive one type of procedure after screening positive. For instance, we identified people who received both laser and IVT, or laser and vitrectomy.

Our analyses showed that many people do not receive any relevant hospital procedures within the first four years after screening positive. Thus, models should not (as some have [107, 266]) assume that people are immediately referred for treatment. It is notable that the key driver of costs can be attendances at the hospital eye service without the receipt of any procedures. Thus, it is important that decision models accurately estimate the frequency and cost of hospital attendances, and that good quality data are available to provide such estimates.

We found that costs unrelated to DR in the month in which treatment was received did not, on average, change the magnitude of costs significantly. However, the inclusion of unrelated costs introduced a great deal of uncertainty into the estimate of average costs. Given our extended process of reviewing procedure codes, it seems reasonable to assert that this additional uncertainty simply represents the addition of noise to the data. Based on our findings, we would recommend that analysts exclude unrelated costs unless there is good reason to believe that they may relate to the intervention being evaluated.

The pathways identified by the analyses presented in this chapter are used in the design of the structure of the decision model described in Chapter 7. The estimates of treatment frequencies and costs are used as parameters in our model and could be used in future by other model developers.

Chapter 7

The cost-effectiveness of risk-based screening for diabetic retinopathy: development of a decision model

Summary

This chapter describes the development of a decision analytic cost-effectiveness model to evaluate a risk-based screening programme for diabetic retinopathy. We designed a state-transition-based individual sampling model that incorporated key screening and treatment pathways. We compare the variable-interval risk-based screening programme being evaluated in the ISDR trial with annual screening (current practice) and with a stratified biennial screening programme. The model incorporates the ISDR risk calculation engine and its structure is based around the states defined by the screening programme's disease grading system. The parameters for the model are drawn from earlier chapters of this thesis and from published literature. The model went through several iterations. Ultimately, the simulation proved too computationally demanding to execute within the time horizon of this project and we describe results from a reduced simulation. Our findings indicate that the ISDR risk-based screening programme is cost-saving and outcome-improving compared with current practice. There is a high level of uncertainty due to between-patient heterogeneity and random variation in parameter values and event probabilities. We provide recommendations for modelling in the context of risk-based screening.

7.1 Background

Current practice in screening for diabetic eye disease in the UK involves people with diabetes over the age of 12 being invited for screening annually. This programme has been in place since 2007. In Chapter 1 we outlined proposed changes to the NHS Diabetic Eye Screening Programme (NDESP) and other alternative forms of screening programme that might be introduced in the future. Chapters 3 through 6 provided new evidence regarding the possible costs and outcomes of screening for diabetic retinopathy (DR). In this chapter, we draw together this evidence and — taking lessons from previous studies identified in Chapter 2 — develop a model to evaluate the cost-effectiveness of alternative screening programmes.

Photographic screening has clearly been demonstrated to be effective through the identification of people who would benefit from early treatment for diabetic eye disease. By receiving treatment early, people can avoid the sight loss that can result from proliferative diabetic retinopathy. The prevalence of diabetes in England is growing and predicted to increase to 9.7% by 2035 [267]. This is creating pressure on screening programmes around the country and has lead to calls for improving the efficiency of screening for diabetic eye disease.

Recently, there have been several studies published that have shown that most people without any retinopathy are at a very low risk of developing referable eye disease within one year. Many have suggested that the use of extended screening intervals (beyond one year) for low-risk people could reduce resource use without risk to patients. Some analyses have evaluated the implications of extending the screening interval for diabetic retinopathy [30, 107]. All have demonstrated that screening intervals of more than one year are likely to be cost-effective – at least for groups at lower risk.

Our focus is on the evaluation of risk-based screening, which we introduced in Chapter 1. In Chapter 8 we further explore the role of cost-effectiveness analysis in the context of risk-based screening. In this chapter we describe the development of a decision model to evaluate three alternative screening programmes that are currently being considered for implementation in the UK.

7.1.1 Development of the model

The model described in this chapter has been through several iterations and undergone complete rebuilding and redesign. Here, we briefly outline the gestation of the model from its initial conception to its current form.

As described in Chapter 2, numerous model-based analyses have been published to evaluate the cost-effectiveness of alternative strategies for screening for diabetic retinopathy. Furthermore, studies have been published that simulate out-

comes associated with alternative screening strategies but do not consider costs. In line with the majority of model-based economic evaluations reported in this context, we set out to develop a state transition model using a cohort simulation. This model was completed and designed to evaluate annual screening (current practice) compared with the programme being evaluated in the ISDR trial and another alternative programme that involved people with background retinopathy being screened every year and people with no retinopathy being screened every two years. The combination of policy developments and a review of the findings of our model encouraged us to take an alternative approach.

In January 2016 — subsequent to the first version of our model being completed — the NSC published a new recommendation [62]. It recommended that screening for people with a low risk of sight loss should be extended to two years. The definition of 'low risk' applies to people who have had two consecutive screening results showing no retinopathy. Thus, an important new comparator — distinct from those included in our model — was being considered by policymakers. The initial design of our model did not allow for differential pathways according to multiple historic screening outcomes (i.e. two consecutive results of no retinopathy). We therefore judged it necessary to adapt our model to evaluate the new programme being recommended by the NSC. This required a structural overhaul to the model.

Having completed the cohort model that evaluated the new NSC recommendation, and prepared results, it became clear that the findings were not adequately reflecting the benefits of risk-based screening. This is because the value of risk-based screening relies on the fact that risk profiles change over time, and risk-based screening is able to detect these changes in risk profiles. Thus, the modelling of fixed cohorts that are defined only in terms of their disease state became untenable. Rather, it became clear that the model would need to incorporate the risk calculation engine and thus model progression, costs, and outcomes at the individual level. This change did not require revisions to the structure of the model, but rather to its operation.

In line with previous research (see Chapter 2), the model was developed in Microsoft Excel. This was ideal for the cohort model and was feasible for the implementation of individual-level simulations. However, the risk calculation engine is computationally expensive and Microsoft Excel was not capable of making the necessary computations. This chapter therefore describes the limitations of Microsoft Excel in this context.

7.2 Methods

7.2.1 Setting

The context for the analysis is the Liverpool Diabetic Eye Screening Programme. We discuss the generalisability of our findings later in the chapter. In Liverpool, eligibility for the screening programme is identified through GP practices. People are invited to attend a screening clinic at one of seven local health centres. Screening is carried out by trained screeners and involves the capture of two digital photographs, as described in greater detail in Chapter 5. For the purpose of the NDESP, 'R0M0' gradings for retinopathy grading and a maculopathy grading are recorded, as defined in Chapter 1.

7.2.2 Comparators

We evaluated three alternative policies: i) annual screening (current practice), ii) biennial screening for people with no retinopathy at two consecutive attendances and annual screening for everybody else, and iii) ISDR stratification, as implemented in the ISDR trial.

Annual

Annual screening represents current practice in England. All people with diabetes over the age of 12 are invited to attend screening annually. In Liverpool, there is some deviation from this standard, and some people are invited to attend screening after six months. However, this practice varies and there is no empirical basis on which to reliably model this form of personalisation. Thus, we assume that everybody in the 'Annual' pathway is invited to attend screening once every year.

Biennial

Recently, the UK National Screening Committee recommended a stratified screening programme based on historic screening outcomes. The NSC recommended that people with background retinopathy should continue to be invited to screening every year. The recommendation specifies a low-risk group, defined as people whose screening outcome shows no presence of disease at two consecutive visits. This low-risk group is recommended to be invited to screening every two years.

This recent recommendation has not been implemented in practice, but represents an important comparator for current practice. Previous modelling work, which constituted part of the basis for the NSC's recommendation, has compared the cost-effectiveness of such a biennial programme with current practice and shown it to be cost-effective.

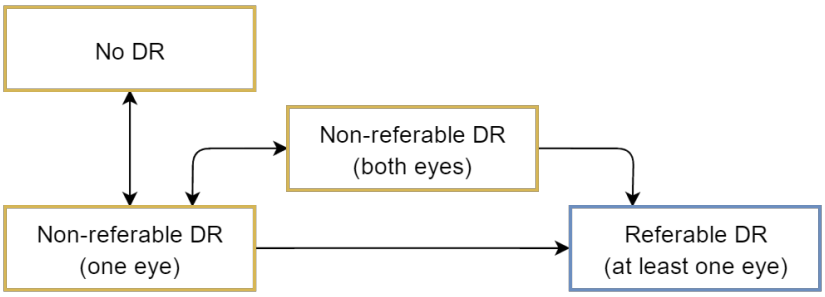

Figure 7.1: Risk calculation engine Markov model structure

ISDR

The ISDR programme uses the risk calculation engine (RCE), which is reported in more detail elsewhere [61]. The RCE was based on a Markov model, as shown in Figure 7.1.

The RCE uses individuals' characteristics to estimate the probability that they will screen positive at 6, 12, or 24 months. Using a 2.5% risk threshold, which was identified through work with patient representatives, the RCE allocates individuals, following a negative screen result, to either 6-, 12-, or 24-month recall, according to the longest recall period at which the individual's risk does not exceed 2.5%.

The structure of the RCE Markov model, and its parameters, are built into the decision analytic model reported in more detail below.

7.2.3 Model structure

The model structure was primarily determined by the grading system used by the NDESP. However, guidance and inspiration was also drawn from previously published models described in Chapter 2.

We developed a microsimulation state transition ('semi-Markov') model based on diabetic retinopathy disease states, as identified through photographic screening, with time- and event-dependant transition probabilities. The patient-level simulation (an individual sampling model) tracks individuals' characteristics through time, so that their changing risk of disease onset could be used to determine their pathways through the model. The simulation incorporates both disease states and events that determine individuals' pathways through the model, which are observed in discrete units of time. Parameters used in the analysis are drawn from both published literature and local data.

Disease states

Disease progression in diabetic retinopathy is well understood, but inconsistently defined. The primary purpose of our model was to inform NDESP policy. There-

	M0	**M1**
R0	R0M0: no retinopathy	NA
R1	R1M0: background retinopathy	R1M1: background retinopathy with maculopathy
R2	R2M0: pre-proliferative retinopathy	R2M1: pre-proliferative retinopathy with maculopathy
R3	R3M0: proliferative retinopathy	R3M1: proliferative retinopathy with maculopathy

Table 7.1: Description of R0M0 states. NA = not applicable.

fore, disease progression was defined in the terms used by the NDESP, namely, 'R0M0' gradings. These are described in Table 7.1. Recently, the NSC introduced a distinction between R3a (proliferative retinopathy) and R3s (stable treated DR). This distinction is not explicitly included in our model's disease states but is represented through differential event pathways, as described below. Levels of visual acuity are not explicitly modelled except for progression to severe vision loss, which we define as visual acuity below 6/60 (Snellen) and refer to as 'Blind'.

In order to account for the biennial programme described above, we implemented a tunnel state for R0M0 gradings to which individuals transition when they have a consecutive R0M0 grading. Thus, we label states as 'R0M0 [2]' if the individual has received at least two consecutive no disease gradings and 'R0M0 [1]' if they have received only one no disease grading or haven't yet been screened.

The model states were further expanded to account for individuals who had background retinopathy in either one eye or both eyes. This is an important distinction because individuals with background retinopathy in both eyes have been shown to be at greater risk of progressing to sight-threatening diabetic retinopathy. Practically, this distinction was also important because it ensured that the pre-referral disease states in the decision model could be easily mapped to the disease states of the RCE Markov model. We labelled disease states as 'R1M0 [R0M0]' for people with background retinopathy in one eye and 'R1M0 [R1M0]' for people with background retinopathy in both eyes. The structure of our disease model is shown in Figure 7.2.

It is possible for retinopathy levels to regress as well as progress. This can occur as a result of treatment but also — particularly in the early stages of the disease — as a result of good glycaemic control. Therefore our model allows for progression and regression across all states, except from blindness and death. This corresponds to 22 possible transitions between disease states, plus transitions to and from the 'R0M0 [2]' tunnel state, plus eight possible transitions to blindness and nine possible transitions to death. The R0M0 and R1M0 states shown in Figure 7.2 correspond to states used within the RCE. All descriptions shown in

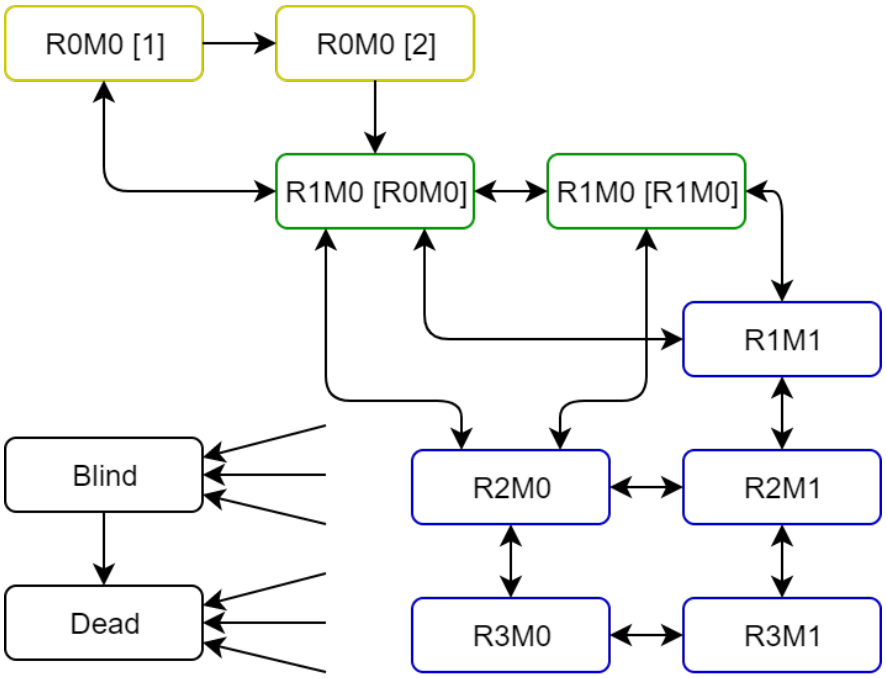

Figure 7.2: State transition model disease states

Figure 7.2 correspond to the individual's retinopathy and maculopathy state in their worse eye, except for the two R1M0 states, which specify retinopathy level in both eyes.

Event pathways

In addition to disease states and pathways, individuals' progression through the model also depends on a series of events. It is important to include these event processes in the model in order to fully account for costs and to adjust progression rates as appropriate. These events differ depending on whether the person is in screening or in follow-up. All individuals are in the screening programme, except for those who have experienced a screen-positive event and are therefore in follow-up and may receive treatment. However, those who receive treatment that is successful may be referred back to the screening programme. Therefore, our model allows for individuals to be in either screening or follow-up whether they have received treatment or not.

Event pathways differ depending on whether a person is in screening or follow-up. Event pathways (but not their probabilities) are assumed to be equivalent for patients regardless of whether or not they have previously received treatment. It is the conclusion of the event pathways that determines whether a person moves between screening or follow-up states and pre-treatment or post-treatment states.

People in screening may or may not attend a screening appointment. If they do, this could be associated with several standard screening outcomes and may re-

sult in either referral to the hospital eye service for follow-up or back to screening. Figure 7.3 shows the event pathways for people in screening.

Those in follow-up may or may not attend an appointment in any one cycle. If they do attend an appointment then they may or may not receive a treatment. In either case, they may or may not be referred back to screening. There are three key treatments for diabetic retinopathy — laser, injections, and vitrectomy — as described in Chapter 6. Each possible treatment is incorporated as a transition event for those individuals in follow-up. Figure 7.4 shows the event pathways for people in follow-up.

Individuals who go on to receive treatment are subject to different progression rates. Therefore, it is necessary to divide the model further based on the stages of the treatment pathway. Our model allows for individuals to be in one of four groups of patients: i) those in screening who have not received treatment, ii) those in screening who have previously received treatment, iii) those in follow-up who have not yet received treatment and iv) those in follow-up who have previously received treatment.

The comparators for our analysis — as described above — are similarly included as an additional event process, applied to all people whose screening outcome is negative and who are referred back to screening. These event pathways are outlined in Figure 7.5.

Operation

The model was developed using Microsoft Excel (version 1907 for Office 365). The model was designed to run over a lifetime horizon, with simulations running for each person until they died within the model, at which point the simulation moved on to the next patient. The model operates on a monthly cycle, with transitions and events occurring at the end of each cycle. A monthly cycle was adopted in order to allow for flexibility in the screening recall period and to mirror the monthly analyses reported in Chapter 6. Both costs and quality-adjusted life years (QALYs) were discounted at 3.5%. For the present analysis, costs were evaluated from an NHS perspective.

7.2.4 Data

Several packages of work within the ISDR programme are used to inform the modelling study presented in this chapter, with multiple data sources that are described in more detail in earlier chapters.

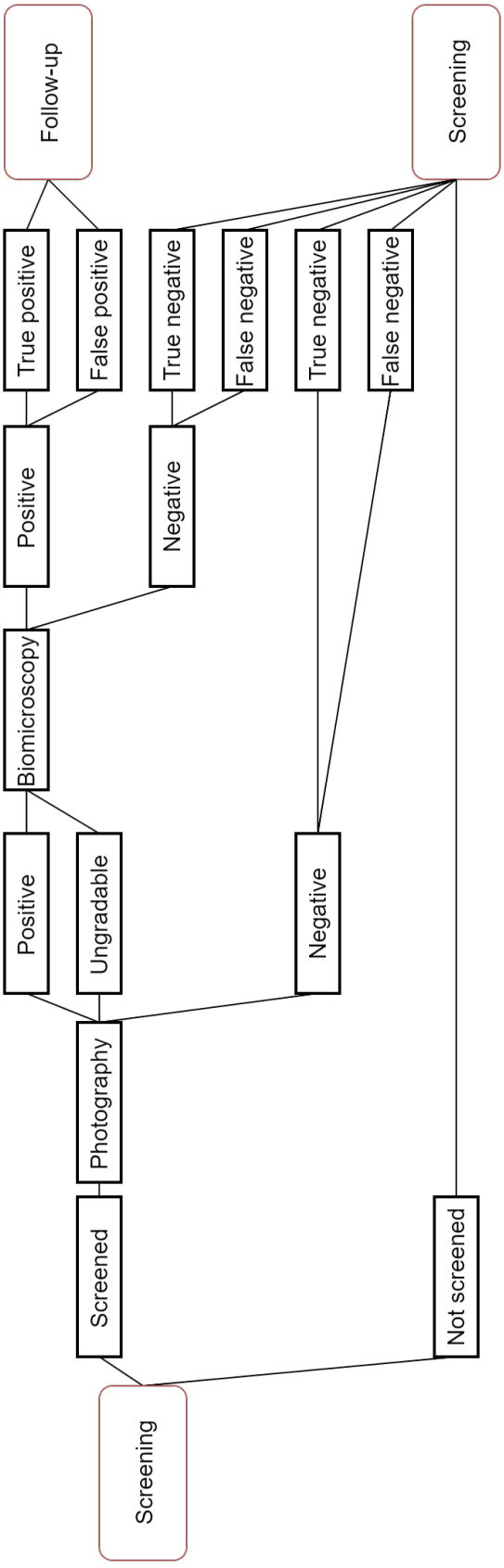

Figure 7.3: Screening event pathways

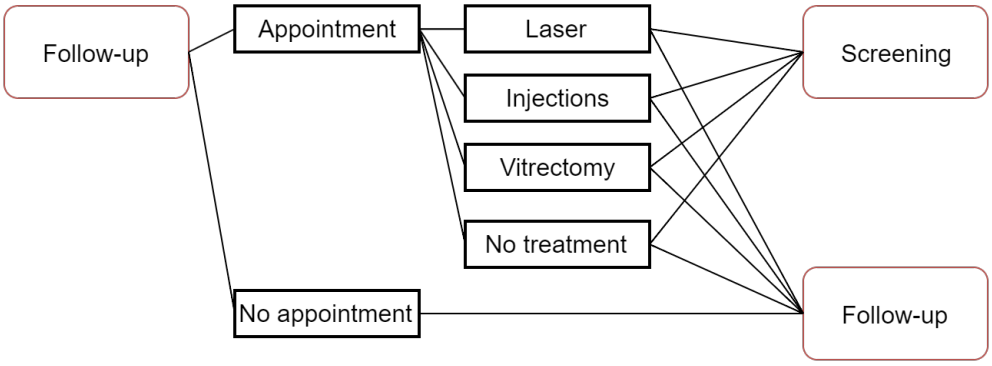

Figure 7.4: Follow-up event pathways

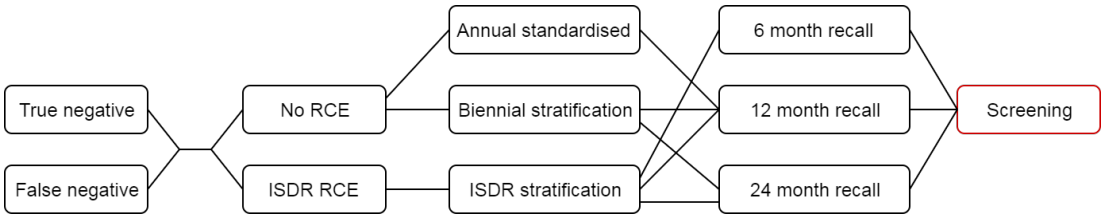

Figure 7.5: Intervention pathways. RCE = risk calculation engine.

Baseline characteristics

The model is set-up to simulate a random sample of 10,000 people with non-referable disease. Individuals are randomly drawn (with replacement) from a sample of 8,111 screening attendances in Liverpool, England, with complete data in 2013. Each individual's lifetime pathway is simulated under each comparator before the model moves on to the next randomly selected individual. The baseline sample was extracted from the ISDR clinical data warehouse. The characteristics of the sample are shown in Table 7.2.

Model parameters

The model principally relies on a set of baseline disease progression rates that would be observed without early intervention. Progression between the pre-referral disease states of R0M0 and R1M0 are determined by the risk calculation engine, as is progression to the first stages of referable disease (R2M0 and R1M1). Table 7.3 describes the hazard rates provided by the RCE, according to baseline Markov state. In order to estimate individual risk (r) in each cycle (t), it is necessary to compute the matrix exponential as described by Eleuteri et al. [61], such that

$$r(t) = \exp(Qt^{0.9}) \tag{7.1}$$

Parameter	**Mean (SD)**
Markov state	
1 (R0M0)	78.54%
2 (R1M0[R0M0])	14.02%
3 (R1M0[R1M0])	7.45%
Male	58.00%
Age in years	63.65 (12.41)
Duration of diabetes in years	7.28 (5.33)
HbA1c	54.24 (15.58)
Total cholesterol	4.15 (0.98)
Systolic blood pressure	131.11 (14.10)

Table 7.2: Baseline cohort characteristics. SD = standard deviation.

where

$$Q = \begin{pmatrix} -\lambda_{11} & \lambda_{12} & 0 & 0 \\ \lambda_{21} & -\lambda_{21} - \lambda_{23} - \lambda_{24} & \lambda_{23} & \lambda_{24} \\ 0 & \lambda_{32} & -\lambda_{32} - \lambda_{34} & \lambda_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{7.2}$$

Microsoft Excel does not have a function to compute the matrix exponential. Therefore, we computed the terms in their power series up to 25, at which point we found their sum to approximately converge to the exponential.

Disease progression rates for individuals with referable disease were derived from the literature, as shown in Table 7.4.

Additionally, all individuals within the model have a baseline risk of severe vision loss and of death. National life tables for England were used to elicit age- and sex-specific mortality rates for each cycle of the model, and standardised mortality ratios for people with diabetes and for people with diabetes who are blind, as shown in Table 7.5. Estimates for the risk of vision loss in this population for reasons other than retinopathy or maculopathy were not available, so we assumed this to be a rare (1/100,000) event.

The progression rates in Table 7.4 and Table 7.5 are revised if an individual receives treatment. The model is designed to be flexible, allowing for the inclusion of treatments that impact retinopathy or maculopathy progression and regression at all levels. However, we only included treatment effects for which estimates were available in the literature, as shown in Table 7.6. The duration of effect for laser is assumed to be 12 months and for injections it is assumed to be 6 months. After these durations, unless another treatment is received, individuals revert to their baseline level of risk. Vitrectomy is assumed to only be indicated for people

	From	To	Log-Baseline	Age	Duration	HbA1c	TC	SBP
λ_{12}	R0M0	R1M0[R0M0]	-1.38	0.0045	0.028	0.01	-0.037	0.0041
λ_{21}	R1M0[R0M0]	R0M0	0.19	0.0058	-0.017	-0.0019	0.015	-0.00069
λ_{23}	R1M0[R0M0]	R1M0[R1M0]	-0.74	-0.011	0.026	0.0062	-0.036	-0.0022
λ_{24}	R1M0[R0M0]	R1M1 / R2M0	-4.28	0.024	-0.011	0.0055	0.023	0.0034
λ_{32}	R1M0[R1M0]	R1M0[R0M0]	-0.042	0.0084	-0.042	-0.01	0.08	-0.0031
λ_{34}	R1M0[R1M0]	R1M1 / R2M0	-2.48	-0.014	0.0042	0.016	0.034	0.005

Table 7.3: RCE hazard rates per baseline state [61]. TC = total cholesterol. SBP = systolic blood pressure.

From	To	Mean*	SE*	Distribution	Derivation	Source
R1M1	R1M0	0.000	0.000	Beta	$\alpha = 1$, $\beta = 9999$	Assumed†
R1M1	R2M1	0.009	0.003	Beta	$\alpha = 10$, $\beta = 11$, 6 months	Assumed, informed by [109]
R2M0	R1M0	0.000	0.001	Beta	$\alpha = 69$, $\beta = 403$, 25 years	[268]
R2M0	R3M0	0.010	0.003	Beta	$\alpha = 760$, $\beta = 470$, 5 years	[269]
R2M0	R2M1	0.002	0.005	Beta	$\alpha = 23$, $\beta = 49$, 14 years	[58]
R2M1	R1M1	0.000	0.000	Beta	$\alpha = 1$, $\beta = 9999$	Assumed†
R2M1	R2M0	0.000	0.000	Beta	$\alpha = 1$, $\beta = 9999$	Assumed†
R2M1	R3M1	0.028	0.003	Beta	$\alpha = 91$, $\beta = 3184$, 6 months	Assumed, informed by [109]
R3M0	R2M0	0.000	0.000	Beta	$\alpha = 1$, $\beta = 9999$	Assumed†
R3M0	R3M1	0.002	0.008	Beta	$\alpha = 15$, $\beta = 23$, 14 years	[58]
R3M1	R2M1	0.000	0.000	Beta	$\alpha = 1$, $\beta = 9999$	Assumed†
R3M1	R3M0	0.000	0.000	Beta	$\alpha = 1$, $\beta = 9999$	Assumed†

Table 7.4: Model parameters: state transition probabilities. SE = standard error. * Shown to three decimal places - the model uses the highest precision supported by Microsoft Excel. †In lieu of published estimates, some regression rates were assumed to be possible but very rare, at a monthly rate of 1 in every 10,000.

	Risk ratio*	SE*	Distribution	Derivation	Source
Risk of severe vision loss					
Due to R3	0.000	0.001	Beta	$\alpha = 26$, $\beta = 1153$, 5 years	[270]
Due M1	0.000	0.001	Beta	$\alpha = 41$, $\beta = 1388$, 5 years	[270]
Due to R3M1	0.001	0.001	Beta	$\alpha = 72$, $\beta = 1031$, 5 years	[270]
Standardised mortality ratios					
Diabetes	1.32	0.005	Normal		[271]
Blindness	3.486	0.236	Normal		[272]

Table 7.5: Model parameters: vision loss and mortality rates. Transition probability. SE = standard error.

Treatment benefit	Risk ratio*	SE*	Distribution	Source
Laser				
Retinopathy progression	0.490	0.069	Beta	[250]
SVL due to R3	0.460	0.158	Beta	[250]
SVL due to M1	0.500		Deterministic	[275]
SVL due to R3M1	0.460	0.158	Beta	Assumed
Injections				
SVL due to R3	0.087	0.194	Beta	[254]
Maculopathy progression	0.329		Deterministic	[276]
SVL due to M1	0.250	0.077	Beta	[277]
SVL due to R3M1	0.250	0.077	Beta	Assumed

Table 7.6: Risk ratios associated with treatment. SE = standard error; SVL = severe vision loss.

who have experienced vision loss and to have no consequences for the progression of disease represented in the model. Photographic screening is assumed to have a sensitivity of 0.878 and a specificity of 0.861, based on estimates provided by Scanlon et al. [273]. Slit lamp biomicroscopy is assumed to have a sensitivity of 0.760 and a specificity of 0.950 [274].

Treatment costs and frequencies are derived from the estimates presented in Chapter 6. The costs of attendance and non-attendance at screening, and the rate of attendance, are derived from the estimates presented in Chapter 5. The per-cycle cost of blindness was based on previous estimates [278]. Health state utility values are derived from Chapter 4.

7.2.5 Analysis

We used the model to estimate the cost-effectiveness of the ISDR risk-based screening programme compared with annual or biennial screening programmes. For the purpose of probabilistic sensitivity analysis, the model was set-up to run 1,000 Monte Carlo simulations to generate a cost-effectiveness plane. The Monte Carlo simulations characterise the uncertainty arising from three sources: i) random variation in the occurrence of probabilistic events, ii) uncertainty in parameter values, and iii) baseline characteristics, as each simulation randomly selects its sample.

7.3 Results

Due to the computational intensity of the model, we were unable to complete the simulations as planned. Figure 7.6 shows the front page of the model, with our

Per person	ISDR vs annual	ISDR vs biennial	biennial vs annual
Incremental cost	-£258	-£95	-£163
Incremental QALYs	0.288	0.008	0.280
ICER	*dominates*	*dominates*	*dominates*

Table 7.7: Cost-effectiveness of alternative screening programmes. ICER = incremental cost-effectiveness ratio; QALY = quality-adjusted life year.

desired inputs and the estimated duration of the simulation, which is in excess of 12 years. This estimate is derived from a moderately high specification personal computer (with a quad-core central processing unit and 16 gigabytes of random access memory). We return to this issue in the Discussion. Here, we describe the results of a probabilistic model with five simulations of 200 patients, which took 16 hours to complete.

At first use, in the ISDR trial, the RCE allocated 10.7% of people to 6-month recall, 8.6% to 12-month recall, and 80.7% to 24-month recall when using the 2.5% threshold. This meant that the annual cost of screening was reduced, though over time the proportion of the cohort being recalled at 24 months declined.

Mean incremental costs and QALYs are shown in Table 7.7. Over the lifetime duration of the simulation, stratified screening according to the ISDR recall allocation cost less on average than annual screening and biennial screening. This is primarily driven by the reduced number of screening episodes overall. On average, the simulations indicated the strong dominance of ISDR recall over biennial screening, and strong dominance of biennial screening over annual screening.

Based on these estimates, for a cohort of 10,000 people, ISDR screening would save around £2.6 million compared with annual screening. Compared with biennial screening, the individualised programme would save around £0.9 million. In terms of QALYs, risk stratification was superior on average to either annual or biennial screening. For a cohort of 10,000, individualised screening would result in 2,877 additional QALYs compared with annual screening and 76 additional QALYs compared with biennial screening. This could be due to the benefits of early intervention delaying severe vision loss for those high risk people allocated to a 6-month recall. However, the magnitude of QALY gains relative to annual screening is likely due to the random occurrence of greater mortality in the annual screening pathways.

There is a high level of uncertainty in these overall findings. Table 7.8 shows total costs, total QALYs, and the proportion of people going blind within each of the five simulations. Within this small number of simulations, all comparators were identified as least costly and most effective in at least one simulation.

By running several reduced-size simulations, we were able to anecdotally observe a lot of variation and uncertainty. A high level of variation between in-

Number of patients	10000
Cost discount rate	3.50%
QALY discount rate	3.50%
ISDR risk threshold	2.50%
☑ Run probabilistic model?	1000
Number of simulations	1000
Duration of simulation per patient	0.000476667
Estimated duration of full simulation (as specified)	
In minutes	6864000
In hours	114400
In days	4767

Run the simulation

Legend

Fixed values	Editable values
Auto-updating values	Data validation
Simulated values	Intentionally empty

Sheet	Description
Model set-up	Used to determine the operation of the model
Patient attributes	Estimates individual risk
Parameters	Lists all parameters and sources used for the model
Patient sample	Sample of patients from Liverpool
Life tables	Mortality rates by age and sex
Random numbers	A sheet of random numbers, used for simulation
Chain_Annual	Specifies probability of each event pathway for people in annual screening
Chain_Biennial	Specifies probability of each event pathway for people in proposed biennial stratified screening
Chain_ISDR	Specifies probability of each event pathway for people in ISDR stratified screening
Simulation	Records results for each patient
Results	Records overall results across all patients

Figure 7.6: Model set-up page

	Simulation 1					
	Annual		Biennial		ISDR	
	Mean	SD	Mean	SD	Mean	SD
Total cost	£253.54	364.48	£378.52	1,401.35	£256.92	419.82
QALYs	4.65	5.46	5.09	6.03	4.82	5.79
Blind	0.00%	0.00	0.50%	0.07	0.00%	0.00
	Simulation 2					
	Annual		Biennial		ISDR	
	Mean	SD	Mean	SD	Mean	SD
Total cost	£221.98	240.86	£229.15	277.83	£210.31	269.81
QALYs	4.66	4.55	4.73	4.73	4.51	4.56
Blind	0.00%	0.00	0.00%	0.00	0.00%	0.00
	Simulation 3					
	Annual		Biennial		ISDR	
	Mean	SD	Mean	SD	Mean	SD
Total cost	£682.83	5,686.35	£724.90	6,549.64	£310.08	538.14
QALYs	4.11	4.75	4.59	4.78	4.97	5.95
Blind	1.00%	0.10	1.00%	0.10	0.00%	0.00
	Simulation 4					
	Annual		Biennial		ISDR	
	Mean	SD	Mean	SD	Mean	SD
Total cost	£754.61	7,375.78	£251.19	368.07	£328.19	966.78
QALYs	4.37	5.12	4.83	5.59	4.98	5.48
Blind	0.50%	0.07	0.00%	0.00	0.50%	0.07
	Simulation 5					
	Annual		Biennial		ISDR	
	Mean	SD	Mean	SD	Mean	SD
Total cost	£681.29	5,982.11	£194.67	285.56	£198.00	307.98
QALYs	4.04	4.35.	3.99	4.70	3.99	4.70
Blind	1.00%	0.10	0.00%	0.00	0.00%	0.00

Table 7.8: Costs and outcomes from five simulations. Highlighted cells represent the lowest cost and highest outcomes within each simulation. SD = standard deviation; QALY = quality-adjusted life year.

dividual patients meant that re-runs of the deterministic model, with random samples of patients, produced highly variable results. In running small numbers of Monte Carlo simulations, we observed high levels of uncertainty in both costs and QALYs over the lifetime horizon.

To demonstrate the between-patient uncertainty, Table 7.9 shows deterministic results for a random sample of ten individuals. There is high variation in both costs and QALYs, and in the comparator associated with the lowest cost or greatest number of QALYs for each individual.

To demonstrate the within-patient uncertainty, Table 7.10 shows ten simulations for one individual (a 55 year-old man with no retinopathy at baseline). Even within this small simulation, all comparators are observed at least twice to be either the least costly or most effective option for this individual.

7.4 Discussion

We found that individualised risk-based screening may be associated with lower costs and improved outcomes compared with annual screening. However, we identified a high level of uncertainty and were not able to fully explore this due to the intractability of our model.

Our tentative findings are broadly in agreement with previous studies. In one of the earliest studies modelling the cost-effectiveness of alternative recall periods, Vijan et al. found support for biennial rather than annual screening [67]. More recently, Scanlon et al. found that low-risk individuals should be screened only every five years, while higher risk individuals could be screened every two or three years [30].

Many previous studies have advocated personalisation on the basis of risk factors, yet an economic evaluation of individualised screening for diabetic retinopathy has not previously been carried out. Some previous studies have questioned the practicality of allocating screening according to personal characteristics [116]. However, this has now been demonstrated as feasible through the ISDR trial [279]. The evaluation of such a programme is now therefore timely.

7.4.1 Strengths and limitations

We developed a flexible decision model with the capacity to simulate complex screening pathways. The model was designed to evaluate individualised risk-based screening with different risk thresholds for screening recall. The development process was thorough and included input from a range of stakeholders, including clinicians, statisticians, mathematicians, and patients. As described in earlier chapters of this thesis, we generated new evidence where candidate parameters for our model would be lacking in the published literature.

	Baseline characteristics			Annual		Biennial		ISDR	
Patient	State	Age	Duration	Cost	QALYs	Cost	QALYs	Cost	QALYs
1	1	63	2	£84	2	£520	14	£427	7
2	1	55	7	£68	2	£411	6	£755	16
3	1	69	14	£110	4	£68	2	£32	1
4	1	71	5	£26	0	£26	1	£42	1
5	3	63	7	£178	6	£120	3	£32	1
6	1	57	11	£490	8	£26	1	£146	7
7	1	52	1	£210	7	£290	6	£146	6
8	1	40	2	£1,321	21	£421	12	£390	11
9	1	51	3	£300	12	£16	0	£190	3
10	1	61	11	£78	2	£48	2	£94	4

Table 7.9: Deterministic cost-effectiveness results for a single individual. Highlighted cells represent the lowest cost and highest outcomes within each simulation. State corresponds to Markov state as described in Table 7.2. Duration = duration of diabetes in years; QALY = quality-adjusted life year.

	Annual		Biennial		ISDR	
Simulation	Cost	QALYs	Cost	QALYs	Cost	QALYs
1	£42	1	£26	0	£94	5
2	£164	2	£136	5	£1,345	8
3	£58	2	£52	1	£42	1
4	£64	3	£16	0	£94	4
5	£42	1	£190	2	£52	1
6	£364	9	£591	9	£110	5
7	£316	7	£52	1	£52	2
8	£132	5	£246	9	£248	7
9	£162	5	£362	7	£84	4
10	£471	10	£272	9	£138	0

Table 7.10: Probabilistic cost-effectiveness results for a single individual. Highlighted cells represent the lowest cost and highest outcomes within each simulation. QALY = quality-adjusted life year.

The complexity of our model, which ultimately rendered it inexecutable, was justified by the level of variation that we anecdotally observed in running reduced simulations. A simpler model, such as that which we initially developed, would not be capable of generating meaningful estimates of the cost-effectiveness of risk-based screening.

While this model was not able to evaluate the cost-effectiveness of risk-based screening compared with alternative programmes, its structure and operational assumptions constitute a strong basis for the construction of a model in a more efficient programming environment.

We were not able to complete the simulations as planned, due to the limitations of Microsoft Excel for making complex calculations. This means that our results, which are based on a greatly reduced simulation, do not provide a robust estimate of the cost-effectiveness of individualised screening compared with annual or biennial screening. Clearly, this is the main limitation of our modelling work. However, there are others that should be noted.

The only differentiation of care in our model was according to individual risk of disease onset. It is possible that treatment effectiveness may differ according to risk. There is scope to differ treatment effectiveness or costs according to risk within the model, but there may be ethical concerns associated with this.

Our approach to modelling did not allow for an accurate characterisation of changes in visual acuity over time. Incremental changes in vision are likely to impact on individuals' health-related quality of life. Therefore, if changes in visual function *within* retinopathy disease states are related to the choice of screening programme, our model will provide biased results. Bias may also arise in our

model due to the way in which the matrix exponential was approximated. In some cases, the power series may not have converged to the exponential, which would result in poorer estimations of individual risk.

7.4.2 Implications

Our study has highlighted some of the challenges associated with the evaluation of risk-based screening using decision modelling. Markov models are a popular approach to the development of risk calculation engines. This lends itself well to the development of a decision analytic state transition model for its evaluation. However, we encountered difficulties in this approach.

It is impractical to evaluate the use of a RCE in screening using cohort state transition methods. This is because of the inherent heterogeneity in the nature of the time dependency of state transition rates. Specifically, the rate of transition from state A to state B is not only a function of previous states but of previous events and of individual characteristics.

Treatment is the primary basis for transition rate adjustment. The probability of receiving treatment is contingent on both previous disease progression and whether or not the individual is screened, which in turn is dependent on individual characteristics. The effectiveness of risk-based screening, compared with standardised recall, depends on the right people being referred to treatment. This means that it is important to identify *which* patients, as opposed to just *what proportion* of patients receive treatments. So, in practice, a cohort state transition model cannot adequately identify whether a reduced screening interval would result in earlier treatment for a high risk person or a low risk person.

Decision analytic state transition models are commonly built using Microsoft Excel spreadsheet software. Despite the fact that use of a Markov model in the development of the RCE supported methodological consistency in the decision modelling approach, we found that Microsoft Excel was not suitable for our needs.

Great effort was expended in trying to rationalise the simulation and improve the efficiency of the Visual Basic code. Several changes were made to the operation of the model for the sole purpose of increasing efficiency. These included the automatic termination of simulations once a person died, rather than running the model over a fixed time horizon, and preventing Microsoft Excel from recalculating cells unnecessarily. The full simulation Visual Basic script is provided in Appendix F. These changes brought significant efficiency savings, as the first version of the model was estimated to take 458 years to complete.

Our planned analysis required the model to simulate in excess of one billion cycles, which would be demanding for any software or programming language. However, the failure of our model was not due simply to the sheer number of simulations. The major computational burden of the model was the need to

repeatedly estimate the exponential of a matrix, which could not be rationalised any further. Our research has demonstrated that it is not feasible to incorporate a risk calculation engine into a decision analysis in Microsoft Excel if the risk calculation engine requires the computation of a matrix exponential.

Using a spreadsheet as the basis for modelling was also problematic in light of recent policy changes. We found our model to lack the necessary flexibility to allow for the simple incorporation of an alternative comparator.

It is not possible to derive clear policy implications from the research reported in this chapter. Despite the inconclusiveness of our modelling work, there is now a substantial evidence base supporting the extension of screening intervals beyond one year for people with diabetes at low risk of developing STDR. However, the optimum means by which to identify those of low or high risk, and appropriate methods for evaluation, are still under development.

Chapter 8

Individualised cost-effectiveness analysis

Summary

In this chapter, we argue that the evaluation of risk-based screening justifies a new perspective on cost-effectiveness analysis. To support the development of risk-based screening, we outline a simple framework for individualised cost-effectiveness analysis (iCEA). The basis for iCEA in risk-based screening is defined on the grounds of the relationship between individual risk of disease onset and the cost-effectiveness of screening. The framework can be used to estimate an optimal risk threshold for screening eligibility at the population level or the optimal recall period for an individual with a given level of risk. We present a stylised example in order to demonstrate the value and implications of the use of iCEA in risk-based screening for diabetic retinopathy.

8.1 The development of screening and CEA

Risk-based screening involves the differentiation of care according to individual risk. This is implicitly because of — and justified by — heterogeneity in the cost-effectiveness of screening for individuals with different levels of risk of disease onset. The ISDR study is concerned with using information about an individual's risk to differentiate the frequency of recall for screening for diabetic retinopathy. A possible outcome in the near future is that individuals will be stratified into a number of groups based on their level of risk as estimated by a risk calculation engine. This type of programme was evaluated in Chapter 7 and the potential for this type of risk-based screening to be cost-effective compared with standardised screening was demonstrated. We can also conceive of taking this further. As outlined in Chapter 1, screening programmes could be truly individualised on a per-person basis and specifically tailored to the individual. In this chapter we consider the role of cost-effectiveness analysis in this context and the methodological adjustments that may be necessary. Chapter 9 then explores some of the ethical challenges that arise as a consequence.

We adopt the view that the purpose of cost-effectiveness analysis is not simply to evaluate that which can be observed in two or more pre-defined groups, but to directly inform that which might be considered for evaluation. That is, cost-effectiveness analysis (CEA) should not only be used to determine whether a particular group of individuals should or should not receive an intervention. Rather, CEA should be able to address the decision problem of *who* should receive an intervention.

As screening moves from standardisation through to stratification and individualisation (as introduced in Chapter 1), it is important to consider whether our approaches to cost-effectiveness analysis should adapt. For all approaches, standardised CEA *could* be used. For example, a stratified programme could be compared with a standardised programme in the context of a clinical trial, with comparison of costs and consequences in aggregate. However, most economic evaluations that inform health technology assessment (HTA) are now model-based and therefore not confined to that which can be evaluated within a trial. This is an important point when we consider that risk-based screening could vary in practice from constant screening for the entire asymptomatic population to no screening for anybody ever. It is impossible to evaluate every potentially cost-effective iteration of a risk-based screening programme within an observational study. However, decision modelling methods may offer the means of evaluating this level of differentiation. If so, the efficiency of screening programmes might be improved by taking into account the heterogeneous effects of individualised screening.

In this section we outline the basis for differentiating CEA in the context of

the different stages of screening programme development.

8.1.1 Standardised CEA

Standardised screening is amenable to standardised CEA. Traditionally, screening programmes have been evaluated along the lines of randomised controlled trials (RCTs), in which a fixed group of people are randomised to either receive screening or not. Indeed, the NSC's terms of reference state that, "There should be evidence from high quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity" [280]. Model-based analyses might use observational data comparing those who have received screening with those who have not. In this case, standardised CEA is also appropriate.

Such analyses take the population that is or ought to be eligible (or considered for eligibility) for screening as given and fixed. The defining characteristic of standardised CEA is that it estimates cost-effectiveness at the aggregate. It takes the average costs and outcomes observed in a 'treatment' group (e.g. those in receipt of screening) and compares them with a 'control' group (e.g. those not in receipt of screening) in order to estimate incremental costs and outcomes. These estimates can then be used to calculate a measure of cost-effectiveness known as the incremental cost-effectiveness ratio (ICER). This can be shown algebraically as

$$ICER = \frac{C_1 - C_0}{E_1 - E_0}, \tag{8.1}$$

where C_1 and E_1 are the costs and outcomes observed in the treatment group and C_0 and E_0 are the costs and outcomes observed in the control group.

A difficulty associated with the ICER approach is that it gives results in the form of a ratio, which can be challenging to analyse statistically. A simple extension is the net benefit approach, which can express cost-effectiveness in terms of either net health benefit (NHB) [281] or net monetary benefit (NMB) [282] as

$$NHB = (E_1 - E_0) - (C_1 - C_0)/\lambda, or \tag{8.2}$$

$$NMB = \lambda(E_1 - E_0) - (C_1 - C_0), \tag{8.3}$$

where λ is the willingness to pay for a unit of the outcome (E). A simple decision rule can then be employed such that the intervention is only provided if it provides a positive net benefit; for example, where $NMB > 0$. Clearly, an added difficulty of this approach is that the decision maker must know λ [283]. Though this may in practice be a difficulty, in principle a decision maker *should* know the value of λ in a given context as part of a more sophisticated decision rule [284].

In the context of screening for diabetic retinopathy, standardised CEA has been used to evaluate a variety of screening procedures including systematic vs opportunistic screening [237], alternative screening strategies [285], and automated grading [98].

8.1.2 Stratified CEA

The standardised approach to CEA — and, indeed, the standardised approach to screening — is only appropriate where there is no substantial amount of heterogeneity. Where there is heterogeneity, failing to take account of it may lead to systematically inefficient resource allocation decisions [39]. In the context of recurrent screening programmes, a simple basis for stratification might be on the basis of disease progression. For example, in diabetic eye screening, individuals with some signs of non-referable disease might be recalled to screening with more frequency than individuals with no signs of disease. Risk-based screening might be based on risk cohorts, such as low-, medium-, and high-risk subgroups. In this case, limited use criteria can be determined using stratified cost-effectiveness analysis, as developed by Coyle et al. [286].

Stratified CEA estimates the cost-effectiveness of an intervention within subgroups, such that

$$ICER_i = \frac{C_{1i} - C_{0i}}{E_{1i} - E_{0i}}, \tag{8.4}$$

where $ICER_i$ is the within-subgroup cost-effectiveness for sub-group i. E_{1i} and E_{0i} are the outcomes for subgroup i in the treatment group and the comparator group respectively, and C_{1i} and C_{0i} are the costs for subgroup i in the treatment group and the comparator group respectively.

Practically, stratified CEA does not require any major adjustment to the form of standard clinical and economic evaluation. Its principal requirement is that data must be available from a sufficiently large sample to detect differences in costs and outcomes within subgroups, and that those alternative interventions can be tested within subgroups. Subgroups might be defined in terms of individual characteristics such as age or sex, and many stratified cost-effectiveness analyses are conducted on this basis.

Both standardised and stratified CEA are based on sample means and may be the product of a number of possible prognoses and treatment pathways. Stratified CEA allows for a revised net benefit decision rule, such that

$$TNB = \sum_i NB_i \quad \forall_i \quad where \quad NB_i > 0 \tag{8.5}$$

where NB_i is the net benefit within subgroup i and TNB is the total net monetary benefit. That is, limited use criteria can be defined such that the intervention should only be provided to those subgroups in which net benefit exceeds zero.

There is a growing recognition that economic evaluation should take heterogeneity into account and conduct stratified analyses [287, 288], including in the context of screening specifically [289].

There are some examples in the literature of cost-effectiveness analyses of screening programmes that take account of individuals' risk factors. These studies implicitly adopt a stratified CEA approach. Using a Markov model to evaluate colorectal cancer screening, Dan et al. found that selective screening, based primarily on an individual's age, was more effective than standardised screening [290]. Similarly, Lansdorp-Vogelaar et al. found a small benefit associated with individualised colonoscopy screening [291]. Also using decision modelling methods, Round et al. found that the most cost-effective screening method for gestational diabetes mellitus depended on a woman's individual risk of disease [292]. Similarly, Aus et al. found that screening intervals for prostate cancer should be individualised based on prostate-specific antigen levels [293].

When using stratified CEA, it is unlikely that a large number of sub-groups could be analysed. Stratified analyses can therefore fail to fully account for heterogeneity due to the often broad nature of subgroups, and this may lead to suboptimal decisions [294]. This is because the evaluation of each group requires the availability of incremental cost and outcome estimates and, as the number of subgroups within the population increases, the size of these subgroups diminishes. This is likely to result in parameters derived from small populations that will introduce high levels of uncertainty.

There is an additional limitation to stratified CEA in the context of risk-based screening, relating to the definition of the subgroups. Subgroup risk thresholds might be determined on the basis of various factors such as acceptability, safety, affordability, or cost-effectiveness. Indeed, risk estimates now inform clinical decision-making in many contexts. However, to define clinically meaningful subgroups based on individual risk is a difficult task.

The probability of disease onset that an individual faces could vary infinitely between zero and one, and a risk calculation engine could be developed that is able to estimate risk with a level of precision that approaches infinity. Given the ratio scale properties of individual risk, we do not believe it to be intuitively possible to determine meaningful risk subgroups. Any threshold between low-, medium-, or high-risk will at best be imprecise in its association with treatment benefit. It is possible that very similar people (i.e. either side of a threshold) may receive very different care, while very different people (i.e. at the extremes of a risk group) may receive the same care.

The application of clinically meaningless risk thresholds has ethical implications that are discussed in Chapter 9. In the following section, we demonstrate a means of identifying optimality within an individualised risk-based screening programme that does not require arbitrary definition of subgroups.

8.1.3 Individualised CEA

Personalisation is increasingly recognised as an important part of health care, and economists are beginning to develop means of incorporating it in to economic evaluation. Often termed 'precision medicine', the differentiation of care on the basis of individuals' characteristics has been identified as having the potential to confer benefits to patients as well as cost savings [295]. However, there has been very limited discussion of the methodological implications for economic evaluation of individualised health care. Here, we introduce a simple framework for individualised cost-effectiveness analysis (iCEA). The notion of iCEA has been briefly discussed in earlier work [296–298], but the theoretical and practical basis for such an approach has never been described.

Individualised cost-effectiveness analysis involves the estimation of expected costs and outcomes (and their combination) at the individual level, rather than the aggregate for a population or subgroup. It can be implemented as a generalisation of stratified CEA. Individualised CEA involves the use of stratified CEA under two real-world conditions. Firstly, in individualised CEA the number of potential subgroups tends to infinity, such that

$$\lim_{i \to \infty} \tag{8.6}$$

Secondly, the size of subgroups tends to 1, such that

$$\lim_{N_i \to 1} \tag{8.7}$$

where N_i is the size of subgroup i.

Both standardised and stratified CEA involve the estimation of incremental costs and effects based on the difference in the sample mean of two groups. They depend on — or at least assume the possibility of — observation of a counterfactual. This counterfactual will usually be based on a randomised controlled trial. Individualised CEA cannot satisfy this requirement because of the two conditions outlined above. One can never observe the counterfactual costs and outcomes where the sample size is equal to one, and therefore cannot estimate incremental effects using traditional approaches. On this basis, the fundamental problem of causal inference cannot be addressed and iCEA appears to be impossible in practical terms. However, as we outline below, the use of further assumptions enables the analyst to operationalise an iCEA approach.

8.2 The nature of individualisation

The presence of uncertainty has become central to the estimation of cost-effectiveness, and estimates are usually made on the basis of expected values within a framework of decision analysis. This notion is maintained in iCEA, which does not preclude the consideration of mutually exclusive prognoses for an individual. For instance, in the context of screening, any individual might screen positive or negative and the outcome cannot be predetermined. In this section we outline the means by which CEA can be individualised and the basis for individualisation according to risk.

8.2.1 Individualisation factors

Whether or not an intervention is cost-effective for an individual is determined by innumerable factors whether biological, social, or behavioural. The relevant effects of all of these can be summarised according to the top-level factors that inform the decision of whether or not an intervention is deemed cost-effective. These are:

- Probabilities,
- Costs,
- Outcomes, and
- Willingness to pay.

Each of these might be heterogeneous across individuals. The net monetary benefit of an intervention for individual i might thus be estimated as[1]

$$NMB_i = \lambda_i(E_{1i} - E_{0i}) - (C_{1i} - C_{0i}) \tag{8.8}$$

Individuals are defined by an infinite number of characteristics. Some of these will be observable and will influence the expected costs and outcomes associated with a treatment. The relationship between individual characteristics and the four factors of cost-effectiveness can be estimated and used to predict the expected cost-effectiveness of treatment for an individual.

It may be possible in some cases that, with a large enough sample and little heterogeneity, the expected cost-effectiveness of treatment could be estimated for

[1]Willingness to pay for the outcome of interest is not generally considered to be a source of heterogeneity and such considerations are often dispatched to the realm of equity analysis. Equation 8.8 clearly shows that such simple interpersonal 'equity' concerns can easily be considered a matter of efficiency analysis.

each individual and all relevant determinants accounted for. Though this seems unlikely.

Randomised trials are not likely to be large enough to elicit such relationships. Regardless, randomised controlled trials cannot observe a true counterfactual and are thus subject to the 'fundamental problem of causal inference' [299]. It may therefore be necessary to use non-randomised observational data from large cohorts to elicit the individual-level determinants of costs and outcomes.

Although the challenge of observing the counterfactual remains, economists have developed a suite of methods that seek to support causal inference. There are now methods by which incremental costs and effects might be estimated for each individual. The analyst might for example use conditional average (CATE) or person-centred (PeT) treatment effects [300]. Widely used regression-based approaches to CEA could prove valuable in carrying out individualised CEA [301].

Conceptually, CATE aligns with stratified CEA insofar as they estimate treatment effects conditional on an individual exhibiting a specific characteristic, which might be the basis for inclusion in a subgroup. For example, CATE could estimate effects for men and effects for women. The development of PeT effects has focussed on evaluating the role of individual preferences and 'passive personalisation' [302, 303]. The purpose of PeT effects are to provide estimates that account for all observed and unobserved heterogeneity and selection into treatment. They could therefore, in principle, be used to inform a multi-factorial individualisation whereby all factors — probabilities, costs, outcomes, and willingness to pay — vary at the individual level.

The use of aggregate data from epidemiological or other observational studies is common in economic evaluation in health care, often being used for parameters in decision analytic models. It would be possible to carry out multi-factorial iCEA using aggregate level data. For example, multiple trials may have been carried out which might demonstrate a relationship between outcomes and age, and between costs and employment status. These relationships could be incorporated into an iCEA using either CATE or PeT effects to predict cost-effectiveness at the individual level. Only limited work has been done to incorporate econometric analyses into CEA, and the development of such tools and their incorporation into iCEA is beyond the scope of this thesis.

In some cases, it may be desirable to individualise CEA on the basis of a single factor. This may be due to practical or ethical reasons. In this case, only one factor is assumed to vary across individuals and to determine cost-effectiveness at the individual level. In this thesis we focus on the single factor of individual risk, which is a probability within the estimation of cost-effectiveness. In Chapter 1 we outlined the relevance of this factor to screening programmes. In Chapter 9

we present an argument that unifactorial iCEA is an equitable approach to the optimisation of risk-based screening.

Central to this approach is the link between individual risk of disease onset and the cost-effectiveness of screening for that disease, which has been implicitly recognised in practice but not explicitly outlined in the academic literature. Below we demonstrate the nature of this relationship and present a basis on which to use iCEA in risk-based screening.

8.2.2 The relationship between individual risk and cost-effectiveness

Within screening programmes, an individual's screening outcome is not random; the probability of screening positive is dependent on a set of (observed or unobserved) risk factors. Through analysis of these risk factors it is now possible, in many cases, to estimate an individual's risk of developing a disease within a given period of time. If we seek to maximise an individual's (expected) health, the extent to which a screening programme can be beneficial depends on the effectiveness of treatment or care following a true positive screening result. If an individual is more likely to screen positive — assuming they are identified early enough — it is more likely that they will receive the intended benefits of screening. As such, it is clear that individuals with different risk levels will experience heterogeneous benefits from screening.

The expected incremental costs and outcomes of screening, at the individual level, are dependent on at least two binary probabilities; whether the individual has a given disease and whether they are screened positive or negative. Accounting for the specificity (true negative rate) and sensitivity (true positive rate) of the screening test, this means that there will be four possible outcomes for an individual who is screened and two for an individual who is not. Figure 8.1 shows these six possible screening pathways, which are generalisable to most screening programmes.

For a given population or individual, ICER calculations in the evaluation of screening interventions are dependent upon at least three probabilities, i) the probability of a positive disease state, ii) the false negative rate and, iii) the false positive rate. For the screening intervention in Figure 8.1, the probability of each pathway for the individual, depending on whether they are screened or not, is:

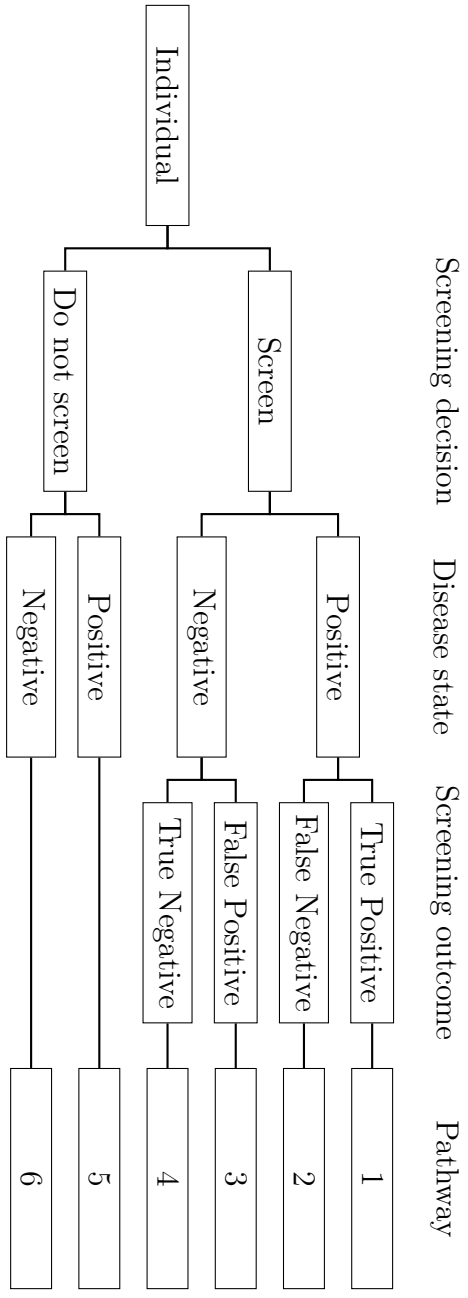

Figure 8.1: Screening pathways

$$Prob(1) = r(1-\beta) \tag{8.9}$$
$$Prob(2) = r\beta \tag{8.10}$$
$$Prob(3) = \alpha(1-r) \tag{8.11}$$
$$Prob(4) = (1-r)(1-\alpha) \tag{8.12}$$
$$Prob(5) = r \tag{8.13}$$
$$Prob(6) = 1-r \tag{8.14}$$

where r is the probability that the disease (e.g. STDR) is present, α is the false positive rate and β is the false negative rate of the screening intervention. The probability that an individual has the disease (r) corresponds to the incidence or prevalence of the disease (depending on the nature of the intervention) in the population being considered for screening. In clinical trials, r is only observed at the sample level. In modelling studies, r is usually estimated at the population level. Our primary assertion in this chapter is that, given recent developments in the estimation of risk, r can be estimated for an individual. This would be in the form of a hazard rate, as described in Chapter 1 and implemented in Chapter 7.

In practice, the values derived from equations 8.9–8.14 may not be known, but they are nevertheless the underlying probabilities in defining an individual's pathway. The expected incremental cost of screening ($C_{1i} - C_{0i}$) and the expected incremental effect ($E_{1i} - E_{0i}$) are dependent on these probabilities. As such, C_{1i}, C_{0i}, E_{1i}, and E_{0i} are defined as:

$$C_{1i} = \sum_{k=1}^{4}(Prob(k) \times C(k)) \tag{8.15}$$
$$C_{0i} = \sum_{k=5}^{6}(Prob(k) \times C(k)) \tag{8.16}$$
$$E_{1i} = \sum_{k=1}^{4}(Prob(k) \times E(k)) \tag{8.17}$$
$$E_{0i} = \sum_{k=5}^{6}(Prob(k) \times E(k)) \tag{8.18}$$

where $k = 1, 2 \ldots 6$ are the possible pathways shown in Figure 8.1.

In order to clearly demonstrate the relationship between individual risk and the cost-effectiveness of screening, we adopt three simplifying assumptions:

1. A true positive outcome (i.e. Pathway 1 in Figure 8.1 and Equation 8.9) is the only outcome of value.

2. Individual risk is equal to the true probability that disease is present.

3. Individualisation can be based only on individual risk (r_i), such that the cost of screening (C_S) and the benefit of a true positive screening outcome (E_S) is constant across individuals.

Table 8.1 shows the expected costs and outcomes associated with each pathway under these assumptions. The expected incremental cost and effect of screening is estimated by

$$C_1 - C_0 = \sum_{k=1}^{4}(Prob(k) \times C(k)) - \sum_{k=5}^{6}(Prob(k) \times C(k)) \tag{8.19}$$

$$E_1 - E_0 = \sum_{k=1}^{4}(Prob(k) \times E(k)) - \sum_{k=5}^{6}(Prob(k) \times E(k) \tag{8.20}$$

Simplified, and using the expressions shown in Table 8.1, these expressions allow for an ICER to be estimated for an individual as

$$ICER_i = \frac{C_S}{(1-\beta)r_i E_S} \tag{8.21}$$

where C_S is the cost of screening, β is the false negative rate of the screening test, r_i is the risk of disease onset for individual i and E_S represents the value (here assumed equal to 1) of a true positive screening outcome. This expression can be presented in terms of net monetary benefit as

$$NMB_i = \lambda(1-\beta)r_i E_S - C_S \tag{8.22}$$

Here we assume a true positive screen (E_S) to be beneficial and fixed across individuals with different risk levels. C_S only includes the cost of screening; any cost associated with risk estimation and the cost of treatment are not included. Under our assumptions, the costs and outcomes in Pathway 3 and Pathway 4 are equivalent, meaning that the false positive rate (α) does not influence the ICER. This means that the relationship between individual risk (r) and the expected incremental benefit of screening is positive and linear. Similarly, if we assume that the cost of screening is positive and constant across risk levels, a higher level of risk (*ceteris paribus*) will be associated with a lower ICER. As such, the relationship between an individual's risk of developing a disease and the expected cost-effectiveness of screening them must be positive. Under the assumptions used in this example, the cost-effectiveness of screening monotonically increases with risk.

Note that equations 8.21 and 8.22 are not necessarily generalisable, and could simplify to different expressions if the cells in Table 8.1 were changed. For example, a negative effect could be associated with a false positive result. It is the process of defining them that is generalisable. They are presented only to illustrate the relationship between the variables under our assumptions.

Pathway	**Probability**	**Cost**	**Expected cost**	**Outcome**	**Expected outcome**
Screening					
1	$r_i(1-\beta)$	C_S	$r_i(1-\beta)C_S$	E_S	$r_i(1-\beta)E_S$
2	$r_i\beta$	C_S	$r_i\beta C_S$	0	$r_i\beta 0$
3	$(1-r_i)\alpha$	C_S	$(1-r_i)\alpha C_S$	0	$(1-r_i)\alpha 0$
4	$(1-r_i)(1-\alpha)$	C_S	$(1-r_i)(1-\alpha)C_S$	0	$(1-r_i)(1-\alpha)0$
Total	1		C_1		E_1
No screening					
5	r_i	0	$r_i 0$	0	$r_i 0$
6	$1-r_i$	0	$(1-r_i)0$	0	$(1-r_i)0$
Total	1		C_0		E_0

Table 8.1: Alternative screening pathways and expected costs and outcomes

8.3 Optimising risk-based screening

The expressions presented can be used to individualise screening decisions in at least two ways: i) by setting an optimal threshold for screening; above which an individual is offered screening and below which they are not and ii) by setting an optimal recall period for each individual. These approaches are equivalent in theory, if not in practice.

8.3.1 Setting a threshold

The use of risk-based thresholds and patient stratification is well-established [304]. However, the potential to allocate variable screening on an individualised basis is a new development. We can use the net benefit approach to estimate whether, at a given level of willingness to pay per true positive screening, it is cost-effective to screen an individual with a given level of risk. In order to do this, we need simply solve equation 8.22 for r_i where net monetary benefit equals zero, such that

$$r_i = \frac{-C_S}{\lambda(\beta - 1)E_S} \tag{8.23}$$

For our example, in which the net benefit of screening monotonically increases in risk, equation 8.23 indicates the minimum level of risk at which individuals should be screened, and thus defines a threshold.

In an optimised programme, the time to next screen would be decided following each negative screening outcome, rather than a screening frequency defined at an arbitrary time point. The results of the screening test can be used to inform a revised estimate of an individual's risk level. The incremental cost-effectiveness of subsequent screens can be characterised in the same way as the first; the decision process is the same, but the inputs may have changed. The relationship rationalises to that already discussed.

8.3.2 Allocating recall

For some programmes it will be more useful to estimate an optimal recall period for an individual at a given point in time, rather than to define a minimum level of risk for population screening. In order to achieve this, it is first necessary to decide on the lowest practical screening interval; a clinic may have the administrative capacity to recall individuals on a weekly, monthly or yearly basis. Maintaining our earlier assumptions, and assuming that individuals' hazard ratios are estimated based on a time at risk (T) of one year and that screening clinics are capable of recalling individuals for screening on a weekly basis, we can

estimate the recall time at which net monetary benefit becomes positive. This is simply a matter of estimating the cost-effectiveness of screening an individual at various periods of recall, accounting for the fact that risk increases in time. The optimal recall period, in terms of cost-effectiveness, can be found by solving for $NMB = 0$ at any given level of individual risk. The optimal recall period for an individual with risk r can be established in three steps:

1. Assuming a constant average rate of disease onset, convert the individual's estimated level of risk to an instant rate using the formula $= -\frac{ln(1-r)}{T}$,
2. Convert the individual's instant rate back to the risk associated with the minimum feasible interval (f) using the formula $\hat{r} = 1 - \exp^{-r\frac{p}{T/f}}$,
3. Select the recall period (p) such that net monetary benefit is at its lowest possible positive value, where r is replaced with $\hat{r}$.

Because $r \sim \hat{r}$, the risk threshold associated with the optimal recall period will be approximate to that found at the programme level.

8.3.3 A simulated example

We present a stylised application of our framework for the optimisation of risk-based screening using aggregate data relating to the NDESP. We use reported figures from a CEA of systematic photographic screening compared with opportunistic screening [237], though we assume that the study compared systematic screening with no screening for the sake of the demonstration. We assume that attendance in the study was 100% (in fact, only four in every five people attended screening). The primary outcome is cost per true positive screen. These examples are simplified, and as such the findings are not designed to inform policy. Rather, they demonstrate the potential value of using such an approach for the optimisation of risk-based screening.

Table 8.2 shows figures provided by James et al. [237] for screening for diabetic retinopathy, which we use to populate Equation 8.22. The CEA is individualised based on the single factor of individual risk of disease onset, in the way described above. In order to demonstrate the iCEA approach — including probabilistic sensitivity analysis (PSA) — we use the parameters in Table 8.2 to simulate 100 sets of net benefit estimates for individual risk values from 0 to 1 and willingness to pay threshold values from £0–500.

Figure 8.2 shows the resulting relationship between individual risk and the cost-effectiveness of screening when the parameters in Table 8.2 are used to populate Equation 8.21. Only risk levels up to 0.1 and ICERs up to £5,000 are shown. We can see that the higher the level of individual risk, the lower the resulting cost per true positive screening result. From Figure 8.2 it is clear that in this example

	Mean	Standard error
α	0.49	0.016
β	0.02	0.003
C_S	£21.00	2.000*
E_S	1.00†	NA
$ICER$	£209.00	NA

Table 8.2: Example model parameters. * Assumed. †The outcome is true positive screening result.

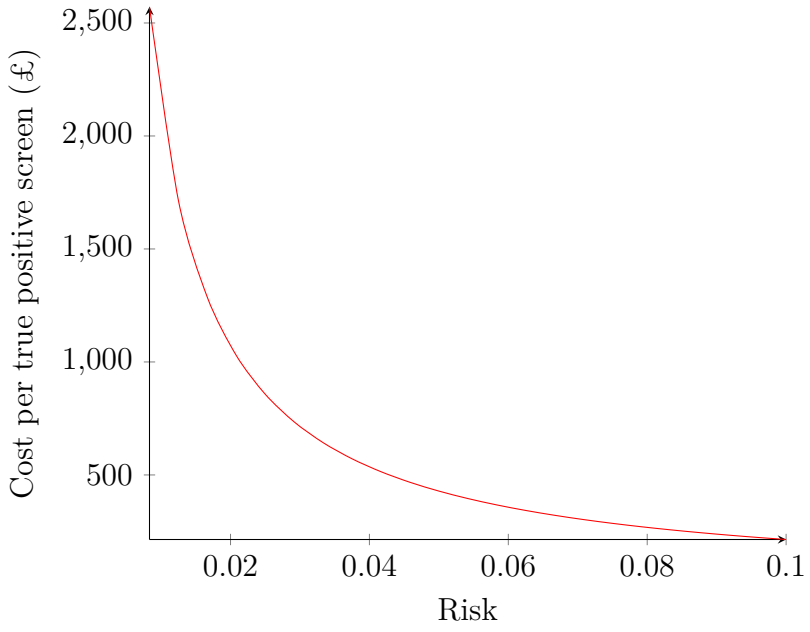

Figure 8.2: The relationship between individual risk and cost-per-outcome

the cost-effectiveness of screening is very sensitive to differences in individual risk when risk is very low. This is primarily because we have not accounted for errors in the estimation of risk, thus implying almost perfect identification of those likely to benefit from screening.

Figure 8.2 highlights that, for a given level of willingness to pay threshold, there will be a corresponding minimum level of risk at and above which it is cost-effective to screen an individual, and below which the expenditure would not be deemed cost-effective. To our knowledge, no previous attempts have been made to estimate this figure, either in the context of DR or any other type of screening. Figure 8.2 therefore demonstrates the potential for optimality in risk-based screening. For a given set of cost and outcome estimates, and for a given willingness to pay, there is likely to be a level of risk at which the net benefit of screening becomes positive.

We can use the net benefit approach to estimate whether, at a given level of willingness to pay per true positive screening result, it is cost-effective to screen

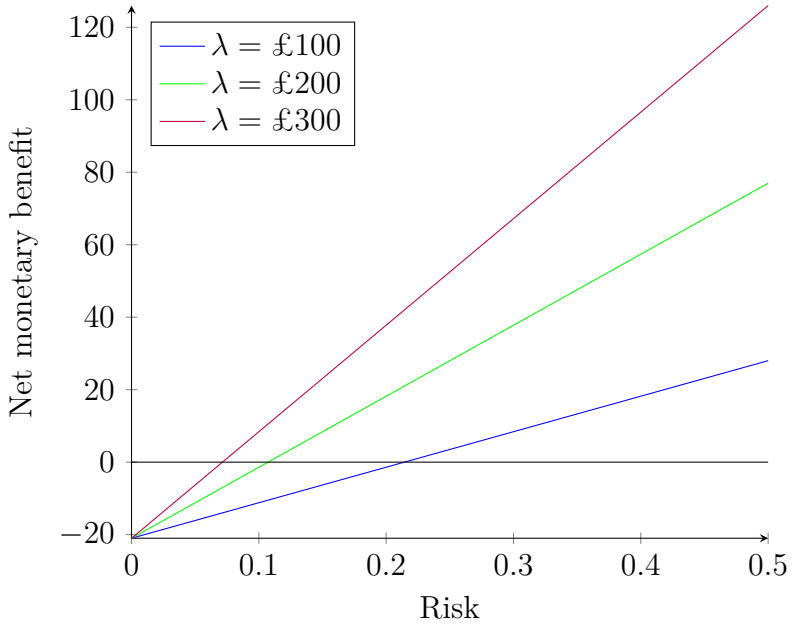

Figure 8.3: The relationship between individual risk and net benefit

an individual with a given level of risk. Figure 8.3 shows the relationship between individual risk and net benefit. Here, for the purpose of the demonstration, we present willingness to pay values for a true positive screening outcome of £100, £200, and £300. As can be seen in Equation 8.22, the slope is defined by the willingness to pay and the intercept, in our case, is defined by the cost.

As described above, by solving for r at our assumed λs, we obtain the minimum level of risk at which individuals should be screened. Where $\lambda = £100$, the minimum level of individual risk is 0.214. At £200 and £300 the thresholds are 0.107 and 0.071, respectively. James et al. find that, for systematic DR screening, the incremental cost per true positive is £209 [237]. Where $\lambda = £209$, it would be cost-effective to screen anybody with a risk level of 0.103 or higher. This assumes that the time period used in the estimation of the hazard rate matches the screening interval.

DR screening can also be varied in terms of the frequency of its administration, with people offered individualised recall periods depending on their level of risk. Figure 8.4 shows the expected net benefit associated with alternative recall periods for individuals with risk levels of 0.05, 0.1 and 0.2. In this example, $\lambda = £209$, $T = 1$, and $f = 52$. The optimal recall period, in terms of cost-effectiveness, can be found by solving for $NMB = 0$ at any given level of individual risk. The relationship that this produces is shown in Figure 8.5

Figures 8.2–8.5 show deterministic relationships. Using the PSA it is possible to demonstrate a relationship between individual risk, willingness to pay per true positive screening outcome, and the probability of cost-effectiveness. The resulting three-dimensional cost-effectiveness acceptability surface is shown in

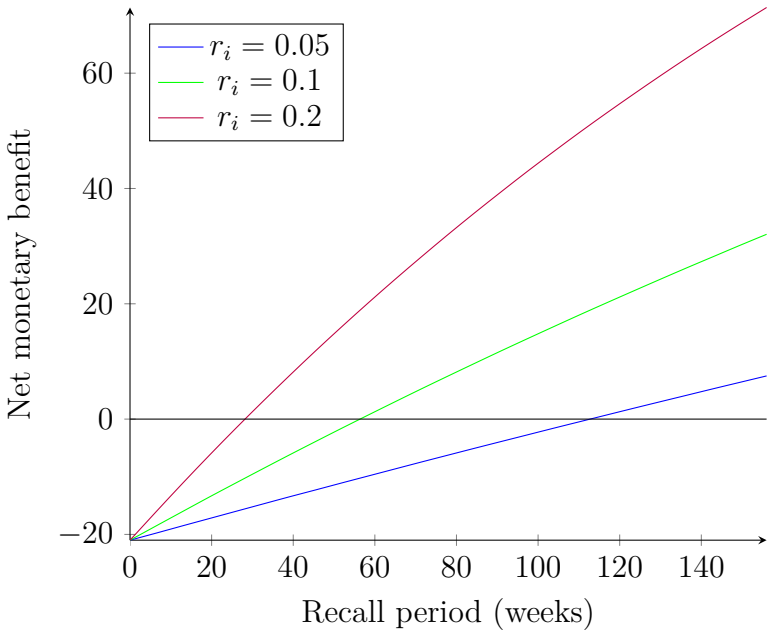

Figure 8.4: The relationship between recall period and net benefit

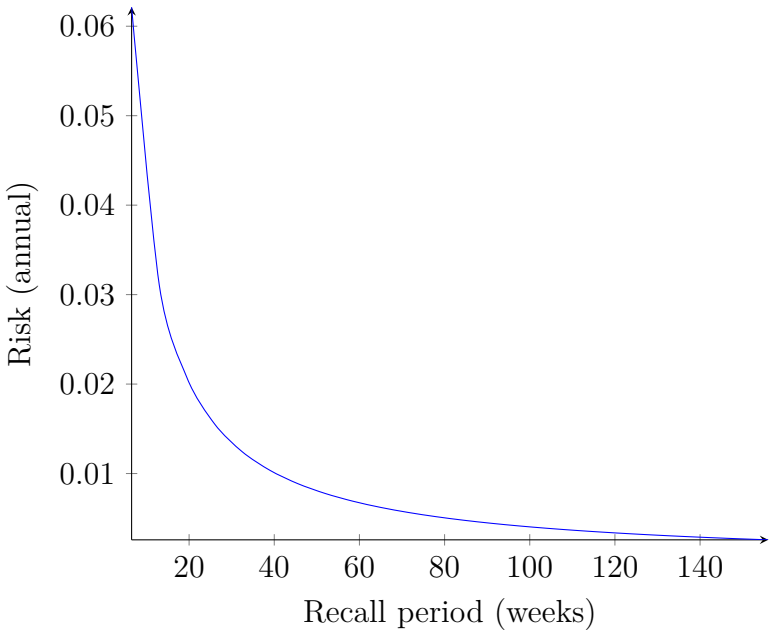

Figure 8.5: The relationship between optimal recall period and individual risk

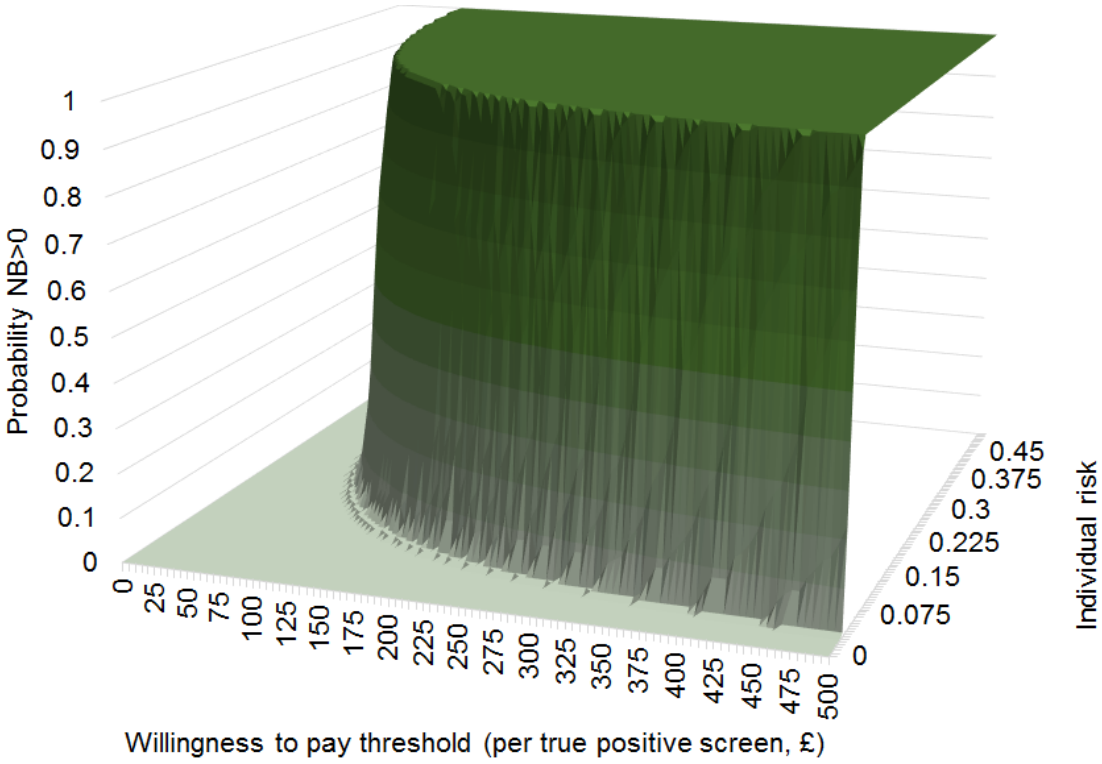

Figure 8.6: Cost-effectiveness-risk-acceptability surface

Figure 8.6.

8.4 Discussion

The relationships presented in this chapter are intuitive. The higher an individual's level of risk, the more likely it is that they will screen positive and therefore benefit as intended from the screening programme. This means that the expected cost per outcome for an individual with a high risk is low, while for those at a low risk it may be very high.

Decision modelling techniques are routinely used in the evaluation of screening programmes, as demonstrated in Chapter 2. We believe that it is these methods that offer the best opportunities for the incorporation of individual risk data into iCEA. However the acceptability of such an approach, to patients and clinicians, is still to be demonstrated.

The framework that we have set out is by no means revolutionary and is, implicitly, how many analysts think about stratified cost-effectiveness analyses. However, the application to risk-based screening is timely and novel. We have demonstrated a simple basis on which to design a risk-based screening programme that seeks to maximise efficiency.

The evaluation of precision medicine and the application of iCEA may be perceived to be highly demanding of data. Yet we have demonstrated that,

with a few defensible assumptions, an optimal risk-based screening regime can be identified. Decision-makers need only know the average expected costs and benefits of a true positive screening outcome.

Nevertheless, our simplifying assumptions do introduce limitations. Screening programmes may involve more complex pathways than those described in this chapter. Decision-makers might also value outcomes other than a true positive screening result. For instance, most screening programmes seek to minimise false positives. More fundamentally, one might expect systematically different costs and outcomes for people at different levels of risk. This would mean that the expressions provided in this chapter would not identify optimal thresholds or recall periods. Furthermore, the process of optimally setting recall periods does not account for expected attendance. If an individual is expected not to attend, it may be optimal to invite them for recall at an earlier date based on the expectation that they will not attend.

In this chapter, we have presented a simple framework for individualised cost-effectiveness analysis and demonstrated how this can form the basis for setting optimised risk-based screening programmes. This would involve inviting people at high risk to attend screening more regularly than people with low risk, on the basis that this is cost-effective. Policy-makers should consider the appropriateness of such an approach and future research should explore its feasibility and acceptability. In Chapter 9, we consider whether such an approach could be justified on ethical grounds.

Chapter 9

The ethics of risk-based screening

Summary

The potential to differentiate the provision of screening based on individuals' risks of disease onset raises methodological challenges, as described in earlier chapters. In this chapter, we identify the ethical issues that it might also raise. In particular, we consider fairness criteria generally accepted for the allocation of health care resources in the UK. We argue that the allocation of screening based on individual risk is justifiable within standard ethical frameworks, focussing on the role of need as the basis for resource allocation. This requires some important redefinition of need as it relates to screening (rather than treatment). We explore the implications of allocating screening based on what we term 'screening need' and consider the role of individual risk in this process.

9.1 Introduction

In the preceding chapters we have demonstrated the potential to differentiate screening according to individual risk, on the basis of expected cost-effectiveness at the individual level. However, as outlined in Chapter 8, cost-effectiveness analysis to date has been conducted on the basis of aggregate incremental costs and effects at the population or sample level. An ethical basis for individualised cost-effectiveness analysis (iCEA) in the context of risk-based screening has not previously been outlined.

The need for such ethical inquiry is made clear by the simple assertion that in *some* situations iCEA *might* not be expected to be ethical. For example, iCEA could be used to indicate that a person should not receive treatment because they are of an age at which their life expectancy is not sufficient to justify a particular life-saving treatment, despite being part of a patient group for which treatment is cost-effective in the aggregate and despite the possibility that the individual's realised outcome may be cost-effective. Such a scenario may reasonably be judged to be ethically problematic, even if not self-evidently unjust. Indeed, many ethical arguments against the QALY-maximisation approach to resource allocation fail in reference to real-world scenarios precisely because cost-effectiveness analyses are conducted at the aggregate [305, 306]. Government guidance on evaluation in the UK states that:

> "On grounds of equity in appraisal QALY values... are based on average values from representative samples of the population (who differ in their incomes, preferences, age, states of health and other circumstances). These values are used when analysing and planning the provision of assets, goods and services at a population or sub-population level." [307]

By abandoning the safety of evaluation in the aggregate, iCEA becomes vulnerable to criticism on ethical grounds.

We will consider the specific case of iCEA with respect to individual risk in the context of risk-based screening. At present, this is one of very few feasible applications of an iCEA approach to economic evaluation and thus worthy of special consideration.

Screening programmes are increasingly moving from a 'one-size-fits-all' design towards stratification and individualisation. Such programmes may involve the use of a risk calculation engine (RCE) to estimate individual risk of disease onset and to determine screening eligibility or allocation. There are a number of clinical areas in which such developments are taking place. Recently, in the UK, the National Screening Committee (NSC) has recommended that people with diabetes identified as being at low risk of sight loss be invited to screening after 2 years

rather than annually, while those not identified as such should continue to be screened every year. This represents a form of risk stratification. Similarly, The American Cancer Society recommends annual breast screening for women aged 45-54, but biennial screening for women aged 55 and older. It has been suggested that, in the future, stratification will be further tailored to the individual [43].

Some commentators have proposed that risk-based screening might be ethically preferable. For example, in the case of prostate cancer screening, risk stratification has been described as "perhaps the most ethical approach" [308]. However, most of the arguments in favour of risk stratified screening have — to date — focussed on the maximisation of benefit and minimisation of harm, as have debates about the ethics of screening more generally [309, 310]. This leaves other ethical challenges unconfronted and does not outline a clear ethical basis on which to determine screening eligibility within a risk-based programme. It can be taken as given that risk stratified screening *could* result in greater benefit and/or reduced harm, as well as reduced expenditure, by targeting screening at those more likely to benefit. Indeed, this is why such programmes are being proposed. However, we do not accept that this potential for health benefit is a sufficient justification of risk stratified screening on ethical grounds.

In this chapter, we argue that risk stratification can be used to allocate screening according to need, allowing for egalitarian, prioritarian, or utilitarian allocations as preferred by decision-makers. We discuss the notion of need in this context, presenting the case for a reconceptualisation of individual need for screening as likelihood of benefit. Screening need — as described below — represents a practical basis for resource allocation that can result in a fairer and non-discriminatory allocation of resources. The discussion does not apply to the specific ethical challenges associated with particular screening possibilities — especially prenatal screening for genetic disorders — and relates primarily to established screening programmes provided by national health services. In such cases, an affordable stratification process could improve technical efficiency and allocate resources more equitably.

9.2 Screening need (as distinct from treatment need)

There is no consensus on the most appropriate definition of need in the context of health care. In this chapter we focus on need as capacity to benefit, which is a widely accepted interpretation in health care. In relation to Bradshaw's taxonomy, we are principally interested in normative need and to a lesser extent comparative need; felt need and expressed need are not the subject of this discussion. A popular interpretation of health care need in resource allocation is as

'capacity to benefit from cost-effective care' [311]. In England, the NHS seeks to make health care available according to need, and makes decisions according to guidance from the National Institute for Health and Care Excellence (NICE). As this guidance is based in large part on cost-effectiveness, we may understand capacity to benefit from cost-effective care to be the principal notion of need operating in UK policy-making.

People are only determined to 'need' health care if it might be expected to improve their health, and therefore have a capacity to benefit in this sense. This is because health is the source of moral weight in this discussion; nobody 'needs' ineffective health care, even if they demand it. Likewise, people are not generally said to need health care for purposes other than health gain. As such, the concept of need in health care is inextricably linked to the potential for health improvement (or prevention of ill health). Such improvements are the result of treatment, which is why we discuss people needing particular treatment for particular diseases. Here we provide basis for understanding need for screening in a different way.

9.2.1 The purpose of screening

Screening differs from treatment and other preventive care in a number of ways, and it is the combination of these factors that calls for a reinterpretation of need in this context.

Screening is of no direct health benefit. This makes it difficult to define screening need in terms of necessity [312]. For a person with a life threatening disease for which curative treatment is available, that treatment might reasonably be said to be necessary. In this sense, treatment is needed. However, if in the current health care context a person is only likely to be offered the treatment following a positive screening outcome, it does not follow that screening is necessary. This is because screening could — in principle — be bypassed with no loss of treatment benefit.

There is no capacity for additional benefit from screening beyond that which results from (early) treatment. Only that which directly addresses the individual's (potential) health shortfall should be considered to be necessary. Accounting for all context-specific indirect determinants of the receipt of care plainly renders 'necessity' as meaningless. The individual's ability to get dressed, their willingness to leave the house, and the availability of hospital-bound public transport cannot be said to be needed in terms of 'necessity' – there are substitutes. The necessity of screening can only be derived through the insistence of screening as part of an effective treatment pathway.

At the individual level, whether or not screening *could* lead to health benefits is unknown before screening takes place. That is, if a person has a screen-negative

disease state it is not possible for them to experience health benefits. Rather, screening serves to identify people who could and who could not benefit from treatment. Therefore, individuals will necessarily be eligible for screening despite having zero capacity for health benefit. This limits the usefulness of 'capacity to benefit' as an interpretation of screening need when screening eligibility is determined at the individual level, as may be the case in risk-based screening. At the individual level, capacity for health benefit is unknown until after screening has taken place. Then, capacity for health benefit from screening is dependent on a true positive screening outcome, generally characterised in a binary way. Thus, there will be people who attend screening who do not have a need for treatment and yet might still have been considered in need of screening. Need for screening should therefore be understood as an instrumental capacity to benefit that is intermediate and only partially related to the notion of treatment need. An individual with no capacity to benefit from treatment might still reasonably be judged to need screening while that capacity to benefit remains unknown.

Though subject to limitations, the interpretation of need as capacity to benefit from cost-effective care is pertinent to screening. Need cannot simply be understood as an absence of something. Current ill health cannot be used to define need, either at the level of the individual or the population. Furthermore, an interpretation that allows for some potential health benefit to be foregone (while maintaining equitability) is necessary because screening is delivered to asymptomatic individuals. It would not be efficient to allocate screening to the majority who stand to gain extremely small potential benefits from screening due to a very low probability of disease onset. Such an approach would be unethical if this — as it would under a fixed health budget — deprived others of cost-effective care.

Need for screening involves a comparison of what might occur in the future (with and without health care) rather than a comparison of current states. This is in the context of an unknown capacity to benefit at the individual level. To have a capacity to benefit from screening, an individual needs to have the characteristics necessary to achieve a true positive screening outcome. Screening need should therefore be understood as the likelihood of this being the case – as likelihood of benefit.

9.2.2 Disaggregating need

By allocating screening according to incidence (or prevalence) of disease in a given population, need is only estimated in the aggregate. Screening programmes are (and will continue to be) increasingly targeted at individuals rather than populations. To use Geoffrey Rose's aetiological distinction [313], we are increasingly able to allocate screening according to determinants of individual cases. To under-

stand need at the individual level and to go 'beyond the mean' is preferable both in terms of efficiency and equity, as discussed in the remainder of this chapter.

In their classic principles of screening, Wilson and Jungner incorporated the requirement that effective treatment be available for those screening positive [19]. We interpret this to mean effective *in the aggregate*. The current discussion relates to existing screening programmes and disregards the possibility that a screening programme might exist without effective follow-up. One implication of this is that — where we do not have information about individual risk of disease onset — the existence of a screening programme creates an apparent unmet need in the entire asymptomatic population. This is because although the population average risk level (incidence) may be sufficient to allow for cost-effective screening, many people may have a very small probability of screening positive. The majority of the population are likely to have no capacity to benefit from screening. This is not a useful understanding of individual need and cannot inform a fair or efficient allocation of resources. By disaggregating need it is possible to allocate screening only to those deemed to have a (sufficient) need for it.

9.2.3 Quantifying individual need

In order to inform resource allocation decisions, both treatment need and screening need require quantification. As discussed in more detail below, we focus here on the achievement of technical efficiency within a screening programme. We are not concerned with allocation to different populations or with treatments beyond the screening process.

Treatment need should be quantified in terms of a measure of health benefit, such as quality-adjusted life years (QALYs). However, as outlined above, capacity to benefit in these terms remains unknown at the individual level for a population eligible for screening. The purpose of screening is not to satisfy this need but to assess it; screening is a means of obtaining more information about a person's health state. An individual's intermediate capacity to benefit from screening should therefore be understood as the likelihood that they will have a true positive screening outcome following a screening procedure.

Risk calculation engines can be designed to estimate individual risk in a number of ways; for example, lifetime risk of disease onset or the probability of a positive screen event. Such estimates should be understood as approximations of an individual's screening need.

Where a screening programme exists (and effective follow-up is available), individuals might have a low or a high need for screening. This can be determined independently of the magnitude of health benefits associated with treatment following a positive screen. For example, an individual estimated to have a 10% risk of disease onset (within a given time frame) has a greater screening need than an

individual with a 2% risk of disease onset. This is independent of the magnitude of health benefit associated with treatment, and only requires that some health benefit is possible. Following screening, it could be that the individual identified as having a 2% risk needs treatment and the individual at 10% risk does not. This does not undermine the fact that — at the point at which screening eligibility was determined — the higher risk individual had a greater need for screening.

This example highlights the importance of information in the determination of need and that the available information not only determines the magnitude of need (e.g. in relation to the level of risk) but also the provision to which that need relates (i.e. screening or treatment). The example demonstrates that screening need can be estimated in isolation from capacity for health benefit from treatment, and that the two may not align in interpersonal comparisons. This characteristic of screening need has a number of advantages, as we will now discuss.

9.3 Screening need and technical efficiency

The consideration of technical efficiency invokes a number of specific ethical viewpoints, particularly utilitarian and welfarist approaches. However, this discussion does not presuppose the appropriate distribution of resources for screening and seeks only to provide a basis on which to define the appropriate distribution: in terms of screening need. For example, a prioritarian approach may be adopted by allocating disproportionately greater resources to those with the greatest screening need (i.e. at highest risk).

We assume the context of determining need to be that of high-level resource allocation decisions, rather than determinations at the physician or patient level. Furthermore, our discussion relates to questions of efficiency within a risk-based screening programme and not the evaluation of alternative screening procedures. The ideas expressed may only translate partially to other contexts. Within a risk-based screening programme, a threshold level of risk could be defined based on the minimum level of screening need necessary to justify screening (in terms of cost-effectiveness). This kind of risk-based screening programme would allocate screening to all people with capacity to benefit from cost-effective care. That is, those with a risk level below the threshold are not sufficiently likely to screen positive to justify screening.

9.3.1 More with less

Individualised risk-based screening could be more efficient and more equitable than either a standardised programme or a stratified programme.

Discrimination in health care lacks a clear definition. In the context of priority setting, allocating care on the basis of capacity to benefit has been characterised as discriminatory [314]. We reject this interpretation and rather define discrimination as the allocation of health care according to personal characteristics that are unrelated to an individual's level of need.

As previously explained, the determination of an individual's level of need is dependent on the information available. Risk-based screening provides an opportunity to make decisions on the basis of a more accurate estimation of need. In this context, using specific personal characteristics as the basis for allocating care may be judged to be discriminatory where they would not previously. This can be illustrated by an example.

Consider a screening programme with a minimum age threshold of 65 years. This threshold is determined by studies showing that screening and treatment is — on average — only cost-effective for over-65s. Whatever the disease, it is unlikely that age is the only risk factor. If RCEs are available, it is probable that there are many identifiable pairs of 60 year olds and 70 year olds who have equivalent levels of risk. The 60 year olds likely have a greater life expectancy and therefore a greater capacity to benefit from treatment. There is no apparent basis — either ethical or economic — for determining that the 70 years olds should have a greater right to screening than the 60 year olds (as they would in this example) if their risk is demonstrably equivalent. Each has the same screening need, while the ineligible party may have a greater capacity to benefit from treatment. This example reflects the case of the NHS abdominal aortic aneurysm screening programme. If individuals' risk of abdominal aortic aneurysm could be estimated, an age-based threshold for eligibility might reasonably be deemed discriminatory.

As well as having the potential to be less discriminatory, risk-based screening also has the potential to make screening programmes more efficient. Individuals with a very low probability of screening positive have a very low need for screening, while those with a high probability have a high need. Those with no probability of screening positive, even if there may be other tangential unintended benefit (for example screening people without diabetes for diabetic eye disease), have no need for the screening in question.

In a risk-based screening programme, it would be possible to allocate screening to all people who have a sufficiently high level of risk to justify screening. If the current budget is maintained, this would lead to a greater number of true positive screening outcomes and therefore the achievement of improved health outcomes. It may also be possible to reduce the budget by carrying out fewer unnecessary screening tests.

9.3.2 Limitations as a basis for efficient allocation

There are a number of conditions that must hold in order for risk-based screening to achieve improvements in technical efficiency. First, there must be sufficient heterogeneity in the risk of disease onset. Second, the risk calculation engine must be sufficiently accurate. Third, the individualisation process must be affordable. Finally, there must not be unintended consequences of individualisation (for example, impact on attendance rates) that could undermine technical efficiency. If these conditions do not hold, the usefulness of this approach is limited. Context-specific research will be necessary to determine whether screening need can be used as a basis for efficient allocation of resources.

The simple identification of screening need as individual risk does not address the question of how much need ought to be met. In practice this question will be *what ought the threshold level of risk be?* This discussion relates to decisions made within a screening programme, rather than decisions between screening programmes. Because the health impact of screening is entirely dependent on the effectiveness of treatment — and the individual that could benefit from that treatment — the distinction between screening need and treatment need cannot be applied so simply to decisions about to which screening programme resources should be allocated. In terms of achieving efficiency, the distinction can be used to achieve technical efficiency; that is, achieving the same outcome (a given number of true positive screen events) with fewer resources (fewer screening episodes). It is less informative for questions of productive efficiency — for example, whether to use a different type of screening test or add a reminder service — or allocative efficiency – for example, whether to screen for diabetic eye disease or for bowel cancer.

9.4 Fairness of allocation based on screening need

Justice and fairness are paramount concerns, but the way in which relevant criteria ought to apply in screening differs to the ways they ought to apply to treatment. High-level resource allocation decisions do not routinely consider individuals' characteristics, and decisions are made based on aggregate estimates of cost-effectiveness. This approach avoids some forms of discrimination (except, most notably, in allocations between interventions for specific socioeconomic groups) and is adopted by NICE in England [315]. Screening programmes do not adhere to this condition and routinely — and explicitly — discriminate based on personal characteristics. Herein lies an implicit acknowledgement of a moral distinction between screening need and treatment need, though there has to date remained a lack of coherent criteria for rationing.

'Need' invokes distributional considerations that limit the use of a 'capacity to

benefit' interpretation of need. If the allocation of resources was based on capacity to benefit from them then it is possible (perhaps probable) that resources would be allocated in a way that was not deemed fair. A crude example is that the rich may have a greater capacity to benefit from treatment due to being better able to convert health care into health and thus greater resources might be allocated to the diseases of the rich in a way deemed to be unfair. However, as implicitly acknowledged by NSC screening programmes in England, these distributional concerns should not apply to screening as they should to treatment.

9.4.1 'NICE' discrimination

One of the principal concerns when designing risk-based screening programmes should be to avoid discrimination, despite the apparent widespread use of discriminatory criteria in existing screening programmes. By discrimination we refer to the practice of reducing or otherwise altering individuals' access to the benefits of health care according to their personal characteristics. Possible bases for discrimination include race, gender, age, and lifestyle choices. These personal characteristics may or may not bear relation to the effectiveness or cost-effectiveness of treatment. The important point is that discrimination occurs when access to health care differs *because* of these factors, despite the potential for these groups to gain the intended benefits. For example, it may be considered discriminatory to refuse breast screening to women over 60 simply because they have a reduced capacity to benefit due to a shorter life expectancy than younger women.

It would not be considered discriminatory to refuse to offer diabetic eye screening to people without diabetes, even if unintended benefits (for example identification of other non-diabetic eye diseases) are possible. This is because people without diabetes have zero likelihood of screening positive and therefore zero screening need.

Allocation of screening according to individual risk is non-discriminatory. The probability that an individual will screen positive is a supra-personal characteristic that should not hold moral weight with regard to the allocation of health care resources. Individual characteristics or behaviours might influence their level of risk, but the characteristics themselves are not directly being used to determine eligibility. It has also been suggested that risk-based allocations of resources may be just in terms of Norman Daniels's prudential lifespan account [316, 317]. That is, people may be more likely to choose to have preventive measures available to them — for example, by buying insurance — if their risk is high [318].

Screening need can be determined independently of health effect, which is the source of moral weight [319]. In order to maintain non-discriminatory allocations of resources, eligibility for risk-based screening should only be on the basis of screening need. Decision-makers should not bear the same moral responsibility

to provide screening to all those with unmet treatment need, because the role of screening is to provide information rather than direct health improvement. At the individual level, this information only holds value if a person screens positive.

Risk stratification has been characterised as profiling when it has the potential to result in less equitable outcomes [320]. So long as the conditions outlined above hold true, stratification of screening based on individual risk can only result in fairer allocations of resources. As such, risk stratification in screening should not be characterised as profiling. Horizontal equity is improved as individuals with equivalent screening need receive equivalent care, while vertical equity is improved by the proper reallocation of resources according to screening need.

9.4.2 Non-Pareto improvements

If screening allocation is determined exclusively on the non-discriminatory basis of individual risk, proportionality can be maintained in the distribution of expected health care. That is, the expected health benefit of screening can be equal across all those who are offered screening. In this sense, both equality of outcome and equality of opportunity are achieved.

However, to move to a screening programme of this nature from current screening programmes will necessarily result in non-Pareto improvements. That is, some people who are currently eligible for screening will no longer be eligible. Some of these people will forego health benefits as a result. While Pareto-optimality has been identified as a very limited ethical basis for the allocation of resources, it remains a practical and ethical challenge to deprive people of health care for which they might previously have been eligible.

9.5 Discussion

Screening programmes are beginning to acknowledge the potential value of stratification using risk calculation engines. There are important ethical implications of this and further investigation is required. Individual risk should be understood as an approximation of an individual's need for screening, and as such screening should be allocated on this basis. To allocate resources in this way is entirely within the currently accepted ethical rules governing such decisions. These depend on the specialness of health care, explanations of which, it is worth noting, have been questioned [321].

The purpose of screening, whether risk-based or otherwise, is to provide information in order to make better treatment allocation decisions. It may be possible to go so far as to identify a moral difference between treating the ill as opposed to treating the potentially ill, and characterising screening need as an example of the latter. The generalisability of our approach to other preventive care settings

may rely on such a distinction, though we consider this to be beyond the scope of this thesis. Nevertheless, such a distinction is not necessary for the purposes of identifying who should and who should not be screened for a given disease in order to achieve technical efficiency, which is the current policy challenge.

Risk-stratified screening programmes raise additional ethical issues, including data privacy concerns and the need to obtain adequate consent [322]. These will will require consideration alongside screening need. It is also important to consider that risk assessment — even if performed automatically without additional human interaction — might loosely be considered to be some form of intervention, and perhaps a form of screening in itself. Therefore, there may be further ethical challenges to consider in this regard.

Policy-makers should give greater consideration to the current ethical basis for screening programmes. It is notable that the ethical basis for the allocation of resources in screening programmes in the UK is not clearly specified. Recognition of the underlying principles currently adopted will inform the appropriateness of allocating resources on the basis of screening need in the way that we have proposed.

Future research should evaluate the extent to which individual risk of disease onset can be accurately and cost-effectively estimated in the context of specific screening programmes, and consideration should be given to other ethical challenges associated with risk-based screening.

Chapter 10

Discussion and conclusions

Summary

The way in which health screening programmes are organised is changing. In the near future, screening will cease to be a one-size-fits-all service, with both eligibility for screening and the nature of screening being altered according to individuals' characteristics. This change is being driven by both demand-side and supply-side pressures. On the demand side, the incidence of chronic and long-term illnesses such as diabetes, cancer, heart disease, and dementia, is growing. On the supply side, more sophisticated analytical tools, greater computing power, and more extensive data collection create the potential for precision medicine. Together, these pressures make individualised screening programmes both more attractive and more feasible.

These new possibilities present new challenges, some of which we sought to address in the research reported in this thesis. In particular, the basis for setting risk thresholds — above which an individual should be invited to screening and below which they should not — needs to be defined, creating a plethora of practical, theoretical, methodological, and ethical challenges.

The NHS Diabetic Eye Screening Programme has become a test bed for more sophisticated screening allocation, with the conduct of the ISDR study. It is possible that the NDESP could become the world's first national individualised risk-based universal screening programme. Thus, it is an ideal context in which to conduct our research and consider the broader decision problem of risk-based screening.

In this final chapter, we briefly summarise the key findings from our research, what the study has contributed, and what this means for research, policy, and future work in this area.

10.1 Summary of findings

The first study reported in this thesis was a review of model-based economic evaluations in diabetic retinopathy. We found that cohort state transition modelling was the most popular approach. We also saw that models tended to be built using TreeAge software. The structures of models used to evaluate interventions for diabetic retinopathy are inconsistent. Two key groups were observed – models based on visual acuity and models based on the level of retinopathy. Furthermore, within each of these groups there is a wide range of states that form the foundation of model structures.

We found that some models implemented unrealistic assumptions about the nature of disease progression or treatment pathways. These limitations are probably borne out of the data that are available to analysts. The selection of parameters, especially health state utility values, is rarely systematic. There is also a reliance on baseline disease progression rates from landmark studies conducted decades ago. This is not easy to overcome because it is no longer possible (in the UK, for instance) to observe disease progression in a population that does not receive screening.

In light of some of the limitations identified in the evidence base, we designed several new studies to collect cost and outcomes data that can be used in model-based economic evaluations. We collected new data on health-related quality of life from people attending screening for diabetic retinopathy. In this study, we found evidence that individuals with background retinopathy tend to report poorer health-related quality of life than people with no retinopathy. This is likely to relate to factors other than retinopathy, though some of the difference could be explained by responses to vision-related questions on the HUI3.

We found that the HUI3 was more sensitive than the EQ-5D-5L to the differences between people with and without background retinopathy. In general, the HUI3 constituted a more sensitive scale in this cohort. However, the HUI3 resulted in poorer quality data, with a relatively high level of missingness. This is presumably because the HUI3 questionnaire is longer than the EQ-5D-5L, with more complicated descriptors.

Model-based economic evaluations in diabetic retinopathy should identify health state utility values in a systematic way, informed by all available evidence. To support this ambition, we conducted a meta-analysis of published values. We found a great deal of inconsistency in HSUV studies with respect to the groups being assessed. As with model-based economic evaluations, there was a conflict between HSUVs derived from groups defined in terms of their level of disease and groups defined in terms of their level of visual acuity. And, again, there were many alternative groupings within these categories. A further challenge was that some studies reported values related to the better-seeing eye while

others reported values related to the worse-seeing eye. This inconsistency made it difficult to pool values. We found a small and statistically insignificant impact associated with background and pre-proliferative DR. The impact of PDR and of being blind was larger and statistically significant. Our model also demonstrated the importance of methodological choices in determining HSUVs, and that these choices can provide more explanatory power than disease level.

Our research also provided new estimates for resource use and costs associated with the screening and treatment of diabetic retinopathy. Prior to this study, up-to-date estimates of the cost of screening were not available. Our costing study found that the average cost of an appointment in Liverpool is £32.03. We also estimated that the cost of an attended appointment is £26.14, while the cost of a non-attendance is £15.97. If a societal perspective is adopted, there are further costs incurred by the patient and through productivity losses, which together sum to £8.62 per appointment attended.

Treatment for DR can involve a variety of pathways, with most including laser procedures. We found that many people do not receive treatment within the first four years after screening positive. Because of this, routine attendance at the hospital eye service is likely to be a key driver of costs rather than the provision of treatments. Our analyses show that, while laser is still the dominant treatment strategy, the use of intravitreal injections is growing. We also show that people who receive injections receive more procedures than people who receive laser, with injections being associated with a higher cost per person.

We developed a new decision analytic model to evaluate the cost-effectiveness of alternative screening programmes for diabetic retinopathy. This model incorporated the ISDR risk calculation engine and maintained flexibility while representing the complexity of disease and treatment pathways within DR. The model was too computationally expensive within the capabilities of Microsoft Excel and, as a result, the model could not be run to its desired specification. A reduced simulation showed that individualised risk-based screening may be cost-saving and outcome-improving compared with current practice.

Within the context of the ISDR study, and the decision model designed to evaluate it, a threshold level of risk was set at 2.5%. This was determined on the basis of acceptability by the patient group involved in the study. We conceived that this threshold could instead be set on the basis of cost-effectiveness. To support this process, we outlined a simple framework for individualised cost-effectiveness analysis. We demonstrated that it is possible to estimate an optimal threshold under some simplifying assumptions. This framework can be used to estimate a programme-level threshold or to estimate a recall period for an individual based on their risk of disease onset.

If we *can* set a threshold for screening eligibility on the basis of individual risk

then the question is raised of whether we *should*. To this end, we outlined the concept of 'screening need' and argued that the output of risk calculation engines constitutes an approximation of this. Risk-based screening is justifiable on the grounds that it allocates screening according to need in a way that is in line with current practice in the allocation of treatment. Further, it is more easily justified than current screening programmes, which discriminate on the basis of personal characteristics.

10.2 Contributions of this thesis

In addition to the findings of this thesis contributing to the evidence base, we have made numerous novel contributions to the literature, which will be valuable to the research community. These can be broadly categorised as relating to the curation of existing evidence, the generation of new data, the development of theory, and the development of methodology.

10.2.1 Curation

The review and synthesis of published literature and data is an important part of evidence generation. As part of our research, we conducted two literature reviews, one of decision models and one of health state utility values. We extracted and organised a large amount of information from these studies, which will be of value to other researchers in future. For example, researchers can browse our findings from Chapter 4 to identify the most suitable HSUVs for their context. Without the information provided by our study, researchers would need to conduct their own review. This curation thus constitutes a valuable contribution to the research community.

10.2.2 New data

Our research involved the creation of new data, relating to the costs and outcomes of screening and treatment for DR. A large sample of people attending screening for diabetic retinopathy completed questionnaires about their health-related quality of life and costs associated with their visit. In both cases, these new data serve to fill a gap in the evidence base that could prove informative to future evaluative research.

10.2.3 Theory

Risk-based screening is a novel area of research and, as such, has nascent theoretical foundations. Our work has strengthened the theory underlying risk-based

screening in practice and provided grounds for its support. Specifically, we justified risk-based screening in terms of its capacity to improve the cost-effectiveness of existing screening programmes. This assertion requires some adjustment to our understanding of cost-effectiveness analysis, as outlined in this thesis. We also opened a conversation about the ethics of risk-based screening and developed the notion of screening need as an ethical basis for its implementation in the UK.

10.2.4 Methodology

Our work has furthered several distinct areas of methodology. By comparing EQ-5D-5L and HUI3 data in this population, we identified several methodological issues and demonstrated the importance of the choice of HRQoL instrument. Our systematic review of HSUVs was at the forefront of methodological development with respect to literature searching, data extraction and synthesis, and modelling methods. This work has already informed recommendations in this space [188, 323, 324]. At the time of the research, we were not aware of any decision analytic models that incorporate a (separately developed) risk calculation engine into their structure and operation. Our work provides numerous lessons for analysts working in this area and seeking to develop new methods in this context. We have also provided a practical method for the identification of thresholds in risk-based screening. This methodology requires further development and testing but constitutes a foundation for the definition of risk-based screening programmes in the future.

10.3 Implications

10.3.1 For research

A theme running through our research is the importance of methodological choices. With respect to the elicitation of health state utility values, our findings suggest that researchers should be cautious of using the HUI3 in this context due to poor completion rates. Our systematic review of HSUVs highlighted the importance of research design. It is unlikely that consensus will ever be reached on the most appropriate methodology for eliciting HSUVs in this context. Therefore, it is very important that researchers give careful consideration to the methods adopted. This applies both to the creation of new HSUV estimates and the selection of estimates to be used as parameters in a decision model.

Our research showed that identifying costs can be particularly challenging in a climate of tendering within a publicly-funded health care system, where information is considered commercially sensitive. Nevertheless, we demonstrated the potential value of disaggregating costs. In particular, a bottom-up costing

methodology facilitated the estimation of a cost of non-attendance, which has not previously been available. Such estimates can be informative in the design of services and this highlights the value in this research beyond simply identifying a more precise or accurate estimate of the headline figure. Our research showed that both screening and treatment pathways have changed over time in a way that impacts on cost estimates. It is therefore important that researchers use up-to-date information in evaluative studies.

Cohort simulations are the most popular approach to modelling in the evaluation of interventions for diabetic retinopathy. However, cohort simulations should not be used when evaluating risk-based screening programmes. This is because the value of risk-based screening lies in the distinction between individuals. Where models rely on the simulation of a cohort, individuals within the cohort cannot be adequately differentiated by different states or pathways in a way that would be feasible. Furthermore, models that seek to incorporate a complex risk calculation engine should not be built in Microsoft Excel due to the software's mathematical limitations. We highlighted a great deal of inconsistency in the modelling methods used in the context of DR. Diabetic eye disease would be a good candidate for the development of a reference model (see, for example, [325]), which incorporates both disease progression and visual acuity and which is built in a flexible programming language that allows for the incorporation of risk calculation engines.

Individualised cost-effectiveness analysis is theoretically feasible in the context of risk-based screening. The framework should be tested in a variety of settings to better understand the assumptions necessary to design optimised risk-based screening programmes. Both economists and ethicists should consider the implications of risk-based screening and the application of individualised cost-effectiveness analysis in this context.

10.3.2 For policy

The successful conduct of the ISDR trial demonstrates that risk-based screening can be feasible. However, its practicality in other settings, particularly beyond screening for diabetic eye disease, needs to be explored. For such a programme to be practical, key infrastructure including data systems and research capacity will need to be in place. We have framed risk-based screening as part of a process of the development of screening from a one-size-fits-all service to one that is optimised on the basis of individual risk. Policy-makers should consider this supposed policy trajectory and its implications in different contexts.

We have demonstrated that risk-based screening is likely to be cost-effective in the context of diabetic retinopathy. We have also argued that risk-based screening could improve the fairness of screening programmes in the UK. Policy-makers

should consider the extent to which our reasoning aligns with current policy objectives.

The ISDR study, and hence most of the data used to generate the findings reported in this thesis, are specific to Liverpool, England. The population studied may not be representative of the wider UK or international population. According to the Public Health Outcomes Framework, life expectancy and healthy life expectancy — for both men and women in Liverpool — are both in the lowest 25th percentile for England. There is a high level of unemployment and a large gap in the employment rate between those with a long-term health condition and the overall employment rate. Furthermore, across all areas in England, Liverpool exhibits the lowest proportion of adults meeting the recommended '5-a-day' consumption of fruit and vegetables. Rates of physical activity are low with only 13.2% of people utilising outdoor space for exercise or health reasons, compared with an English average of 17.9%. However, the rate of recorded diabetes (6%) is similar to the average for England (6.4%). The rate of sight loss due to diabetic retinopathy in Liverpool is around 2.2% according to the Public Health Outcomes Framework for England. Such factors need to be considered when generalising our findings to other settings.

We were unable to identify certain local arrangements as part of our research. These may prove important in determining whether risk-based screening is cost-effective. Decision-makers at the national and local level need to work together to understand regional heterogeneities and how these could affect the efficiency and fairness of screening services.

10.4 Conclusion

This thesis has explored the potential for risk-based screening in the context of diabetic retinopathy. People with diabetic retinopathy report lower levels of health-related quality of life, with a major impact associated with severe sight loss. Screening can prevent or postpone sight loss and thus benefit patients. By estimating individuals' risks of disease onset, it is possible to identify people who are most likely to benefit from intervention. Using this information, resources can be reallocated in a risk-based screening programme such that people with higher risk are prioritised over people with lower risk. This reallocation of resources can improve outcomes by identifying disease earlier and providing effective early intervention. Risk-based screening can also be cost-saving if a high proportion of people are found to be at low risk and have their screening eligibility limited accordingly. An optimised risk-based screening programme can be designed on the basis of cost-effectiveness. We have demonstrated that risk-based screening could improve outcomes and save costs in the context of diabetic retinopathy and

that this represents a fair and equitable policy.

Bibliography

1. Sampson C, James M, Broadbent D, and Harding S. Stratifying the NHS Diabetic Eye Screening Programme: Into the Unknown? Diabetic Medicine: A Journal of the British Diabetic Association 2016;33:1612–4. DOI: 10.1111/dme.13192.

2. Sampson CJ, James M, and Whynes D. The Relationship between Individual Risk and Cost-Effectiveness in Screening Interventions. 2013. DOI: 10.2139/ssrn.2362371.

3. Sampson CJ, Tosh JC, Cheyne CP, Broadbent D, and James M. Health State Utility Values for Diabetic Retinopathy: Protocol for a Systematic Review and Meta-Analysis. Systematic Reviews 2015;4:15. DOI: 10/f7gbfx.

4. Sampson C, James M, Van Eker D, Szmyt D, and Harding S. Health-Related Quality of Life of People Attending Screening for Diabetic Retinopathy within a Trial Setting. European Journal of Ophthalmology 2016;26:E100–101.

5. James M, Sampson C, Van Eker D, and Harding S. Micro-Costing Diabetic Eye Screening: Estimation of Personal Expense, Attendance and Health Care Resource Use. European Journal of Ophthalmology 2016;26:E100–101.

6. Harding SP, Sampson CJ, Criddle TM, Broadbent DM, and James M. The Costs of Screening for Sight-Threatening Diabetic Retinopathy. Investigative Ophthalmology & Visual Science 2019;60:4802–2. URL: https://iovs.arvojournals.org/article.aspx?articleid=2744254 (visited on 09/02/2019).

7. Harding SP, Sampson C, Cheyne C, et al. 10 Year Observational Study of Patient Related Outcomes during Expansion to Full Population Coverage of a Systematic Diabetes Eye Care Programme – the Liverpool Diabetic Eye Study. Investigative Ophthalmology & Visual Science 2016;57. URL: https://iovs.arvojournals.org/article.aspx?articleid=2559561 (visited on 09/02/2019).

8. Sampson C, James M, Fisher A, and Harding S. Cost-Effectiveness of a Risk-Based Screening Programme for Diabetic Retinopathy: A Modelling Approach. European Journal of Ophthalmology May-June 2015;25. URL: https://insights.ovid.com/european-ophthalmology/ejoop/2015/05/000/cost-effectiveness-risk-based-screening-programme/44/00008369 (visited on 09/02/2019).

9. Sampson C, James M, Whynes D, Eleuteri A, and Harding S. OP60 Optimising Risk-Based Screening: The Case Of Diabetic Eye Disease. International Journal of Technology Assessment in Health Care 2018/ed;34:21–1. DOI: 10.1017/S0266462318001046.

10. Diabetes UK. Facts And Stats. 2016. URL: https://www.diabetes.org.uk/Documents/Position%20Statements/DiabetesUK_Facts_Stats_Oct16.pdf (visited on 08/28/2019).

11. Klein R, Klein BE, Moss SE, Davis MD, and DeMets DL. The Wisconsin Epidemiologic Study of Diabetic Retinopathy: II. Prevalence and Risk of Diabetic Retinopathy When Age at Diagnosis Is Less than 30 Years. Archives of Ophthalmology 1984;102:527–32.

12. Klein R, Klein BE, Moss SE, Davis MD, and DeMets DL. The Wisconsin Epidemiologic Study of Diabetic Retinopathy. III. Prevalence and Risk of Diabetic Retinopathy When Age at Diagnosis Is 30 or More Years. Archives of Ophthalmology 1984;102:527–32.

13. Mathur R, Bhaskaran K, Edwards E, et al. Population Trends in the 10-Year Incidence and Prevalence of Diabetic Retinopathy in the UK: A Cohort Study in the Clinical Practice Research Datalink 2004-2014. BMJ Open 2017;7:e014444. DOI: 10.1136/bmjopen-2016-014444.

14. Yau JWY, Rogers SL, Kawasaki R, et al. Global Prevalence and Major Risk Factors of Diabetic Retinopathy. Diabetes Care 2012;35:556–64. DOI: 10/gfzgfw.

15. Wilkinson CP, Ferris FL, Klein RE, et al. Proposed International Clinical Diabetic Retinopathy and Diabetic Macular Edema Disease Severity Scales. Ophthalmology 2003;110:1677–82. DOI: 10.1016/S0161-6420(03)00475-5.

16. Early Treatment Diabetic Retinopathy Study Research Group. Grading Diabetic Retinopathy from Stereoscopic Color Fundus Photographs—An Extension of the Modified Airlie House Classification. Ophthalmology 1991;98:786–806. DOI: 10.1016/S0161-6420(13)38012-9.

17. Harding SP, Broadbent DM, Neoh C, White MC, and Vora J. Sensitivity and Specificity of Photography and Direct Ophthalmoscopy in Screening for Sight Threatening Eye Disease: The Liverpool Diabetic Eye Study. BMJ 1995;311:1131–5. DOI: 10/bzqrh4.

18. Vijan S, Hofer TP, and Hayward RA. Estimated Benefits of Glycemic Control in Microvascular Complications in Type 2 Diabetes. Annals of Internal Medicine 1997;127:788–95.

19. Wilson JMG and Jungner YG. Principles and Practice of Mass Screening for Disease. 1968.

20. UK National Screening Committee. Screening in the UK: Making Effective Recommendations 2015 to 2016. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/733227/Screening_in_the_UK___making_effective_recommendations_2015_to_2016.pdf (visited on 08/28/2019).

21. Public Health England. Criteria for Appraising the Viability, Effectiveness and Appropriateness of a Screening Programme. 2015. URL: https://www.gov.uk/government/publications/evidence-review-criteria-national-screening-programmes.

22. Butler JRG. Economic Evaluations of Screening Programs: A Review of Methods and Results. Clinica Chimica Acta; International Journal of Clinical Chemistry 2002;315:31–40. DOI: 10.1016/s0009-8981(01)00718-5.

23. Donaldson C. Using Economics to Assess the Place of Screening. Journal of Medical Screening 1994;1:124–8, discussion 128–9. URL: https://www.ncbi.nlm.nih.gov/pubmed/8790501.

24. Cairns J and Shackley P. Sometimes Sensitive, Seldom Specific: A Review of the Economics of Screening. Health Economics 1993;2:43–53. URL: https://www.ncbi.nlm.nih.gov/pubmed/8269046.

25. Scanlon PH. The English National Screening Programme for Sight-Threatening Diabetic Retinopathy. Journal of Medical Screening 2008;15:1–4. DOI: 10.1258/jms.2008.008015.

26. UK National Screening Committee. Essential Elements in Developing a Diabetic Retinopathy Screening Programme. Workbook 2004;4:1–79.

27. Public Health England. Diabetic Eye Screening: Care Pathway. 2015. URL: https://www.gov.uk/government/publications/diabetic-eye-screening-care-pathway.

28. Lund SH, Aspelund T, Kirby P, et al. Individualised Risk Assessment for Diabetic Retinopathy and Optimisation of Screening Intervals: A Scientific Approach to Reducing Healthcare Costs. The British Journal of Ophthalmology 2016;100:683–7. DOI: 10/f8jq4q.

29. Scotland G, McKeigue P, Philip S, et al. Modelling the Cost-Effectiveness of Adopting Risk-Stratified Approaches to Extended Screening Intervals in the National Diabetic Retinopathy Screening Programme in Scotland. Diabetic Medicine: A Journal of the British Diabetic Association 2016;33:886–95. DOI: 10.1111/dme.13129.

30. Scanlon PH, Aldington SJ, Leal J, et al. Development of a Cost-Effectiveness Model for Optimisation of the Screening Interval in Diabetic Retinopathy Screening. Health Technology Assessment 2015;19:1–116. DOI: 10.3310/hta19740.

31. Leese GP, Stratton IM, Land M, et al. Progression of Diabetes Retinal Status within Community Screening Programs and Potential Implications for Screening Intervals. Diabetes Care 2015;38:488–94. DOI: 10/f63262.

32. Taylor-Phillips S, Mistry H, Leslie R, et al. Extending the Diabetic Retinopathy Screening Interval beyond 1 Year: Systematic Review. The British Journal of Ophthalmology 2016;100:105–14. DOI: 10/f74tt3.

33. Echouffo-Tcheugui JB, Ali MK, Roglic G, Hayward RA, and Narayan KM. Screening Intervals for Diabetic Retinopathy and Incidence of Visual Loss: A Systematic Review. Diabetic Medicine: A Journal of the British Diabetic Association 2013;30:1272–92. DOI: 10.1111/dme.12274.

34. Aspelund T, Thornórisdóttir O, Olafsdottir E, et al. Individual Risk Assessment and Information Technology to Optimise Screening Frequency for Diabetic Retinopathy. Diabetologia 2011;54:2525–32. DOI: 10.1007/s00125-011-2257-7.

35. Younis N, Broadbent DM, Vora JP, Harding SP, and Liverpool Diabetic Eye Study. Incidence of Sight-Threatening Retinopathy in Patients with Type 2 Diabetes in the Liverpool Diabetic Eye Study: A Cohort Study. The Lancet 2003;361:195–200. URL: https://www.ncbi.nlm.nih.gov/pubmed/12547541.

36. Younis N, Broadbent DM, Harding SP, and Vora JP. Incidence of Sight-Threatening Retinopathy in Type 1 Diabetes in a Systematic Screening Programme. Diabetic Medicine: A Journal of the British Diabetic Association 2003;20:758–65. URL: https://www.ncbi.nlm.nih.gov/pubmed/12925058.

37. Agardh E and Tababat-Khani P. Adopting 3-Year Screening Intervals for Sight-Threatening Retinal Vascular Lesions in Type 2 Diabetic Subjects without Retinopathy. Diabetes Care 2011;34:1318–9. DOI: 10.2337/dc10-2308.

38. Kravitz RL, Duan N, and Braslow J. Evidence-Based Medicine, Heterogeneity of Treatment Effects, and the Trouble with Averages. The Milbank Quarterly 2004;82:661–87. DOI: 10/c85gcx.

39. Stevens W and Normand C. Optimisation versus Certainty: Understanding the Issue of Heterogeneity in Economic Evaluation. Social Science & Medicine 2004;58:315–20. DOI: 10/d4xdd4.

40. Sculpher MJ. Subgroups and Heterogeneity in Cost-Effectiveness Analysis. PharmacoEconomics 2008;26:799–806. DOI: 10/fg3x3v.

41. Ramaekers BLT, Joore MA, and Grutters JPC. How Should We Deal with Patient Heterogeneity in Economic Evaluation: A Systematic Review of National Pharmacoeconomic Guidelines. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2013;16:855–62. DOI: 10/f2xdsm.

42. Bradbury A and Olopade OI. The Case for Individualized Screening Recommendations for Breast Cancer. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology 2006;24:3328–30. DOI: 10.1200/jco.2006.05.8586.

43. Brawley OW. Risk-Based Mammography Screening: An Effort to Maximize the Benefits and Minimize the Harms. Annals of Internal Medicine 2012;156:662–3. DOI: 10.7326/0003-4819-156-9-201205010-00012.

44. Ten Haaf K, Bastani M, Cao P, et al. A Comparative Modeling Analysis of Risk-Based Lung Cancer Screening Strategies. JNCI: Journal of the National Cancer Institute. DOI: 10.1093/jnci/djz164.

45. Rainey L, van der Waal D, Jervaeus A, et al. Are We Ready for the Challenge of Implementing Risk-Based Breast Cancer Screening and Primary Prevention? The Breast 2018;39:24–32. DOI: 10.1016/j.breast.2018.02.029.

46. Goldstein BA, Navar AM, Pencina MJ, and Ioannidis JP. Opportunities and Challenges in Developing Risk Prediction Models with Electronic Health Records Data: A Systematic Review. Journal of the American Medical Informatics Association: JAMIA 2016. DOI: 10/gffmc5.

47. Win AK, Macinnis RJ, Hopper JL, and Jenkins MA. Risk Prediction Models for Colorectal Cancer: A Review. Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology 2012;21:398–410. DOI: 10/gfzgd7.

48. Gray EP, Teare MD, Stevens J, and Archer R. Risk Prediction Models for Lung Cancer: A Systematic Review. Clinical Lung Cancer 2016;17:95–106. DOI: 10/f8fxvx.

49. The Diabetes Control and Complications Trial Research Group. The Effect of Intensive Treatment of Diabetes on the Development and Progression of Long-Term Complications in Insulin-Dependent Diabetes Mellitus. The New England Journal of Medicine 1993;329:977–86. DOI: 10/fxfkdk.

50. Davis MDD, Fisher MRR, Gangnon REE, et al. Risk Factors for High-Risk Proliferative Diabetic Retinopathy and Severe Visual Loss: Early Treatment Diabetic Retinopathy Study Report #18. Investigative Ophthalmology & Visual Science 1998;39:233–52.

51. Marshall G, Garg SK, Jackson WE, Holmes DL, and Chase HP. Factors Influencing the Onset and Progression of Diabetic Retinopathy in Subjects with Insulin-Dependent Diabetes Mellitus. Ophthalmology 1993;100:1133–9. DOI: 10/gfzgf8.

52. Stratton IM, Kohner EMM, Aldington SJ, et al. UKPDS 50: Risk Factors for Incidence and Progression of Retinopathy in Type II Diabetes over 6 Years from Diagnosis. Diabetologia 2001;44:156–63. DOI: 10.1007/s001250051594.

53. Agardh CD, Agardh E, and Torffvit O. The Association between Retinopathy, Nephropathy, Cardiovascular Disease and Long-Term Metabolic Control in Type 1 Diabetes Mellitus: A 5 Year Follow-up Study of 442 Adult Patients in Routine Care. Diabetes Research and Clinical Practice 1997;35:113–21.

54. UK Prospective Diabetes Study Group. Tight Blood Pressure Control and Risk of Macrovascular and Microvascular Complications in Type 2 Diabetes: UKPDS 38. BMJ 1998;317:703–13. DOI: 10/dtrxj5.

55. Gallego PH, Craig ME, Hing S, and Donaghue KC. Role of Blood Pressure in Development of Early Retinopathy in Adolescents with Type 1 Diabetes: Prospective Cohort Study. BMJ 2008;337:a918–a918. DOI: 10/cbpgz8.

56. Chaturvedi N, Sjolie AK, Stephenson JM, et al. Effect of Lisinopril on Progression of Retinopathy in Normotensive People with Type 1 Diabetes. The EUCLID Study Group. EURODIAB Controlled Trial of Lisinopril in Insulin-Dependent Diabetes Mellitus. The Lancet 1998;351:28. DOI: 10.1016/s0140-6736(97)06209-0.

57. Keech A, Simes RJ, Barter P, et al. Effects of Long-Term Fenofibrate Therapy on Cardiovascular Events in 9795 People with Type 2 Diabetes Mellitus (the FIELD Study): Randomised Controlled Trial. The Lancet 2005;366:1849–61. DOI: 10/fbjhfn.

58. Klein R, Klein BE, Moss SE, and Cruickshanks KJ. The Wisconsin Epidemiologic Study of Diabetic Retinopathy: XVII. The 14-Year Incidence and Progression of Diabetic Retinopathy and Associated Risk Factors in Type 1 Diabetes. Ophthalmology 1998;105:1801–15. DOI: 10/dx3fzk.

59. Van der Heijden AAWA, Walraven I, van 't Riet E, et al. Validation of a Model to Estimate Personalised Screening Frequency to Monitor Diabetic Retinopathy. Diabetologia 2014;57:1332–8. DOI: 10.1007/s00125-014-3246-4.

60. Soto-Pedre E, Pinies JA, and Hernaez-Ortega MC. External Validation of a Risk Assessment Model to Adjust the Frequency of Eye-Screening Visits in Patients with Diabetes Mellitus. Journal of Diabetes and Its Complications 2015. DOI: 10/f7bcqk.

61. Eleuteri A, Fisher AC, Broadbent DM, et al. Individualised Variable-Interval Risk-Based Screening for Sight-Threatening Diabetic Retinopathy: The Liverpool Risk Calculation Engine. Diabetologia 2017. DOI: 10.1007/s00125-017-4386-0.

62. UK National Screening Committee. The UK NSC Recommendation on Diabetic Retinopathy Screening in Adults. 2016. URL: http://legacy.screening.nhs.uk/diabeticretinopathy.

63. Solomon SD, Chew E, Duh EJ, et al. Diabetic Retinopathy: A Position Statement by the American Diabetes Association. Diabetes Care 2017;40:412–8. DOI: 10.2337/dc16-2641.

64. Javitt JC, Canner JK, Frank RG, Steinwachs DM, and Sommer A. Detecting and Treating Retinopathy in Patients with Type I Diabetes Mellitus. A Health Policy Model. Ophthalmology 1990;97:483–94, discussion 494–5.

65. Dasbach EJ, Fryback DG, Newcomb PA, Klein R, and Klein BE. Cost-Effectiveness of Strategies for Detecting Diabetic Retinopathy. Medical Care 1991;29:20–39.

66. Javitt JC, Aiello LP, Chiang Y, Ferris FL, Canner JK, and Greenfield S. Preventive Eye Care in People With Diabetes Is Cost-Saving to the Federal Government: Implications for Health-Care Reform. Diabetes Care 1994;17:909–17. DOI: 10/c62dfm.

67. Vijan S, Hofer TP, and Hayward RA. Cost-Utility Analysis of Screening Intervals for Diabetic Retinopathy in Patients with Type 2 Diabetes Mellitus. JAMA: The Journal of the American Medical Association 2000;283:889–96.

68. Davies R, Roderick P, Canning C, and Brailsford S. The Evaluation of Screening Policies for Diabetic Retinopathy Using Simulation. Diabetic Medicine: A Journal of the British Diabetic Association 2002;19:762–70. DOI: 10.1046/j.1464-5491.2002.00773.x.

69. Brailsford SC, Gutjahr WJ, Rauner MS, and Zeppelzauer W. Combined Discrete-Event Simulation and Ant Colony Optimisation Approach for Selecting Optimal Screening Policies for Diabetic Retinopathy. CMS Books in Mathematics/Ouvrages de Mathematiques de la SMC 2007;4:59–83. DOI: 10.1007/s10287-006-0008-x.

70. Tung TH, Shih HC, Chen SJ, Chou P, Liu CM, and Liu JH. Economic Evaluation of Screening for Diabetic Retinopathy among Chinese Type 2 Diabetics: A Community-Based Study in Kinmen, Taiwan. Journal of Epidemiology / Japan Epidemiological Association 2008;18:225–33.

71. Rein DB, Wittenborn JS, Zhang X, et al. The Cost-Effectiveness of Three Screening Alternatives for People with Diabetes with No or Early Diabetic Retinopathy. Health Services Research 2011;46:1534–61. DOI: 10.1111/j.1475-6773.2011.01263.x.

72. Chalk D, Pitt M, Vaidya B, and Stein K. Can the Retinal Screening Interval Be Safely Increased to 2 Years for Type 2 Diabetic Patients without Retinopathy? Diabetes Care 2012;35:1663–8. DOI: 10.2337/dc11-2282.

73. Mehlsen J, Erlandsen M, Poulsen PL, and Bek T. Individualized Optimization of the Screening Interval for Diabetic Retinopathy: A New Model. Acta Ophthalmologica 2012;90:109–14. DOI: 10.1111/j.1755-3768.2010.01882.x.

74. Day TE, Ravi N, Xian H, and Brugh A. Sensitivity of Diabetic Retinopathy Associated Vision Loss to Screening Interval in an Agent-Based/Discrete Event Simulation Model. Computers in Biology and Medicine 2014;47:7–12. DOI: 10.1016/j.compbiomed.2014.01.007.

75. Harding SP. RP-PG-1210-12016 - Introducing Personalised Risk Based Intervals in Screening for Diabetic Retinopathy: Development, Implementation and Assessment of Safety, Cost-Effectiveness and Patient Experience. National Institute for Health Research Programme Grants for Applied Research, 2013. URL: http://www.ccf.nihr.ac.uk/PGfAR/about/Pages/Abstract.aspx?ID=7790.

76. Broadbent DM, Sampson CJ, Wang A, et al. Individualised Screening for Diabetic Retinopathy: The ISDR Study. Rationale, Design and Methodology for a Randomised Controlled Trial Comparing Annual and Personalised Variable-Interval Risk-Based Screening. BMJ Open 2019;9. DOI: 10.1136/bmjopen-2018-025788.

77. Petrou S and Gray A. Economic Evaluation Using Decision Analytical Modelling: Design, Conduct, Analysis, and Reporting. BMJ 2011;342:d1766. DOI: 10/cw48vr.

78. Cooper K, Brailsford SC, and Davies R. Choice of Modelling Technique for Evaluating Health Care Interventions. The Journal of the Operational Research Society 2007;58:168–76. DOI: 10.1057/palgrave.jors.2602230.

79. Standfield L, Comans T, and Scuffham P. Markov Modeling and Discrete Event Simulation in Health Care: A Systematic Comparison. International Journal of Technology Assessment in Health Care 2014;30:165–72. DOI: 10/gfzggb.

80. Philips Z, Bojke L, Sculpher M, Claxton K, and Golder S. Good Practice Guidelines for Decision-Analytic Modelling in Health Technology Assessment: A Review and Consolidation of Quality Assessment. PharmacoEconomics 2006;24:355–71. DOI: 10/btnfww.

81. Caro JJ. Response to Letter to the Editor Regarding Discretely Integrated Condition Event Simulation for Pharmacoeconomics. PharmacoEconomics 2016:1–2. DOI: 10/gfzgfq.

82. Toumi M, Beck E, Sherman S, Mohseninejad L, and Aballéa S. Discretely Integrated Condition Event Simulation for Pharmacoeconomics. PharmacoEconomics 2016:1–2. DOI: 10/gfzgft.

83. Caro JJ. Discretely Integrated Condition Event (DICE) Simulation for Pharmacoeconomics. PharmacoEconomics 2016. DOI: 10.1007/s40273-016-0394-z.

84. Hollman C, Paulden M, Pechlivanoglou P, and McCabe C. A Comparison of Four Software Programs for Implementing Decision Analytic Cost-Effectiveness Models. PharmacoEconomics 2017:1–14. DOI: 10/gbspkn.

85. Brennan A, Chick SE, and Davies R. A Taxonomy of Model Structures for Economic Evaluation of Health Technologies. Health Economics 2006;15:1295–310. DOI: 10.1002/hec.1148.

86. Ackerman SJ. Benefits of Preventive Programs in Eye Care Are Visible on the Bottom Line. A New Nationwide Effort to Improve Eye Care for People with Diabetes Gets Backing from a Study on the Cost-Effectiveness of Screening for Retinopathy. Diabetes Care 1992;15:580–1. DOI: 10.2337/diacare.15.4.580.

87. Fendrick AM, Javitt JC, and Chiang YP. Cost-Effectiveness of the Screening and Treatment of Diabetic Retinopathy. What Are the Costs of Underutilization? International Journal of Technology Assessment in Health Care 1992;8:694–707. DOI: 10/b4nft6.

88. Crijns H, Casparie AF, and Hendrikse F. Toekomstige Behoefte Aan Oogzorg Voor Patiënten Met Diabetes Mellitus, Kosten En Effectiviteit. Nederlands Tijdschrift voor Geneeskunde 1995;139:1336–41.

89. Tennvall GR, Rosenqvist U, and Nilsson LG. [It is worth concentrating on prevention of complications. An example from diabetic care]. Lakartidningen 1997;94:2817–21.

90. Wu SY, Sainfort F, Tomar RH, et al. Development and Application of a Model to Estimate the Impact of Type 1 Diabetes on Health-Related Quality of Life. Diabetes Care 1998;21:725–31. DOI: 10/d49h7v.

91. Crijns H, Casparie AF, and Hendrikse F. Continuous Computer Simulation Analysis of the Cost-Effectiveness of Screening and Treating Diabetic Retinopathy. International Journal of Technology Assessment in Health Care 1999;15:198–206. URL: https://www.ncbi.nlm.nih.gov/pubmed/10407606.

92. Palmer AJ, Weiss C, Sendi PP, et al. The Cost-Effectiveness of Different Management Strategies for Type I Diabetes: A Swiss Perspective. Diabetologia 2000;43:13–26. DOI: 10/cvbjdg.

93. Sharma S, Brown GC, Brown MM, Hollands H, and Shah GK. The Cost-Effectiveness of Grid Laser Photocoagulation for the Treatment of Diabetic Macular Edema: Results of a Patient-Based Cost-Utility Analysis. Current Opinion in Ophthalmology 2000;11:175–9.

94. Brown GC, Brown MM, and Sharma S. How Often Should Patients with Diabetes Be Screened for Retinopathy? JAMA: The Journal of the American Medical Association 2000;284:438, author reply 439. URL: https://www.ncbi.nlm.nih.gov/pubmed/10904496.

95. Hoskins Jr HD. How Often Should Patients with Diabetes Be Screened for Retinopathy? JAMA: The Journal of the American Medical Association 2000;284:438–9. URL: https://www.ncbi.nlm.nih.gov/pubmed/10904497.

96. Polak BCP, Crijns H, Casparie AF, and Niessen LW. Cost-Effectiveness of Glycemic Control and Ophthalmological Care in Diabetic Retinopathy. Health Policy 2003;64:89–97.

97. Whited JD, Datta SK, Aiello LM, et al. A Modeled Economic Analysis of a Digital Tele-Ophthalmology System as Used by Three Federal Health Care Agencies for Detecting Proliferative Diabetic Retinopathy. Telemedicine Journal and E-Health: The Official Journal of the American Telemedicine Association 2005;11:641–51. DOI: 10/b3hvv2.

98. Scotland GS, McNamee P, Philip S, et al. Cost-Effectiveness of Implementing Automated Grading within the National Screening Programme for Diabetic Retinopathy in Scotland. The British Journal of Ophthalmology 2007;91:1518–23. DOI: 10/crnsjs.

99. Mitchell P, Annemans L, Gallagher M, et al. Cost-Effectiveness of Ranibizumab in Treatment of Diabetic Macular Oedema (DME) Causing Visual Impairment: Evidence from the RESTORE Trial. The British Journal of Ophthalmology 2012;96:688–93. DOI: 10/gb9q8d.

100. Dewan V, Lambert D, Edler J, Kymes S, and Apte RS. Cost-Effectiveness Analysis of Ranibizumab plus Prompt or Deferred Laser or Triamcinolone plus Prompt Laser for Diabetic Macular Edema. Ophthalmology 2012;119:1679–84. DOI: 10.1016/j.ophtha.2012.01.049.

101. Brady CJ, Villanti AC, Gupta OP, Graham MG, and Sergott RC. Tele-Ophthalmology Screening for Proliferative Diabetic Retinopathy in Urban Primary Care Offices: An Economic Analysis. Ophthalmic Surgery, Lasers & Imaging Retina 2014;45:556–61. DOI: 10.3928/23258160-20141118-11.

102. Chan CKW, Gangwani RA, McGhee SM, Lian J, and Wong DSH. Cost-Effectiveness of Screening for Intermediate Age-Related Macular Degeneration during Diabetic Retinopathy Screening. Ophthalmology 2015;122:2278–85. DOI: 10.1016/j.ophtha.2015.06.050.

103. Wolowacz S, Pearson I, Shannon P, et al. Development and Validation of a Cost-Utility Model for Type 1 Diabetes Mellitus. Diabetic Medicine: A Journal of the British Diabetic Association 2015;32:1023–35. DOI: 10/gfzgfg.

104. Wu B, Li J, and Wu H. Strategies to Screen for Diabetic Retinopathy in Chinese Patients with Newly Diagnosed Type 2 Diabetes: A Cost-Effectiveness Analysis. Medicine 2015;94:e1989. DOI: 10/gfzgfv.

105. Kamble S, Schulman KA, and Reed SD. Cost-Effectiveness of Sensor-Augmented Pump Therapy in Adults with Type 1 Diabetes in the United States. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2012;15:632–8. DOI: 10/f365tg.

106. Javitt JC, Canner JK, and Sommer A. Cost Effectiveness of Current Approaches to the Control of Retinopathy in Type I Diabetics. Ophthalmology 1989;96:255–64.

107. Kawasaki R, Akune Y, Hiratsuka Y, Fukuhara S, and Yamada M. Cost-Utility Analysis of Screening for Diabetic Retinopathy in Japan: A Probabilistic Markov Modeling Study. Ophthalmic Epidemiology 2015;22:4–12. DOI: 10.3109/09286586.2014.988876.

108. Scotland GS, McNamee P, Fleming AD, et al. Costs and Consequences of Automated Algorithms versus Manual Grading for the Detection of Referable Diabetic Retinopathy. The British Journal of Ophthalmology 2010;94:712–9. DOI: 10/cmtprr.

109. Royle P, Mistry H, Auguste P, et al. Pan-Retinal Photocoagulation and Other Forms of Laser Treatment and Drug Therapies for Non-Proliferative Diabetic Retinopathy: Systematic Review and Economic Evaluation. Health Technology Assessment 2015;19:v–xxviii, 1–247. DOI: 10.3310/hta19510.

110. Sharma S, Hollands H, Brown GC, Brown MM, Shah GK, and Sharma SM. The Cost-Effectiveness of Early Vitrectomy for the Treatment of Vitreous Hemorrhage in Diabetic Retinopathy. Current Opinion in Ophthalmology 2001;12:230–4.

111. Pershing S, Enns EA, Matesic B, Owens DK, and Goldhaber-Fiebert JD. Cost-Effectiveness of Treatment of Diabetic Macular Edema. Annals of Internal Medicine 2014;160:18–29. DOI: 10/gbfj7f.

112. Brandle M, Zhou H, Smith BRK, et al. The Direct Medical Cost of Type 2 Diabetes. Diabetes Care 2003;26:2300–4. DOI: 10.2337/diacare.26.8.2300.

113. Rachapelle S, Legood R, Alavi Y, et al. The Cost-Utility of Telemedicine to Screen for Diabetic Retinopathy in India. Ophthalmology 2013;120:566–73. DOI: 10.1016/j.ophtha.2012.09.002.

114. Sculpher MJ, Buxton MJ, Ferguson BA, Spiegelhalter DJ, and Kirby AJ. Screening for Diabetic Retinopathy: A Relative Cost-Effectiveness Analysis of Alternative Modalities and Strategies. Health Economics 1992;1:39–51.

115. Eastman RC, Javitt JC, Herman WH, et al. Model of Complications of NIDDM. I. Model Construction and Assumptions. Diabetes Care 1997;20:725–34.

116. Jones S and Edwards RT. Diabetic Retinopathy Screening: A Systematic Review of the Economic Evidence. Diabetic Medicine: A Journal of the British Diabetic Association 2010;27:249–56. DOI: 10.1111/j.1464-5491.2009.02870.x.

117. Stein JD, Newman-Casey PA, Kim DD, Nwanyanwu KH, Johnson MW, and Hutton DW. Cost-Effectiveness of Various Interventions for Newly Diagnosed Diabetic Macular Edema. Ophthalmology 2013;120:1835–42. DOI: 10.1016/j.ophtha.2013.02.002.

118. Javitt JC and Aiello LP. Cost-Effectiveness of Detecting and Treating Diabetic Retinopathy. Annals of Internal Medicine 1 Pt 2 1996;124:164–9. DOI: 10/gfzgfn.

119. Tosh J and Wailoo A. Review of Software for Decision Modelling. Decision Support Unit, 2008. URL: http://www.nicedsu.org.uk/PDFs%20of%20reports/softwarereport-final.pdf.

120. World Health Organization. Preamble to the Constitution of the World Health Organization as Adopted by the International Health Conference. 2. 1946:100.

121. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal 2013. 2013.

122. Sullivan SM, Wells G, and Coyle D. What Guidance Are Economists Given on How to Present Economic Evaluations for Policymakers? A Systematic Review. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2015;18:915–24. DOI: 10.1016/j.jval.2015.06.007.

123. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-Effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. JAMA: The Journal of the American Medical Association 2016;316:1093–103. DOI: 10/gfxjrc.

124. Sampson CJ. Identifying Objects of Value at the End of Life. In: *Care at the End of Life: An Economic Perspective*. Ed. by Round J. Switzerland: Adis, 2016:103–22. DOI: 10.1007/978-3-319-28267-1_8.

125. Richardson J, McKie J, and Bariola E. Multiattribute Utility Instruments and Their Use A2 - Culyer, Anthony J. In: *Encyclopedia of Health Economics*. San Diego: Elsevier, 2014:341–57. DOI: 10.1016/B978-0-12-375678-7.00505-8.

126. EuroQol Group. EuroQol - a New Facility for the Measurement of Health-Related Quality of Life. Health Policy 1990;16:199–208.

127. Herdman M, Gudex C, Lloyd A, et al. Development and Preliminary Testing of the New Five-Level Version of EQ-5D (EQ-5D-5L). Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation 2011;20:1727–36. DOI: 10.1007/s11136-011-9903-x.

128. Szende A, Oppe M, and Devlin N. EQ-5D Value Sets: Inventory, Comparative Review and User Guide. Ed. by Szende A, Oppe M, and Devlin N. Vol. 2. EuroQol Group Monographs. Springer Netherlands, 2006. DOI: 10.1007/1-4020-5511-0.

129. Oppe M, Devlin NJ, van Hout B, Krabbe PF, and de Charro F. A Program of Methodological Research to Arrive at the New International EQ-5D-5L Valuation Protocol. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2014;17:445–53. DOI: 10.1016/j.jval.2014.04.002.

130. Devlin NJ, Shah KK, Feng Y, Mulhern B, and van Hout B. Valuing Health-Related Quality of Life: An EQ-5D-5L Value Set for England. Health Economics 2018;27:7–22. DOI: 10.1002/hec.3564.

131. Feng Y, Devlin N, and Herdman M. Assessing the Health of the General Population in England: How Do the Three- and Five-Level Versions of EQ-5D Compare? Health and Quality of Life Outcomes 2015;13:171. DOI: 10.1186/s12955-015-0356-8.

132. Janssen MF, Pickard AS, Golicki D, et al. Measurement Properties of the EQ-5D-5L Compared to the EQ-5D-3L across Eight Patient Groups: A Multi-Country Study. Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation 2012:1717–27. DOI: 10.1007/s11136-012-0322-4.

133. Espallargues M, Czoski-Murray CJ, Bansback NJ, et al. The Impact of Age-Related Macular Degeneration on Health Status Utility Values. Investigative Ophthalmology & Visual Science 2005;46:4016–23. DOI: 10.1167/iovs.05-0072.

134. Feeny D, Furlong W, Boyle M, and Torrance GW. Multi-Attribute Health Status Classification Systems. Health Utilities Index. PharmacoEconomics 1995;7:490–502. DOI: 10.2165/00019053-199507060-00004.

135. Furlong W, Feeny D, Torrance G, et al. Multiplicative Multi-Attribute Utility Function for the Health Utilities Index Mark 3 (HUI3) System: A Technical Report. Centre for Health Economics and Policy Analysis (CHEPA), McMaster University, Hamilton, Canada, 1998. URL: http://ideas.repec.org/p/hpa/wpaper/199811.html.

136. Shah KK, Mulhern B, Longworth L, and Janssen MF. Views of the UK General Public on Important Aspects of Health Not Captured by EQ-5D. The Patient - Patient-Centered Outcomes Research 2017. DOI: 10.1007/s40271-017-0240-1.

137. Tosh JC, Longworth LJ, and George E. Utility Values in National Institute for Health and Clinical Excellence (NICE) Technology Appraisals. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2011;14:102–9. DOI: 10/c8k2fq.

138. Bunce C and Wormald R. Leading Causes of Certification for Blindness and Partial Sight in England & Wales. BMC Public Health 2006;6:58. DOI: 10.1186/1471-2458-6-58.

139. Congdon N, O'Colmain B, Klaver CCW, et al. Causes and Prevalence of Visual Impairment among Adults in the United States. Archives of Ophthalmology 2004;122:477–85. DOI: 10.1001/archopht.122.4.477.

140. Gothwal VK and Bagga DK. Utility Values in the Visually Impaired: Comparing Time-Trade off and VisQoL. Optometry and Vision Science: Official Publication of the American Academy of Optometry 2013;90:843–54. DOI: 10.1097/OPX.0b013e318291063a.

141. Crewe JM, Morlet N, Morgan WH, et al. Quality of Life of the Most Severely Vision-Impaired. Clinical & Experimental Ophthalmology 2011;39:336–43. DOI: 10.1111/j.1442-9071.2010.02466.x.

142. Knauer C and Pfeiffer N. The Value of Vision. Graefe's Archive for Clinical and Experimental Ophthalmology = Albrecht Von Graefes Archiv Fur Klinische Und Experimentelle Ophthalmologie 2008;246:477–82. DOI: 10.1007/s00417-007-0668-4.

143. Sharma S, Oliver-Fernandez A, Liu W, Buchholz P, and Walt J. The Impact of Diabetic Retinopathy on Health-Related Quality of Life. Current Opinion in Ophthalmology 2005;16:155–9. DOI: 10.1097/01.icu.0000161227.21797.3d.

144. Li R, Zhang P, Barker LE, Chowdhury FM, and Zhang X. Cost-Effectiveness of Interventions to Prevent and Control Diabetes Mellitus: A Systematic Review. Diabetes Care 2010;33:1872–94. DOI: 10.2337/dc10-0843.

145. Mohamed Q, Gillies MC, and Wong TY. Management of Diabetic Retinopathy: A Systematic Review. JAMA: The Journal of the American Medical Association 2007;298:902–16. DOI: 10.1001/jama.298.8.902.

146. Brown MM. Utility Values Associated with Blindness in an Adult Population. The British Journal of Ophthalmology 2001;85:327–31. DOI: 10.1136/bjo.85.3.327.

147. Brown MM, Brown GC, Sharma S, and Shah G. Utility Values and Diabetic Retinopathy. American Journal of Ophthalmology 1999;128:324–30. DOI: 10.1016/S0002-9394(99)00146-4.

148. Brown MM, Brown GC, Sharma S, and Landy J. Health Care Economic Analyses and Value-Based Medicine. Survey of Ophthalmology 2003;48:204–23. DOI: 10.1016/s0039-6257(02)00457-5.

149. Sharma S, Brown GC, Brown MM, et al. Converting Visual Acuity to Utilities. Canadian Journal of Ophthalmology. Journal Canadien D'ophtalmologie 2000;35:267–72. DOI: 10/gfzgfm.

150. Lloyd A, Nafees B, Gavriel S, Rousculp MD, Boye KS, and Ahmad A. Health Utility Values Associated with Diabetic Retinopathy. Diabetic Medicine: A Journal of the British Diabetic Association 2008;25:618–24. DOI: 10.1111/j.1464-5491.2008.02430.x.

151. Poku E, Brazier J, Carlton J, and Ferreira A. Health State Utilities in Patients with Diabetic Retinopathy, Diabetic Macular Oedema and Age-Related Macular Degeneration: A Systematic Review. BMC Ophthalmology 2013;13:74. DOI: 10/f5km9x.

152. Heintz E, Wiréhn AB, Peebo BB, Rosenqvist U, and Levin LÅ. QALY Weights for Diabetic Retinopathy—A Comparison of Health State Valuations with HUI-3, EQ-5D, EQ-VAS, and TTO. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2012;15:475–84. DOI: 10.1016/j.jval.2011.11.031.

153. Wolowacz SE, Briggs A, Belozeroff V, et al. Estimating Health-State Utility for Economic Models in Clinical Studies: An ISPOR Good Research Practices Task Force Report. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2016;19:704–19. DOI: 10.1016/j.jval.2016.06.001.

154. National Institute for Health and Care Excellence. Position Statement on Use of the EQ-5D-5L Valuation Set. 2017.

155. Van Hout B, Janssen M, Feng YS, et al. Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2012;15:708–15. DOI: 10.1016/j.jval.2012.02.008.

156. Manning WG and Mullahy J. Estimating Log Models: To Transform or Not to Transform? Journal of Health Economics 2001;20:461–94. URL: https://www.ncbi.nlm.nih.gov/pubmed/11469231.

157. Glick HA, Doshi JA, Sonnad SS, and Polsky D. Economic Evaluation in Clinical Trials. Oxford University Press, 2014.

158. Glick HA, Doshi JA, Sonnad SS, and Polsky D. Glmdiag. 2008. URL: http://www.uphs.upenn.edu/dgimhsr/glmdiagnostic_004.do.

159. StataCorp. Stata Statistical Software. College Station, TX: StataCorp LP, 2015.

160. Mander A. BATPLOT: Stata Module to Produce Bland-Altman Plots Accounting for Trend. Boston College Department of Economics, 2012. URL: https://ideas.repec.org/c/boc/bocode/s448703.html.

161. Tosh J, Brazier JE, Evans P, and Longworth L. A Review of Generic Preference-Based Measures of Health-Related Quality of Life in Visual Disorders. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2012;15:118–27. DOI: 10/b9t572.

162. McClure NS, Sayah FA, Ohinmaa A, and Johnson JA. Minimally Important Difference of the EQ-5D-5L Index Score in Adults with Type 2 Diabetes. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2018. DOI: 10/gfbwzh.

163. Tsilimbaris MK, Kontadakis GA, Tsika C, Papageorgiou D, and Charoniti M. Effect of Panretinal Photocoagulation Treatment on Vision-Related Quality of Life of Patients with Proliferative Diabetic Retinopathy. Retina 2013;33:756–61. DOI: 10.1097/IAE.0b013e31826b0c06.

164. Scanlon PH, Loftus J, Starita C, and Stratton IM. The Use of Weighted Health-Related Quality of Life Scores in People with Diabetic Macular Oedema at Baseline in a Randomized Clinical Trial. Diabetic Medicine: A Journal of the British Diabetic Association 2014;32:97–101. DOI: 10/gd89sb.

165. Kaltenthaler E, Tappenden P, Paisley S, and Squires H. NICE DSU Technical Support Document 13: Indentifying and Reviewing Evidence to Inform the Conceptualisation and Population of Cost-Effectiveness Models. 2011.

166. National Institute for Health and Clinical Excellence. Guide to the Methods of Technology Appraisal. National Institute for Health and Clinical Excellence, 2008:1–54. URL: www.nice.org.uk.

167. Brazier JE, Green C, and Kanis JA. A Systematic Review of Health State Utility Values for Osteoporosis-Related Conditions. Osteoporosis International: A Journal Established as Result of Cooperation Between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 2002;13:768–76. DOI: 10.1007/s001980200107.

168. Bremner KE, Chong CAKY, Tomlinson G, Alibhai SMH, and Krahn MD. A Review and Meta-Analysis of Prostate Cancer Utilities. Medical Decision Making: An International Journal of the Society for Medical Decision Making 2007;27:288–98. DOI: 10.1177/0272989x07300604.

169. Doth AH, Hansson PT, Jensen MP, and Taylor RS. The Burden of Neuropathic Pain: A Systematic Review and Meta-Analysis of Health Utilities. Pain 2010;149:338–44. DOI: 10.1016/j.pain.2010.02.034.

170. McLernon DJ, Dillon J, and Donnan PT. Health-State Utilities in Liver Disease: A Systematic Review. Medical Decision Making: An International Journal of the Society for Medical Decision Making 2008;28:582–92. DOI: 10/bzm95x.

171. Tengs TO and Lin TH. A Meta-Analysis of Utility Estimates for HIV/AIDS. Medical Decision Making: An International Journal of the Society for Medical Decision Making 2002;22:475–81.

172. Peasgood T, Herrmann K, Kanis JA, and Brazier JE. An Updated Systematic Review of Health State Utility Values for Osteoporosis Related Conditions. Osteoporosis International: A Journal Established as Result of Cooperation Between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 2009;20:853–68. DOI: 10/dp4bv6.

173. Sturza J. A Review and Meta-Analysis of Utility Values for Lung Cancer. Medical Decision Making: An International Journal of the Society for Medical Decision Making 2010;30:685–93. DOI: 10/cpz2vz.

174. Peasgood T, Ward SE, and Brazier J. Health-State Utility Values in Breast Cancer. Expert Review of Pharmacoeconomics & Outcomes Research 2010;10:553–66. DOI: 10/d98tdq.

175. Post PN, Stiggelbout AM, and Wakker PP. The Utility of Health States after Stroke: A Systematic Review of the Literature. Stroke; a Journal of Cerebral Circulation 2001;32:1425–9.

176. Lung TWC, Hayes AJ, Hayen A, Farmer A, and Clarke PM. A Meta-Analysis of Health State Valuations for People with Diabetes: Explaining the Variation across Methods and Implications for Economic Evaluation. Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation 2011;20:1669–78. DOI: 10.1007/s11136-011-9902-y.

177. Wyld M, Morton RL, Hayen A, Howard K, and Webster AC. A Systematic Review and Meta-Analysis of Utility-Based Quality of Life in Chronic Kidney Disease Treatments. PLoS Medicine 2012;9:e1001307. DOI: 10/gfzgb4.

178. Bermingham SL and Ashe JF. Systematic Review of the Impact of Urinary Tract Infections on Health-Related Quality of Life. BJU International 11 Pt C 2012;110:E830–6. DOI: 10.1111/j.1464-410x.2012.11337.x.

179. Doyle S, Lloyd A, Moore L, Ray J, and Gray A. A Systematic Review and Critical Assessment of Health State Utilities. PharmacoEconomics 2012;30:1133–43. DOI: 10.2165/11599420-000000000-00000.

180. Van Litsenburg RRL, Kunst A, Huisman J, Ket JCF, Kaspers GJL, and Gemke RJBJ. Health Status Utilities in Pediatrics: A Systematic Review of Acute Lymphoblastic Leukemia. Medical Decision Making: An International Journal of the Society for Medical Decision Making 2014;34:21–32. DOI: 10/f5m2fq.

181. Mohiuddin S and Payne K. Utility Values for Adults with Unipolar Depression: Systematic Review and Meta-Analysis. Medical Decision Making: An International Journal of the Society for Medical Decision Making 2014;34:666–85. DOI: 10/f57dz5.

182. Beaudet A, Clegg J, Thuresson PO, Lloyd A, and McEwan P. Review of Utility Values for Economic Modeling in Type 2 Diabetes. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2014;17:462–70. DOI: 10.1016/j.jval.2014.03.003.

183. Djalalov S, Rabeneck L, Tomlinson G, Bremner KE, Hilsden R, and Hoch JS. A Review and Meta-Analysis of Colorectal Cancer Utilities. Medical Decision Making: An International Journal of the Society for Medical Decision Making 2014;34:809–18. DOI: 10.1177/0272989x14536779.

184. Kinney MR, Burfitt SN, Stullenbarger E, Rees B, and DeBolt MR. Quality of Life in Cardiac Patient Research: A Meta-Analysis. Nursing Research 1996;45:173–80. URL: https://www.ncbi.nlm.nih.gov/pubmed/8637799.

185. Dijkers M. Quality of Life after Spinal Cord Injury: A Meta Analysis of the Effects of Disablement Components. Spinal Cord 1997;35:829–40.

186. Meacock R, Dale N, and Harrison MJ. The Humanistic and Economic Burden of Systemic Lupus Erythematosus : A Systematic Review. PharmacoEconomics 2013;31:49–61. DOI: 10/gfzgdx.

187. Brazier JE, Papaioannou D, and Cantrell A. Identifying and Reviewing Health State Utility Values for Populating Decision Models. In: ed. by Shemilt I, Mugford M, Vale L, Marsh K, and Donaldson C. Wiley-Blackwell, 2010:93–105. DOI: 10.1002/9781444320398.ch8.

188. Peasgood T and Brazier J. Is Meta-Analysis for Utility Values Appropriate Given the Potential Impact Different Elicitation Methods Have on Values? PharmacoEconomics 2015;33:1101–5. DOI: 10.1007/s40273-015-0310-y.

189. Moher D, Liberati A, Tetzlaff J, Altman DG, and PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. BMJ 2009;339:b2535. DOI: 10/fmvxsp.

190. Papaioannou D, Brazier JE, and Paisley S. NICE DSU Technical Support Document 9: The Identification, Review and Synthesis of Health State Utility Values from the Literature. 2011:1–64.

191. Papaioannou D, Brazier JE, and Paisley S. Systematic Searching and Selection of Health State Utility Values from the Literature. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2013;16:686–95. DOI: 10/f2xhrm.

192. Saramago P, Manca A, and Sutton AJ. Deriving Input Parameters for Cost-Effectiveness Modeling: Taxonomy of Data Types and Approaches to Their Statistical Synthesis. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2012;15:639–49. DOI: 10/gfzgb5.

193. Kaltenthaler E, Tappenden P, and Paisley S. Reviewing the Evidence to Inform the Population of Cost-Effectiveness Models within Health Technology Assessments. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2013;16:830–6. DOI: 10/f2xdsg.

194. Zechmeister-Koss I, Schnell-Inderst P, and Zauner G. Appropriate Evidence Sources for Populating Decision Analytic Models within Health Technology Assessment (HTA): A Systematic Review of HTA Manuals and Health Economic Guidelines. Medical Decision Making: An International Journal of the Society for Medical Decision Making 2014;34:288–99. DOI: 10/f5zddq.

195. Cooper NJ, Coyle D, Abrams K, Mugford M, and Sutton A. Use of Evidence in Decision Models: An Appraisal of Health Technology Assessments in the UK since 1997. Journal of Health Services Research & Policy 2005;10:245–50. DOI: 10.1258/135581905774414187.

196. Bagos PG. Meta-Analysis in Stata Using Gllamm. Research Synthesis Methods 2015;6:310–32. DOI: 10.1002/jrsm.1157.

197. Ahola AJ, Saraheimo M, Forsblom C, et al. Health-Related Quality of Life in Patients with Type 1diabetes—Association with Diabetic Complications (the FinnDiane Study). Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association 2010;25:1903–8. DOI: 10.1093/ndt/gfp709.

198. Alva M, Gray A, Mihaylova B, and Clarke P. The Effect of Diabetes Complications on Health-Related Quality of Life: The Importance of Longitudinal Data to Address Patient Heterogeneity. Health Economics 2014;23. 00000:487–500. DOI: 10.1002/hec.2930.

199. Bagust A and Beale S. Modelling EuroQol Health-Related Utility Values for Diabetic Complications from CODE-2 Data. Health Economics 2005;14:217–30. DOI: 10.1002/hec.910.

200. Clarke P, Gray A, and Holman R. Estimating Utility Values for Health States of Type 2 Diabetic Patients Using the EQ-5D (UKPDS 62). Medical Decision Making: An International Journal of the Society for Medical Decision Making 2002;22. 00453:340–9. DOI: 10.1177/0272989X0202200412.

201. Clarke PM, Simon J, Cull CA, and Holman RR. Assessing the Impact of Visual Acuity on Quality of Life in Individuals with Type 2 Diabetes Using the Short Form-36. Diabetes Care 2006;29:1506–11. DOI: 10.2337/dc05-2150.

202. Grandy S, Fox KM, and SHIELD Study Group. Change in Health Status (EQ-5D) over 5 Years among Individuals with and without Type 2 Diabetes Mellitus in the SHIELD Longitudinal Study. Health and Quality of Life Outcomes 2012;10:99. DOI: 10.1186/1477-7525-10-99.

203. Hahl J, Hämäläinen H, Sintonen H, Simell T, Arinen S, and Simell O. Health-Related Quality of Life in Type 1 Diabetes without or with Symptoms of Long-Term Complications. Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation 2002;11:427–36. DOI: 10/bbw35r.

204. Hammerschmidt T, Zeitler HP, and Leidl R. A Utility-Theoretic Approach to the Aggregation of Willingness to Pay Measured in Decomposed Scenarios: Development and Empirical Test. Health Economics 2004;13:345–61. DOI: 10/d5vjcq.

205. Hayes A, Arima H, Woodward M, et al. Changes in Quality of Life Associated with Complications of Diabetes: Results from the ADVANCE Study. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2016;19:36–41. DOI: 10/f8chxj.

206. Hunger M, Schunk M, Meisinger C, Peters A, and Holle R. Estimation of the Relationship between Body Mass Index and EQ-5D Health Utilities in Individuals with Type 2 Diabetes: Evidence from the Population-Based KORA Studies. Journal of Diabetes and Its Complications 2012;26:413–8. DOI: 10/f38zx9.

207. Kiadaliri AA, Gerdtham UG, Eliasson B, Gudbjörnsdottir S, Svensson AM, and Carlsson KS. Health Utilities of Type 2 Diabetes-Related Complications: A Cross-Sectional Study in Sweden. International Journal of Environmental Research and Public Health 2014;11:4939–52. DOI: 10/f22m22.

208. Lee WJ, Song KH, Noh JH, Choi YJ, and Jo MW. Health-Related Quality of Life Using the EuroQol 5D Questionnaire in Korean Patients with Type 2 Diabetes. Journal of Korean Medical Science 2012;27:255–60. DOI: 10/f3wmv6.

209. Lin JC and Yu JH. Assessment of Quality of Life among Taiwanese Patients with Visual Impairment. Journal of the Formosan Medical Association = Taiwan yi zhi 2012;111:572–9. DOI: 10/f4d65g.

210. Peasgood T, Brennan A, Mansell P, Elliott J, Basarir H, and Kruger J. The Impact of Diabetes-Related Complications on Preference-Based Measures of Health-Related Quality of Life in Adults with Type I Diabetes. Medical Decision Making: An International Journal of the Society for Medical Decision Making 2016. DOI: 10/f87h2f.

211. Real FJ, Brown GC, Brown HC, and Brown MM. The Effect of Comorbidities upon Ocular and Systemic Health-Related Quality of Life. The British Journal of Ophthalmology 2008;92:770–4. DOI: 10/bb9sw8.

212. Solli O, Stavem K, and Kristiansen IS. Health-Related Quality of Life in Diabetes: The Associations of Complications with EQ-5D Scores. Health and Quality of Life Outcomes 2010;8:18. DOI: 10/d7bj6x.

213. Javanbakht M, Abolhasani F, Mashayekhi A, Baradaran HR, and Jahangiri noudeh Y. Health Related Quality of Life in Patients with Type 2 Diabetes Mellitus in Iran: A National Survey. PLoS ONE 2012;7:e44526. DOI: 10/gfzggd.

214. Tung TH, Chen SJ, Lee FL, Liu JH, Lin CH, and Chou P. A Community-Based Study for the Utility Values Associated with Diabetic Retinopathy among Type 2 Diabetics in Kinmen, Taiwan. Diabetes Research and Clinical Practice 2005;68:265–73. DOI: 10/ct58wf.

215. Hannula V, Hautala N, Sintonen H, and Falck A. Health-Related Quality of Life – Using the 15D Instrument – of Young Adults with Type 1 Diabetes since Childhood. Population-Based Oulu Cohort Study of Diabetic Retinopathy. Acta Ophthalmologica 2014;92:205–8. DOI: 10/f5xhxv.

216. Polack S, Alavi Y, Rachapalle Reddi S, Kulothungan V, and Kuper H. Utility Values Associated with Diabetic Retinopathy in Chennai, India. Ophthalmic Epidemiology 2015;22:20–7. DOI: 10/gfzgfh.

217. Smith D, Johnson E, Russell A, et al. PDB43 Associations among Visual Impairment, Stage of Diabetic Retinopathy and Quality of Life (EQ-5D). Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2005;8:365. DOI: 10/b27x7k.

218. Sullivan PW and Ghushchyan VH. EQ-5D Scores for Diabetes-Related Comorbidities. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2016. DOI: 10/f9wpcd.

219. Huang ES, Brown SES, Ewigman BG, Foley EC, and Meltzer DO. Patient Perceptions of Quality of Life with Diabetes-Related Complications and Treatments. Diabetes Care 2007;30:2478–83. DOI: 10/dx2xc4.

220. Sakamaki H, Ikeda S, Ikegami N, et al. Measurement of HRQL Using EQ-5D in Patients with Type 2 Diabetes Mellitus in Japan. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2006;9:47–53. DOI: 10/cs366v.

221. Loftus JV, Sultan MB, Pleil AM, and Macugen 1013 Study Group. Changes in Vision- and Health-Related Quality of Life in Patients with Diabetic Macular Edema Treated with Pegaptanib Sodium or Sham. Investigative Ophthalmology & Visual Science 2011;52:7498–505. DOI: 10/bpwdn8.

222. Knudsen MS, Thomas S, Gallagher M, and Mitchell P. PSS30 Assessment of Utility Loss From Diabetic Macular Edema Based on Restore Trial. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2011;14:A507. DOI: 10/cwpfwf.

223. Fenwick EK, Xie J, Ratcliffe J, et al. The Impact of Diabetic Retinopathy and Diabetic Macular Edema on Health-Related Quality of Life in Type 1 and Type 2 Diabetes. Investigative Ophthalmology & Visual Science 2012;53:677–84. DOI: 10.1167/iovs.11-8992.

224. Fenwick EK, Xie J, Pesudovs K, et al. Assessing Disutility Associated with Diabetic Retinopathy, Diabetic Macular Oedema and Associated Visual Impairment Using the Vision and Quality of Life Index. Clinical & Experimental Optometry: Journal of the Australian Optometrical Association 2012;95:362–70. DOI: 10.1111/j.1444-0938.2012.00742.x.

225. Venkataraman K, Wee HL, Leow MKS, et al. Associations between Complications and Health-Related Quality of Life in Individuals with Diabetes. Clinical Endocrinology 2013;78:865–73. DOI: 10/f4vxhz.

226. Lee BS, Kymes SM, Nease Jr RF, Sumner W, Siegfried CJ, and Gordon MO. The Impact of Anchor Point on Utilities for 5 Common Ophthalmic Diseases. Ophthalmology 2008;115:898–903.e4. DOI: 10/fq66tk.

227. Brown GC. Vision and Quality-of-Life. Transactions of the American Ophthalmological Society 1999;97:473–511. URL: https://www.ncbi.nlm.nih.gov/pubmed/10703139.

228. Szabo SM, Beusterien KM, Pleil AM, et al. Patient Preferences for Diabetic Retinopathy Health States. Investigative Ophthalmology & Visual Science 2010;51:3387–94. DOI: 10/cfxdk8.

229. Ohsawa I, Ishida T, Oshida Y, Yamanouchi K, and Sato Y. Subjective Health Values of Individuals with Diabetes in Japan: Comparison of Utility Values with the SF-36 Scores. Diabetes Research and Clinical Practice 2003;62:9–16. DOI: 10/ctcwhn.

230. Brown MM, Brown GC, Sharma S, Landy J, and Bakal J. Quality of Life with Visual Acuity Loss from Diabetic Retinopathy and Age-Related Macular Degeneration. Archives of Ophthalmology 2002;120:481–4.

231. Sharma S, Oliver-Fernandez A, Bakal J, Hollands H, Brown GC, and Brown MM. Utilities Associated with Diabetic Retinopathy: Results from a Canadian Sample. The British Journal of Ophthalmology 2003;87:259–61. DOI: 10/bd6mdn.

232. Brown GC, Brown MM, Sharma S, Brown H, Gozum M, and Denton P. Quality of Life Associated with Diabetes Mellitus in an Adult Population. Journal of Diabetes and Its Complications 2000;14:18–24. DOI: 10.1016/s1056-8727(00)00061-1.

233. Gonder JR, Walker VM, Barbeau M, et al. Costs and Quality of Life in Diabetic Macular Edema: Canadian Burden of Diabetic Macular Edema Observational Study (C-REALITY). Journal of Ophthalmology 2014;2014:939315. DOI: 10/gb6xtg.

234. Public Health England. Professional Briefing of KPI Data: Q3 (October to December) 2015 to 2016. 2016192. 2016.

235. Department of Health. Download Appendix NSRC04: NHS Trust and PCT Combined Reference Cost Schedules. 2011. URL: https://www.gov.uk/government/publications/nhs-reference-costs-2009-2010.

236. Department of Health. Download NSRC4 NHS Trusts and PCTs Combined Reference Cost Schedules 2010-11. 2011. URL: https://www.gov.uk/government/publications/2010-11-reference-costs-publication.

237. James M, Turner DA, Broadbent DM, Vora J, and Harding SP. Cost Effectiveness Analysis of Screening for Sight Threatening Diabetic Eye Disease. BMJ 2000;320:1627–31.

238. Yeo ST, Edwards RT, Luzio SD, et al. Diabetic Retinopathy Screening: Perspectives of People with Diabetes, Screening Intervals and Costs of Attending Screening. Diabetic Medicine: A Journal of the British Diabetic Association 2012;29:878–85. DOI: 10.1111/j.1464-5491.2012.03637.x.

239. Public Health England. Screening KPI Data: Q3 (October to December) 2015 to 2016. 2016. URL: https://www.gov.uk/government/publications/nhs-screening-programmes-kpi-reports-2015-to-2016.

240. James M and Sampson C. ISDR Visit Questionnaire. 2015. URL: http://dirum.org/instruments/details/90.

241. Ridyard CH and Hughes DA. Development of a Database of Instruments for Resource-Use Measurement: Purpose, Feasibility, and Design. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research 2012;15:650–5. DOI: 10.1016/j.jval.2012.03.004.

242. Curtis L and Burns A. Unit Costs of Health and Social Care 2017. 2017. DOI: 10.22024/UniKent/01.02/65559.

243. The Campbell and Cochrane Economics Methods Group. CCEMG - EPPI-Centre Cost Converter v.1.6. 2019. URL: https://eppi.ioe.ac.uk/costconversion/.

244. Phan ADT, Koczman JJ, Yung CW, Pernic AA, Doerr ED, and Kaehr MM. Cost Analysis of Teleretinal Screening for Diabetic Retinopathy in a County Hospital Population. Diabetes Care 2014;37:e252–e253. DOI: 10/gfxbc8.

245. Byrne MM, Parker DF, Tannenbaum SL, et al. Cost of a Community-Based Diabetic Retinopathy Screening Program. Diabetes Care 2014;37:e236–7. DOI: 10.2337/dc14-0834.

246. Garoon RB, Lin WV, Young AK, Yeh AG, Chu YI, and Weng CY. Cost Savings Analysis for a Diabetic Retinopathy Teleretinal Screening Program Using an Activity-Based Costing Approach. Ophthalmology Retina 2018;2:906–13. DOI: 10/gfxbdh.

247. Scotland G and Bryan S. Why Do Health Economists Promote Technology Adoption Rather Than the Search for Efficiency? A Proposal for a Change in Our Approach to Economic Evaluation in Health Care. Medical Decision Making: An International Journal of the Society for Medical Decision Making 2016. DOI: 10.1177/0272989X16653397.

248. Waqar S, Bullen G, Chant S, Salman R, Vaidya B, and Ling R. Cost Implications, Deprivation and Geodemographic Segmentation Analysis of Non-Attenders (DNA) in an Established Diabetic Retinopathy Screening Programme. Diabetes & Metabolic Syndrome: Clinical Research & Reviews 2012;6. 00022:199–202. DOI: 10/gfxbdc.

249. The Royal College of Ophthalmologists. Diabetic Retinopathy Guidelines. 2013.

250. Evans JR, Michelessi M, and Virgili G. Laser Photocoagulation for Proliferative Diabetic Retinopathy. In: *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd, 1996. DOI: 10.1002/14651858.CD011234.pub2.

251. Luo D, Zheng Z, Xu X, et al. Systematic Review of Various Laser Intervention Strategies for Proliferative Diabetic Retinopathy. Expert Review of Medical Devices 2015;12:83–91. DOI: 10.1586/17434440.2014.953057.

252. Virgili G, Parravano M, Menchini F, and Evans JR. Anti-Vascular Endothelial Growth Factor for Diabetic Macular Oedema. In: *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd, 1996. DOI: 10.1002/14651858.CD007419.pub4.

253. Grover DA, Li T, and Chong CCW. Intravitreal Steroids for Macular Edema in Diabetes. In: *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd, 1996. DOI: 10.1002/14651858.CD005656.pub2.

254. Martinez-Zapata MJ, Martí-Carvajal AJ, Solà I, et al. Anti-Vascular Endothelial Growth Factor for Proliferative Diabetic Retinopathy. In: *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd, 1996. DOI: 10.1002/14651858.CD008721.pub2.

255. Do DV, Wang X, Vedula SS, et al. Blood Pressure Control for Diabetic Retinopathy. In: *Cochrane Database of Systematic Reviews*. Vol. 1. John Wiley & Sons, Ltd, 2015:CD006127. DOI: 10.1002/14651858.CD006127.pub2.

256. NHS Digital and Department of Health. OPCS Classifications of Interventions and Procedures. The Stationery Office, 2006.

257. Joint Formulary Committee. British National Formulary (Online). London: BMJ Group, Pharmaceutical Press, and RCPCH Publications, 2018.

258. Stewart MW. Treatment of Diabetic Retinopathy: Recent Advances and Unresolved Challenges. World Journal of Diabetes 2016;7:333–41. DOI: 10.4239/wjd.v7.i16.333.

259. Virgili G, Parravano M, Evans JR, Gordon I, and Lucenteforte E. Anti-vascular Endothelial Growth Factor for Diabetic Macular Oedema: A Network Meta-analysis. Cochrane Database of Systematic Reviews 2018. DOI: 10.1002/14651858.CD007419.pub6.

260. Simunovic MP and Maberley DAL. Anti-Vascular Endothelial Growth Factor Therapy for Proliferative Diabetic Retinopathy: A Systematic Review and Meta-Analysis. Retina 2015;35:1931–42. DOI: 10.1097/IAE.0000000000000723.

261. Li X, Zarbin MA, and Bhagat N. Anti-Vascular Endothelial Growth Factor Injections: The New Standard of Care in Proliferative Diabetic Retinopathy? Management of Diabetic Retinopathy 2017;60:131–42. DOI: 10.1159/000459699.

262. Cheung N, Cheung CMG, Talks SJ, and Wong TY. Management of Diabetic Macular Oedema: New Insights and Global Implications of DRCR Protocol V. Eye 2019:1–4. DOI: 10.1038/s41433-019-0738-y.

263. Parikh R, Ross JS, Sangaralingham LR, Adelman RA, Shah ND, and Barkmeier AJ. Trends of Anti-Vascular Endothelial Growth Factor Use in Ophthalmology Among Privately Insured and Medicare Advantage Patients. Ophthalmology 2017;124:352–8. DOI: 10.1016/j.ophtha.2016.10.036.

264. Prescott G, Sharp P, Goatman K, et al. Improving the Cost-Effectiveness of Photographic Screening for Diabetic Macular Oedema: A Prospective, Multi-Centre, UK Study. The British Journal of Ophthalmology 2014;98:1042–9. DOI: 10.1136/bjophthalmol-2013-304338.

265. Lian J, McGhee SM, Gangwani RA, et al. The Impact of a Co-Payment on the Cost-Effectiveness of Screening for Diabetic Retinopathy. Journal of Public Health 2015:fdv168. URL: http://jpubhealth.oxfordjournals.org/content/early/2015/11/14/pubmed.fdv168.abstract.

266. Kirkizlar E, Serban N, Sisson JA, Swann JL, Barnes CS, and Williams MD. Evaluation of Telemedicine for Screening of Diabetic Retinopathy in the Veterans Health Administration. Ophthalmology 2013;120:2604–10. DOI: 10.1016/j.ophtha.2013.06.029.

267. Public Health England. Diabetes Prevalence Model. 2016243. London, 2016. URL: https://www.gov.uk/government/publications/diabetes-prevalence-estimates-for-local-populations (visited on 08/25/2019).

268. Klein R, Knudtson MD, Lee KE, Gangnon R, and Klein BEK. The Wisconsin Epidemiologic Study of Diabetic Retinopathy: XXII the Twenty-Five-Year Progression of Retinopathy in Persons with Type 1 Diabetes. Ophthalmology 2008;115:1859–68. DOI: 10/dcvsmm.

269. Early Treatment Diabetic Retinopathy Study Research Group. Fundus Photographic Risk Factors for Progression of Diabetic Retinopathy: ETDRS Report Number 12. Ophthalmology 5, Supplement 1991;98:823–33. DOI: 10.1016/S0161-6420(13)38014-2.

270. Early Treatment Diabetic Retinopathy Study Research Group. Early Photocoagulation for Diabetic Retinopathy: ETDRS Report Number 9. Ophthalmology 5, Supplement 1991;98:766–85. DOI: 10.1016/s0161-6420(13)38011-7.

271. NHS Digital. National Diabetes Audit, 2015-16 Report 2a: Complications and Mortality (Complications of Diabetes). 2017. URL: https://files.digital.nhs.uk/pdf/4/t/national_diabetes_audit__2015-16__report_2a.pdf (visited on 08/28/2019).

272. Trautner C, Icks A, Haastert B, Plum F, Berger M, and Giani G. Diabetes as a Predictor of Mortality in a Cohort of Blind Subjects. International Journal of Epidemiology 1996;25. 00013:1038–43. DOI: 10.1093/ije/25.5.1038.

273. Scanlon PH, Malhotra R, Thomas G, et al. The Effectiveness of Screening for Diabetic Retinopathy by Digital Imaging Photography and Technician Ophthalmoscopy. Diabetic Medicine: A Journal of the British Diabetic Association 2003;20:467–74.

274. Prasad S, Kamath GG, Jones K, Clearkin LG, and Phillips RP. Effectiveness of Optometrist Screening for Diabetic Retinopathy Using Slit-Lamp Biomicroscopy. Eye Pt 5 2001;15:595–601. DOI: 10/d9wgcb.

275. Early Treatment Diabetic Retinopathy Study Research Group. Photocoagulation for Diabetic Macular Edema. Early Treatment Diabetic Retinopathy Study Report Number 1. Early Treatment Diabetic Retinopathy Study

Research Group. Archives of Ophthalmology 1985;103:1796–806. DOI: 10.1001/archopht.1985.01050120030015.

276. Ip MS, Domalpally A, Hopkins JJ, Wong P, and Ehrlich JS. Long-Term Effects of Ranibizumab on Diabetic Retinopathy Severity and Progression. Archives of Ophthalmology 2012;130. 00167:1145–52. DOI: 10.1001/archophthalmol.2012.1043.

277. Varma R, Bressler NM, Doan QV, et al. Visual Impairment and Blindness Avoided with Ranibizumab in Hispanic and Non-Hispanic Whites with Diabetic Macular Edema in the United States. Ophthalmology 2015;122. 00013:982–9. DOI: 10.1016/j.ophtha.2014.12.007.

278. Meads C and Hyde C. What Is the Cost of Blindness? The British Journal of Ophthalmology 2003;87:1201–4. URL: https://www.ncbi.nlm.nih.gov/pubmed/14507746.

279. Appelbe D, Broadbent D, Mobayen-Rahni M, et al. Introducing Personalised Risk Based Intervals in Screening for Diabetic Retinopathy: Development, Implementation and Assessment of Safety, Cost-Effectiveness and Patient Experience (ISDR): A Case Study in the Use of Automated Systems in Trials. Trials Suppl 2 2015;16:O59. DOI: 10.1186/1745-6215-16-S2-O59.

280. Public Health England. Evidence Review Criteria: National Screening Programmes. 2015. URL: https://www.gov.uk/government/publications/evidence-review-criteria-national-screening-programmes/criteria-for-appraising-the-viability-effectiveness-and-appropriateness-of-a-screening-programme.

281. Stinnett AA and Mullahy J. Net Health Benefits: A New Framework for the Analysis of Uncertainty in Cost-Effectiveness Analysis. Medical Decision Making: An International Journal of the Society for Medical Decision Making 2 Suppl 1998;18:169–79.

282. Tambour M, Zethraeus N, and Johannesson M. A Note on Confidence Intervals in Cost-Effectiveness Analysis. International Journal of Technology Assessment in Health Care 1998;14:467–71. DOI: 10/d9swqw.

283. Gafni A and Birch S. Incremental Cost-Effectiveness Ratios (ICERs): The Silence of the Lambda. Social Science & Medicine 2006;62:2091–100. DOI: 10.1016/j.socscimed.2005.10.023.

284. Culyer T. Introduction: Ought NICE to Have a Cost-Effectiveness Threshold. Cost-Effectiveness Thresholds: Economic and Ethical 2002.

285. Porta M, Rizzitiello A, Tomalino M, et al. Comparison of the Cost-Effectiveness of Three Approaches to Screening for and Treating Sight-Threatening Diabetic Retinopathy. Diabetes & Metabolism 1999;25:44–53.

286. Coyle D, Buxton MJ, and O'Brien BJ. Stratified Cost-Effectiveness Analysis: A Framework for Establishing Efficient Limited Use Criteria. Health Economics 2003;12:421–7. DOI: 10.1002/hec.788.

287. Cao Q, Buskens E, Hillege HL, Jaarsma T, Postma M, and Postmus D. Stratified Treatment Recommendation or One-Size-Fits-All? A Health Economic Insight Based on Graphical Exploration. The European Journal of Health Economics: HEPAC: Health Economics in Prevention and Care 2019;20:475–82. DOI: 10.1007/s10198-018-1013-z.

288. Lavelle TA, Kent DM, Lundquist CM, et al. Patient Variability Seldom Assessed in Cost-Effectiveness Studies. Medical Decision Making: An International Journal of the Society for Medical Decision Making 2018;38:487–94. DOI: 10.1177/0272989X17746989.

289. Vervaart MAM, Burger EA, and Aas E. Acknowledging Patient Heterogeneity in Colorectal Cancer Screening: An Example from Norway. Nordic Journal of Health Economics 2018;6:83-98–98. DOI: 10.5617/njhe.4881.

290. Dan YY, Chuah BYS, Koh DCS, and Yeoh KG. Screening Based on Risk for Colorectal Cancer Is the Most Cost-Effective Approach. Clinical Gastroenterology and Hepatology: The Official Clinical Practice Journal of the American Gastroenterological Association 2012;10:266–71.e1–6. DOI: 10.1016/j.cgh.2011.11.011.

291. Lansdorp-Vogelaar I, van Ballegooijen M, Zauber AG, et al. Individualizing Colonoscopy Screening by Sex and Race. Gastrointestinal Endoscopy 2009;70:96–108, 108.e1–24. DOI: 10/d35fqt.

292. Round JA, Jacklin P, Fraser RB, Hughes RG, Mugglestone MA, and Holt RIG. Screening for Gestational Diabetes Mellitus: Cost-Utility of Different Screening Strategies Based on a Woman's Individual Risk of Disease. Diabetologia 2011;54:256–63. DOI: 10/ckcr52.

293. Aus G, Damber JE, Khatami A, Lilja H, Stranne J, and Hugosson J. Individualized Screening Interval for Prostate Cancer Based on Prostate-Specific Antigen Level: Results of a Prospective, Randomized, Population-Based Study. Archives of Internal Medicine 2005;165:1857–61. DOI: 10.1001/archinte.165.16.1857.

294. Vickers A, Bennette C, Steineck G, et al. Individualized Estimation of the Benefit of Radical Prostatectomy from the Scandinavian Prostate Cancer Group Randomized Trial. European Urology 2012;62:204–9. DOI: 10/f2kwnn.

295. Gavan SP, Thompson AJ, and Payne K. The Economic Case for Precision Medicine. Expert Review of Precision Medicine and Drug Development 2018;3. 00016:1–9. DOI: 10.1080/23808993.2018.1421858.

296. Karnon J and Qizilbash N. Economic Evaluation alongside N-of-1 Trials: Getting Closer to the Margin. Health Economics 2001. 00018. DOI: 10.1002/1099-1050(200101)10:1<79::AID-HEC567>3.0.CO;2-Z.

297. Bala MV and Zarkin GA. Pharmacogenomics and the Evolution of Healthcare. PharmacoEconomics 2004;22:495–8. DOI: 10.2165/00019053-200422080-00002.

298. Bala MV and Zarkin GA. On Pharmacogenomics and Cost-Effectiveness Analysis at the Individual Level. PharmacoEconomics 2005;23:527–7. DOI: 10.2165/00019053-200523050-00011.

299. Holland PW. Statistics and Causal Inference. Journal of the American Statistical Association 1986;81:945–60. DOI: 10/gfzgfs.

300. Basu A. Economics of Individualization in Comparative Effectiveness Research and a Basis for a Patient-Centered Health Care. Journal of Health Economics 2011;30:549–59. DOI: 10.1016/j.jhealeco.2011.03.004.

301. Hoch JS, Briggs AH, and Willan AR. Something Old, Something New, Something Borrowed, Something Blue: A Framework for the Marriage of Health Econometrics and Cost-Effectiveness Analysis. Health Economics 2002;11:415–30. DOI: 10.1002/hec.678.

302. Basu A. Estimating Person-Centered Treatment (PeT) Effects Using Instrumental Variables: An Application to Evaluating Prostate Cancer Treatments. Journal of Applied Econometrics 2014;29:671–91. DOI: 10.1002/jae.2343.

303. Basu A, Jena AB, Goldman DP, Philipson TJ, and Dubois R. Heterogeneity in Action: The Role of Passive Personalization in Comparative Effectiveness Research. Health Economics 2014;23:359–73. DOI: 10.1002/hec.2996.

304. Pauker SG and Kassirer JP. The Threshold Approach to Clinical Decision Making. The New England Journal of Medicine 1980;302:1109–17. DOI: 10/bsxjqn.

305. Kawachi I. QALYs and Justice. Health Policy 1989;13:115–20. DOI: 10.1016/0168-8510(89)90066-3.

306. Cubbon J. The Principle of QALY Maximisation as the Basis for Allocating Health Care Resources. Journal of Medical Ethics 1991;17:181–4. DOI: 10.1136/jme.17.4.181.

307. HM Treasury. The Green Book. 00047. 2018. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685903/The_Green_Book.pdf (visited on 08/16/2019).

308. Roobol MJ and Carlsson SV. Risk Stratification in Prostate Cancer Screening. Nature Reviews. Urology 2013;10:38–48. DOI: 10/f4gsg9.

309. Degeling C, Rock M, and Rogers WA. Testing Relationships: Ethical Arguments for Screening for Type 2 Diabetes Mellitus with HbA1C. Journal of Medical Ethics 2012;38:180–3. DOI: 10.1136/medethics-2011-100086.

310. Brownsword R and Earnshaw JJ. The Ethics of Screening for Abdominal Aortic Aneurysm in Men. Journal of Medical Ethics 2010;36:827–30. DOI: 10.1136/jme.2010.035519.

311. Culyer A. Need–Is a Consensus Possible? Journal of Medical Ethics 1998;24:77–80. DOI: 10.1136/jme.24.2.77.

312. Culyer AJ. Need: An Instrumental View. In: *Principles of Health Care Ethics*. John Wiley & Sons, Ltd, 2006:231–8. DOI: 10.1002/9780470510544.ch30.

313. Rose G. Sick Individuals and Sick Populations. International Journal of Epidemiology 1985;14:32–8. DOI: 10/cjtjm4.

314. Hadorn DC. The Problem of Discrimination in Health Care Priority Setting. JAMA: The Journal of the American Medical Association 1992;268:1454–9. DOI: 10.1001/jama.1992.03490110092037.

315. Cookson R. Justice and the NICE Approach. Journal of Medical Ethics 2015;41:99–102. DOI: 10.1136/medethics-2014-102386.

316. Daniels N. Justice between Adjacent Generations: Further Thoughts. The Journal of Political Philosophy 2008;16:475–94. DOI: 10.1111/j.1467-9760.2008.00320.x.

317. Daniels N. Justice between Age Groups: Am I My Parents' Keeper? The Milbank Memorial Fund Quarterly. Health and Society 1983;61:489–522. DOI: 10.2307/3349870.

318. Walker T. Ageing, Justice and Resource Allocation. Journal of Medical Ethics 2016. DOI: 10/f8qqwh.

319. Daniels N. Just Health: Meeting Health Needs Fairly. Vol. 0. Cambridge University Press, 2007. 397 pp.

320. Braithwaite RS, Stevens ER, and Caplan A. Is Risk Stratification Ever the Same as 'Profiling'? Journal of Medical Ethics 2016. DOI: 10.1136/medethics-2015-103047.

321. Rumbold B. On Engster's Care-Justification of the Specialness Thesis about Healthcare. Journal of Medical Ethics 2016:medethics–2015–102799. DOI: 10/gfzgf2.

322. Hall AE, Chowdhury S, Pashayan N, Hallowell N, Pharoah P, and Burton H. What Ethical and Legal Principles Should Guide the Genotyping of Children as Part of a Personalised Screening Programme for Common Cancer? Journal of Medical Ethics 2014;40:163–7. DOI: 10/f5t8nk.

323. Petrou S, Kwon J, and Madan J. A Practical Guide to Conducting a Systematic Review and Meta-Analysis of Health State Utility Values. PharmacoEconomics 2018;36:1043–61. DOI: 10.1007/s40273-018-0670-1.

324. Aceituno D, Pennington M, Iruretagoyena B, Prina MA, and McCrone P. Health State Utility Values in Schizophrenia: Protocol for a Systematic Review and Meta-Analysis. Evidence-Based Mental Health 2019. DOI: 10.1136/ebmental-2019-300089.

325. Afzali HHA, Bojke L, and Karnon J. Improving Decision-Making Processes in Health: Is It Time for (Disease-Specific) Reference Models? Applied Health Economics and Health Policy 2019:1–4. DOI: 10.1007/s40258-019-00510-6.

326. Chin MH, Drum ML, Jin L, Shook ME, Huang ES, and Meltzer DO. Variation in Treatment Preferences and Care Goals among Older Patients with Diabetes and Their Physicians. Medical Care 2008;46:275–86. DOI: 10.1097/mlr.0b013e318158af40.

327. Coffey JT, Brandle M, Zhou H, et al. Valuing Health-Related Quality of Life in Diabetes. Diabetes Care 2002;25:2238–43. DOI: 10.2337/diacare.25.12.2238.

328. Da Mata AR, Álvares J, Diniz LM, et al. Quality of Life of Patients with Diabetes Mellitus Types 1 and 2 from a Referal Health Centre in Minas Gerais, Brazil. Expert Review of Clinical Pharmacology 2016;9:739–46. DOI: 10.1586/17512433.2016.1152180.

329. Godshalk AN, Brown GC, Brown HC, and Brown MM. The Power of Hope: Being a Doctor Is More than Relying Solely on the Numbers. The British Journal of Ophthalmology 2008;92:783–7. DOI: 10.1136/bjo.2008.141663.

330. Huang ES, Shook M, Jin L, Chin MH, and Meltzer DO. The Impact of Patient Preferences on the Cost-Effectiveness of Intensive Glucose Control in Older Patients with New-Onset Diabetes. Diabetes Care 2006;29:259–64. DOI: 10/fn3mkx.

331. Kontodimopoulos N, Pappa E, Chadjiapostolou Z, Arvanitaki E, Papadopoulos AA, and Niakas D. Comparing the Sensitivity of EQ-5D, SF-6D and 15D Utilities to the Specific Effect of Diabetic Complications. The European Journal of Health Economics: HEPAC: Health Economics in Prevention and Care 2010;13:111–20. DOI: 10/b8x473.

332. Kontodimopoulos N, Pappa E, Tadros C, and Niakas D. Dimensional Sensitivity of the 15D Multiattribute Utility Instrument in Diabetic Retinopathy. Applied Research in Quality of Life 2013;9:413–27. DOI: 10/f5zx3w.

333. Lloyd AJ, Loftus J, Turner M, Lai G, and Pleil A. Psychometric Validation of the Visual Function Questionnaire-25 in Patients with Diabetic Macular Edema. Health and Quality of Life Outcomes 2013;11:10. DOI: 10/f45w7d.

334. Morgan CL, McEwan P, Morrissey M, Peters JR, Poole C, and Currie CJ. Characterization and Comparison of Health-Related Utility in People with Diabetes with Various Single and Multiple Vascular Complications. Diabetic Medicine: A Journal of the British Diabetic Association 2006;23:1100–5. DOI: 10/bs4vrf.

335. Papadopoulos AA, Kontodimopoulos N, and Niaka D. Clinical Determinants of Health-Related Quality of Life in an Elderly Type 2 Diabetes Mellitus Population. In: 2009.

336. Sakthong P, Charoenvisuthiwongs R, and Shabunthom R. A Comparison of EQ-5D Index Scores Using the UK, US, and Japan Preference Weights in a Thai Sample with Type 2 Diabetes. Health and Quality of Life Outcomes 2008;6:71. DOI: 10/b4j2ck.

337. Yeo ST, Edwards RT, Fargher EA, Luzio SD, Thomas RL, and Owens DR. Preferences of People with Diabetes for Diabetic Retinopathy Screening: A Discrete Choice Experiment. Diabetic Medicine: A Journal of the British Diabetic Association 2012;29:869–77. DOI: 10/gfzgf4.

Appendices

Appendix A

Model review: structures

Study and country	Comparators	Perspective	Model type	States / events	Time horizon	Cycle length	Software
Javitt et al. (1989) [106] USA	Screening, follow-up and treatment as recommended; no screening or treatment	Government	ISM	Background DR; DMO; PDR; SVL; CAL	Lifetime	1 year	Turbo Pascal
Dasbach et al. (1991) [65] USA	Alternative screening modalities (ophthalmoscopy; nonmydriatic fundus photography; mydriatic fundus photography) at alternative intervals (biannual* ; annual)	Government	CST	Low-risk DR; high-risk DR; treated; blind	Lifetime	1 year	NR
Fendrick et al. (1992) [87] Sweden	Annual screening; no screening	NR	ISM	Background DR; DMO; PDR; SVL; CAL	60 years	1 year	NR

Sculpher et al. (1992) [114] UK	GP ophthalmoscopy; ophthalmic optician ophthalmoscopy; non-mydriatic photography; ophthalmoscopy + non-mydriatic photography; selective referral of high-risk people	Societal	DT	Screening outcome (true/false positive/negative)	NR	NA	NR
Javitt and Aiello (1996) [118] USA	Systematic screening and treatment; no screening	Health insurer	ISM	Background DR; DMO; PDR; SVL; CAL/blind	Lifetime	2 months	NR
Wu et al. (1998) [90] USA	Intensive treatment; conventional treatment	HMO	CST	No DR; any DR	10 years	1 year	Microsoft Excel

Crijns et al. (1999) [91] The Netherlands	Alternative screening frequencies	NR	CST	No DR; DR excluding both DMO and PDR; PDR; adequate vision (<20/40); poor central and/or peripheral vision (20/40 or worse, at least 20/200); blindness (less than or equal to 20/200)	Lifetime	3 months	NR
Palmer et al. (2000) [92] Switzerland	Conventional/intensive insulin therapy with no screening; annual screening with treatment	Health insurer	CST	No DR; background DR; PDR; blind	Lifetime	1 year	IMIB TOM
Sharma et al. (2000) [93] USA	Grid laser therapy; no laser	Insurer	DT	Treatment; complications	40 years	NA	TreeAge
Vijan et al. (2000) [67] USA	Alternative screening intervals	Third party payer	CST	No DR; DR1; DR2; DR3; PDR; DMO; blind	Lifetime	NR	NR
Sharma et al. (2001) [110] USA	Early or deferred vitrectomy	Insurer	CST	Visual acuity (G1–G5)	Lifetime (55 years)	1 year	TreeAge

Polak et al. (2003) [96] The Netherlands	Standard/intensive glycaemic control; alternative screening intervals	NR	CST	NR	Lifetime	3 months	NR
Whited et al. (2005) [97] USA	Digital teleophthalmology; ophthalmoscopy	Government	DT	PDR; high-risk PDR; PRP; SVL	1 year	NA	TreeAge
Scotland et al. (2007) [98] Scotland	Manual grading; automated grading	NHS	DT	No/mild DR; observable DR; technical failures	1 year	NA	TreeAge
Scotland et al. (2010) [108] Scotland	Manual grading; automated grading	NR	DT	NR	20 years	NA	NR
Dewan et al. (2012) [100] USA	Sham injection + laser; ranibizumab + prompt laser; ranibizumab + deferred laser; triamcinolone + deferred laser	Payer	ISM	NR	10 years	1 month	TreeAge

Mitchell et al. (2012) [99] UK	Laser; ranibizumab; laser + ranibizumab	Healthcare payer	CST	BCVA <26 letters; BCVA 26–35; BCVA 36-45; BCVA 46–55; BCVA 56–65; BCVA 66–75; BCVA 76–85; BCVA 86–100	15 years	3 months	NR
Rachapelle et al. (2013) [113] India	No screening; alternative screening frequencies (once-in-a-lifetime; twice-in-a-lifetime; 5-yearly; 3-yearly; 2-yearly; annual)	Provider; societal	CST	No DR; non-STDR; STDR; CSMO; Blind from DR	25 years	1 year	TreeAge
Stein et al. (2013) [117] USA	FLP; FLP + ranibizumab; FLP + bevacizumab; FLP + triamcinolone	NR	CST	20/25; 20/32–20/40; 20/50–20/63; 20/80–20/100; 20/125–20/160; 20/200	25 years	1 year	TreeAge
Brady et al. (2014) [101] USA	Tele-ophthalmology screening; no screening	Payer	DT	PDR; no PDR; no treatment; PRP; vitrectomy with laser; vitrectomy with membrane peel	NR	NA	TreeAge

Pershing et al. (2014) [111] USA	Laser; intraocular triamcinolone; anti-VEGF; laser + triamcinolone; laser + anti-VEGF	Societal	CST	VA 1-6	Lifetime	1 month	TreeAge
Kawasaki et al. (2015) [107] Japan	Screening programme; no screening programme	Payer	CST	NPDR; severe NPDR; PDR; high-risk PDR; CSMO (high VA; low VA); stabilised DR (high VA; low VA); blind	Lifetime (50 years)	1 year	TreeAge
Royle et al. (2015) [109] UK	Early PRP for NPDR; current practice	NHS + PSS	CST	Moderate NPDR; severe NPDR; early PDR; HR-PDR; severe PDR; (with/without CSMO; with/without treatment); SVI/blindness	30 years	6 months	Microsoft Excel

Scanlon et al. (2015) [30] UK	Alternative screening recall periods (6; 12; 24; 36; 60 months)	NHS + PSS	CST	R0M0|R0M0; R1M0|R0M0; R1M0|R1M0; R2/3M0|R0/1M0; R2/3M0|R2/3M0; RxM0|RxM1; RxM1|RxM1	Lifetime	6 months	Microsoft Excel
Wolowacz et al. (2015) [103] UK	Hypothetical	NHS + PSS	ISM	PDR; blind	50 years	1 year	Microsoft Excel
Wu et al. (2015) [104] China	No screening; alternative screening intervals (1–5 years)	Healthcare	DES	No DR; NPDR; PDR; DMO; blind	Lifetime (100 years)	Continuous	NR

Table A.1: Details of the modelling structures. DT = decision tree; CST = cohort state transition; ISM = individual sampling model; DES = discrete event simulation. * Dasbach et al. [65] purport to evaluate biannual screening, but the results imply that they evaluated biennial screening

Appendix B

Model review: analyses

Study	Disease progression	Costs / resource use	Health outcomes	Uncertainty assessment	Validation
Javitt et al. (1989) [106] USA	WESDR	Medicare reimbursement	NA	One-way	NR
Dasbach et al. (1991) [65] USA	DRS; WESDR	Local data	NA	One-way	NR
Fendrick et al. (1992) [87] Sweden	DRS; ETDRS; WESDR	NR	NA	One-way	NR
Sculpher et al. (1992) [114] UK	Local data	NHS reimbursement	NA	One-way	NR
Javitt and Aiello (1996) [118] USA	WESDR	Medicare reimbursement	NR	One-way	NR
Wu et al. (1998) [90] USA	DCCT	Local HMO	Local data with mapping	NR	External
Crijns et al. (1999) [91] The Netherlands	WESDR	"official medical charges"	NA	NR	NR
Palmer et al. (2000) [92] Switzerland	DCCT	Swiss tariffs	NA	One-way	NR
Sharma et al. (2000) [93] USA	ETDRS	Medicare reimbursement	Literature (single study)	One-way	NR
Vijan et al. (2000) [67] USA	DRS; ETDRS	Medicare reimbursement	Literature (blindness only, single study)	One-way; PSA	NR

Sharma et al. (2001) [110] USA	DRVS	Medicare reimbursement	Literature (single study)	One-way	NR
Polak et al. (2003) [96] The Netherlands	WESDR	"medical charges"	NA	NR	NR
Whited et al. (2005) [97] USA	Literature	Assumed	NA	One-way; PSA	NR
Scotland et al. (2007) [98] Scotland	Local data	Assumed	NA	One-way; PSA	NR
Scotland et al. (2010) [108] Scotland	Local data	Assumed	Literature	PSA	NR
Dewan et al. (2012) [100] USA	Local data	Medicare	NA	One-way	External
Mitchell et al. (2012) [99] UK	WESDR	Reference costs	Trial	One-way; PSA	NR
Rachapelle et al. (2013) [113] India	Literature	Local data	Local TTO	One-way; PSA	NR
Stein et al. (2013) [117] USA	DRCRnet	Medicare reimbursement and expenditures	Literature (single study) [148]	One-way; PSA	NR
Brady et al. (2014) [101] USA	Local data	Medicare fee schedule	NA	One-way SA; PSA	NR

Pershing et al. (2014) [111] USA	Expert opinion	Medicare reimbursement and expenditures	Literature (single study)	One-way; PSA	None (calibration of disease progression)
Kawasaki et al. (2015) [107] Japan	Literature	NR	Literature (single study)	One-way; PSA	External
Royle et al. (2015) [109] UK	ETDRS	Expert opinion; reference costs; literature	literature	Multi-way; PSA	NR
Scanlon et al. (2015) [30] UK	Local data	Local data	Published literature (single study)	One-way; PSA	NR
Wolowacz et al. (2015) [103] UK	DCCT/EDIC; ETDRS	NHS Reference Costs 2008-9 + assumptions	CORE	PSA	Internal and external
Wu et al. (2015) [104] China	Prevalence calibration	Local data	Literature	One-way, two-way; PSA	Internal calibration with epidemiological data

Table B.1: Details of the data sources, uncertainty analysis and validation

Appendix C

HSUV review: extraction form

Study characteristics

Fill in this part of the form once for each study

*Required

1. **First author surname** *

2. **Publication year** *

3. **Article title** *

4. **Publication name** *
   e.g. "Diabetologia"

5. **Study design**
   Mark only one oval.
   - Clinical decision analysis (e.g. RCT)
   - Outcomes study
   - Other:

6. **Interventions / comparators**
   if appropriate

7. **Study sample size**
   Total across all groups

8. **Inclusion / exclusion criteria**

9. **Response rates**

10. **Loss to follow-up**

11. **Missing data**
    What percentage of data were missing and how did the author(s) address this?

12. **Total number of HSUVs reported in paper**

HSUV specifics

Fill in this part of the form once for each separately reported HSUV

13. **Reported HSUV point estimate type**
    Mark only one oval.

    - ◯ Mean
    - ◯ Median

14. **Reported HSUV** *

15. **Reported measure of variance type**
    Mark only one oval.

    - ◯ Standard deviation
    - ◯ Variance
    - ◯ Other:

16. **Reported variance statistic**

17. **Retinopathy state** *
    as described in the study

18. **Maculopathy state**
    as described in the study

19. **Grading system**
    The retinopathy grading system to which this HSUV relates
    Mark only one oval.

- NHS Diabetic Eye Screening Programme (NDESP)
- American Academy of Ophthalmology (AAO)
- Early Treatment Diabetic Retinopathy Study (ETDRS)
- Liverpool Diabetic Eye Study (LDES)
- Scottish Diabetic Retinopathy Grading Scheme (SDRGS)
- Royal College of Ophthalmologists (RCO)
- Other:

20. **Visual function measurement method**
    Mark only one oval.

- LogMAR
- Snellen
- Other:

21. **Visual acuity/function level**

22. **Sample size**
    For reported HSUV

23. **Sample country**

24. **Sample age range**

25. **Other sample specifics**

26. **Valuation method** *

Direct or indirect methods

Mark only one oval.

- ◯ Standard gamble
- ◯ Time trade-off
- ◯ Person trade-off
- ◯ Discrete choice experiment
- ◯ EQ-5D
- ◯ HUI3
- ◯ SF-6D
- ◯ 15D
- ◯ Mapping algorithm
- ◯ Visual analogue scale
- ◯ Other: _____

27. **Valuation source**

Mark only one oval.

- ◯ Patients
- ◯ Public
- ◯ Other: _____

28. **Cooper rank**

Refer to Cooper et al (2005)

Mark only one oval.

- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6

29. **Value set country**

_____

30. **Upper anchor**

Mark only one oval.

- ◯ "full health"
- ◯ Other: _____

31. **Lower anchor**

Mark only one oval.

- ◯ "worst health state imaginable"
- ◯ "dead"
- ◯ Other: _____

32. **Administration method**

Select all that apply

Tick all that apply.

- [ ] Face-to-face interview
- [ ] Telephone
- [ ] Web
- [ ] Self-complete
- [ ] Proxy
- [ ] Other: _____

33. **Study arm**

if applicable

Mark only one oval.

- ◯ Treatment
- ◯ Control
- ◯ Other: _____

34. **Treatment status**

if applicable

Mark only one oval.

- ◯ Pre-treatment
- ◯ Post-treatment
- ◯ Other: _____

35. **Mapped R0M0 value**

Mark only one oval.

- ◯ R0M0
- ◯ R0M1
- ◯ R1M0
- ◯ R1M1
- ◯ R2M0
- ◯ R2M1
- ◯ R3M0
- ◯ R3M1

Appendix D

HSUV review: study details

Citation	Country	Sample size	HSUVs reported	Retinopathy states	Vision states	Valuation methods
Brown (1999) [227]	USA	325	5	Any DR	BCVA (worse eye): $\leq$ 20/40, BCVA (better eye): 86–20/25, 20/30–20/50, 20/70, 20/200–20/400	TTO
Brown et al. (1999) [147]	USA	100	10	Any DR	BCVA (better eye): 20/20–25/25, 20/30–20/50, 20/60–20/100, 20/200–20/400, CF-HM	TTO, SG
Brown et al. (2000) [232]	USA	310	2	Any DR	BCVA (better eye): 20/20–20/25, $\leq$ 20/30	TTO
Brown et al. (2002) [230]	USA	617	4	Any DR	BCVA (better eye): 20/20–25/25, 20/30–20/40, 20/50–20/100, $\leq$ 20/200	TTO
Chin et al. (2008) [326]	USA	473	1	NA	Blind (both eyes)	TTO
Coffey et al. (2002) [327]	USA	2,048	4	NA	Unilateral blindness, bilateral blindness	QWB-SA

da Mata et al. (2016) [328]	Brazil	346	2	No DR, any DR	NA	EQ-5D index
Fenwick et al. (2012) [223]	Australia	577	6	ETDRS / AAO (worse eye): 13-15 / 10/20, 20 / 30, 31-41 / 40	$\geq$ 6/12, $\leq$ 6/12	EQ-5D index
Fenwick et al. (2012) [224]	Australia	203	9	ETDRS / AAO (worse eye): 13-15 / 10/20, 20 / 30, 31-41 / 40, 51 60-80 / 50	BCVA (better eye): $\geq$ 0.18, 0.18–0.3, 0.3–0.48, 0.48–0.78, $\leq$ 0.78	VisQoL index
Godshalk et al. (2008) [329]	USA	247	3	Any DR	NA	TTO
Gonder et al. (2014) [233]	Canada	145	6	DMO	BCVA (worse eye): 20/10–20/80, 20/80–20/200, $\leq$ 20/200	EQ-5D index, EQ VAS
Hannula et al. (2014) [215]	Finland	216	2	No or NPDR, PDR	NA	15D
Heintz et al. (2012) [152]	Sweden	152	108	No DR, BR, PDR, DMO, blind	BCVA (worse eye): 20/10–20/25, 20/32–20/63, 20/80–20/160, $\leq$ 20/200	HUI3, EQ-5D index, EQ VAS, TTO
Huang et al. (2006) [330]	USA	519	1	NA	Both eyes: blind	TTO

Huang et al. (2007) [219]	USA	701	2	Symptomatic, blind	Worse eye: blind	TTO
Javanbakht et al. (2012) [213]	Iran	3,472	4	No DR, any DR	NA	EQ-5D index, EQ VAS
Knudsen et al. (2011) [222]	NR	NR	2	DMO	BCVA (worse eye): 100–76, 25–0	NR
Kontodimopoulos et al. (2010) [331]	Greece	319	6	No DR, any DR	NA	EQ-5D index, SF-6D index, 15D
Kontodimopoulos et al. (2013) [332]	Greece	85	2	NPDR, PDR	NA	15D
Lee et al. (2008) [226]	USA	434	8	NA	perfect vision, unilateral blindness, bilateral blindness	SG
Lloyd et al. (2008) [150]	UK	321	33	No DR, any DR	BCVA (better eye): 6/6–6/9, 6/12–6/18, 6/24–6/36, 6/60–6/120, CF-HM	SG, EQ-5D index, EQ VAS, HUI3

Lloyd et al. (2013) [333]	Multinational	235	4	DMO	NA	EQ-5D index, EQ VAS
Loftus et al. (2011) [221]	Multinational	260	6	Any DM	NA	EQ-5D index
Mitchell et al. (2012) [99]	Multinational	345	8	DMO	BCVA (better eye): 100-86, 85-76, 75-66, 65-56, 55-46, 45-36, 35-26, 25-0	EQ-5D index
Morgan et al. (2006) [334]	Wales	4,502	1	Any DR	NA	EQ-5D index
Ohsawa et al. (2003) [229]	Japan	68	2	NA	Both eyes: blind	SG, VAS
Papadopoulos et al. (2009) [335]	Greece	183	4	No DR, any DR	NA	EQ-5D index, EQ VAS
Polack et al. (2015) [216]	India	249	11	Worse eye: No DR, mild/moderate NPDR, severe NPDR / PDR, blind	Both eyes: $\leq 6/60$	TTO, EQ-5D index
Rachapelle et al. (2013) [113]	India	249	4	No DR, NPDR, STDR, blind	$\leq 6/60$	TTO

Sakamaki et al. (2006) [220]	Japan	220	4	Fukuda: $\geq$ A1, No DR	NA	EQ-5D index, EQ VAS
Sakthong et al. (2008) [336]	Thailand	303	6	No DR, any DR	NA	EQ-5D index
Scanlon et al. (2014) [164]	Multinational	289	3	DMO	BCVA (better eye): $\geq$ 80, $\leq$ 60	EQ-5D index
Sharma et al. (2003) [231]	Canada	221	5	Any DR	BCVA (better eye): $\geq$ 6/7.5, 6/9–6/15, 6/18–6/30, 6/60–6/120, CF-NLP	TTO
Smith et al. (2005) [217]	USA	155	4	Worse eye: NPDR, PDR	BCVA (better eye): $\geq$ 20/40, $\leq$ 0/40	EQ-5D index
Sullivan & Ghushchyan (2016) [218]	USA	20,705	4	No DR, PDR/CSMO	Worse eye: blind	EQ-5D index

Szabo et al. (2010) [228]	Canada	98	11	NA	BCVA (better eye): current vision, 20/20-20/40, 20/50-20/80, 20/100-20/160, 20/200 / CS<21, BCVA (worse eye): current vision, >20/200 / CA≥ 21, <20/200 / CS<21	TTO
Tung et al. (2005) [214]	Taiwan	406	4	Worse eye: No DR, NPDR, PDR, blind	≤ 6/60 (blind)	TTO
Venkataraman et al. (2013) [225]	Singapore	2,601	6	ETDRS: 31 / 63, 41 / 64, 51 / 64, 31 / 63, 41 / 64, 51 / 64	NA	SF-6D index
Wu et al. (1998) [90]	USA	143	6	No DR, any DR	NA	QWB (SF-36 mapping)
Yeo et al. (2012) [238]	Wales	621	2	NA (screening attenders)	NA	EQ-5D index, EQ VAS
Yeo et al. (2012) [337]	Wales	198	2	NA (screening attenders)	NA	EQ-5D index, EQ VAS

Table D.1: Details of health state utility values

Appendix E

ISDR visit questionnaire

Visit questionnaire

Patient initials

Patient E Number

Randomisation Number

(leave blank at baseline - can be applied after randomisation)

Patient date of birth

d	d	m	m	y	y	y	y

Date of visit

d	d	m	m	y	y	y	y

Did the patient self-complete?: Yes / No

This questionnaire is only to be completed by the individual on their first visit to the site.

We would now like to ask you about your visit to the clinic. If you are unsure about an answer, please provide your best estimate.

1. Which modes of transport did you use in travelling to and from the centre today?

Car ☐ Taxi ☐ Bus / train ☐ Hospital transport ☐ Bicycle / on foot ☐

Other_____

2. If using public transport or taxi, what is the total cost of your return travel?

£_____ Not applicable ☐

3. If travelling by car, how many miles is your return journey?

_____miles Not applicable ☐

4. If travelling by car, what is the total cost of parking?

£_____ Not applicable ☐

5. How much time did you spend in total on this visit, including preparation, travel time and attending?

_____hours _____minutes

6. Are you currently in employment?

Yes ☐ No ☐

7. Did you take time off work to attend today?

Yes ☐ No ☐ Not applicable ☐

8. Did a friend, family member or any other person assist you in attending your appointment today?

Yes ☐ No ☐

9. Did they take time off work to do so?

Yes ☐ No ☐ Not applicable ☐

10. How much time did they spend helping you to attend this visit?

_____hours _____minutes Not applicable ☐

Appendix F

Decision model VBA code

```
Sub Full_Simulation()

Dim a, b, c

Dim shPatA As Worksheet, shRes As Worksheet, shSim As Worksheet, shRan As
    Worksheet, shChA As Worksheet, shChB As Worksheet, shChI As Worksheet,
    shPatS As Worksheet, shMod As Worksheet
Set shPatA = ThisWorkbook.Worksheets("Patient attributes")
Set shPar = ThisWorkbook.Worksheets("Parameters")
Set shRes = ThisWorkbook.Worksheets("Results")
Set shSim = ThisWorkbook.Worksheets("Simulation")
Set shRan = ThisWorkbook.Worksheets("Random numbers")
Set shChA = ThisWorkbook.Worksheets("Chain_Annual")
Set shChB = ThisWorkbook.Worksheets("Chain_Biennial")
Set shChI = ThisWorkbook.Worksheets("Chain_ISDR")
Set shPatS = ThisWorkbook.Worksheets("Patient sample")
Set shMod = ThisWorkbook.Worksheets("Model set-up")

Application.ScreenUpdating = False
Application.Calculation = xlCalculationManual
Application.DisplayStatusBar = False
Application.EnableEvents = False

' Delete results from 'Results'
shRes.Range("G5:AG1004").ClearContents

indexE = 0

Do While indexE < Range("Sim_probs")

    a = Now()

    ' Delete simulations from 'Simulation'
    shSim.Range("B7:K1175").ClearContents

    indexI = 0

    Do While indexI < Range("Sim_patients")

        shRan.Calculate
        shPar.Calculate

        ' Delete patient characteristics in 'Patient attributes'
        shPatA.Range("C3:C15").ClearContents
        ' Delete patient characteristics in 'Chain_Annual'
```

```
        shChA.Range("E56:L56").ClearContents
        ' Delete risk and transitions in 'Chain_Annual'
        shChA.Range("AQ57:CG1256").ClearContents
        ' Delete risk and transitions in 'Chain_Biennial'
        shChB.Range("AQ57:CG1256").ClearContents
        ' Delete risk and transitions in 'Chain_ISDR'
        shChI.Range("AQ57:CG1256").ClearContents

        Application.Calculate

        ' Assign patient characteristics from 'Patient sample' to 'Chain_Annual
            '

        P = Application.WorksheetFunction.RandBetween(1, 8111)

        shChA.Range("E56:L56").Value = shPatS.Range("A2").Offset(P, 0).Resize
            (1, 8).Value
        shChA.Calculate

        indexA = 0

        Do While shChA.Range("CJ1260").Value = 0

            shRan.Calculate

            ' Assign patient characteristics from 'Chain_Annual' to 'Patient
                attributes'

            shPatA.Range("C3:C15").Value = Application.Transpose(shChA.Range("
                E56").Offset(indexA, 0).Resize(1, 13).Value)
            shPatA.Calculate

            ' Assign risk and transitions from 'Patient attributes' to '
                Chain_Annual'

            shChA.Range("AQ57").Offset(indexA, 0).Resize(1, 43).Value =
                Application.Transpose(shPatA.Range("C16:C58").Value)
            shChA.Calculate

            indexA = indexA + 1

        Loop

        ' Assign pay-offs and patient ID from 'Chain_Annual' to 'Simulation'

        shSim.Range("B7").Offset(indexI, 0).Resize(1, 4).Value = shChA.Range("
            CK1260:CN1260").Value

        indexB = 0

        Do While shChB.Range("CJ1260").Value = 0

            shRan.Calculate

            ' Assign patient characteristics from 'Chain_Biennial' to 'Patient
                attributes'

            shPatA.Range("C3:C15").Value = Application.Transpose(shChB.Range("
                E56").Offset(indexB, 0).Resize(1, 13).Value)
            shPatA.Calculate
```

```
            ' Assign risk and transitions from 'Patient attributes' to '
                Chain_Biennial'

            shChB.Range("AQ57").Offset(indexB, 0).Resize(1, 43).Value =
                Application.Transpose(shPatA.Range("C16:C58").Value)
            shChB.Calculate

            indexB = indexB + 1

        Loop

        ' Assign pay-offs and patient ID from 'Chain_Biennial' to 'Simulation'
        shSim.Range("F7").Offset(indexI, 0).Resize(1, 3).Value = shChB.Range("
            CL1260:CN1260").Value

        indexC = 0

        Do While shChI.Range("CJ1260").Value = 0

            shRan.Calculate

            ' Assign patient characteristics from 'Chain_ISDR' to 'Patient
                attributes'

            shPatA.Range("C3:C15").Value = Application.Transpose(shChI.Range("
                E56").Offset(indexC, 0).Resize(1, 13).Value)
            shPatA.Calculate

            ' Assign risk and transitions from 'Patient attributes' to '
                Chain_ISDR'

            shChI.Range("AQ57").Offset(indexC, 0).Resize(1, 43).Value =
                Application.Transpose(shPatA.Range("C16:C58").Value)
            shChI.Calculate

            indexC = indexC + 1

        Loop

        ' Assign pay-offs and patient ID from 'Chain_ISDR' to 'Simulation'

        shSim.Range("I7").Offset(indexI, 0).Resize(1, 3).Value = shChI.Range("
            CL1260:CN1260").Value

        indexI = indexI + 1

    Loop

    Application.Calculate

    ' Assign results from 'Simulation' to 'Results'

    shRes.Range("G5").Offset(indexE, 0).Resize(1, 2).Value = Application.
        Transpose(shSim.Range("C3:C4").Value)
    shRes.Range("I5").Offset(indexE, 0).Resize(1, 2).Value = Application.
        Transpose(shSim.Range("D3:D4").Value)
    shRes.Range("K5").Offset(indexE, 0).Resize(1, 2).Value = Application.
        Transpose(shSim.Range("E3:E4").Value)
    shRes.Range("M5").Offset(indexE, 0).Resize(1, 2).Value = Application.
        Transpose(shSim.Range("F3:F4").Value)
```

```
    shRes.Range("O5").Offset(indexE, 0).Resize(1, 2).Value = Application.
        Transpose(shSim.Range("G3:G4").Value)
    shRes.Range("Q5").Offset(indexE, 0).Resize(1, 2).Value = Application.
        Transpose(shSim.Range("H3:H4").Value)
    shRes.Range("S5").Offset(indexE, 0).Resize(1, 2).Value = Application.
        Transpose(shSim.Range("I3:I4").Value)
    shRes.Range("U5").Offset(indexE, 0).Resize(1, 2).Value = Application.
        Transpose(shSim.Range("J3:J4").Value)
    shRes.Range("W5").Offset(indexE, 0).Resize(1, 2).Value = Application.
        Transpose(shSim.Range("K3:K4").Value)
    shRes.Range("Y5").Offset(indexE, 0).Resize(1, 9).Value = shSim.Range("L3:T3
        ").Value

    Application.Calculate

    indexE = indexE + 1

    b = Now()
    c = (b - a) / Range("Sim_patients")
    shMod.Range("C12") = c

Loop

Application.Calculation = xlCalculationAutomatic
Application.ScreenUpdating = True
Application.DisplayStatusBar = True
Application.EnableEvents = True

ThisWorkbook.Close Savechanges:=True

End Sub
```