DEVELOPMENT OF EVIDENCE SYNTHESIS METHODS FOR HEALTH POLICY DECISION MAKING – A CHAIN OF EVIDENCE APPROACH

Clare Louise Gillies BSc MSc

Abstract

This project comprises a critical exploration and development of methods for the synthesis of evidence, using a *chain of evidence* approach, from diverse, yet inter-related, sources. The methodologies were explored through the development of a comprehensive decision model to assess different health policies in respect to screening for type 2 diabetes mellitus (T2DM). Four strategies were compared which were, no screening (current policy), screening for T2DM alone allowing for early diagnosis and treatment of the condition, and two strategies whereby both impaired glucose tolerance (IGT) and T2DM were screened for, allowing for early treatment of T2DM and for either lifestyle or pharmacological interventions to be applied to those with IGT in an attempt to delay the onset of T2DM.

The comprehensive decision model developed here was innovative when compared to current published models in a number of ways. Firstly the entire model was encompassed within a single flexible framework, which has a number of advantages, and secondly as much of the available data as was feasible to use, was incorporated into the model inputs. A number of methodological issues and techniques were explored during the development of the comprehensive decision model. These included mixed treatment comparison analyses, assessment of baseline risk on intervention effects and the use of individual patient data. A number of sensitivity analyses and model extensions were carried out to assess the parameters with most influence on model results, and to adapt the model to different screening scenarios.

The results of the model provide evidence that a screening strategy for IGT and T2DM, followed by appropriate treatment and interventions appears to be a cost-effective screening strategy. Uncertainty still surrounds the cost-effectiveness of screening for T2DM alone and further research is required. Running decision models within a Bayesian, comprehensive decision modelling framework, allows for model flexibility and has advantages over more conventional modelling techniques.

Contents

1. INTRODUCTION 1

1.1 AIMS OF THE THESIS 1
1.2 T2DM AND IGT, A GROWING HEALTH PROBLEM 1
1.3 ASSESSING POTENTIAL SCREENING POLICIES FOR IGT AND T2DM 2
1.4 METHODOLOGICAL ISSUES CONCERNING DECISION MODELS 2
1.5 OVERVIEW OF THE THESIS 5

2. INTRODUCTION TO THE METHODOLOGY 7

2.1 CHAPTER OVERVIEW 7
2.2 TYPES OF MODELS UTILISED 7
2.2.1 Decision trees 7
2.2.2 Markov models 8
2.2.3 Comprehensive decision models 10
2.2.4 A review of current comprehensive decision models 11
2.2.5 Problems with current decision models 14
2.3 METHODOLOGIES UTILISED WITHIN THE MODEL 15
2.3.1 Bayesian methods 15
2.3.2 Markov Chain Monte Carlo and WinBUGS 17
2.3.3 Prior and sampling distributions 18
2.3.4 Meta-analysis 22
2.3.5 Meta-regression 25
2.3.6 Mixed treatment comparisons 26
2.4 ASSESSING CLINICAL AND COST-EFFECTIVENESS 28
2.4.1 Quality adjusted life years and utilities 28
2.4.2 Numbers needed to treat 29
2.4.3 Ascertaining costs 29
2.4.4 Incremental cost-effectiveness ratio 29
2.4.5 The cost-effectiveness plane 30
2.4.6 Cost-effectiveness acceptability curve 32
2.4.7 Adjusting costs from different years 34
2.4.8 Discounting 34
2.5 ASSESSING THE EFFECTIVENESS OF SCREENING TESTS 35
2.6 SUMMARY 36

3. OVERVIEW OF CLINICAL ISSUES 36

3.1 CHAPTER OVERVIEW 36
3.2 CLINICAL DEFINITIONS OF GLUCOSE TOLERANCE STATUS 36
3.2.1 Normal glucose tolerance 36
3.2.2 Impaired glucose tolerance and impaired fasting glucose 36
3.2.3 Type 2 diabetes mellitus 37
3.2.4 Other forms of diabetes 38
3.2.5 Metabolic syndrome 38
3.3 NATURAL HISTORY OF IGT, IFG AND T2DM 39
3.4 DIAGNOSTIC AND SCREENING TESTS FOR GLUCOSE TOLERANCE 40
3.4.1 Screening tests 40
3.4.2 Diagnosis of IGT and T2DM 41
3.5 CURRENT STUDIES 41
3.5.1 STAR 42
3.5.2 ADDITION 42
3.5.3 UKPDS 43
3.5.4 DECODE 44
3.6 SUITABILITY OF IGT AND T2DM FOR A SCREENING HEALTH PROGRAMME 44
3.7 SUMMARY 46

4. INTERVENTIONS FOR THE PREVENTION OR DELAY OF TYPE 2 DIABETES MELLITUS 47

4.1 CHAPTER OVERVIEW 47
4.2 CURRENT LITERATURE AND REVIEWS 47
4.3 SYSTEMATIC REVIEW 48
4.3.1 Literature search 48
4.3.2 Study characteristics 51
4.3.3 Studies not included in the meta-analyses 52
4.3.4 Data extraction 56
4.4 ANALYSES AND RESULTS 58
4.4.1 Meta-analyses of the intervention trials 58
4.4.2 Results of the meta-analyses 61
4.4.3 Factorial trials 66
4.4.4 Mixed treatment comparison (MTC) analyses 67
4.5 EXPLORATION OF HETEROGENEITY, STUDY QUALITY AND PUBLICATION BIAS 71
4.5.1 Assessment of heterogeneity 71
4.5.2 Study quality and publication bias 77
4.5.3 Adverse events 80
4.6 DISCUSSION 81
4.7 SUMMARY 84

5. MODELLING THE SCREENING AND INTERVENTION PATHWAY 86

5.1 CHAPTER OVERVIEW 86
5.2 EXISTING WORK 86
5.2.1 Previous models that have considered screening tests for IGT and/or T2DM 86
5.2.2 Previous models that have considered interventions for T2DM prevention 89
5.2.3 Previous models that have considered screening for T2DM 94
5.2.4 How the model developed for this thesis will add to current knowledge 97
5.3 THE PROPOSED MODEL 98
5.3.1 Structuring the decision tree 99
5.3.2 Outline of the decision tree 100
5.3.3 Structuring the Markov model 102
5.3.4 Outline of the Markov model 102
5.4 CHAPTER SUMMARY 104

6. DEVELOPMENT OF A PRIMARY MODEL 105

6.1 CHAPTER OVERVIEW 105
6.2 SCREENING DATA AND PREVALENCES 106
6.3 TRANSITION PROBABILITIES 114
6.3.1 NGT to undiagnosed IGT 116
6.3.2 Undiagnosed IGT to diagnosed IGT 116
6.3.3 Undiagnosed IGT to undiagnosed T2DM 117
6.3.4 Diagnosed IGT to screen diagnosed T2DM 122
6.3.5 Undiagnosed T2DM to clinically diagnosed T2DM 127
6.3.6 Mortality rates 127
6.3.7 Calculating the number in each state at each cycle 129
6.4 COSTS 131
6.4.1 Cost of screening 131
6.4.2 Cost of interventions 132
6.4.3 Costs of undiagnosed T2DM 132
6.4.4 Costs of diagnosed T2DM 133
6.4.5 Standardising costs to 2006 134
Table 6.8: Standardisation of costs to 2006 using price inflation indices 134
6.5 UTILITIES 135
6.6 MODEL OUTCOMES 142
6.7 RUNNING THE MODEL 144
6.8 RESULTS 146
6.9 DISCUSSION 153
6.10 CHAPTER SUMMARY 155

7: MODEL CHECKING AND SENSITIVITY ANALYSES 156

7.1 Chapter overview 156
7.2 Convergence of model parameters 156
7.2.1 Graphics utilised *156*
7.2.2 Results of convergence checks *158*
7.3 Distributions of model data 165
7.4 Prior distributions 165
7.5 Economic sensitivities 167
7.5.1 Discounting costs and benefits *167*
7.5.2 Sensitivity analyses for costs *171*
7.6 Clinical sensitivities 178
7.6.1 Test sensitivity *178*
7.6.2 Prevalence of glucose tolerance status *179*
7.6.3 Multi-way sensitivity analyses *183*
7.6.4 Models with populations of either 100% T2DM or 100% IGT *185*
7.7 Chapter summary 187

8. MODEL EXTENSIONS 188

8.1 Chapter Overview 188
8.2 Utilising the mixed treatment comparison analysis in the decision model 188
8.3 Modelling different ethnic cohorts 190
8.3.1 Adapting the model for different ethnicities *190*
8.3.2 Results for different ethnic cohorts *192*
8.4 Compliance to screening 195
8.4.1 Adapting the model to assess compliance to screening *195*
8.4.2 Results for varying screening compliance rates *198*
8.5 Compliance to interventions 198
8.5.1 Adapting the model to assess compliance to interventions *198*
8.5.2 Results for varying intervention compliance rates *199*
8.6 Re-screening 201
8.6.1 Adapting the model to assess the effects of more than one screening *201*
8.6.2 Results for additional screenings *203*
8.7 Alternative sources of diagnostic data 205
8.8 Value of information 210
8.8.1 Methods for assessing value of information *210*
8.8.2 Results of value of information analyses *212*
8.9 Discussion and chapter summary 214

9. APPLYING THE METHODOLOGY TO OTHER CLINICAL EXAMPLES 215

9.1 Chapter overview 215
9.2 Screening cholesterol levels and intervening to reduce risk of CHD 215
9.2.1 CHD *215*
9.2.2 Identifying individuals at high risk *215*
9.2.3 Intervening to prevent CHD *216*
9.2.4 Modelling the information *216*
9.3 Screening and prevention of colorectal cancer 218
9.3.1 Colorectal cancer *218*
9.3.2 Screening for precursors of colon cancer *218*
9.3.3 Interventions *219*
9.3.4 Modelling the information *219*
9.4 Comparison of the three models 220
9.5 Chapter Summary 222

10: DISCUSSION AND CONCLUSIONS 223

10.1 Summary 223
10.1.1 Methodological summary *223*
10.1.2 Clinical summary *224*
10.2 Discussion and limitations of this work 226
10.2.1 Methodological discussion *226*
10.2.2 Clinical discussion *229*

10.3 FURTHER WORK 230
10.4 CONCLUSIONS 232
APPENDIX 1: SEARCH STRATEGIES **233**
APPENDIX 2: FORM USED TO CALCULATE JADAD QUALITY SCORE **240**
APPENDIX 3: WINBUGS CODE **242**
APPENDIX 4: DIAGNOSTIC GRAPHS **279**
APPENDIX 5: PUBLICATIONS, PRESENTATIONS AND POSTERS ASSOCIATED WITH THIS THESIS **288**
BIBLIOGRAPHY **275**

1. INTRODUCTION

1.1 Aims of the thesis

The aim of this thesis was to review, critically appraise, and where appropriate develop, methods for modelling the screening/treatment/disease pathway of chronic health conditions using comprehensive decision models, with the complete disease pathway from screening through to treatment, further complications and death, incorporated within a single framework. The principle clinical example for this thesis was that of screening for impaired glucose tolerance (IGT) and type 2 diabetes mellitus (T2DM), although other clinical settings of major health importance were also considered, and the methodological issues surrounding them discussed. To develop the comprehensive decision model, with a view to assessing both clinical and cost-effectiveness, evidence had to be combined from a number of sources, resulting in a number of methodological issues that needed to be resolved. A background to the clinical and methodological issues that justify the need for this project are discussed in sections 1.2, 1.3 and 1.4 of this Chapter, with further details given in Chapters 2 and 3.

1.2 T2DM and IGT, a growing health problem

An estimated 171 million people worldwide in 2000, had diabetes and cases are projected to double by 2030 (Wild et al., 2004). Individuals with diabetes have a life expectancy that may be shortened by as much as 15 years (Donnelly et al., 2000) and incur around 5% of total NHS resources and 10% of hospital in-patient resources (Department of Health, 2001). Currently there is no systematic or structured screening policy for T2DM in the United Kingdom. One approach to screening would be to screen for T2DM only, which will allow for early diagnosis and treatment. This may be important as early detection and treatment may prevent future microvascular and macrovascular complications associated with T2DM. It is estimated that around 50% of individuals with diabetes are currently undiagnosed (King et al., 1998) and at presentation around 20-30% of individuals have already developed complications (The DECODE study group, 1999). An alternative screening approach would be to lower the threshold of the screening test and to screen for both IGT and T2DM together. IGT can be thought of as a precursor to T2DM, therefore screening for both conditions allows

for early diagnosis of T2DM and for interventions to be administered to persons identified with IGT, in order to attempt to delay or prevent the onset of T2DM. A more detailed description of the clinical definitions and issues, which are referred to throughout this thesis, is given in Chapter 3.

1.3 Assessing potential screening policies for IGT and T2DM

As no definitive trials assessing the effectiveness of screening for T2DM or IGT have been carried out (Waugh et al., 2007, Davies et al., 2004), assessment of such policies have so far been conducted through simulation studies. A number of decision models, reviewed in Chapter 5, have been developed that have assessed the clinical and cost-effectiveness of either screening for T2DM or interventions to prevent T2DM, however conclusive evidence is still not available for the cost-effectiveness of such strategies. In addition, the previous models that have assessed interventions assume an already identified IGT population, and do not include the costs and clinical issues involved with the identification of such individuals by screening. Also models have so far either assessed either screening for T2DM or interventions for T2DM prevention, but not a comparison of the two policies. The limitations and conclusions of previous models are discussed in detail in Chapter 5. The comprehensive decision model developed for this thesis has directly addressed the limitations of previous models, to produce a more comprehensive assessment of screening and intervention policies for T2DM.

1.4 Methodological issues concerning decision models

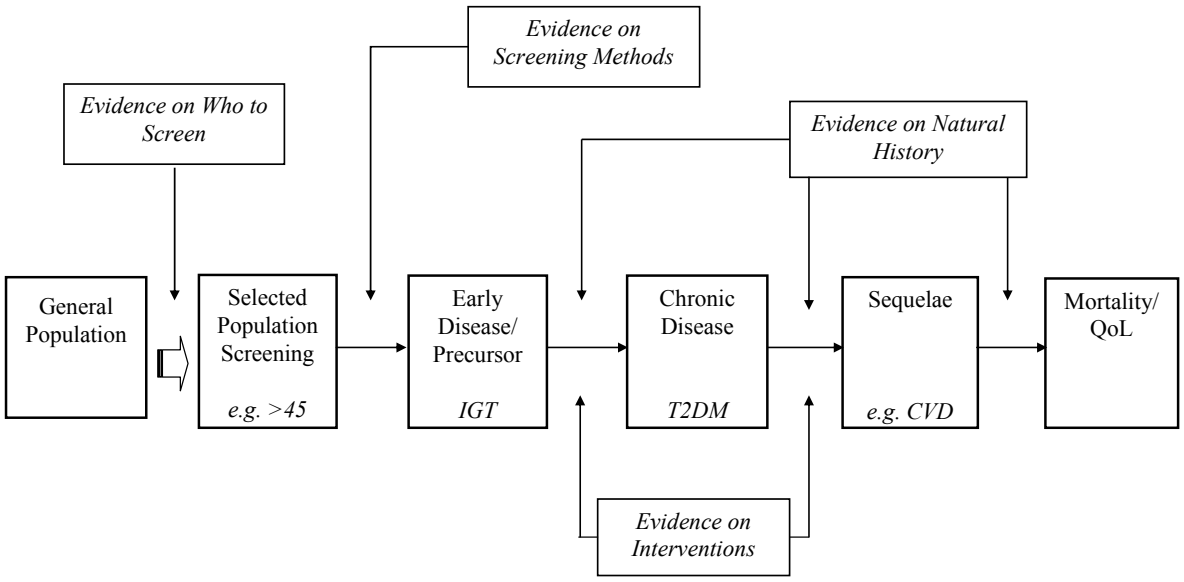
To assess the clinical and cost-effectiveness of potential screening and intervention strategies for IGT and T2DM, a decision model needs to be developed which addresses a number of inter-related questions. These are:

i) Who to screen for early stage disease or a precursor?
ii) How to screen them?
iii) What interventions to use for those individuals identified?

Answering all three questions simultaneously from one single study is impossible, and therefore evidence from a variety of sources needs to be synthesized within a coherent and flexible modelling framework. A decision model uses mathematical relationships to define a series of possible consequences, which would flow from a set of alternative options being evaluated. Appropriate sources of uncertainty and correlation need to be accounted for in the model, and clinical evidence needs to be integrated with evidence on costs and utilities (Cooper et al., 2004). Decision models are discussed in general, along with a review of comprehensive decision models, in Chapter 2.

The information on screening, precursors to disease, interventions and disease progression can be thought of as forming a 'chain of evidence' (Ades, 2002). Figure 1.1 illustrates the components in such a chain both generally and specifically for IGT and T2DM.

Figure 1.1: **Illustration of how the screening/disease pathway allows for a chain of evidence approach to the decision model**



Whilst the use of a 'chain of evidence' synthesis approach yields an overall assessment of the clinical effectiveness of different public health policies, evidence on cost and utilities also needs to be included within the modelling framework, in order for the most cost-effective policy to be identified (Cooper et al., 2004). To allow for appropriate sources of uncertainty and correlation in the model inputs to be accounted for, a

Bayesian approach to fitting the decision model was utilised for this thesis, specifically using Markov Chain Monte Carlo (MCMC) methodology as detailed in Chapter 2.

Advantages of using a Bayesian approach for the decision model are that all parameter uncertainty can be allowed for in the model, pertinent information that would be excluded from a more traditional analysis can be included, correlations induced by the same study contributing to more than one part of the model can be accounted for, and the model has the ability to be extended to accommodate more complex scenarios, such as mixed treatment comparisons. In addition a Bayesian framework allows for probability statements to be made directly regarding quantities of interest, for example the probability that intervention A is superior to intervention B (Sutton and Abrams, 2001). Advantages of using a Bayesian, comprehensive approach are discussed further in Chapter 2.

Three methodological issues regarding evidence synthesis commonly occur in a setting such as this. Firstly heterogeneity in the data available for the model may be present. This may include clinical heterogeneity, such as patient or treatment differences between studies, methodological heterogeneity, such as differences in the statistics chosen to report results or methodologies used, and statistical heterogeneity, such as sampling error, both within and between studies. Secondly there may be both direct and indirect evidence regarding quantities of interest available and finally both individual patient data (IPD) and published summary evidence may be available to inform model parameters.

For example, when considering evidence on interventions for T2DM prevention, a number of issues needed to be addressed. Firstly studies are heterogeneous in terms of their population composition for age, ethnicity and baseline risk of developing diabetes. This raises methodological issues with respect to both generalisability to, and consideration of, the effectiveness of interventions in specific subgroups. Therefore information had to be combined and interpreted carefully. Intervention studies also considered a variety of lifestyle and pharmacological interventions both individually

and in combination, whereby direct comparisons could only be made between some forms of interventions using mixed treatment comparison methods (Caldwell et al., 2006). Mixed treatment comparison (MTC), is an expansion of a standard pair-wise meta-analyses, say for trials comparing A vs. B, to an analysis that includes trials that may for example, compare A vs. B, B vs. C, or A vs. C (Lu and Ades, 2004). Indirect treatment comparison is a subgroup of mixed treatment comparisons, and whereas a complete mixed treatment comparison could include all trials that considered A or B or C, indirect comparisons would be restricted to those only containing A, B and C. Published evidence concerning transition rates from different glucose tolerance status were also diverse in terms of ethnic population and age range considered. All the methodological issues concerning evidence synthesis are discussed in more detail, in relation to the comprehensive decision model, in future chapters.

1.5 Overview of the thesis

This thesis comprises a critical exploration, and where appropriate development of, methods for the synthesis of evidence, using a chain of evidence approach, within a comprehensive decision modelling framework. How to contend with common issues that arise when combining heterogeneous evidence sources are explored and discussed. The model developed uses the highly relevant example, considering current health issues and policies, of screening for IGT and T2DM, although screening for two other chronic conditions will also be considered briefly to assess the generalisability of the methods (chapter 8).

The model developed compares three active screening strategies comprising of:

i) A one-off screening for T2DM, allowing for early diagnosis and treatment of the condition.

ii) A screening for both IGT and T2DM and intervening with lifestyle interventions in those diagnosed with IGT, and early diagnosis and treatment for those with T2DM.

iii) As for ii) but using pharmacological interventions.

All three active screening strategies were compared against a fourth strategy of no screening (current practice). The full pathway from screening to intervention and treatment for T2DM, all the way through to death, is modelled. This is the first model to directly compare the two alternative approaches of either screening for T2DM alone or screening for both IGT and T2DM together. By carrying out a number of sensitivity analyses the essential elements that affect the cost and clinical effectiveness of different screening policies can be fully understood.

To briefly outline the thesis, Chapters 2 and 3 will provide a more detailed introduction to the clinical terminology and issues surrounding this project. Chapter 4 will outline a comprehensive systematic review and meta-analysis that was undertaken to assess interventions for the prevention or delay of T2DM. Chapters 5 and 6 will summarise the structure of the comprehensive decision model, followed by a detailed description of the data used for the primary model and how it was developed. Chapter 7 will discuss the thorough model checks and sensitivity analyses that were carried out on the primary model, and Chapter 8 outlines a number of extensions that were made to the model. Chapter 9 will describe how the methodologies developed for this case study of screening for T2DM and IGT could be applied to other clinical examples, and finally Chapter 10 will give an overview of the conclusions of the decision model, with a discussion on the more significant methodological issues encountered, as well as opportunities for further work.

Overall, the model developed and described in this thesis is the most comprehensive ever attempted in the field of T2DM screening, and maybe even that of decision modelling in general. As much of the published data relevant to the model was incorporated as possible, many of the methods utilised are original, developed specifically for this example, and the general approach took into consideration, and attempted to overcome, common issues that occur frequently in evidence synthesis and decision models applied to common chronic conditions. The model was hybrid in that it incorporated both a decision tree and a Markov model, and included previously published models from the UKPDS trial within its framework. Although the model was complex in structure, the methods utilised and developed are both applicable and accessible for use in future comprehensive decision models, for a variety of chronic health conditions.

2. INTRODUCTION TO THE METHODOLOGY

2.1 Chapter overview

A variety of statistical techniques were utilised and developed during the compilation of this thesis. This chapter gives a brief introduction to the techniques used, describing what they are, why they were chosen and outlining any assumptions or limitations of each methodology. Firstly the terms used to define the model compiled for this thesis are described. Secondly the methodologies utilised within the model, such as meta-analysis, meta-regression and mixed treatment comparisons are discussed. Thirdly information on how the cost-effectiveness of different screening strategies was incorporated into the model and how health economic techniques can be used to effectively interpret the results is outlined, and finally a brief description of the terminology used when assessing the efficiency of screening tests is provided. The novel aspect of this thesis is the simultaneous use of the methodologies discussed here, within a single, coherent framework.

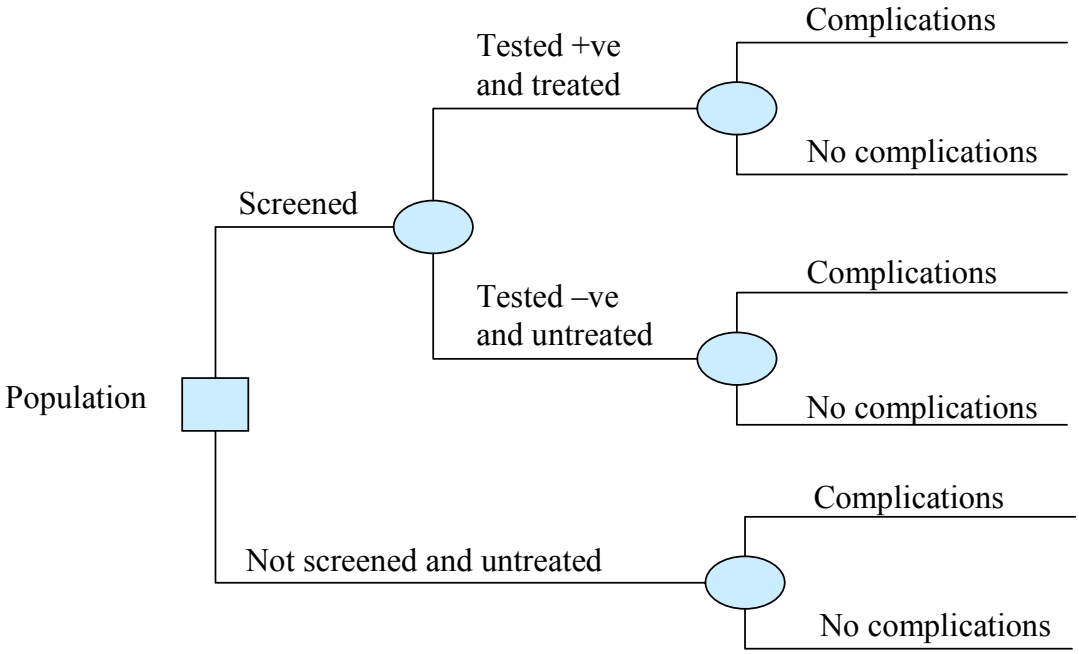
2.2 Types of Models Utilised

2.2.1 Decision trees

The screening stage of the comprehensive decision model was structured in the form of a decision tree. A decision tree flows from left to right beginning with an initial clinical choice or decision, indicated by a box, on a defined cohort of patients (Drummond et al., 2005). As a result of the decision made, there will be outcomes of given prior probabilities, so for this model the decision was whether to screen or not, and the outcome was the probability of being diagnosed with IGT or T2DM, dependent on the screening strategy adopted. The main drawback of decision trees is that they can quickly become unwieldy as the number of clinical decisions and outcomes you wish to model increase (Briggs et al., 2006), also as probabilities are needed for each branch of the tree, as the number of braches increase the ability to obtain reliable data decreases. An example of a decision tree is given in figure 2.1 and illustrates a model whereby the impact of screening a population for a disease is assessed by comparing two strategies

of either screening or not screening. The probabilities of testing negative or positive if screened, along with complication rates if the disease is left unidentified or if it is treated, can then be applied at each of the relevant nodes of the decision tree. The model would therefore allow a comparison of complication rates in an unscreened and hence untreated population, compared to a scenario where screening and treatment takes place. The structure of the decision tree utilised for this model is described in chapters 5 and 6. The use of decision trees in practice are varied as they are easily applied to numerous clinical situations, for example Grau et. al (2007) utilised a decision tree to model the clinical effectiveness of different treatments for micro-organisms, and Todorova et al. (2007) assessed the cost-effectiveness of different treatments for gestational diabetes (Grau et al., 2007, Todorova et al., 2007).

Figure 2.1: A diagrammatic example of a decision tree



2.2.2 Markov models

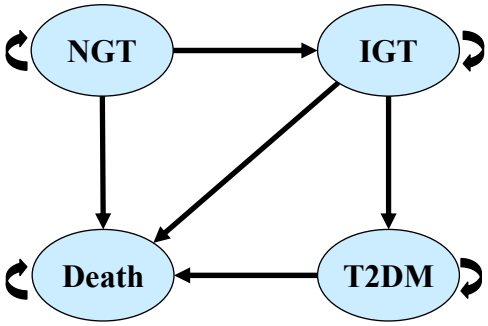
Markov models have been extensively used in cost-effectiveness analysis as they enable long-term outcomes to be modelled (Spiegelhalter et al., 2004). A Markov model assumes that in each cycle (often representing one year), an individual is in one of a finite set of states, with a probability of transferring to another state by the next cycle. The transition probabilities govern both the direction and speed of transitions between disease states (Briggs et al., 2006), with absorbing states, such as death, having transition probabilities to other states of 0. Utilities and annual costs can then be attached to each state, at each cycle, and hence the long-term cost and clinical

effectiveness of different scenarios estimated. Markov models were used in the decision model developed in this thesis, to model long-term glucose tolerance status, development of T2DM, and the associated complications, of a cohort of individuals for each screening strategy. The model is outlined in detail in Chapter 5.

Markov models have some very useful properties. They are dynamic, probabilistic and can address problems in which events and decisions are occurring, subject to chance, over time (Eddy, 2006). Also the use of discrete states and the notion that people progress between states fits in well with our society's notion of diseases and how they are classified. Markov states and transitions are useful model components on which to attach costs and utilities, although definition of, and characterisation of states in terms of costs and utilities needs to be clinically meaningful. An additional advantage of Markov models is that the basic mathematical structure is easy to understand and interpret.

Although Markov models generally offer greater flexibility than decision trees, they do have some limitations. Markov models are restrictive in that they have a '*memoryless*' property. Once a simulated individual has moved from one state to another, the model will have '*no memory'* of where the patient has come from or the timing of that transition (Briggs et al., 2006). It can also be argued that the use of discrete states may be too simplistic, as health states can often be measured on a continuous scale (Eddy, 2006). Furthermore the transition rates from one state to the next may be difficult to quantify, especially when numerous metabolic factors may influence the transition.

Figure 2.2: A diagrammatic example of a Markov model



* States are normal glucose tolerance (NGT), impaired glucose tolerance (IGT) and Type 2 diabetes mellitus (T2DM)

A simple example of a Markov model, whereby movement between glucose tolerance states is modelled over time, is given in figure 2.2. Four Markov states are modelled whereby transition rates determine the movements between states at each cycle, or the probability of remaining within a state. Death is an absorbing state in this example, as no movement is allowed out of this state.

2.2.3 Comprehensive decision models

A decision model uses mathematical relationships to define a series of possible consequences, that would flow from a set of alternative options (Briggs et al., 2006). Each consequence has an expected cost and outcome, enabling the clinical and cost effectiveness of each of the options to be evaluated. Decision models can be structured using a variety of methodologies, with both decision trees and Markov models being utilised for the model described in this thesis. The model compiled here was defined as a *comprehensive* decision model for two reasons. Firstly, where appropriate, all sources of available data were included in the model. For example where interventions for the prevention of diabetes were considered, a full systematic review and meta-analysis was carried out to identify all relevant trials. Secondly all parts of the model were carried out within a single framework. That is, for example, if a meta-analysis of several studies was needed for one of the model parameters, the meta-analysis was carried out within the decision model, as opposed to separately with the pooled effect size added as data into the decision model. This would only be possible with a Bayesian approach. Using a

single model framework facilitates the inclusion of uncertainty and correlations within the model. The main advantages can be summarised as (Cooper et al., 2003):

i) The incorporation of more appropriate parameter uncertainty by allowing for the fact that both the overall population effect of μ and between-study precision τ^2 in the meta-analyses have both been estimated by the data.

ii) The ability to make direct probability statements and thus direct answers to the question of interest (e.g. Bayesian meta-analysis can give a probability that the effect is above or below a particular value).

iii) External evidence from expert opinions can easily be incorporated within the modelling framework.

iv) The actual posterior distributions from the meta-analyses are used, as opposed to making assumptions of normality (or some other parametric form) which is necessary for a classical analysis.

v) Correlation, where one study may contribute to more than one part of the model, can be accounted for more easily.

2.2.4 A review of current comprehensive decision models

Conventional methods to developing Markov models and decision trees would use a two-stage approach, whereby firstly a series of meta-analyses would be performed to obtain model parameter estimates, and these would then be entered into the model (Cooper et al., 2004). An alternative approach is to use a comprehensive framework, whereby a single, coherent model is developed, incorporating all the analysis. This has many advantages as discussed in section 2.2.3. Despite the benefits of using a

comprehensive approach, few such models have been developed in practice (Cooper et al., 2004). The models that have been described in the literature will be detailed and appraised in this section.

The first publications to advocate a Bayesian, comprehensive approach for modelling the assessment of health technologies were by Eddy et al. (Eddy, 1989, Eddy, 1990). They described a Bayesian method for interpreting, adjusting and combining evidence from a number of diverse sources, to estimate a probability distribution for a parameter of interest (Eddy, 1990). They applied the methodology to a number of basic examples, including the assessment of the effect of thrombolytic agents on 1-year survival post heart attack (Eddy, 1989). These methods have been expanded in more recent years, with more complex models being fitted. A review of the literature found eight comprehensive decision models (Cooper et al., 2003, Cooper et al., 2002, Matchar et al., 1997, Parmigiani et al., 1997, Samsa et al., 1999, Sendi et al., 1999, Spiegelhalter and Best, 2003, Nixon and Duffy, 2002). The two models by Cooper and colleagues assessed the cost-effectiveness of taxane use in advanced breast cancer, and the use of prophylactic antibiotics during caesarean sections to reduce the risk of wound infection. Both models were developed in WinBUGS, using a Bayesian, Markov Chain Monte Carlo (MCMC) approach (this software and these methodologies are described in detail in section 2.3).

The taxane/breast cancer model (Cooper et al., 2003) consisted of four health states, and pooled estimates of transition probabilities were obtained from random effects meta-analyses, run within the model, with data from between 1 and 4 studies being combined for each transition. Limitations of this study were the relatively small number of studies used to estimate transition rates and the fact that correlations, where one study provided information for multiple parts of the model, were not accounted for. The antibiotics/caesarean model (Cooper et al., 2002), used a similar framework to the above example, although for this example, a decision tree format was used. A published systematic review of prophylactic antibiotic treatment on infectious complications in women undergoing caesareans was utilised and re-analysed within the model, using Bayesian methods. Both these models clearly highlighted the advantages of using a

comprehensive approach, particularly in the flexibility of the framework and the ability to include all parameter uncertainty.

Nixon and Duffy (2002) used MCMC methods to combine information from studies addressing different but clinically related questions, specifically the relationships between breast cancer, tamoxifen and genetic susceptibility. Meta-analyses were utilised for data inputs, where they had been previously published, otherwise single studies were used. The model was complex in structure containing numerous states, and a hierarchal structure. Spiegelhalter et al (2003) described a Markov model, consisting of five Markov states to assess the cost-effectiveness of total hip replacement, using multiple sources of evidence (Spiegelhalter and Best, 2003). Although many of the model inputs were taken from single studies, the analysis was thorough in that a number of sensitivity analyses were carried out to investigate the modelling of heterogeneity and uncertain model parameters.

Three of the comprehensive decision models in the literature assessed treatment and/or prevention of stroke (Matchar et al., 1997, Parmigiani et al., 1997, Samsa et al., 1999). The model by Parmigiani was by far the most detailed and complex, whereby two approaches were used, Bayesisan inference and resampling techniques, to model the four states of healthy, transient ischaemic attack, stroke and death, and the transition rates between these states. The methodologies used were complex and therefore not easily transferable to other clinical examples.

The final comprehensive decision model found in the literature assessed the cost-effectiveness of azithromycin for preventing *Myobacterium avium* complex in HIV positive patients (Sendi et al., 1999). They used a Bayesian approach to develop a Markov model consisting of four health states, to assess the clinical and cost implications of the *Myobacterium avium* complex. The model was comprehensive in that all the analyses were conducted within one framework, but the data sources utilised were limited.

Whilst comprehensive decision models improve on conventional decision models, they do not always utilise all available sources of data. The model developed here incorporates all available data sources where feasible, and whilst the model is more complex than any previous model, the methodologies used are easily transferable to other clinical examples. A comprehensive decision model has never been applied to the clinical area of screening for T2DM.

2.2.5 Problems with current decision models

To demonstrate how the model developed in this thesis will build and improve on current methods commonly used to fit decision models, it is important to understand the flaws and limitations that often exist in current published models.

Criticisms that commonly arise for decision models:

i) Model inputs often consist of just one estimate from a readily available source.

ii) In reference to non-comprehensive models, where more than one source is used to estimate a model parameter, the mean and standard error from a meta-analysis are entered into the model, therefore not all uncertainty is included, and assumptions are made on the distribution of the estimate.

iii) Correlation between model inputs are not always considered.

iv) A classical analysis may comprise of a rigid framework, which is difficult to adapt to different scenarios, or to model complex issues, for example the use of indirect evidence to inform model parameters.

v) Model assumptions and model structure are often not clearly reported, although good practice guidelines have recently been published (Philips et al., 2006).

vi) The implications of combining data from different populations, with different demographics and different risk of disease/complications are often not fully considered.

vii) Model parameters are often kept uniform over the time horizon of the model due to ease of implementation, when in fact it would be more realistic if they were allowed to change over time.

viii) Model checks and sensitivity analyses are often limited, and not fully undertaken. It is vital to check issues such as model convergence, the effects of prior distributions, and model data inconsistency, if the conclusions of the model are to be considered credible.

The model developed for this thesis will aim to address all of these issues, and by doing so produce a comprehensive decision model that can be considered robust, with realistic results that are relevant to a U.K. population. Decision models specific to T2DM and related health issues, all of which were non-comprehensive, are discussed and critically appraised in chapter 5.

2.3 Methodologies utilised within the model

2.3.1 Bayesian methods

The concept of Bayes theorem originates from a posthumous publication by Thomas Bayes in the 18th century. The basic concept is that data is supplemented using external prior beliefs or evidence, in that the likelihood (defined as the support for different values of the study outcome, based solely on data from the new study) is combined with a prior distribution (that is a reasonable opinion concerning the plausibility of different values of the study outcomes, excluding evidence from the current study). Once combined, a final belief is formed termed the posterior distribution. The combining of the two data sources is done using Bayes' theorem, which essentially weights the likelihood from the data with the relative plausibilities defined by the prior distribution (Spiegelhalter et al., 2004). Bayesian methods therefore differ from frequentist ones in that both the data and the model are assumed to be random quantities (Spiegelhalter et al., 1999). Equation 2.1 summarises the Bayesian framework, where θ is the parameter of interest, Y is the data, $p(\theta \mid Y)$ is the posterior distribution of the parameter after including the data, $p(Y \mid \theta)$ is the conditional likelihood of the data given the parameter (i.e. the likelihood function), and $p(\theta)$ is the prior distribution of the parameter of interest.

$$p(\theta \mid Y) = \frac{p(Y \mid \theta)p(\theta)}{p(Y)} \qquad \text{[Equation 2.1]}$$

and $p(Y) = \int p(Y \mid \theta)p(\theta) \quad d\theta$ i.e. an integrating constant

Results from a Bayesian analysis are conventionally reported with 95% credible intervals, which are comparable to the 95% confidence intervals reported for classical analyses. Although credible intervals (CrI) and confidence intervals (CI) are similar, they differ in a number of important ways (Spiegelhalter et al., 2004). The key difference is that they are interpreted differently. There is a 95% possibility that the true θ lies with in a 95% credible interval, but when interpreting confidence intervals it is only correct to say that in a long series of confidence intervals 95% will contain θ. Also, whilst the width of a confidence interval is governed by the standard error of the estimator, the width of credible intervals is determined by the posterior standard deviation.

Advantages of using a Bayesian approach for the decision model are that all parameter uncertainty can be allowed for in the model, pertinent external information that would be much more difficult to include within a classical analysis, can be easily included, and the model has the ability to be extended to accommodate more complex scenarios, such as mixed treatment comparisons A Bayesian framework also allows for probability statements to be made directly regarding quantities of interest, for example the probability that intervention A is superior to intervention B (Sutton and Abrams, 2001). Although it has been argued that these probability statements can be made classically (Burton, 1994), it could not be done with the same ease that is possible with a Bayesian framework, or on a formal basis. A final advantage is that as a predictive distribution for the parameter of interest, based on the posterior distribution (Spiegelhalter et al., 2004), can be obtained, prediction of a new study or generalisation to a new patient population is relatively straight forward.

Disadvantages of Bayesian methods include the fact that there is no automatic, easily interpretable measure of statistical significance from a Bayesian analysis, such as a p-value (Sutton and Abrams, 2001). Also the use of a prior belief means the analysis is no

longer completely objective, and defining a prior distribution can be a difficult task, for which at present there are no guidelines. Where the inclusion of prior beliefs may be inappropriate or indefinable, vague prior distributions may be used, so that the data will effectively dominate the prior distribution. However, defining an appropriate vague prior can be a difficult task in itself, and it has been shown that the choice of 'vague' prior can lead to marked variation in results, particularly for variance parameters (Lambert et al., 2005, Browne and Draper, 2000). It is therefore important when using a Bayesian framework to assess the specification of prior distributions through sensitivity analyses. Commonly used prior distributions are discussed in detail in section 2.3.3.

The principle reason for using a Bayesian approach for this thesis was not to allow inclusion of prior information, but because of the computational advantages of using the Bayesian software package WinBUGS. This provides a very flexible framework, within which complex and computationally intensive models can be specified. WinBUGS is discussed in more detail in section 2.3.2.

2.3.2 Markov Chain Monte Carlo and WinBUGS

The comprehensive decision model was fitted using the software package WinBUGS (Spiegelhalter et al.). WinBUGS uses Gibbs sampling, a particular form of Markov Chain Monte Carlo methodology (MCMC)(Gilks et al., 1996). In broad terms, Markov chains are processes describing trajectories where successive quantities are described probabilistically according to the values of their immediate predecessors. MCMC techniques enable simulation from a distribution by embedding it as a limiting distribution of a Markov chain and simulating from the chain until it reaches equilibrium (Gamerman, 1997). Gibbs sampling generates samples from the conditional posterior densities, which eventually converge to the desired marginal posterior densities. So for example, if the joint posterior distribution is given by $P(\boldsymbol{\theta})$= $P(\theta_1, \theta_2, ..., \theta_p \mid data)$, then let $P(\theta_j \mid \boldsymbol{\theta}_{(-j)}, data)$ represent the full conditional posterior distribution of parameter *j*, given the value of the other parameters and the data. The Gibbs sampler starts with initial values for the parameters $(\boldsymbol{\theta}^0)$ =

$(\theta_1^0, \theta_2^0, ..., \theta_p^0)$, then successive random observations are made from the full conditional posterior distributions $P(\theta_j \mid \boldsymbol{\theta}_{(-j)}, data), j=1...p$.

The sampling then starts as detailed below:

θ_1^1 from $P(\theta_1 \mid \theta_2^0, ..., \theta_p^0, \ data)$

θ_2^1 from $P(\theta_2 \mid \theta_1^1, \theta_3^0, ..., \theta_p^0, \ data)$

……

θ_p^1 from $P(\theta_p \mid \theta_1^1, ..., \theta_{p-1}^1, \ data)$

Hence $(\boldsymbol{\theta}^0) = (\theta_1^0, \theta_2^0, ..., \theta_p^0)$ has been changed to $(\boldsymbol{\theta}^1) = (\theta_1^1, \theta_2^1, ..., \theta_p^1)$. Repeatedly applying the algorithm *m* times will produce a series of observations $\theta_1 = (\theta_1^1, ...\theta_1^m), ..., \theta_p = (\theta_p^1, ..., \theta_p^m)$, which are realisations from a Markov chain with a distribution equivalent to the joint posterior distribution.

WinBUGS is a useful package for fitting a wide range of complex and computationally intensive models. It needs to be used with care though, and a number of issues, including convergence of model parameters and problems of autocorrelation between simulations, need to be assessed (Brooks and Gelman, 2007, Cowles and Carlin, 1996). These are discussed further in Chapter 7, where a full range of model checks carried out on the comprehensive decision model are described and discussed.

2.3.3 Prior and sampling distributions

For the comprehensive decision model a number of distributions were used for specifying both sampling distributions, that is the distribution of individual data points or summary statistics which contribute to the likelihood of the Bayesian model, or prior distributions for the parameters, which define the range of plausible values that a parameter could feasibly take (Spiegelhalter et al., 2004). One of the most difficult

tasks, when carrying out a Bayesian meta-analysis, is the choice of prior distribution for the between study standard deviation (τ). Where external evidence is not being utilised it is usually required that the distribution specified is vague, that is of a density that is sufficiently diffuse and gives a similar prior probability to a wide but plausible range of values. In reality any prior distribution will exert some influence on the posterior distribution, particularly where data is sparse. Therefore the real aim is to find a posterior distribution that has minimal effect on the final inference relative to the data (Bernando and Smith, 1994). For all the meta-analyses carried out for this thesis, sensitivity analyses were carried out to check the influence of the choice of prior distribution on τ, using a range of commonly used distributions. Both the uniform and half normal distributions were used in this thesis as prior distributions for τ, and these are described below along with some of the more unusual prior and sampling distributions utilised within the comprehensive decision model.

Beta distribution

Beta distributions are useful for defining quantities constrained to lie between 0 and 1, and therefore are often used as a prior distribution for an unknown proportion (Spiegelhalter et al., 2004). Where $Y \sim Beta[a,b]$ the distribution will have the properties as outlined in equation 2.2, where Γ represents the gamma function, in that $\Gamma(a) = (a-1)!$, if a is an integer.

$$p(y \mid a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{a-1}; \quad y \in (0,1) \qquad \text{[Equation 2.2]}$$

$$E(Y \mid a,b) = \frac{a}{a+b}$$

$$V(Y \mid a,b) = \frac{ab}{(a+b)^2(a+b+1)}$$

Dirichlet distribution

The dirichlet distribution is the multivariate generalisation of the beta distribution, and it is commonly used as a prior distribution for proportions where there are more than two possible categories (*K*). This distribution was used for the prior distribution on prevalences and test results, as detailed in Chapter 6. Where $Y \sim Dir(\alpha)$ and $\alpha_0 = \sum_{i=1}^{K} \alpha_i$, then the properties of the distribution can be described as in equation 2.3 (Briggs et al., 2003).

$$p(y \mid \alpha_1,..,\alpha_i) = \frac{\Gamma(\alpha_1 + ... + \alpha_i)}{\Gamma(\alpha_i)...\Gamma(\alpha_i)} y_1^{\alpha_1 - 1} ... y_i^{\alpha_i - 1} (1 - y_1 - ... y_i)^{\alpha_i + 1^{-1}} \quad \text{[Equation 2.3]}$$

$$E[Y_i \mid \alpha] = \frac{\alpha_i}{\alpha_0}$$

$$V[Y_i \mid \alpha] = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)}$$

Gamma distribution

Gamma distributions are used for quantities that need to be constrained to positive values, such as costs, and were utilised in the decision model for time taken to carry out a screening test. $Y \sim Gamma[a,b]$ represents a distribution with the properties as outline in equation 2.4 (Spiegelhalter et al., 2004).

$$p(y \mid a,b) = \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by}; \quad y \in (0,\infty) \quad \text{[Equation 2.4]}$$

$$E(Y \mid a,b) = \frac{a}{b}$$

$$V(Y \mid a,b) = \frac{a}{b^2}$$

Half-normal distribution

The half-normal distribution arises by folding a normal distribution around 0, such that if $Y \sim N[0,\sigma^2]$, then $|Y| \sim HN[\sigma^2]$ (Spiegelhalter et al., 2004). The properties of the distribution are given in equation 2.5. It is a useful distribution for expressing support for values near 0, with σ determining the upper range of supported values, and is a distribution often used for standard deviations. This distribution was used when carrying out sensitivity analyses on the prior distributions placed on the between study standard deviation (τ), for the four meta-analyses carried out for the decision model.

$$p(y \mid \sigma^2) = \sqrt{\frac{2}{\pi\sigma^2}} e^{\frac{-y^2}{2\sigma^2}}; \qquad y \in (0,\infty) \qquad \text{[Equation 2.5]}$$

$$E(Y \mid \sigma^2) = \sqrt{\frac{2}{\pi}}\sigma$$

$$V(Y \mid \sigma^2) = \sigma^2\left(1 - \frac{2}{\pi}\right)$$

Log-normal distribution

The log-normal is a distribution over positive values and is used as a sampling distribution for positive observations such as costs, or as a prior distribution for positive parameters such as variances. This distribution was used to model the transition probabilities and the intervention effects. Where $Y \sim LN(\mu, \sigma^2)$, the distribution will have properties as specified in equation 2.6.

$$p(y \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma y} e^{-(\log y - \mu)^2 / 2\sigma^2} \qquad \text{[Equation 2.6]}$$

$$E(Y \mid \mu, \sigma^2) = e^{\mu + \sigma^2/2}$$

$$V(Y \mid \mu, \sigma^2) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$$

Uniform distribution

A uniform distribution, sometimes also known as a rectangular distribution, is a distribution that has constant probability. It is usually adopted for an unknown parameter $Y \sim Unif[a,b]$, and specifies that Y has an equal probability of taking any value between a and b. The distribution is used to express indifference concerning the prior plausibility of a range of values (Spiegelhalter et al., 2004), and was used for the sensitivity analyses carried out on the prior distribution specified for the between study standard deviation (τ). Details of the distribution are given in equation 2.7.

$$p(y \mid a,b) = \frac{1}{b-a}; \quad y \in (a,b) \qquad \text{[Equation 2.7]}$$

$$E(Y \mid a,b) = \frac{a+b}{2}$$

$$V(Y \mid a,b) = \frac{(b-a)^2}{12}$$

2.3.4 Meta-analysis

Meta-analysis can be defined as the statistical analysis of a large collection of results from individual studies for the purpose of integrating the findings (Glass, 1976). Meta-analyses provided an important methodology for the decision model, in that where there was more than one reported estimate for a model parameter, the data could be combined into a pooled estimate for the model.

Meta-analysis models can be defined as either fixed effect or random effect models, depending on how the variability between study results is treated (Sutton et al., 2000). The fixed effect model assumes no heterogeneity between studies, that is, all studies estimate the same true underlying effect size, with the estimates differing only because of random variation. In the random effects model the studies are assumed to estimate

different underlying effect sizes due to differences between studies. A random term for the effect sizes is included in the model to account for the extra variability represented by study heterogeneity. Equation 2.8 shows the random effects model expressed algebraically.

$$y_i \sim N(\theta_i, s_i^2) \quad , \qquad \text{with} \quad \theta_i \sim N(\mu, \tau^2) \qquad \text{[Equation 2.8]}$$

y_i is an estimate of effect size, θ_i is the underlying true effect size and s_i^2 is the estimated variance of y_i for study *i*, μ is the pooled estimate of effect size, and τ^2 is the random effects variance, which represents the between study variance. When τ^2 is 0, the random effects model will reduce to the fixed effect model (Sutton et al., 2000). Compared to fixed effect models, random effects models are more conservative and produce wider confidence intervals, as they give lower weight to larger studies. A random effects meta-analysis may still produce a confidence interval for τ^2 which is too narrow, in that τ^2 is assumed known in the calculation of the standard error of μ (Hardy and Thompson, 1996). This is easily addressed in Bayesian meta-analysis by placing a prior distribution on τ^2, although classical methods are also available which use likelihood based methods to produce a confidence interval for θ which accounts for the fact that τ^2 has had to be estimated from the data (Hardy and Thompson, 1996).

The use of a Bayesian approach to the meta-analyses within the decision model has a number of advantages, including some discussed previously. Bayesian methods allow for the inclusion of uncertainty in all parameters and for all sources of evidence, for direct probability statements of model outcomes to be assessed. (Sutton et al., 2000).

When carrying out a meta-analysis, using either classical or Bayesian methods, a number of issues need to be considered. Firstly the quality of studies included in the meta-analysis needs to be assessed, with the primary concern being that combining study results of poor quality, may lead to biased and misleading pooled estimates (Sutton et al., 2000). The problem can be overcome by restricting the meta-analysis to

sources of high quality data, by referring to a hierarchy of sources of best evidence. Well designed randomised controlled trials are considered the best evidence, with population studies and other sources of evidence more prone to bias, listed further down the hierarchy (Deeks et al., 1996). The quality of randomised controlled trials can also be assessed using a quality score such as the Jadad (Jadad et al., 1996), which rates studies by factors such as adequacy of randomisation and blinding. For the systematic meta-analysis carried out on intervention trials, the meta-analysis was restricted to randomised controlled trials only, all trials were assessed for quality using the Jadad score, and sensitivity analyses were carried out to investigate the effect of study quality on the estimated intervention effect. This is described in detail in Chapter 4.

Another consideration is publication bias. Although the publication of large, high quality studies is not thought to be influenced by the statistical significance of the results, it is thought that the publication of small studies may be more likely, if they show a statistically significant, rather than a non-significant result. This is termed publication bias (Sutton et al., 2000), and its presence is assessed by either visual assessment of a funnel plot, whereby the effect size of each study is plotted against the inverse of its standard error, and an asymmetrical plot around the mean effect size indicates the presence of publication bias, or through a test such as Begg's (Begg and Mazumdar, 1994) or Egger's (Egger et al., 1997). An adjustment can be made to account for publication bias, using the 'trim and fill' method (Duval and Tweedie, 2000). This is a simple rank-based data augmentation technique, whereby the outlying, 'asymetric' part of the funnel plot is trimmed off, the symmetric remainder is then used to estimate the 'true' centre of the funnel plot, and then the trimmed studies and their missing counterparts around the centre, are replaced. The pooled estimate is recalculated using the 'filled' funnel plot. The 'trim and fill' method was utilised in Chapter 4, to assess the implications of possible publication bias on the meta-analyses of intervention trials.

A final important consideration when carrying out a meta-analysis is the issue of heterogeneity. Although the extent of heterogeneity may be measured by estimating τ^2, its interpretation is then specific to a particular treatment effect metric (Higgins and

Thompson, 2002). Measures of the impact of heterogeneity have been developed that are independent of the number of studies used in the meta-analysis and the treatment effect metric, and these include the I^2 statistic (Higgins and Thompson, 2002, Higgins et al., 2003). The I^2 statistic is interpreted as the proportion of total variation in study estimates that is due to heterogeneity, and is utilised for the meta-analyses carried out in Chapter 4.

A random effects meta-analysis will allow for heterogeneity between studies, and the I^2 statistic will quantify it, but neither will explain it. Study estimates of an effect size may vary for a number of reasons. For example the studies may be heterogeneous in terms of age, sex, weight and ethnicity of their participants, follow-up times may vary between studies and the dose or intensity of the treatment may vary. Sources of heterogeneity can be investigated using meta-regression techniques as discussed in section 2.3.5.

2.3.5 Meta-regression

To examine whether heterogeneity between study results can be explained by one or more factors across studies, meta-regression analyses can be performed (Sutton et al., 2000). For example, if there are *k* studies, each with effect sizes $Y_1,\ldots,\ Y_K$, and underlying effect size parameters $\theta_1,\ldots,\ \theta_K$, and there are p known predictor variables, $x_1,\ldots,x_P$, that may be related to the effect size, then the meta-regression model used to assess any interactions is presented in equation 2.9

$$\theta_i = \beta_{0i} + \beta_1 x_{i1} + \ldots.. + \beta_P x_{iP}$$
$$\beta_{0i} \sim N(\beta_0, \tau^2) \qquad \text{[Equation 2.9]}$$

Meta-regression is a useful tool for exploring why treatment effects may differ between studies. As only study level characteristics, such as mean age and mean weight of the study sample, are utilised for the analysis, results need to be interpreted with care, as although a relationship may appear to exist at the study level, this may not be true at the

individual participant level. To truly understand relationships between treatment effect and metabolic variables, individual patient data is required (Riley et al., 2007). Meta-regression analyses also have low power to detect an association between study characteristics and treatment effect, as demonstrated in a simulation study by Lambert et al. (2002) that compared summary patient-level covariates with individual patient data, and concluded that the statistical power of meta-regression was dramatically and consistently lower than that of individual patient data analysis, with little agreement between estimates obtained from the two methods (Lambert et al., 2002). Therefore, although meta-regression analysis is a useful tool for exploring sources of heterogeneity, the results need to be interpreted with care. The issues associated with meta-regression are discussed in an applied context in Chapter 4.

Meta-regression analyses are not suitable for the assessment of interactions between baseline risk of a disease or outcome, and treatment effect on the disease or outcome, due to correlation between the two and the fact that the uncertainty in the baseline risk would not be accounted for. Specialist methods have been developed for assessment of baseline risk effects (Sharp and Thompson, 2000), and these are explored further in Chapter 4, where an assessment of baseline risk was carried out.

2.3.6 Mixed treatment comparisons

Mixed treatment comparison (MTC) meta-analysis is an expansion of a standard pair-wise meta-analyses, say for trials comparing A vs. B, to an analysis that includes trials that may for example, compare A vs. B, B vs. C, or A vs. C (Lu and Ades, 2004). Indirect comparisons are a subset of MTC, and whilst an MTC should theoretically contain all trials that assessed either A or B or C alongside any number of other treatments, indirect comparisons are restricted to those that only compared A, B or C. The advantage of such an analysis is that firstly you strengthen inference concerning the relative efficacy of two treatments, as you are able to include both direct and indirect evidence, and secondly it allows for simultaneous comparison of all treatments, which enables the ‘best’ treatment to be identified (Lu and Ades, 2004).

The key assumption for a fixed effect MTC analysis is that the relative effect of one treatment compared with another is the same across all trials. In a random effects model it is assumed that although the effect sizes may differ between trials, they are from a common population distribution that is the same across trials (Caldwell et al., 2006).

These assumptions are extremely similar to those made for a standard pair-wise meta-analysis. The only difference is that the similarity of the relative effects of treatments is for the entire set of trials, irrespective of which treatments were actually evaluated. To assess this assumption imagine all trials have assessed the same two treatments and decide whether they are sufficiently similar to justify combining them in a meta-analysis. This is an important assumption to consider as different interventions may have been trialled on very different populations, for example treatment B may have been compared against the standard treatment of A in individuals who have an additional complication of hypertension, whereas C may have been compared against A in a relatively young age group. In this case combining all trials would not be a useful exercise and the assumption of similar treatment differences across all trials would probably not hold.

In addition the MTC analysis assumes an additive scale of measurement, in that the relative effect of A vs. C can be predicted from A vs. B and B vs. C. Therefore an appropriate measure of effect, such as the log odds ratio or risk difference, needs to have been chosen (Deeks, 2002). Also a common τ is usually assumed, rather than if individual meta-analyses were carried out for different interventions, with a different τ modelled for each meta-analysis. Finally, if a comprehensive MTC analysis is carried out, containing all relevant studies that contained at least one of the treatments of interest, the size of the network may become a problem, with parts of the network being unconnected. This will result in the unconnected parts of the network having to be excluded from the analyses.

An MTC analysis was carried out in Chapter 4 of this thesis, to enable trials comparing lifestyle interventions against controls, and/or pharmacological interventions against

controls, to be combined, enabling a direct comparison of lifestyle vs. pharmacological interventions. The MTC was also incorporated within the full comprehensive decision model, as a model extension in Chapter 8. A number of MTCs exist in current published literature, with recent examples including studies on stroke prevention and treatments for rheumatoid arthritis (Cooper et al., 2006, Nixon et al., 2007).

2.4 Assessing clinical and cost-effectiveness

2.4.1 Quality adjusted life years and utilities

To assess the clinical effectiveness of the different screening strategies simulated by the model, quality adjusted life years (QALYs) were calculated. A comparison of life years between two health policies will just give an estimate of the difference in mortality between the two, whereas QALYs allow for both morbidity and mortality occurrence to be assessed, this is because QALYs reflect an individual's length of life and their health-related quality of life in a single measure (Briggs et al., 2006). QALYs are a useful statistic in that they allow all health care interventions to be put on a common scale, which allows direct comparisons of very different health policies with very different outcomes. QALYs are calculated by applying health state preference scores, or utility weights, to each life year. Utilities usually take a value between 0 and 1, where 1 represents perfect health and 0 death. Values less than 0 are possible, representing a health state worse than death, though these are rarely used in practice. So for example if T2DM is believed to have a utility of 0.75 and an individual is followed up for five years over the course of a study until death, their QALYs for that time period will be 5 years multiplied by 0.75, which is equivalent to 3.75 QALYs.

Not all QALYs are the same, as utilities may be based on a variety of measures including visual analogue scales, estimates by physicians or researchers, or questionnaire health scores, such as the EuroQol-5D or the Health Utilities Index (Drummond et al., 2005). These health scores measure health-related quality of life and by polling members of a healthy 'reference' population, utility estimates can be assigned to each combination of responses to the questionnaire. The two principle advantages of using health indexes to generate utilities is that the questionnaires are easily completed by study participants, and the utilities they generate represent community or societal preferences (Hunink and Glasziou, 2001).

Arguments against the use of QALYs range from those claiming them to be over-complicated,

with utilities being difficult to estimate, to those claiming QALYs are over simplistic (Drummond et al., 2005), but despite criticism they are still the most frequently used measure in cost-effectiveness analyses (Briggs et al., 2006), and the preferred statistic, when comparing the cost-effectiveness of different health policies, as recommended by NICE (National Institute for Clinical Excellence, 2004).

2.4.2 Numbers needed to treat

The idea of measuring clinical effectiveness by determining number needed to treat (NNT), that is the number needed to be treated with a new treatment versus a standard treatment, for one additional patient to benefit, was first introduced by Laupacis (Laupacis et al., 1988). It is calculated from the absolute risk reduction (*ARR*) between the proportion with an adverse outcome on a new treatment (p_N), compared to the proportion on a standard treatment (p_S), equation 2.10 (Altman, 1998). It is a useful estimate of clinical effectiveness and is discussed further in Chapter 4.

$$ARR = p_N - p_S \qquad \text{[Equation 2.10]}$$

$$NNT = 1/ARR$$

2.4.3 Ascertaining costs

When assessing potential health policies and treatments, it is important to include predicted costs in the evaluation. The National Health Service in the U.K. has a fixed budget, and therefore the policies that provide the maximum health gain for the available, limited resources, need to be identified. In practice when estimating costs, people often collect resource use data, and then apply a unit cost to the different resources, to calculate total costs. Difficulties in identifying and modelling relevant costs are discussed further, with reference to the comprehensive decision model, in Chapter 6.

2.4.4 Incremental cost-effectiveness ratio

The incremental cost-effectiveness ratio (ICER) allows the comparison of two treatments or health policies in terms of both their clinical and cost effectiveness. It is calculated by dividing

the difference in costs between two treatments by the difference in their clinical effectiveness (usually in terms of quality adjusted life years) (equation 2.11).

$$\text{ICER} = \Delta_C / \Delta_E \qquad \text{[Equation 2.11]}$$

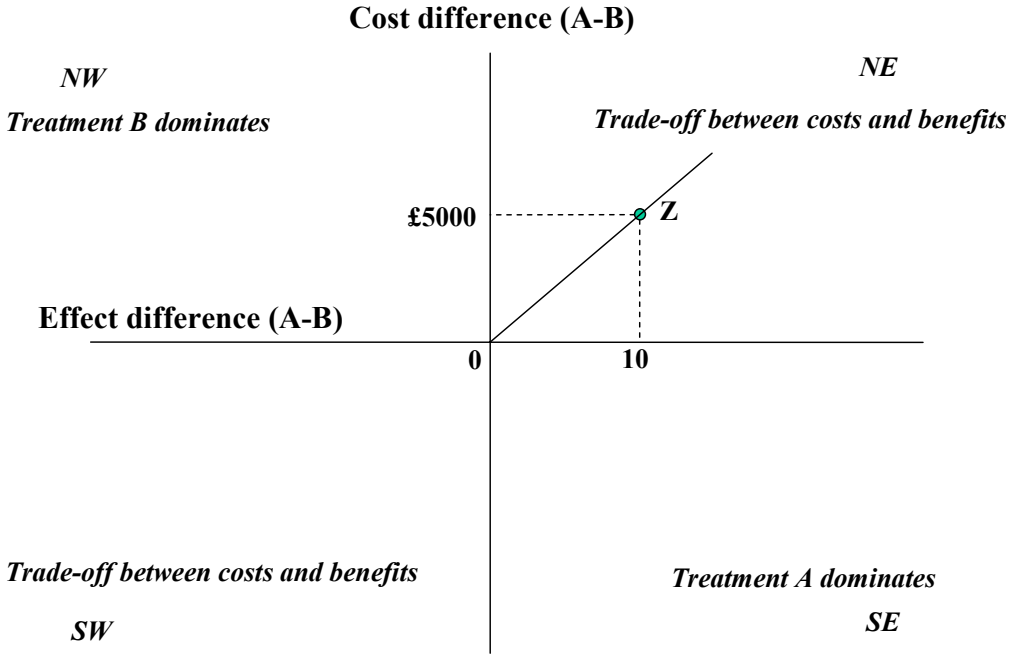
The ICER therefore represents the cost of each additional unit of clinical effect gained, and can thus be used to determine the preferred strategy, depending on how much decision makers are willing to pay to gain an additional unit of effect (Briggs et al., 2006). Calculating a confidence interval for the ICER can be problematic, as because it is a ratio, there is a probability of obtaining a zero or near zero value on the denominator. This suggests that moments of the ICER will be undefined (Heitjan et al., 1999), unless all the simulations of the ICER fall into the same quadrant of the cost-effectiveness plane (described in section 2.4.5). Although solutions have been put forward for constructing a viable confidence interval around the ICER (Severens et al., 1999), for this thesis the ICERs are reported without confidence intervals and instead uncertainty around cost-effectiveness is expressed using cost-effectiveness acceptability curves and probabilities of cost-effectiveness at different willingness-to-pay thresholds, as described in the following two sections.

2.4.5 The cost-effectiveness plane

The cost-effectiveness plane is a plot of the difference in clinical effectiveness between two treatments for each patient, plotted against the difference in costs per patient. The slope of the line joining any point on the plane to the origin is equivalent to the ICER (Briggs et al., 2006). The cost-effectiveness plane, which can be thought of as having four separate quadrants (*NW*, *NE*, *SW*, *SE*), is presented in figure 2.3. A difference in clinical effect greater than 0 indicates that A is better than B (x-axis), similarly a difference in costs greater than 0 indicates A is more costly than B (y-axis).

Consider an example where the position on a cost-effectiveness plane, for a comparison of two treatments A and B, is known with no uncertainty, then a number of conclusions can be drawn. If one treatment is less costly and more beneficial than the other, i.e. the estimate falls into the *NW* or *SE* quadrant, it is said to dominate, and the choice between the two treatments is clear. If the comparison of the two treatments fall into one of the other two quadrants (*NE* or *SW*), then a trade-off between costs and benefits needs to be made, and a decision must be taken on how much policy makers are willing to spend for every unit of clinical effectiveness gained.

Figure 2.3: The cost-effectiveness plane



For the example given in Figure 2.3, it can be seen that treatment A costs £5000 more than treatment B, but gives an additional 10 units, for example quality adjusted life years, of clinical benefit. Thereby the ICER (the slope of the line joining the origin and Z) is £500 per quality adjusted life year, and therefore treatment A appears to be of clinical benefit, with only a small additional cost.

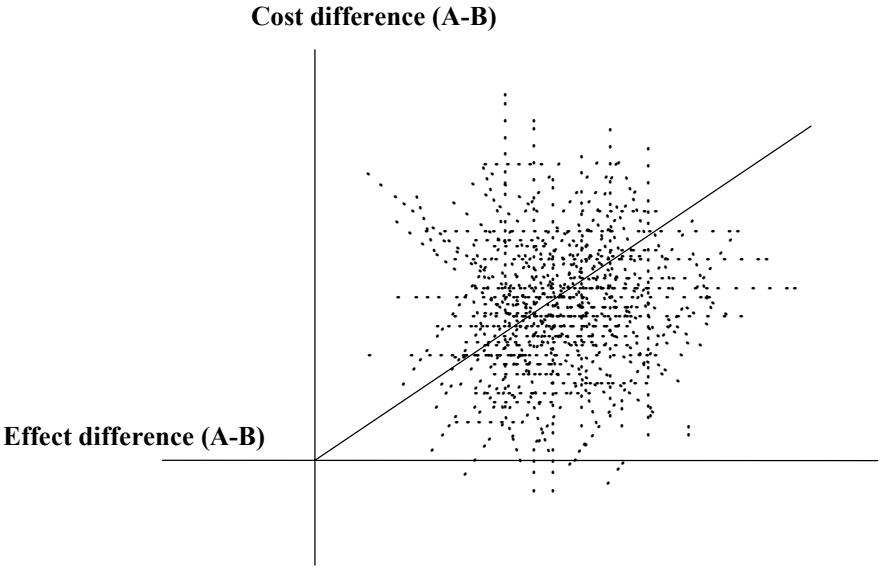
When the willingness-to-pay threshold (λ), that is the costs per QALY that a funding body is prepared to pay for a superior treatment, is known, then the assessment of cost-effectiveness can be expressed as in equation 2.12, in that the superior treatment is deemed cost effective if:

$$\Delta_C / \Delta_E < \lambda$$

$$\text{or} \qquad \lambda\Delta_E - \Delta_C > 0 \qquad \text{[Equation 2.12]}$$

The second equation specified above represents the net monetary benefit (NMB) (Drummond et al., 2005).

Figure 2.4: An estimated joint cost-effectiveness density



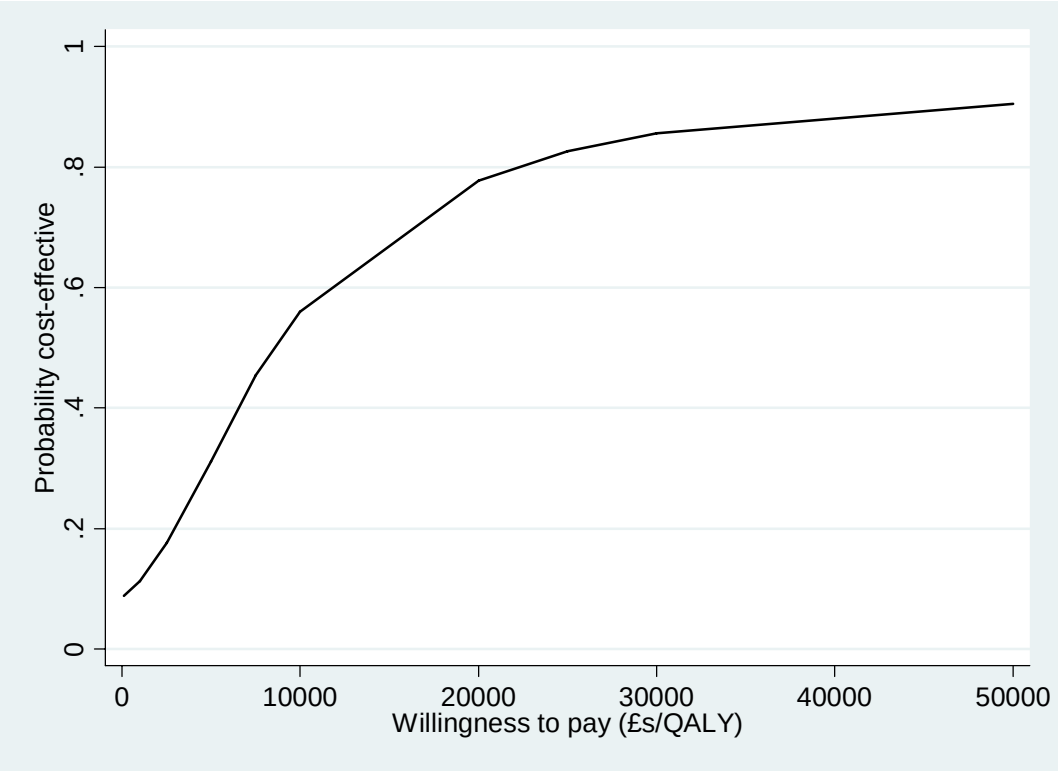
Unfortunately it would be unusual to known the exact position of a treatment comparison on a cost-effectiveness plane and in practice the output of probabilistic models will give a distribution of both cost and clinical effectiveness, and the joint cost-effect distribution (Briggs et al., 2006). Therefore in reality the results from a decision model are likely to produce a plot on the cost-effectiveness plane as given in figure 2.4. A line of best fit can be drawn through the origin, allowing the ICER to be estimated. A further step would be to take the uncertainty into account and plot cost-effectiveness acceptability curves (Drummond et al., 2005), as discussed in section 2.4.6.

2.4.6 Cost-effectiveness acceptability curve

The cost-effectiveness acceptability curve (CEAC) shows the probability that an intervention is more cost-effective than its comparator at different willingness-to-pay values (Drummond et al., 2005, Van Hout et al., 1994). They are calculated by firstly producing a scatter plot, as presented in figure 2.4. A number of lines are then drawn on the plot, each representing a different ICER value the decision maker may be willing to pay for every unit of clinical effectiveness gained. The plotted points below the line will represent estimates where the cost per clinical effectiveness unit gained will be less than what the decision maker is willing to pay, and points above the line will be estimates where the costs are greater than the willingness-to-pay threshold. The percentage of points that fall on or below the line therefore can be interpreted as the probability the intervention is cost-effective compared to its comparator, at that particular willingness-to-pay threshold (λ). This is the same as assessing the number of points with an NMB > 0.

By calculating the probabilities for a number of thresholds, a cost-effectiveness acceptability curve can be constructed (figure 2.5). The graph in turn can then be used to visually assess the cost-effectiveness of a screening strategy compared to a standard treatment and the probability of cost-effectiveness at different willingness-to-pay thresholds can be read from the graph. For example, when considering new treatments and programmes NICE usually use a willingness-to-pay threshold of between £20,000 to £30,000. Therefore any programmes with an ICER less than £20,000 are usually recommended and any between £20,000 to £30,000 considered (National Institute for Clinical Excellence, 2004).

Figure 2.5: An example of a cost-effectiveness acceptability curve



2.4.7 Adjusting costs from different years

Where cost data collected for the model was from different years, the effect of price inflation needed to be accounted for. This can be done by either inflating the data from an earlier year to the chosen year or by deflating the data from a later year, using published cost indices. As the costs utilised were associated with the medical care sector the Hospital and Community Health Service (HCHS) pay and prices index was used (Curtis and Netten, 2006). The use of HCHS index, as presented in equation 2.13, allows medical costs to be measured in a constant currency; that is, in the currency of a fixed (base) year.

$$Base\ year\ price = \frac{HCHS\ index\ base\ year}{HCHS\ index\ actual\ year} \times price\ actual\ year \qquad \text{[Equation 2.13]}$$

2.4.8 Discounting

Discounting is an adjustment to model outcomes, whereby both costs and benefits may be reduced by a certain percent each year, such that the costs and benefits of the first year of the model have a greater weight than costs and benefits in the future. The National Institute of Clinical Excellence (NICE) currently recommend that cost-effectiveness analyses should discount both costs and benefits at a rate of 3.5% per year (National Institute for Clinical Excellence, 2004).

There are a number of reasons why immediate benefits should be seen as preferential to future benefits. Firstly individuals often have a short-term view of life, in that they 'live for today rather than for the future'. Secondly, the future is uncertain, thereby making immediate benefits more advantageous. Thirdly, costs are discounted as, due to positive economic growth in recent decades, individuals may expect to be more wealthy in the future, therefore a pound today is more important than in the future when you are likely to be wealthier (Drummond et al., 2005). When carrying out a cost-effectiveness analysis, it is usual to report both discounted and undiscounted results (Drummond et al., 2005).

2.5 Assessing the effectiveness of screening tests

A series of diagnostic tests are available for identifying the presence of IGT or T2DM. The tests themselves are detailed in chapter 3, but the terminology used to portray the effectiveness of diagnostic tests will be described here. Table 2.1 (adapted from Altman, 1991) shows a general representation of a diagnostic test that has a binary result, negative or positive, based on a binary diagnosis, e.g. presence or absence of a disease.

Table 2.1: A general representation of a diagnostic test

		Disease status		
		Positive	Negative	Total
Test Result	Positive	*a*	*b*	*a + b*
	Negative	*c*	*d*	*c + d*
	Total	*a + c*	*b + d*	*N*

The *sensitivity* of a screening test is defined as the proportion of true positives (diseased) that are correctly identified by the test, that is the probability of testing positive given you have the disease, which from the table is computed as *a/(a + c)*. The *specificity* of a test is the proportion of true negatives (non-diseased) that are correctly identified by the test, in other words the probability of testing negative given you do not have the disease, which using the table is *d/(b + d)*.

In clinical practice it is often more useful to know the probability a test is giving the correct diagnosis. For this predictive probabilities can be used, whereby a *positive predictive value* is the probability of having the disease given you have had a positive screening test result, that is *a/(a + b)* and the *negative predictive value* is the probability of not having the disease given you had a negative screening test result, that is *d/(c + d)* (Altman, 1991). Although they may be more clinically useful, predictive values have the disadvantage that unlike sensitivity and specificity, they are affected by the prevalence of a disease. Therefore if a screening study is carried out the sensitivity and specificity of the test can be generalised to all populations, but the predictive values can only be interpreted in terms of the prevalence of disease in the study sample.

To overcome this, the sensitivity and specificity of a screening test can be used to calculate the predictive values of the test, for any disease prevalence, using the formulae given in equations 2.14 and 2.15 (Altman, 1991). For the comprehensive decision model the positive and negative predictive values were required, and as diagnostic results are more commonly reported using sensitivity and specificity values, the two equations given below, and the prevalence of the screened population, were utilised to specify the first part of the model.

$$\textit{Positive predictive value} = \frac{\textit{sensitivity} \times \textit{prevalence}}{\textit{sensitivity} \times \textit{prevalence} + (1 - \textit{specificity}) \times (1 - \textit{prevalence})}$$

[Equation 2.14]

$$\textit{Negative predictive value} = \frac{\textit{specificity} \times (1 - \textit{prevalence})}{(1 - \textit{sensitivity}) \times \textit{prevalence} + \textit{specificity} \times (1 - \textit{prevalence})}$$

[Equation 2.15]

Receiver operating characteristic (ROC) curves, that is plots of sensitivity against 1-specificity (Altman, 1991), were utilised to obtain optimum cut-offs for screening tests, and these are described in more detail in Chapter 6.

2.6 Summary

This chapter has introduced the statistical theory and methodologies that are utilised throughout this thesis, with more detail given in further chapters where they are employed. A critical review of current comprehensive decision models has been given, to put into context the work that has been carried out for this thesis. Using the methodologies described in this chapter, this work aims to improve on current comprehensive decision models, and produce a fully comprehensive model, whilst developing methodologies that are applicable to other clinical scenarios. The model developed here is the first comprehensive decision model to be developed in the area of diabetes.

3. OVERVIEW OF CLINICAL ISSUES

3.1 Chapter overview

The aim of this chapter is to provide details of the clinical issues relating to IGT and T2DM, which are referred to throughout the thesis. Definitions of the clinical terminology are provided and current screening tests used for IGT and T2DM are described. Important ongoing clinical trials in T2DM are outlined and the growing health problem that is T2DM is discussed, including a brief description of disease progression and possible complications.

3.2 Clinical definitions of glucose tolerance status

Glucose tolerance is measured on a continuous scale, where individuals are either in a fasting or postprandial (that is after a meal or a glucose load) state. Although the scale is continuous, the results are interpreted clinically by categorising the scale. As studies usually report glucose tolerance on a categorical scale and society as a whole has a better understanding, and can interpret the categorical rather than the continuous scale, then this study is restricted to modelling glucose tolerance by category. This has benefits in that it fits easily into the Markov model structure, although it is acknowledged that information is lost, particularly concerning risk of developing further complications associated with T2DM and risk of mortality. The clinical states used for defining glucose tolerance are normal glucose tolerance (NGT), impaired fasting glucose (IFG), IGT, T2DM and other forms of diabetes. These are now all described.

3.2.1 Normal glucose tolerance

Individuals with NGT can be thought of as having a healthy glucose metabolism. NGT, IGT and T2DM are mutually exclusive states, such that as glucose tolerance worsens an individual will pass from NGT to IGT to T2DM. As NGT is the state below which IGT is diagnosed it can be defined as a two-hour plasma glucose level after an oral glucose tolerance test of <7.8mmol/l.

3.2.2 Impaired glucose tolerance and impaired fasting glucose

Impaired glucose regulation, that is either IGT or impaired fasting glucose (IFG), both refer to a metabolic state intermediate between normal glucose homeostasis and diabetes, one in a fasting state and one post-prandial (World Health Organisation, 1999). IGT and IFG are not

interchangeable terms though, and represent distinct forms of abnormal glucose regulation. IGT can be defined as a two-hour plasma glucose level after an oral glucose tolerance test of between 7.8 and 11mmol/l and IFG as a fasting plasma glucose concentration of ≥ 6.1 and <7.0mmol/l. The two conditions are not mutually exclusive and it is possible to be diagnosed as having both.

The determinants of elevated fasting glucose and 2-h plasma glucose in an oral glucose tolerance test (2-HPG) levels differ. Raised hepatic glucose output and a defect in early insulin secretion are characteristic of the former, and peripheral insulin resistance is most characteristic of the latter. Therefore, it is not surprising that the concordance between the categories of IFG and IGT is limited. In all prevalence studies to date only half or less of people with IFG have IGT, and even a lower proportion (20-30%) with IGT also have IFG (Unwin et al., 2002). In the majority of populations studied, IGT is more prevalent than IFG, and there is a difference in phenotype and gender distribution between the two categories. IFG is substantially more common amongst men and IGT slightly more common amongst women. The prevalence of IFG tends to plateau in middle age whereas the prevalence of IGT rises into old age (Unwin et al., 2002).

3.2.3 Type 2 diabetes mellitus

The term diabetes mellitus describes a metabolic disorder characterized by chronic hyperglycaemia with disturbances of carbohydrate, fat and protein metabolism, resulting from defects in insulin secretion, insulin action, or both (World Health Organisation, 1999). T2DM accounts for approximately 85 to 95 percent of all diagnosed cases of diabetes and is usually characterized by insulin resistance, in that target tissues do not use insulin properly (Narayan et al., 2006). The definition of T2DM was revised in the late 1990s after separate recommendations from both the American Diabetes Association (American Diabetes Association, 1997) and the World Health Organisation (World Health Organisation, 1999). Currently an individual with a fasting plasma glucose concentration of 7.0mmol/l and above is classified as having T2DM (previously the cut-off was higher at ≥ 7.8mmol/l).

3.2.4 Other forms of diabetes

In addition to T2DM, there are a number of other clinical forms of diabetes. Type 1 diabetes mellitus results from the destruction of beta cells in the pancreas, leading to absolute insulin deficiency (Narayan et al., 2006). It usually occurs in children and young adults and requires insulin treatment, unlike T2DM which can often initially be controlled by diet alone. Gestational diabetes is first recognised during pregnancy. It does not exclude the possibility that glucose intolerance may pre-date the start of the pregnancy, just that it was previously undiagnosed. Once the pregnancy is over the woman is reclassified as either NGT, IGT or T2DM (World Health Organisation, 1999). Women who have had gestational diabetes are at a higher risk of developing T2DM. Other rare types of diabetes include those caused by genetic conditions (e.g. maturity-onset diabetes of youths), surgery, drug use, malnutrition, infections, and other illnesses (Narayan et al., 2006).

3.2.5 Metabolic syndrome

The metabolic syndrome can be thought of as a cluster of cardiovascular risk factors associated with insulin resistance. It has been defined (NCEP, 2001) as the presence of three or more of the following:

i) Abdominal obesity, waist circumference > 100cm in men and >90cm in women
ii) Triglycerides over 1.68 mmol/l
iii) HDL cholesterol < 1.03 mmol/l in men and < 1.29mmol/l in women.
iv) Blood pressure 130 or more systolic, 85 or more diastolic
v) Fasting glucose > 6.1 mmol/l

Hypertension and diabetes are both common components of the metabolic syndrome with about 40% of persons with the metabolic syndrome having diabetes (Waugh et al., 2007). Individually each component of the metabolic syndrome increases an individuals cardiovascular disease risk, but in combination they become more powerful (Kaplan, 1989). Therefore management of persons with the metabolic syndrome should focus not only on controlling their blood glucose but also on reducing their other risk factors (World Health Organisation, 1999).

If a screening policy for T2DM is introduced in the UK it may produce greater clinical benefit if additional components of the metabolic syndrome are screened for, with a view to treating

them, concurrently. This was not considered for the model developed here, but could provide an extension to the model as briefly discussed in Chapter 9.

3.3 Natural history of IGT, IFG and T2DM

Diabetes is one of the most costly and burdensome chronic diseases of our time and is a condition that is increasing throughout the world, particularly in westernised societies (King et al., 1998), and the prevalence of diabetes is projected to increase dramatically by 2025 (International Diabetes Federation, 2006). The increase in diabetes is associated with lifestyle changes that have led to an increase in obesity and a decrease in physical activity levels.

The complications resulting from diabetes are significant in terms of morbidity and mortality and include damage or failure of the eyes, kidneys and nerves. Individuals with T2DM are also at a significantly higher risk of coronary heart disease, peripheral vascular disease and stroke (American Diabetes Association and National Institute of Diabetes and Digestive and Kidney Diseases, 2002). The World Health Organisation estimates that, in 2001, 959,000 deaths worldwide were attributable to diabetes, accounting for 1.6% of all deaths, although their more recent estimates suggest actual numbers may be triple this (Narayan et al., 2006).

Risk factors for diabetes include a family history of the disease, obesity, physical inactivity, dietary, increasing age, and exposure to diabetes in-utero, with the strongest and most consistent risk factor, across different populations, being obesity and weight gain (Narayan et al., 2006). A study by Lindström et al., whereby a random sample of 4,435 subjects were followed up for 10 years, found a BMI greater than 30kg/m^2 nearly tripled the odds of developing diabetes, odds ratio 2.99 (95% CI: 1.31 to 6.81), and a waist circumference greater than 102cm in men and 88cm in women, gave an increased odds of developing diabetes of 3.86 (1.93 to 7.71) (Lindstrom and Tuomilehto, 2003). Many intervention studies have focused on reducing modifiable risk factors through lifestyle interventions, and these are discussed in detail in Chapter 4.

The transition from the early metabolic abnormalities of IFG and IGT to T2DM may take many years, although it is estimated that up to 70% of these individuals will eventually develop diabetes (Nathan et al., 2007). Although data is limited it appears that individuals with IGT may

have a faster progression to T2DM than IFG individuals (World Health Organisation, 1999) and because IGT has a greater prevalence than IFG in most populations, it is more sensitive for identifying people who will develop diabetes (Unwin et al., 2002). Due to these reasons studies investigating T2DM prevention have focused more on IGT rather than IFG individuals, and as a result the model developed here focuses on IGT also. The model could be extended to include IFG as discussed further in Chapter 10. Individuals with IGT and IFG have a moderate increase in their risk of cardiovascular disease, although the risk is much smaller than that of individuals with T2DM (Nathan et al., 2007).

With prevalence of T2DM increasing, interventions which aim to prevent or delay T2DM need to be considered for future health policies. Obviously not all risk factors for T2DM, such as age and genetic factors, can be modified. Obesity and physical activity are risk factors that could be addressed using intensive lifestyle interventions aimed at improving these risk factors. Intensive lifestyle interventions may encompass a range of initiatives including dietary advice, group exercise and counselling sessions. Pharmacological interventions have also been considered in intervention studies. For example metformin or acarbose may prevent or delay T2DM in individuals at risk of T2DM, by controlling blood glucose levels, thereby addressing the symptoms rather than the underlying cause. The majority of intervention studies have so far been conducted in individuals with IGT, who are known to be at high risk of developing T2DM. The long-term effect on incidence of diabetes complications, if T2DM is delayed, is still unknown, and this issue in relation to modelling screening effects, is discussed further in Chapter 6. Chapter 4 of this thesis discusses a systematic review and meta-analysis of all relevant intervention studies for diabetes prevention, and directly compares the effectiveness of lifestyle against pharmacological interventions.

3.4 Diagnostic and screening tests for glucose tolerance

3.4.1 Screening tests

There are a number of blood tests available which include the oral glucose tolerance test (OGTT), the fasting plasma glucose test (FPG) and the glycated haemoglobin test (HbA_{1c}). Although there is no consensus on the most accurate screening test for detection of diabetes (Bennett et al., 2007), the OGTT is often considered the gold standard of glucose tolerance tests. It is administered in the morning following an overnight fast of between eight to fourteen hours. 75g of anhydrous glucose in 250-300ml of water is consumed over the course of 5 minutes and

blood samples are then collected two hours later. As the OGTT is costly, time-consuming and labour intensive, it is not appropriate to use as an initial screening test for T2DM, but can be used to confirm diabetic status after an initial screening test.

The FPG test also requires a blood sample to be taken after an overnight fast, although is slightly less time consuming than the OGTT. The accuracy of both the OGTT and the FPG test may be affected by the individuals adherence to the overnight fast (Bennett et al., 2007). The HbA_{1c} test has the advantage over the other two blood tests in that it can be taken at any time of day, regardless of food intake, and it is a quick and convenient test, HbA_{1c} levels represent a 2-3month average of blood glucose concentrations.

An alternative screening strategy for diabetes is the use of risk scores, such as the Cambridge risk score (CRS)(Griffin et al., 2000), which is based on data that is often available in general practice records, such as age, weight, family history of diabetes and smoking status. Individuals found to be at high risk based on the risk score, could then be invited to receive a diagnostic test.

3.4.2 Diagnosis of IGT and T2DM

Currently no systematic screening of T2DM is carried out in the UK and a diagnosis of T2DM, is at present, usually made in general practice. If an individual presents with symptoms such as increased thirst and urine volume, recurrent infections, unexplained weight loss, blurred vision, drowsiness, and tingling and numbness in hands and feet, then diagnosis of T2DM can be confirmed by a single FPG test. If no symptoms are present, so for example if a general population were being screened, then diagnosis is made by either two FPG tests on different days, or one FPG and one OGTT. IGT is usually diagnosed using an OGTT test (Waugh et al., 2007). Studies that have modelled the cost and clinical effectiveness of different screening tests, are discussed in detail in Chapter 5.

3.5 Current studies

A substantial amount of research has been carried out in the field of T2DM. This section aims to give some background information on a few of the key studies which provided data for parameters within the decision model. The interventions studies will not be described here as

they are described in detail in Chapter 4, where an account of the systematic review and meta-analysis carried out on these studies is given.

3.5.1 STAR

The Screening those at risk (STAR) study was designed to identify the prevalence of undiagnosed diabetes. The study was conducted in Leicestershire and all individuals aged 40-75 years (Caucasians) or 25-75 years (Non-Caucasians), who had at least one recognised risk factor for diabetes, from 15 general practices, were invited for screening. Risk factors were identified from general practice computer records and included a known history of coronary heart disease, hypertension, dyslipidaemia, cerebrovascular disease, peripheral vascular disease, IGT or IFG, a first degree relative with T2DM and a body mass index greater than 25 kg/m^2. Further risk factors in females included previous gestational diabetes or polycystic ovary syndrome in those that were overweight. Individuals were screened using both a HbA_{1c} and an OGTT test, and those found to have glucose results within the diabetic range were invited for a repeat OGTT to confirm diagnosis.

Results from the STAR study are yet to be published but individual patient data from the study was made available for our model. The STAR results provided data on screening tests and prevalences of IGT and T2DM, which were useful for the decision model, as discussed in Chapter 6 (Davies M et al., 2003).

3.5.2 ADDITION

The ADDITION study is a multi-centre trial in which over 80,000 participants, aged 40-60 years, have had risk scores for T2DM calculated. Over 8,000 at a raised risk of T2DM have so far been invited and attended their GP practices for diabetes screening, with the long-term aim of assessing the clinical benefits of early diagnosis and treatment of T2DM using intensive multi-factorial treatments (Lauritzen et al., 2000). The study is ongoing and in recent years has been extended to clinical practices in the Leicestershire region. As well as assessing the clinical impact of screening for T2DM, the study will also collate important data on prevalence rates, and rates of transition from IGT to T2DM through a series of screenings (Srinivasan et al., 2007). Individual patient data was made available from the Leicester arm of the ADDITION study, for this thesis. It was utilised to provide information on the utilities of T2DM, which were needed for the comprehensive decision model.

3.5.3 UKPDS

The UK Prospective Diabetes Study (UKPDS) was a 20-year trial which recruited 5,102 patients with T2DM in 23 clinical centres based in England, Northern Ireland and Scotland (UK Prospective Diabetes Study (UKPDS) Group, 1991). The primary objective of the trial was to investigate whether tight glucose and blood pressure control in individuals newly diagnosed with T2DM, lowered the risk of diabetes related complications compared to standard treatment. The major results were published in 1998 (UK Prospective Diabetes Study (UKPDS) Group, 1998b, UK Prospective Diabetes Study (UKPDS) Group, 1998a), although subsequent papers on complication rates associated with diabetes (Clarke et al., 2004) and estimated utility rates for diabetes and its complications (Clarke et al., 2002), provided very useful information for populating the decision model. Results from the UKPDS study were utilised to model both costs of T2DM and costs of associated complications, and also to model the utility of T2DM over time. This was done by factoring in the effect of increasing complications associated with duration of T2DM, and the resulting effect of complications on the decrement in the utility values. How this was achieved is described in detail in section 6.5.

3.5.4 DECODE

The DECODE study group (Balkau, 1999, Balkau, 2000) invited researchers in Europe who had carried out population based studies or large studies in occupational groups, of the standard 2-hour OGTT, to participate in a large collaborative data analysis. Only studies with prospective data on mortality and at least 20 sex-specific deaths were analysed. 13 centres provided mortality data, encompassing 25,364 individuals, 1275 who had previously been diagnosed with T2DM. The duration of follow-up was truncated to ten-years to allow comparability between centres. The analysis of data from all 13 centres allowed for assessment of the effect of glucose tolerance status on risk of mortality. The DECODE study provided data for T2DM mortality rates within the decision model.

3.6 Suitability of IGT and T2DM for a screening health programme

Probably the most important clinical issue to address before going forward with this thesis, was whether the conditions of IGT and T2DM meet the requirements necessary for a viable screening policy to be implemented. The UK National Screening Committee (NSC) has compiled a list of criteria for evaluating potential screening programmes, and the full list is available on their website http://www.nsc.nhs.uk/pdfs/criteria.pdf.

Criteria fall under four headings; the condition, the test, the treatment and the screening programme. For the condition requirements include that it is an important health problem, that the epidemiology and natural history is adequately understood, and that a detectable latent period or early symptomatic stage exists. The screening tests need to be simple, safe, validated, precise and acceptable to the public and there should be effective treatment or interventions for individuals identified through screening. There needs to be evidence that a screening programme is effective in reducing morbidity and mortality, and that the benefits of screening outweigh harms and risks.

A full investigation as to whether screening for T2DM and IGT fit all the NSC criteria has previously been carried out (Waugh et al., 2007). It was concluded that of 22 criteria only three were not fulfilled, although a further three were uncertain. Those not met were:

Clinical management of the condition and patient outcomes should be optimised in all healthcare providers prior to participation in a screening programme.

Unfortunately many individuals with T2DM are not well controlled, although measures are being taken to improve this (Waugh et al., 2007).

There should be evidence from high-quality randomised controlled trials (RCTs) that the screening programme is effective in reducing morbidity and mortality.

As yet no RCTs have been carried out assessing the effects of T2DM screening, although the ADDITION study is currently in progress and will report on this. Hopefully simulation studies, such as that being carried out here, can provide an evaluation of the clinical and cost-effectiveness of screening programmes when there is a gap in the clinical data.

Adequate staffing and facilities for testing, diagnosis, treatment and programme management should be available prior to the commencement of the screening programme.

Primary care services and diabetes clinics are already under pressure, therefore a screening policy would need to be brought in slowly.

All three NSC criteria that were not met for the conditions of IGT and T2DM are issues that can be overcome or resolved. Therefore overall IGT and T2DM appear to be suitable conditions for a screening policy, at least in theory, if such a policy could be proved to be clinically and cost-effective. This thesis aims to determine if screening would be preferential to no screening if costs and clinical outcomes are compared, and if so the decision model will aim to identify the optimum screening policy, in terms of the cut-off for glucose tolerance and the target population for screening.

3.7 Summary

This chapter has discussed the clinical issues surrounding this thesis, including the disease pathway from NGT, to IGT and T2DM and subsequent complications, and the current and published research that provided information to populate the comprehensive decision model. The suitability of IGT and T2DM was considered, and whilst current information indicates it is a suitable condition for screening, further information on the effects of early diagnosis and treatment is needed.

4. INTERVENTIONS FOR THE PREVENTION OR DELAY OF TYPE 2 DIABETES MELLITUS

4.1 Chapter overview

For the decision model a pooled statistic, for example a hazard ratio and its associated distribution, that summarised the effect of intervening in IGT individuals to delay T2DM was required. This was considered a key component of the decision model, and therefore both time and effort were taken to estimate this statistic. This chapter will first briefly outline a literature search that was carried out to try and identify such summary statistics in the current published literature. As no relevant statistics were found, a systematic review of intervention studies was carried out for the purpose of this thesis. This review is described here, along with the analyses that were conducted to obtain the statistics necessary for the decision model. The analyses included meta-analyses to obtain pooled estimates of intervention effects, as well as an assessment of the effect of baseline risk on the intervention effect. A mixed treatment comparison analyses was also carried out to enable a direct comparison between different types of interventions, namely lifestyle and pharmacological. The systematic review described in this chapter has been published in the British Medical Journal (Gillies et al., 2007). Further details and an expansion of the analyses are described here.

4.2 Current literature and reviews

For the comprehensive decision model a key parameter was the estimated effect of any interventions on the transition rate from IGT to T2DM. What was required was a relative effect measure, such as a hazard ratio and its associated distribution, that could then be applied to the estimated baseline transition rate from IGT to T2DM. A search of the literature was undertaken to identify all previous reviews or meta-analyses of individuals with IGT, with the aim that these may provide data that could be used for the decision model. Embase, Medline and the Cochrane library were searched using the strategy detailed in appendix 1.1. From the literature search five reviews that had considered interventions for the prevention or delay of T2DM (Angelo et al., 2005, Davies et al., 2004, Norris et al., 2005a, Norris et al., 2005b, Padwal and Laupacis, 2004, Prisant, 2004), and one meta-analysis (Yamaoka and Tango, 2005) were identified. They were generally positive about the benefits of interventions. Norris et al. reported that weight loss interventions produced a significant decrease in T2DM incidence among persons with pre-diabetes. Angelo et al (2005) reported that dietary intervention and

enhanced physical activity are the most effective methods for preventing or delaying the onset of T2DM, whereas pharmacological interventions may be effective in specific high risk individuals, and Padwal and Laupacis (2004) concluded from their review of drug therapies, that currently no single agent could definitively be recommended for diabetes prevention. None of the reviews reported summary statistics on the effectiveness of interventions that could be used for the decision model. The one meta-analysis identified from the literature search combined evidence on lifestyle interventions, but not pharmacological. They concluded lifestyle education compared to controls reduced 1 year incidence of T2DM by nearly 50%, relative risk 0.55 (95% CI: 0.44 to 0.69). As they had included individuals with impaired fasting glucose, their pooled estimate of the intervention effect was not suitable for the comprehensive decision model being developed here. Consequently it was necessary to carry out a further systematic review and meta-analyses of all intervention studies, to obtain the data on intervention effects needed for the model.

4.3 Systematic review

A systematic review was undertaken which aimed to consolidate all the evidence from published intervention trials that had aimed to prevent or delay T2DM. Both lifestyle and pharmacological interventions were considered and meta-analyses of any relevant trials was undertaken to provide summary statistics of intervention effects that could be utilised in the decision model.

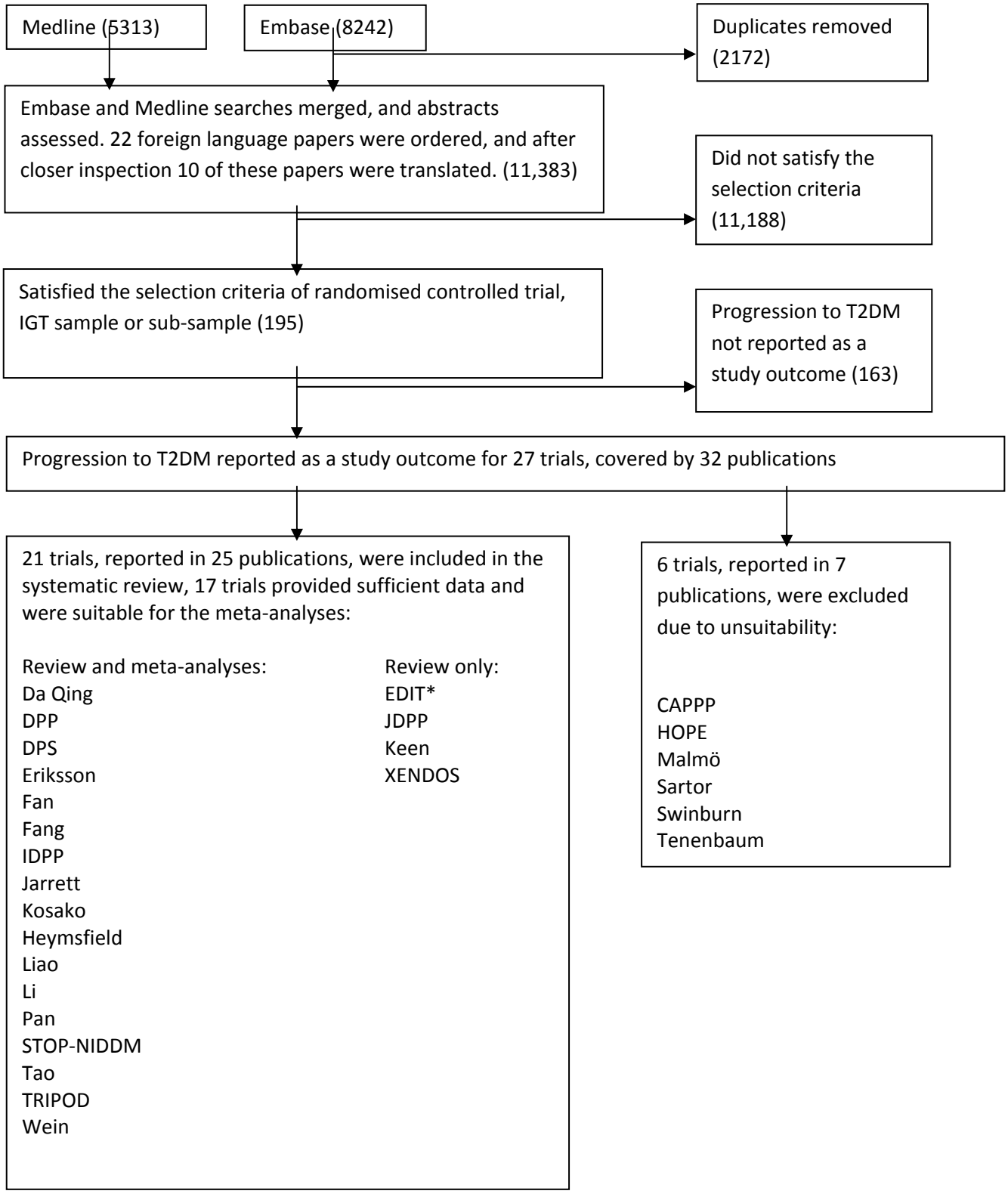
4.3.1 Literature search

Both Medline (1966 to July, week 3, 2006) and Embase (1980 to week 29, 2006) were searched. The search strategies used were developed by combining: (i) phase 1 and 2 of the Cochrane Collaboration's randomised controlled trials filter (Higgins and Green, 2005), (ii) search terms covering both T2DM and prevention, and (iii) clinical terms for IGT. Additionally the Cochrane central register of controlled trials was searched (2006, Issue 2), and expert opinion on relevant trials was sought from Professor Kamlesh Khunti and Professor Jaako Tuomilehto, who are both specialists in the field of diabetes. The references of any articles that met the inclusion criteria, as well as published reviews that considered prevention of T2DM, were also checked. This included a search of the Cochrane library of systematic reviews (2006, Issue 2). The search strategy used is given in detail in appendix 1.2.

Study selection was restricted to randomised controlled trials, to ensure only high quality evidence was included. Studies were selected where an intervention had been applied with the aim of delaying or preventing T2DM in a sample or sub-sample of individuals with IGT. Development of T2DM was a required outcome measure. Trial inclusion was determined by consensus between myself and both my supervisors (Professor Keith Abrams and Dr Paul Lambert). Foreign language papers with relevant titles or English abstracts were assessed jointly by myself, Professor Keith Abrams and a translator. All translators were familiar with medical literature and terminology.

11,383 articles were identified by the Medline and Embase searches (figure 4.1). The titles and abstracts were assessed and the full articles obtained for any that were potentially relevant for this review. 27 of the papers were in English and examination showed they reported 22 trials. A further study, the Early Diabetes Intervention Trial (EDIT) (Holman et al., 2003, Holman et al., 2000), was identified in a published review (Davies et al., 2004). Six of these 22 studies were excluded after obtaining the full papers, either because the treatment allocation process had not been fully randomised (Sartor et al., 1980, Swinburn et al., 2001, Eriksson and Lindgarde, 1991, Eriksson and Lindgarde, 1998), or the primary aim of the administered intervention was not to prevent T2DM (Niklason et al., 2004, Tenenbaum et al., 2004, Yusuf et al., 2001). Ten foreign language papers were assessed, four Chinese, three Japanese, one Spanish, one Russian and one German. Four were excluded as they were discussion papers rather than presenting original findings (Costa, 2002, Anonymous, 1996, Mkrtumian, 2002, Hirose, 2005) and one was excluded as although it met most of our inclusion criteria, T2DM was not a reported outcome (Cao, 2004). Of the five remaining relevant articles two reported results from the Japanese Diabetes Prevention Program (JDPP) (Kuzuya, 2004, Sakane, 2005) and three were results from three separate Chinese studies (Fang et al., 2004, Fan et al., 2004, Tao et al., 2004), and these four trials were included in the review.

Figure 4.1: Flow chart of literature search and meta-analysis



* The EDIT trial was identified from checking existing reviews, not from Medline or Embase

4.3.2 Study characteristics

21 trials met the inclusion criteria for this systematic review (Pan et al., 1997, Knowler et al., 2005, Knowler et al., 2002, Lindstrom et al., 2003a, Eriksson et al., 1999, Fan et al., 2004, Fang et al., 2004, Jarrett et al., 1984, Jarrett et al., 1979, Kosaka et al., 2005, Heymsfield et al., 2000, Liao et al., 2002, Li et al., 1999, Pan et al., 2003, Chiasson et al., 1998, Chiasson et al., 2002, Tao et al., 2004, Buchanan et al., 2002, Wein et al., 1999, Kuzuya, 2004, Sakane, 2005, Keen et al., 1973, Torgerson et al., 2004, Eriksson et al., 2006, Ramachandran et al., 2006). 17 trials, containing 8084 participants provided sufficient information to be included in the meta-analyses (table 4.1). The reasons for not including trials in the meta-analyes are given in section 4.3.3. The trials included were heterogeneous in terms of interventions and participant ethnicity, weight at baseline and age.

Due to the time period covered by the trials, 1979 to 2005, a number of definitions for T2DM and IGT had been used (World Health Organisation Expert Committee, 1994, World Health Organisation Expert Committee, 1980, World Health Orgnisation Expert Committee, 1985, American Diabetes Association, 1997, Alberti and Zimmet, 1998). The majority of definitions are similar, for T2DM they involve a plasma glucose reading two hours after a 75g glucose load of $\geq$ 11.1mmol/l and a fasting plasma glucose level of $\geq$ 7.8mmol/l. For IGT the definition is that of a two hour post glucose load reading of $\geq$ 7.8 and $\leq$ 11.1mmol/l. In 1997 the American Diabetes Association revised the criteria (American Diabetes Association, 1997) and the fasting plasma glucose level was lowered for the definition of T2DM from $\geq$ 7.8mmol/l to $\geq$ 7.0mmol/l. The WHO endorsed this reduction and seven of the more recent studies in this review used this lower threshold (table 4.1). The implication of this is that the more recent studies have a less strict definition of T2DM.

Three relevant studies, Keen et al.(Keen et al., 1973), the EDIT trial (Holman et al., 2003, Holman et al., 2000), and the JDPP trial (Kuzuya, 2004, Sakane, 2005) were not used in the meta-analyses as they reported insufficient data. Contact details for trial authors could only be found for the two most recent trials (EDIT and JDPP), but although further information was requested, none was forthcoming.

Interventions in this review fell into two main types, (i) lifestyle, comprising diet and exercise interventions, and (ii) pharmacological and herbal, comprising oral anti-diabetic agents, the anti-obesity agent orlistat and a Chinese herbal remedy called Jiangtang bushen recipe.

4.3.3 Studies not included in the meta-analyses

Due to a number of reasons not all the studies identified by the literature search were included in the meta-analyses. The Tripod (Buchanan et al., 2002) study and the arm of the DPP trial (Knowler et al., 2005) that assessed troglitazone were omitted from the meta-analyses as, due to safety concerns, troglitazone is no longer a viable intervention for delaying T2DM. Both had shown a significant reduction in development of T2DM with troglitazone. Three trials provided insufficient data for inclusion in the meta-analyses, EDIT (Holman et al., 2003, Holman et al., 2000), Keen (Keen et al., 1973) and JDPP (Sakane, 2005, Kuzuya, 2004). The JDPP trial is ongoing but the preliminary results that have been published report a halving of risk of T2DM in individuals who received diet and exercise advice. The results of the EDIT and Keen trials were less conclusive than the meta-analyses reported here. EDIT (Holman et al., 2003, Holman et al., 2000) found the relative risk of T2DM was significantly reduced by acarbose (0.66, p=0.046), but not metformin (1.09, p=0.70) or combination therapy (0.72, p=0.27). No confidence intervals were reported around these estimates. Keen et al. concluded there was no evidence that either tolbutamide or a carbohydrate restricted diet reduced incidence of T2DM.

Table 4.1: Characteristics of studies included in the review

Trial	Population	Interventions	Definitions of IGT and T2DM
Lifestyle interventions			
DA Qing impaired glucose tolerance and diabetes study, China (1997)	577 Chinese with IGT. All >25 years, 283 males and 247 females	Diet group received individual and group counselling sessions, those with body mass index >25kg/m^2 were encouraged to lose weight. Exercise group were encouraged to increase their daily exercise. Diet and exercise group received both interventions as above. Control group received routine advice	World Health Organisation (WHO) 1985 criteria
Diabetes Prevention Study (DPS) Finland (2003)	522 overweight subjects with IGT, 67% female	Control group received limited advice on diet and exercise, while the intervention group were given tailored, detailed advice on diet, weight reduction and exercise	WHO 1985 criteria
JDPP Japan (2005)	240 participants with IGT, 49% female, mean age 51 years, mean BMI 25 kg/m^2	Control group received standard diet and exercise advice. The intervention group were encouraged to lose weight if necessary, walk for 700 kcals worth per week, and change their diet with the help of a dietician	WHO 1999 criteria
Kosaka Japan (2005)	356 men with IGT, all between 30 and 70 years of age	Control group were advised to lose weight if BMI=>24kg/m^2 and intervention group if BMI>=22kg/m^2 by eating smaller meals and increasing physical activity. Advice repeated every 6 months for controls and 3-4 months for the intervention group	WHO criteria in 1980
Liao USA (2002)	70 Japanese Americans with IGT. 55% female	Intervention group were put on the American Heart Association step 2 diet, plus 1hr endurance exercise 3 times a week. Control group were recommended the less intensive step 1 diet and stretching exercises three times a week	WHO criteria in 1998
Tao China (2004)	60 individuals with IGT. 43% female, aged 34 to 65 years, with a mean age of 51	Both groups received dietary advice. The intervention group also received regular moderately intensive exercise training	WHO 1999 criteria
Wein Australia (1999)	200 women with previous gestational diabetes and currently with IGT	The intervention group received advice on intensive dietary modification, while the controls were given routine advice	WHO 1985 criteria

Table 4.1 continued:

Pharmacological/ herbal interventions			
EDIT UK (2003)	631 subjects, some with IGT. 49% male, 94% white Caucasian	Factorial trial. Subjects randomised to acarbose (50mg three times daily) or placebo and either metformin (500mg three times daily) or placebo	WHO 1985 criteria
Eriksson Finland (2006)	34 individuals with IGT and a first degree relative with T2DM. 35 to 70 years of age, BMI 25 to 35kg/m^2, 74% female	Randomised to either placebo or 2.5mg glipizide daily	WHO criteria in 2006
Fan China (2004)	51 subjects with IGT, over 35 years of age, BMI >19 kg/m^2	All received standard diet and exercise advice. The intervention group additionally took the jiangtang bushen recipe 2-3 times a week	WHO 1999 criteria
Heymsfield USA and Europe (2000)	675 obese adults (120 with IGT), BMI 30-43kg/m^2	Everyone was recommended a low-energy diet then randomised to either placebo or 120mg Orlistat, three times daily	WHO 1985 criteria
Li China (1999)	90 subjects with IGT. All between 30 and 60 years of age	250mg of metformin or placebo three times a day for 12 months	WHO 1985 criteria
Pan China (2003)	261 subjects with IGT, aged 35-70 years, BMI>19 and <=34kg/m^2, 60% female	50mg of acarbose or placebo three times a day	American Diabetes Association (ADA) 1997 criteria
STOP-NIDDM Canada, Germany, Austria, Norway, Denmark, Sweden, Finland, Israel and Spain (2002)	1429 patients, with IGT, 40-70 years, BMI of between 25 and 40kg/m^2	100mg of acarbose or placebo three times a day	WHO 1985 criteria
TRIPOD USA (2002)	266 insulin resistant (167 with IGT), Hispanic women with previous gestational diabetes	400mg troglitazone or placebo once a day, all received standard diet and exercise	IGT diagnosed if the sum of 5 OGTT was >=625mg/dl. T2DM defined by ADA 1997 criteria
XENDOS Sweden (2004)	3277 patients, 694 of who had IGT, age 30-60years, minimum BMI 30kg/m^2	120mg orlistat or placebo three times a day, all patients prescribed a low-cal diet and exercise	WHO 1994

Table 4.1 continued:

Both pharmacological and lifestyle interventions			
Diabetes Prevention Programme (DPP) USA (2002)	3234 subjects with IGT, all 25 years or over. Minimum BMI 24 kg/m^2 (22 in Asians). 32.3% male, 54.7% white.	Four interventions; standard lifestyle recommendations plus placebo, standard lifestyle recommendations plus metformin (850mg twice daily), standard lifestyle recommendations plus troglitazone (400mg daily) and an intensive programme of lifestyle modification. The troglitazone arm was discontinued early due to safety reasons.	ADA 1997 criteria.
Fang China (2004)	178 subjects with IGT, 55% male.	Four interventions; standard prevention education, education and monitoring of diet and exercise, acarbose 25-50mg 3 times a day, or flumamine 125-250mg 3 times a day.	WHO 1985 criteria
Indian Diabetes Prevention Programme (IDDP) India (2006)	531 native Asian Indians with IGT, aged 35-55 years. 21% female.	Four interventions, standard lifestyle advice, lifestyle modification, metformin (500mg twice daily, dropping to 250mg twice daily after a median of 40 days) and fourthly a combination of lifestyle modification and metformin.	WHO 1999 criteria
Jarrett UK (1979)	204 men with IGT.	Factorial trial. Patients received either 50mg phenformin daily or placebo, and were also recommended to either limit their carbohydrate intake to 120g/day or just to limit sucrose (table sugar).	IGT defined as a) survey blood glucose 6.1-11.0mmol/l and b) OGTT peak blood glucose>=10mmol/l and 2hr blood glucose 6.7-11.0mmol/l; or 2 values>10mmol/l; or peak blood glucose>=10.0and mean 2hr glucose>=6.7 T2DM defined as 2 successive 2-hour post glucose blood glucose levels>11.1; 3 non-successive 2hr tests>11.1; the development of symptoms and elevated glucose
Keen UK (1982)	241 subjects with IGT	Factorial trial. Patients were randomised to tolbutamide (0.5g twice daily) or placebo and either dietary teaching to restrict carbohydrate intake to 120g/day or advice to restrict table sugar.	IGT defined as blood glucose levels between 6.7-11.1mmol/l 2 hours after an oral glucose load of 50g. T2DM defined as a 2 hour post load glucose reading >11.1mmol/l

4.3.4 Data extraction

From the trials which could be included in the meta-analyses, myself and a colleague (Professor Kamlesh Khunti) independently extracted data concerning progression to T2DM, body mass index (BMI) at baseline, and age. Disagreements were resolved through discussion. Meta-analyses were conducted on the log hazard ratio scale. Not all the trials reported the necessary summary statistics directly, hence some transformation of and estimation from the reported data was necessary (Parmar et al., 1998, Clayton and Hills, 1993).

Where the hazard ratio (HR) and a confidence interval were reported, as in the DPS, STOP-NIDDM and Wein studies, the *SE* of the *Ln(HR)* was obtained using equation 4.1. *UppCI* and *LowCI* are the value for the upper and lower ends of the confidence interval for *Ln(HR),* Φ^{-1} is the inverse cumulative normal distribution function and α is the percent for the confidence interval.

$$SE[Ln(HR)] \approx \frac{[UppCI - LowCI]}{2\Phi^{-1}\left(1-\frac{\alpha}{2}\right)} \qquad \text{[Equation 4.1]}$$

One study (Xendos) reported only a HR along with a p-value from the log-rank test, but no confidence interval. The SE of the *Ln(HR)* was calculated using the formula given in equation 4.2, where *p* is the reported (two-sided) p-value.

$$SE[Ln(HR)] \equiv \frac{[Ln(HR)]}{\Phi^{-1}\left(1-\frac{p}{2}\right)} \qquad \text{[Equation 4.2]}$$

Six studies (Heymsfield, Jarrett, Kosako, Liao, Li and Pan) did not report results in the form of hazard ratios but instead reported the percentage that developed T2DM. Incidence rates were calculated using information on person years of follow-up (number of cases of diabetes divided by total person years of follow-up). Incidence rate ratios (IRRs) could then be calculated to compare the two intervention groups. Where person years of follow-up was not reported it was

estimated by assuming drop-outs, deaths and development of T2DM had occurred on average half-way through the trial and therefore these individuals added half the trial length to the total person years of follow-up. Those who continued to the end of the trial attributed the full trial length to the total**.** The IRRs could therefore be estimated**.** IRRs can be considered approximations of HRs, although both are modelled under different assumptions. For HRs calculated from a Cox regression model (Collett, 1994), the only assumption is that the HR is constant over time, whereas IRRs additionally assume constant hazards in each of the comparison groups (Clayton and Hills, 1993). This is not an ideal scenario but is the only solution where results are reported as described above. The IRRs were transformed to the log scale for the meta-analyses, and the standard error of the estimated Ln(IRR)s was calculated using equation 4.3**,** where d_T and d_C were the numbers who developed T2DM in the treatment and control arms respectively.

$$SE[Ln(IRR)] \approx \sqrt{\frac{1}{d_T} + \frac{1}{d_C}} \qquad \text{[Equation 4.3]}$$

When data was extracted from the Da Qing trial, the fact that they had assessed three different lifestyle interventions needed to be accounted for, as entering all three intervention effects in to the lifestyle meta-analysis would result in multiple use of the same control group. To adjust for this the number of cases of diabetes and estimated person years for the control group was divided by three, the number of interventions from the trial, and the IRRs then calculated, effectively using a proportion of the control group. Furthermore, as the Da Qing study was randomised at the clinic level, the consequential clustering effect was adjusted for by reanalysing the reported data by fitting a Poison regression model, with clinic included as a random effect (γ) (Equation 4.4). As number of events (E) is modelled, rather than rate, person years of follow-up are entered as an offset in the linear predictor (Y), t representing treatment group (control or intervention), j clinic and the intervention effect is represented by β, with x as a dummy variable taking the value 0 for the control group and 1 for the intervention group.

$$\begin{aligned} &E_{tj} \sim Poisson(\mu_{tj}) \\ &Ln(\mu_{tj}) = Ln(Y_{tj}) + \beta x_t + \gamma_j \\ &\gamma_j \sim N(0, \sigma^2) \end{aligned} \qquad \text{[Equation 4.4]}$$

All studies included in the meta-analyses were assessed for quality, using the Jadad score (Jadad et al., 1996). The Jadad score is a simple tool that assesses the quality of a clinical trial by giving a total mark out of 5 for the three following questions;

i) Was the study described as randomised?

ii) Was the study described as double blind?

iii) Was there a description of withdrawals and dropouts?

A more detailed description of the Jadad score is given in appendix 2. As it is an important aspect of quality not included in the Jadad score, concealment of allocation was also assessed as an additional measure of trial quality.

4.4 Analyses and Results

A number of analyses were carried using data from the identified studies, including meta-analyses (methodology described in section 2.3.4) of the intervention trials, using both a Bayesian and a traditional frequentist approach, a mixed treatment comparison analysis (methodology described in section 2.3.6), an assessment of sources of study heterogeneity through meta-regression analyses (methodology described in section 2.3.5) and an assessment of baseline risk on the effectiveness of interventions. The analyses were carried out using either maximum likelihood estimation (MLE) in Stata (StataCorp, 2001), or MCMC methodologies in WinBUGS (Spiegelhalter et al., 2000).

4.4.1 Meta-analyses of the intervention trials

Meta-analyses were carried out to combine the hazard ratios (*HR*) from each study into pooled estimates. As the trials varied considerably in terms of participants and intervention duration, random effect models were fitted to allow for the presence of between study heterogeneity (equation 4.5, *i* represents study, HR_i is an estimate of effect size, θ_i is the true effect size and s_i^2 is the observed variance of y_i for study *i*, μ is the pooled estimate of effect size, and τ^2 represents the between-study variance) (Sutton et al., 2000).

$$Ln(HR_i) \sim N(\theta_i, s_i^2) \qquad \text{[Equation 4.5]}$$

$$\theta_i \sim N(\mu, \tau^2)$$

Meta-analyses were carried out separately for lifestyle interventions, anti-diabetic agents and anti-obesity agents in both Stata and WinBUGS (code in Appendix 3.1). The lifestyle interventions were further stratified by whether the intervention consisted of diet alone, exercise alone, or a combination of diet and exercise regimes. The pharmacological intervention troglitazone was not included in the meta-analyses as this drug has been withdrawn from a number of markets worldwide due to problems of liver toxicity (Knowler et al., 2005). The data used for the meta-analyses is given in table 4.2. Sensitivity analyses, using both half normal and uniform distributions, were carried out to check whether the prior distribution placed on the between study variance (τ^2), was influencing the results of the meta-analyses, with the results discussed in section 4.4.2. Distributions commonly used for τ^2, and the importance of carrying out sensitivity analyses on this prior, are discussed in detail in Chapter 2.

Additionally the pooled hazard ratios from the meta-analyses, together with the pooled hazards of developing T2DM from the control arms of the trials, were used to estimate the difference intervening would make in 5 year probability of developing T2DM (*P*) and the associated number needed to treat, under the assumption of a constant hazard (equation 4.6, where μ represents the log *HR* of T2DM from the meta-analyses, α the baseline hazard of T2DM in the control groups of the intervention trials and δ the absolute difference in 5 year probability of T2DM). The WinBUGS code for this is given in Appendix 3.2, more detail on NNT is given in Chapter 2. The results of the meta-analyses and NNT calculations are presented in section 4.4.2.

Control groups	$P_C \equiv 1 - \exp(-5\alpha)$	[Equation 4.6]
Intervention group	$P_T = 1 - \exp[-5\alpha.\exp(\mu)]$	
Absolute difference	$\delta = \kappa_C - \kappa_T$	
NNT	$\eta = \frac{1}{\delta}$	

Table 4.2: Information used for the meta-analyses

Trial and intervention	Log hazard ratio (SE)*	Mean BMI (kg/m^2)	Mean age (years)	Average follow-up (years)	Baseline risk†	Quality assessment**				
						Concealed allocation	Randomised	Blinded	Dropouts	Jadad score
Lifestyle										
Da Qing, Diet	-0.45 (0.22)	25.8	45.6	4.51*	15.7	No	1	0	1	2
Jarrett , Diet	-0.17 (0.39)	26.2	56.7	4.39*	2.6	No	1	0	1	2
Wein, Diet	-0.46 (0.30)	25.4	38.7	4.25	7.1	No	1	0	1	2
Da Qing, Exercise	-0.64 (0.23)	25.8	45.3	4.62*	15.7	No	1	0	1	2
Tao, Diet and exercise	-1.20 (0.57)	25.4	51.0	2.58	17.0	No	1	0	1	2
Da Qing, Diet and exercise	-0.49 (0.23)	26.3	45.5	4.52*	15.7	No	1	0	1	2
DPP, Diet and exercise	-0.87 (0.11)	34.0	50.4	2.80	11.0	No	1	0	0	1
DPS, Diet and exercise	-0.92 (0.22)	31.2	55.0	3.20	7.4	No	1	0	0	1
Fang, Diet and exercise	-0.29 (0.39)	25.0	48.0	3.88*	10.0	No	2	0	1	3
IDDP, Diet and exercise	-0.47 (0.20)	25.8	45.9	2.50	18.3	No	1	0	1	2
Kosaka, Diet and exercise	-1.24 (0.60)	23.8	51.5*	3.64	2.6	No	1	0	1	2
Liao, Diet and exercise	-0.66 (1.22)	26.1	54.0	1.83*	3.1	No	1	0	1	2
Pharmacological/herbal										
Fang, Acarbose	-1.31 (0.55)	24.8	48.7	4.14*	10.0	No	2	0	1	3
Pan, Acarbose	-0.51 (0.48)	25.7	54.5	0.37	30.0	No	1	2	0	3
STOP, Acarbose	-0.29 (0.09)	30.9	54.5	3.30	12.6	Yes	2	0	1	3
Fang, Flumamine	-0.84 (0.49)	25.0	48.7	4.06*	10.0	No	2	0	1	3
Eriksson, Glipizide	-1.74 (1.10)	28.1	56.5.	1.32	23.8	No	1	2	1	4
DPP, Metformin	-0.37 (0.10)	34.0	50.6	2.80	11.0	No	1	1	0	2
IDDP, Metformin	-0.43 (0.20)	25.8	45.9	2.50	18.3	No	1	0	1	2
Li, Metformin	-0.72 (0.71)	26.2	49.5	0.92	7.1	No	1	1	1	3
Jarrett, Phenformin	0.01 (0.39)	26.2	56.7	4.36*	2.6	No	1	1	1	3
Heymsfield, Orlistat	-0.95 (0.35)	35.8	44.1	1.59	4.8	No	1	1	0	2
Xendos, Orlistat	-0.73 (0.31)	37.3	43.0	2.78	5.9	Yes	1	2	0	3
Fan, Jiangtang bushen recipe	-1.14 (1.15)	25.5	56.0	0.90*	13.3	No	2	0	1	3

*estimated from, or a transformation of, the original data † incidence of T2DM per 100 person years in the control group

** Quality was assessed by allocation of concealment and the Jadad score. The Jadad score comprised of randomisation and blinding (both marked out of 2), and the description of withdrawals and dropouts (marked out of 1). The total Jadad score was therefore out of 5.

4.4.2 Results of the meta-analyses

The results of the meta-analyses and sensitivity analyses are given in table 4.3. The MLE models were fitted in Stata and the MCMC models in WinBUGS. All the MLE meta-analyses, apart from the one trial that had assessed a herbal intervention, provided overwhelming evidence to support the benefit of interventions to prevent or delay T2DM. The pooled effect for all forms of lifestyle interventions gave a hazard ratio of 0.51 (95% confidence interval, 0.44 to 0.60), p<0.001 (figure 4.2), indicating a relative 49% reduction in risk of developing T2DM. When diet, exercise, and diet and exercise in combination were considered separately, they all showed a similar reduction in risk, hazard ratios 0.67 (0.49 to 0.92), p=0.013, 0.49 (0.32 to 0.74), p=0.001, and 0.49 (0.40 to 0.59), p<0.001, respectively.

Both forms of pharmacological intervention, oral anti-diabetic agents and the anti-obesity agent, also showed a highly significant benefit of intervention compared to control, hazard ratios 0.70 (0.62 to 0.79), p<0.001, and 0.44 (0.28 to 0.69), p<0.001, respectively (figure 4.3). The one trial that assessed a herbal intervention had a hazard ratio favourable to the intervention, although this was non-significant, 0.32 (0.03 to 3.07), p=0.323.

The pooled estimates were robust to varying the distribution of the between study standard deviation (τ) (Table 4.3). All the MCMC estimates had greater uncertainty around the pooled hazard ratio compared to the MLE analyses, due to their inclusion of uncertainty around τ, and also because a larger value of τ increases the uncertainty around the pooled estimate. For a number of the MLE analyses τ was estimated as 0. This is the minimum allowed by Stata, although in fact the data may have supported a value < 0. For the MCMC analyses, as a positive value of τ is sampled for each iteration, the mean value estimated must always be > 0.

Where several studies were combined in the meta-analyses, for example the analyses of lifestyle interventions, varying the prior distribution of τ had minimal impact on the pooled hazard ratio. The meta-analyses where only two studies were combined were more sensitive to the prior distribution on τ. This is to be expected as where there are fewer studies to estimate the variance, the prior distribution on τ will have a greater impact, and therefore problems occur if vague priors are used for meta-analyses with few studies (Browne and Draper, 2000, Lambert et al., 2005). The largest prior used in the sensitivity analyses of $\tau \sim Uniform(0,10)$, covers an extremely large range of possible values for the hazard ratio, as demonstrated by (Spiegelhalter

et al., 2004). From table 4.3 it can be seen that the pooled estimates from the MLE and MCMC analyses, were most similar for the overall pooled effect of lifestyle interventions, which were the meta-analyses containing the most number of studies. For the comprehensive decision model, the prior distribution used for the intervention meta-analyses was $\tau \sim Uniform(0,2)$.

Table 4.3: Results of the meta-analyses

Intervention	Study N	Model	Hazard ratio (95% CI/CrI)	Between study standard deviatio (95% CrI)
Anti-diabetic agents	9	*MLE*	0.70 (0.62, 0.79)	0.00
	9	*MCMC, τ ~ Uniform(0,2)*	0.66 (0.47, 0.83)	0.20 (0.014, 0.656)
	9	*MCMC, τ ~ Uniform(0,10)*	0.66 (0.47, 0.82)	0.19 (0.003, 0.656)
	9	*MCMC, τ ~ Normal(0,1)I(0,)*	0.67 (0.49, 0.82)	0.17 (0.001, 0.588)
Anti-obesity agent	2	*MLE*	0.44 (0.28, 0.69)	0.00
	2	*MCMC, τ ~ Uniform(0,2)*	0.56 (0.10, 1.86)	0.70 (0.023, 0.550)
	2	*MCMC, τ ~ Uniform(0,10)*	0.44 (0.01, 81.08)	2.33 (0.036, 9.039)
	2	*MCMC, τ ~ Normal(0,1)I(0,)*	0.52 (0.13, 1.33)	0.52 (0.018, 1.749)
Lifestyle (all)	12	*MLE*	0.51 (0.44, 0.60)	0.08
	12	*MCMC, τ ~ Uniform(0,2)*	0.53 (0.43, 0.64)	0.16 (0.015, 0.438)
	12	*MCMC, τ ~ Uniform(0,10)*	0.53 (0.43, 0.65)	0.16 (0.015, 0.438)
	12	*MCMC, τ ~ Normal(0,1)I (0,)*	0.53 (0.43, 0.64)	0.17 (0.012, 0.423)
Diet only	3	*MLE*	0.67 (0.49, 0.92)	0.00
	3	*MCMC, τ ~ Uniform(0,2)*	0.74 (0.31, 1.57)	0.43 (0.016, 1.604)
	3	*MCMC, τ ~ Uniform(0,10)*	0.69 (0.18, 2.84)	0.73 (0.014, 4.864)
	3	*MCMC, τ ~ Normal(0,1)I (0,)*	0.73 (0.34, 1.29)	0.35 (0.010, 0.235)
Exercise only	2	*MLE*	0.49 (0.32, 0.74)	0.00
	2	*MCMC, τ ~ Uniform(0,2)*	0.57 (0.09, 1.89)	0.77 (0.03, 1.90)
	2	*MCMC, τ ~ Uniform(0,10)*	0.44 (0.01, 119.9)	2.64 (0.065, 9.089)
	2	*MCMC, τ ~ Normal(0,1)I (0,)*	0.53 (0.12, 1.41)	0.59 (0.017, 0.454)
Diet and exercise combined	7	*MLE*	0.49 (0.40, 0.59)	0.10
	7	*MCMC, τ ~ Uniform(0,2)*	0.50 (0.37, 0.70)	0.25 (0.014, 0.810)
	7	*MCMC, τ ~ Uniform(0,10)*	0.50 (0.36, 0.71)	0.27 (0.017, 0.845)
	7	*MCMC, τ ~ Normal(0,1)I (0,)*	0.50 (0.37, 0.69)	0.24 (0.013, 0.675)

Figure 4.2: Meta-analysis of the effect of lifestyle interventions on the risk of developing T2DM

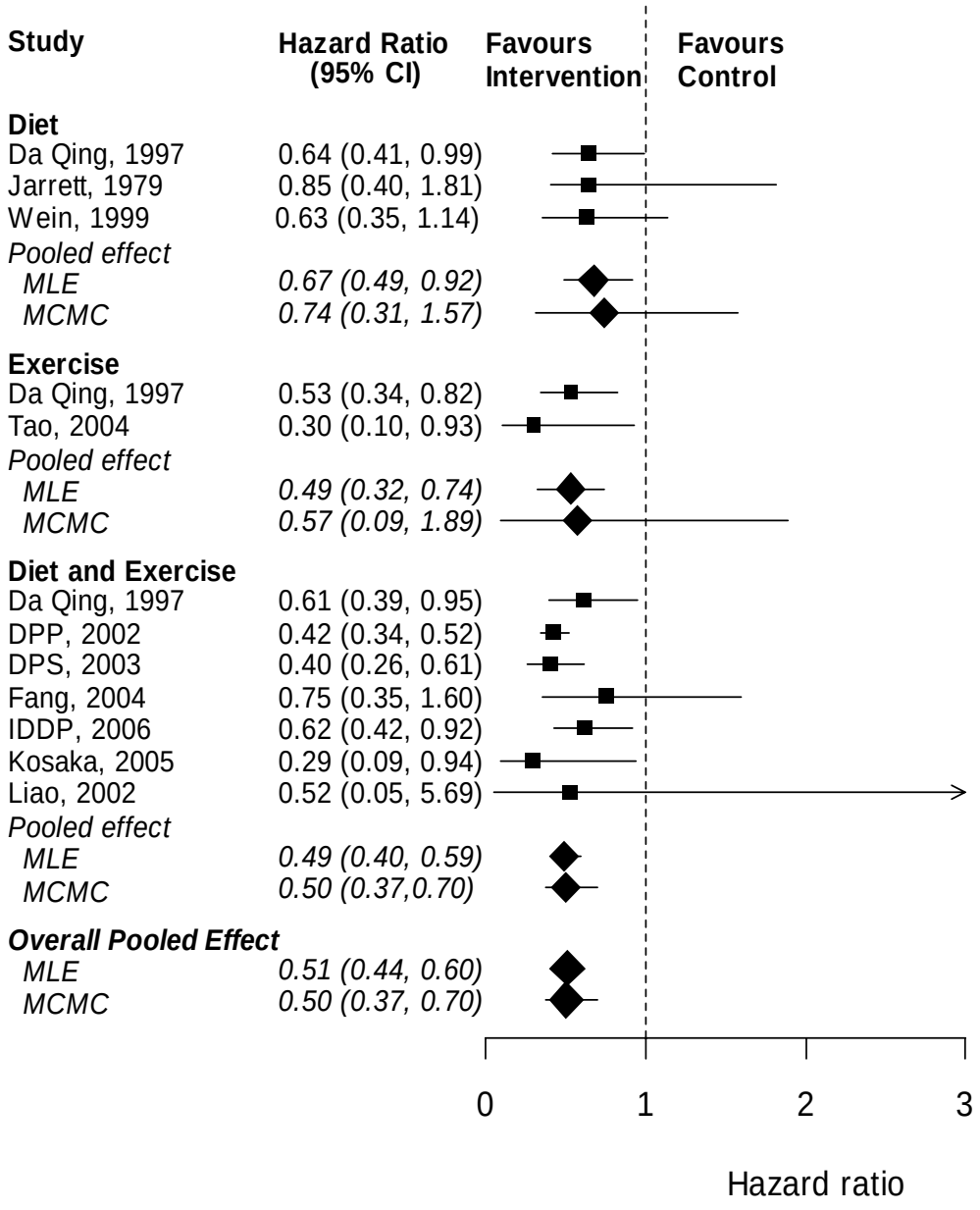
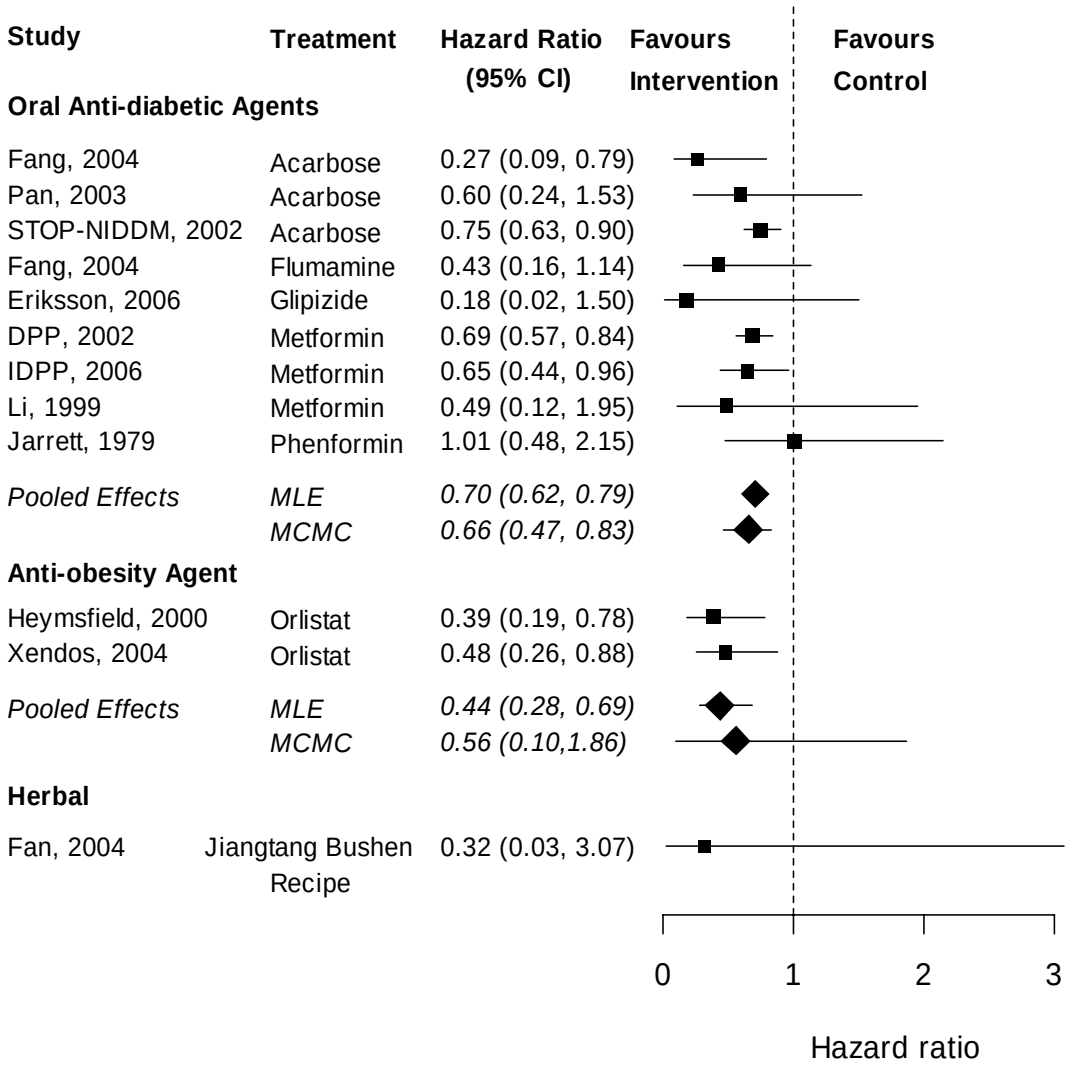


Figure 4.3: Meta-analyses of the effect of pharmacological and herbal interventions on the risk of developing T2DM



Combining the baseline rate of T2DM in the control arms of all 17 trials gave a probability of developing T2DM over 5 years of 37.1%, which is in line with previously reported estimates (de Vegt et al., 2001, Edelstein et al., 1997b). Using the pooled hazard ratios from the meta-analyses, as detailed in equation 4.6, the absolute difference in probability of T2DM in terms of percentage points would be -16 (95% credible interval, -12 to -20) by the lifestyle intervention, -9 (-6 to -12) by oral anti-diabetic agents, -18 (-13 to -24) for orlistat and -20 (-38 to 12) for the jiangtang bushen recipe (table 4.4). These were used to calculate numbers needed to treat (NNT), where NNTB infers a benefit, that is the number needed to be treated with the intervention compared to the control treatment to prevent or delay one case of T2DM, and NNTH infers a harming effect of the intervention, that is the number needed to be treated by the control treatment compared to intervention, to prevent or delay one case of T2DM (Altman, 1998). NNT were, for lifestyle 6.4 (95% credible interval, NNTB 5.0 to NNTB 8.4), oral anti-diabetic agents 10.8 (NNTB 8.1 to NNTB 15.0), orlistat 5.4 (NNTB 4.1 to NNTB 7.6) and jiangtang bushen recipe 4.0 (NNTH 16.9 to NNTB 24.8).

Table 4.4: Results from the number needed to treat analysis

Intervention	5-year probability of developing T2DM in *all* control group (P_C)	5-year probability of developing T2DM in the intervention group (P_T)	Absolute difference (δ)	Number needed to treat (η)
Anti-diabetic agents	0.37 (0.27, 0.48)	0.27 (0.20, 0.37)	0.09 (0.06, 0.12)	10.8 (8.1, 15.0)
Anti-obesity agents	0.37 (0.27, 0.48)	0.19 (0.12, 0.26)	0.18 (0.13, 0.25)	5.4 (4.1. 7.6)
Lifestyle	0.37 (0.27, 0.48)	0.21 (0.15, 0.29)	0.16 (0.12, 0.20)	6.4 (5.0, 8.4)
Herbal remedy	0.37 (0.27, 0.48)	0.17 (0.03, 0.49)	0.20 (-0.12, 0.38)	4.0 (-16.9, 24.8)

* values in parenthesis are 95% CrI

4.4.3 Factorial trials

The factorial trials present in the review were interesting in that the results they provided meant they were dealt with differently in the standard meta-analyses compared to the mixed treatment comparisons. The trial by Jarrett et al. (Jarrett et al., 1979) used a factorial design, assessing one pharmacological and one lifestyle intervention. For the standard meta-analyses it was assumed there was no interaction between treatments, and groups were combined, so that all individuals were entered into both the pharmacological and lifestyle meta-analyses. As the lifestyle and pharmacological meta-analyses were carried out separately, then there was no

issues over double counting, where individuals were used in both. For the MTC analyses the four groups, placebo and control diet, placebo and intensive diet, phenformin and control diet and phenformin and intensive diet, were entered separately.

The IDDP trial (Ramachandran et al., 2006) included an arm that combined treatments in that individuals were randomised to a control, lifestyle intervention, metformin and finally a combination of lifestyle and metformin. For the standard meta-analyses the hazard ratios reported for the three arms assessing individual treatments were used. For the MTC analyses incidence rates per 100 person years were calculated for each of the four treatment groups. This second method has the limitation that constant incidence rates must be assumed.

The Da Qing trial (Pan et al., 1997) also included an arm that combined treatments. Treatment groups were control, diet, exercise and then diet and exercise combined. For the standard meta-analyses incidence rate ratios were used for each of the three lifestyle interventions compared to the controls, allowing for the multiple use of the control arm. For the MTC all three treatment groups were combined and compared against controls. The MTC analysis is discussed further in section 4.4.4.

4.4.4 Mixed treatment comparison (MTC) analyses

To enable a direct comparison between lifestyle and anti-diabetic interventions a mixed treatment comparison (MTC) model was fitted in WinBUGS (appendix 3.3). MTC meta-analysis is a generalization of standard pair wise meta-analysis for A vs. B trials, to more complex data structures such as A vs. B, B vs. C and A vs. C, as described in section 2.3.6. MTC facilitates simultaneous comparisons of all treatments to allow the selection of the best treatment. The model allows for correlation between treatment arms from the same study (ρ_{ij}) (Lu and Ades, 2004), α represents the log hazard rate in the control group, and γ the difference in log hazard between an intervention and the controls, *i* represents study and *t* intervention. Thus the model can be specified as in equation 4.7, where α_i is the study effect, τ is the between study standard deviation, d_{β_i} is the baseline treatment in study *i* and d_{Ti} is the comparator treatment:

$$
\begin{aligned}
&Ln(HR_i) \sim N(\theta_i, s_i^2) \\
&\theta_i = \alpha_i + \delta_{it} \\
&\delta_i \sim N(\mu_i, \tau^2) \\
&\mu_i = d_{\beta i} - d_{Ti}
\end{aligned}
\qquad \text{[Equation 4.7]}
$$

with vague prior distributions specified as:

$$
\begin{aligned}
&\alpha_i \sim N(0,100^2) \\
&\tau \sim Unif(0,2) \\
&d_{\beta i} \sim N(0,100^2) \\
&d_{Ti} \sim N(0,100^2)
\end{aligned}
$$

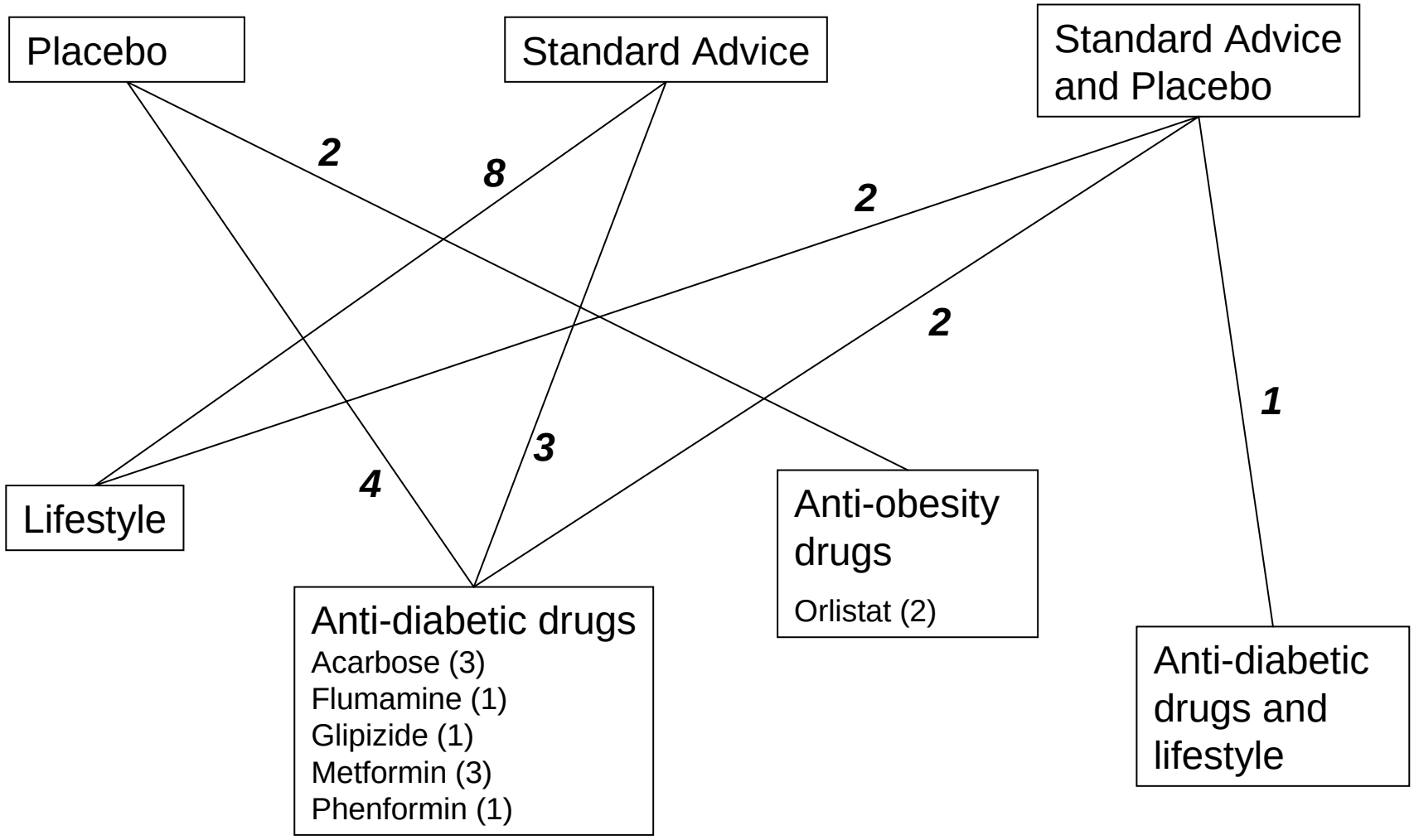
The graph in figure 4.4 represents the network of treatment comparisons made by the 17 trials included in the meta-analyses. For example eight trials compared standard advice against a lifestyle intervention, and two trials compared orlistat with placebo. Where the evidence is linked in such a manner, i.e. a pathway can be traced between all treatments, then indirect comparisons can be made between treatments that were not compared directly within a trial. For the purposes of the MTC, the three control groups (placebo, standard advice and both placebo and standard advice) in graph 4.4 were combined.

As different baselines (placebo, standard lifestyle advice etc.) were combined in the model, an initial assessment, using maximum likelihood estimation, was carried out to make sure baseline risk did not differ between forms of controls. For example trials involving placebos might have been associated with patient groups at greater risk of T2DM. Three separate meta-analyses were carried out to combine the log incidence rate of diabetes in control groups that had received standard lifestyle advice, those that had received placebo and those that had received both. The comparison showed all three types of baseline groups had similar baseline risk, standard lifestyle advice (11 studies) pooled log incidence rate per 100 person years 2.48 (95% CI: 2.37 to 2.59), placebo (6 studies) 2.46 (2.36 to 2.57) and control groups receiving both standard advice and placebo (5 studies) 2.44 (2.38 to 2.50). All had significant within group heterogeneity, but between group heterogeneity was non-significant, p=0.327. With the control groups combined in the model, an indirect comparison of lifestyle against anti-diabetic drugs

resulted in a non-significant hazard ratio of 1.061 (95% CrI: 0.704 to 1.461). Therefore there is currently no evidence to support one intervention being more effective than the other.

The above model also allowed for an assessment of the presence of an interaction effect for when interventions were administered in combination. Lifestyle and anti-diabetic treatments were administered together in both the Jarrett and IDDP trials. The hazard ratio of lifestyle and anti-diabetic agents in combination, versus controls, compared to the effect of lifestyle and the effect of anti-diabetic agents simply added together (when added on the log scale), was 1.406 (95% CrI: 0.6794 to 2.538). The credible interval is wide as only two studies assessed the effect of giving lifestyle and anti-diabetic treatments in combination. This suggests that although the null value of 1 is within the 95% credible interval, a synergistic effect if treatments are administered in combination, cannot be ruled out.

Figure 4.4: Network of type 2 diabetes prevention trials



4.5 Exploration of heterogeneity, study quality and publication bias

Further analyses, as described below, were conducted separately for lifestyle interventions, oral anti-diabetic agents, orlistat and jiangtang bushen recipe, although not all analyses could be carried out for the last two categories due to the small number of trials.

4.5.1 Assessment of heterogeneity

Between study heterogeneity was quantified by the I^2 statistic, as described in Chapter 2, which measures the proportion of inconsistency in individual studies that cannot be explained by chance (Higgins et al., 2003), and explored through meta-regression models (Sutton et al., 2000), where the study level covariates mean age, mean body mass index and length of follow-up were individually assessed (WinBUGS code detailed in appendix 3.4). A generic meta-regression model is given in equation 4.8, where Y_i represents the intervention effect, *i* study, *x* the covariate of interest, θ_i is the true effect size when *x*=0, s_i^2 is the observed variance of Y_i, μ is the pooled estimate of effect size, and τ^2 represents the between-study variance.

$$Y_i \sim N(\theta_i + \beta x_i, s_i^2) \qquad \text{[Equation 4.8]}$$

$$\theta_i \sim N(\mu, \tau^2)$$

The I^2 statistics indicated that 0% of the variation in the anti-obesity agents and the oral anti-diabetic agents meta-analyses, and just 8.8% in the lifestyle meta-analysis, was due to between study heterogeneity. The results of the meta-regression analyses are reported in table 4.5 and figures 4.5 and 4.6, where the size of the circle represents the SE of the Log hazard ratio and the weight given to each study estimate. For lifestyle interventions each 1 kg/m^2 increase in the mean body mass index of a study at baseline, led to a decrease in the hazard ratio of -7.3% (-13.6 to -0.9), p=0.029. This provides evidence that as the average trial body mass index at baseline increased, the effectiveness of the lifestyle intervention also increased, meaning that lifestyle interventions were more effective in trials that recruited participants with higher body mass index values. From graph 4.5 a) though, it can be seen that two studies with high mean BMI, appeared to have high leverage, and hence were possibly influencing this result.

The between study standard deviation (τ), which was estimated as 0 for anti-diabetic agents in the MLE meta-analysis, and 0.08 for lifestyle interventions (table 4.3), was reduced to 0 for lifestyle interventions in the meta-regression analyses, when some of the between study heterogeneity was accounted for (table 4.5). For anti-diabetic agents, the between study standard deviation could not be reduced further, and was estimated as 0 for the meta-regression analyses.

Table 4.5: Results of the meta-regression analyses

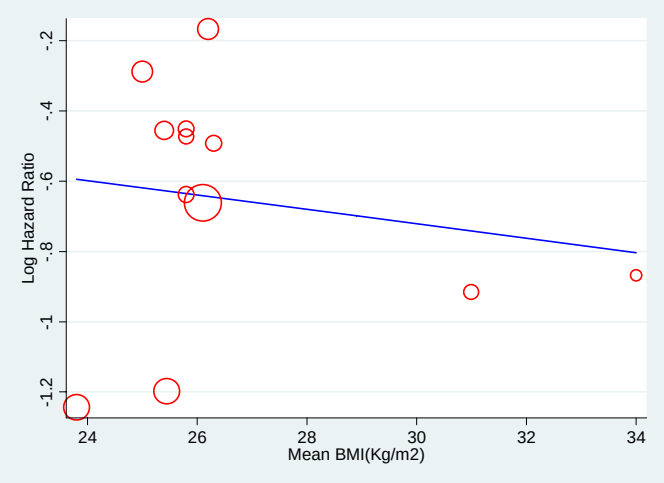
Intervention	**Covariate**	**Coefficient* (95% CI)**	**Percentage change in the hazard ratio (95% CI)**	τ	**P-value**
Lifestyle	Mean BMI (kg/m^2)	-0.04 (-0.08, -0.01)	-7.3 (-13.6, -0.9)	0	0.029
	Mean age (yrs)	-0.03 (-0.07, 0.01)	-3.8 (-8.6, 1.0)	0	0.106
	Follow-up (yrs)	0.15 (-0.04, 0.35)	9.5 (-3.3, 30.0)	0	0.108
Anti-diabetic agents	Mean BMI (kg/m^2)	0.02 (-0.03, 0.07)	1.9 (-4.1, 7.8)	0	0.482
	Mean age (yrs)	0.03 (-0.02, 0.08)	1.5 (-1.4, 4.4)	0	0.257
	Follow-up (yrs)	0.08 (-0.17, 0.33)	13.1 (-30.0, 56.3)	0	0.495

* the coefficient represents the change in the log hazard ratio for a 1 unit increase in the covariate.

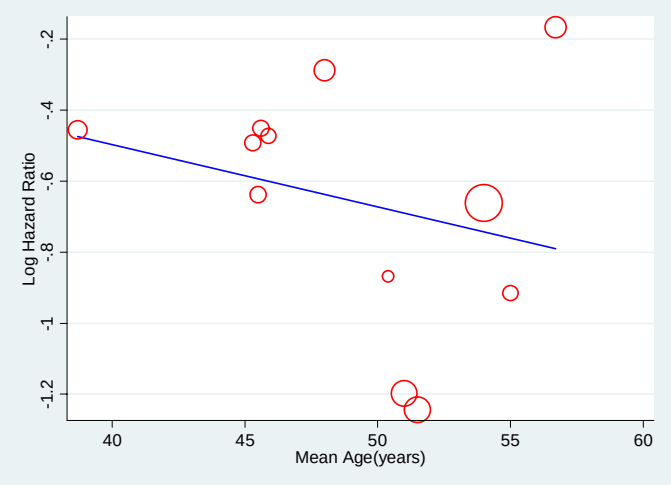
The meta-regression analyses were carried out using aggregated patient data, in that it was the study level mean BMI at baseline, mean age and mean length of follow-up that were utilised in the analyses, and not individual patient data. This is problematic, as discussed in section 2.3.5, in that not only do meta-regression analyses lack power to identify interactions between study level covariates and estimated intervention effects, interpreting results can lead to incorrect conclusions. Problems of 'ecological bias' may be introduced, in that although a relationship may appear to exist at the study level, it may not be found to be true if individual patient data was assessed (Lambert et al., 2002). So although it appears intervention effects increase with the mean BMI of a study, if individual patients within a trial were examined it may be found intervention effects actually decrease with BMI, or in fact no relationship exists. Methods have recently been developed that combine the use of both IPD and summary data to assess sources of heterogeneity (Jackson et al., 2006).

Figure 4.5: Association between study level covariates and the effectiveness of lifestyle interventions

a) Mean study BMI (Kg/m^2)



b) Mean age of study participants (years)



c) Mean length of follow-up (years)

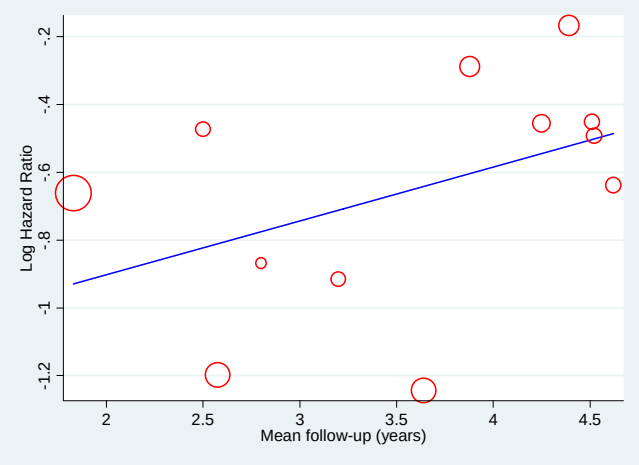
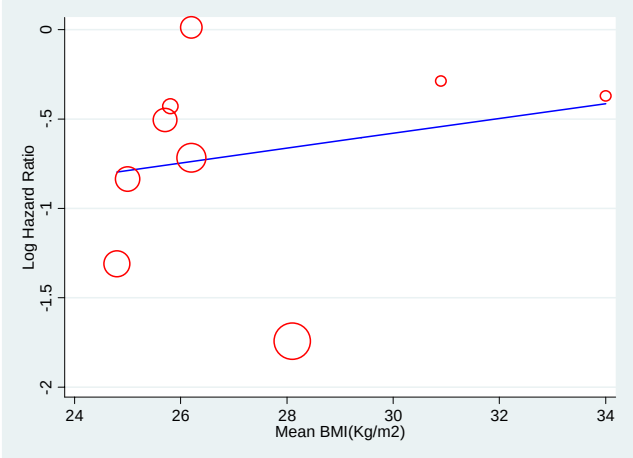
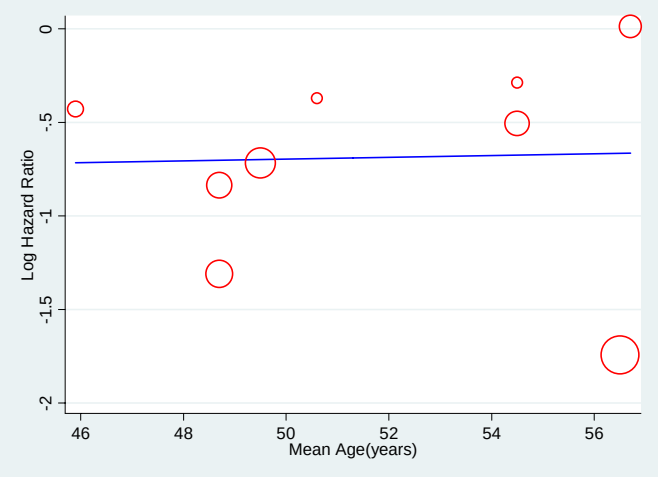


Figure 4.6: Association between study level covariates and the effectiveness of anti-diabetic interventions

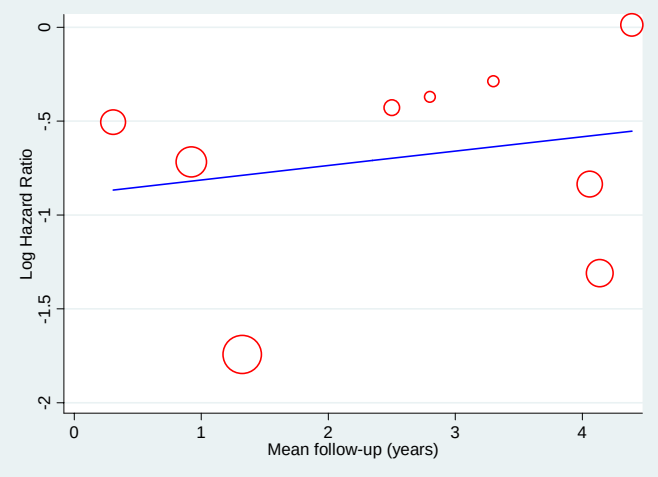
a) Mean study BMI (Kg/m^2)



b) Mean age of study participants (years)



c) Mean length of follow-up (years)



Due to heterogeneity in the trial populations, the underlying rate of development of T2DM varied between trials. The effect of this baseline risk on effectiveness of interventions was assessed by fitting a Bayesian meta-analysis model, accounting for both the uncertainty in the baseline risk and the inherent correlation between the baseline risk and hazard ratios (Sharp and Thompson, 2000), the code for which is given in appendix 3.5. The model was fitted as given in Equation 4.9, where d_i^t and d_i^c represents the number of cases of T2DM in both the control group and intervention group for each study (*i*), *y* represents the person-years of follow-up, and γ the natural log of the incidence rate ratio at the mean rate ($\bar{\alpha}$), where the $\bar{\alpha}$ varies from one iteration to the next. The effect of baseline is represented by β.

$$d_i^c \sim Poisson(\mu_i^c) \text{ and } d_i^t \sim Poisson(\mu_i^t)$$

$$Ln(\mu_i^c) = Ln(y_i^c) + \alpha_i \qquad \text{[Equation 4.9]}$$

$$Ln(\mu_i^t) = Ln(y_i^t) + \alpha_i + \delta_i$$

$$\delta_i = \gamma_i + \beta(\alpha_i - \bar{\alpha})$$

Baseline risk of T2DM varied greatly between trials, from 2.6 to 30.0 cases per 100 person years (table 4.2). Assessments of the data showed no indication of an interaction between the underlying baseline risk and the effect of lifestyle interventions (figure 4.7), with only a small change in the log hazard ratio for a one unit increase in the log baseline risk, and the 95% credible intervals containing the null value of zero, -0.01 (95% CrI: -0.31 to 0.34). The assessment of baseline risk for anti-diabetic agents showed more of an interaction with their effectiveness (figure 4.8), although again the null value of 0 was contained within the credible interval, -0.13 (95% CrI: -0.72 to 0.43).

Sensitivity analyses, using maximum likelihood estimation, were undertaken to assess the effect of different definitions of IGT and T2DM used by different trials, by removing the studies using the newer, lower threshold for fasting plasma glucose from the meta-analyses. Removing the trials that had used the newer diagnosis criteria for IGT or T2DM had minimal effect on the results, with the

pooled hazard ratio and 95% confidence interval for anti-diabetic drugs changing to 0.66 (0.46 to 0.94) and for lifestyle interventions to 0.55 (0.45 to 0.66).

Figure 4.7: Effect of baseline risk on the effectiveness of lifestyle interventions

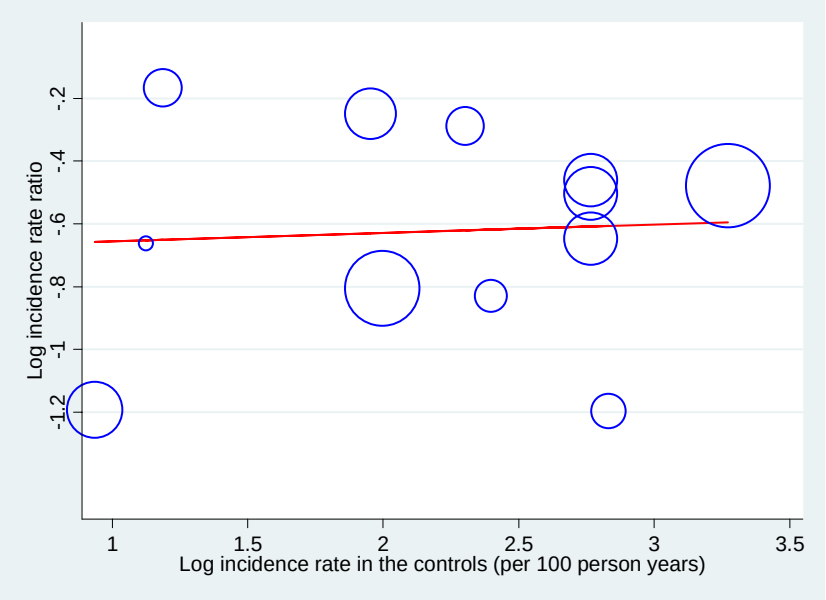
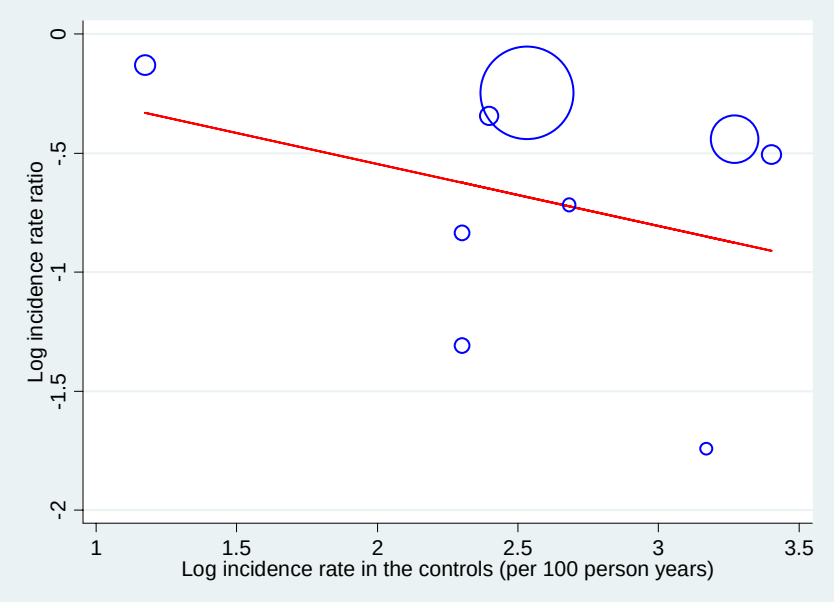


Figure 4.8: Effect of baseline risk on the effectiveness of anti-diabetic agents



To investigate individual study influence on the hazard ratio, again using maximum likelihood estimation methods, the effect of removing each study individually from the meta-analyses was examined. Results of the pooled intervention effects, with each study removed, are presented in table 4.6. It can be seen that no single study was greatly influencing the meta-analyses results.

Table 4.6: Pooled intervention effects with each study individually removed

Lifestyle Interventions		Anti-diabetic interventions	
Study removed	Pooled intervention effect (95% CI)	Study removed	Pooled intervention effect (95% CI)
Da Qing	0.49 (0.41, 0.60)	DPP	0.69 (0.57, 0.83)
DPP	0.56 (0.48, 0.68)	Eriksson	0.70 (0.62, 0.79)
DPS	0.53 (0.45, 0.62)	Fang	0.71 (0.63, 0.80)
Fang	0.51 (0.44, 0.59)	IDPP	0.70 (0.61, 0.81)
IDDP	0.50 (0.43, 0.58)	Jarrett	0.69 (0.62, 0.78)
Jarrett	0.50 (0.43, 0.57)	Li	0.70 (0.61, 0.79)
Kosako	0.52 (0.44, 0.60)	Pan	0.70 (0.61, 0.80)
Liao	0.52 (0.44, 0.61)	STOP-NIDDM	0.66 (0.57, 0.78)
Tao	0.52 (0.45, 0.61)	*All studies*	*0.70 (0.62, 0.79)*
Wein	0.51 (0.44, 0.60)		
All studies	*0.51 (0.44, 0.60)*		

4.5.2 Study quality and publication bias

Impact of study quality was considered using both the overall Jadad score (as reported in table 4.2), as well as each component separately. A cumulative analysis approach was used by firstly meta-analysing all the highest scoring, followed by inclusion of those which scored one point less, continuing until all studies were included. The pooled effect sizes at each step were then compared. Concealment of allocation was assessed by removing the trials that had reported this from the meta-analyses. The three meta-analyses, lifestyle, anti-diabetic and anti-obesity, varied minimally when studies that had scored low on the Jadad were omitted, or when the individual components of the Jadad score were assessed individually through sensitivity analyses. For example, for the anti-diabetic meta-analysis removing the two studies that only scored 2 on the Jadad score, gave a pooled estimate of 0.71 (95% CI: 0.61 to 0.84). Removing the lowest scoring lifestyle studies, that is the two studies which only scored 1 on the Jadad score, gave a pooled intervention effect of 0.60 (0.50 to 0.72). Concealment of allocation had potentially only been carried out by two studies, STOP-NIDDM (Chiasson et al., 2002) and XENDOS (Torgerson et al., 2004), although this was difficult to assess in many studies due to poor reporting.

Publication bias, as discussed in chapter 2, was assessed using both Begg's and Egger's tests (Sutton et al., 2000). Publication bias was not identified for the lifestyle meta-analysis, Begg's test p=0.945 and Egger's test p=0.340. For the anti-diabetic meta-analysis Begg's test was statistically significant, p=0.012 and Egger's tests, although not statistically significant, still indicated a problem may be present, p=0.058, especially as this test has low power. The funnel plots for both forms of interventions, which allow for a visual assessment of publication bias, are presented in figures 4.9 and 4.10, and also indicate publication bias may be a problem for the anti-diabetic studies. Pseudo confidence interval limits are plotted to assist in interpreting the funnel plot, these are plotted at the +/- z x SE of theta, where z is the standard Normal variate for the 95% confidence level.

To assess the implications of publication bias on the estimated intervention effect, a 'trim and fill' methodology was used, as described in more detail in section 2.3.4. This estimates the number of missing studies by adding studies to make the funnel plot symmetrical. Using the *metatrim* command in STATA, three additional studies were added to the meta-analysis. The adjusted intervention effect was estimated as 0.72 (95% CI: 0.60 to 0.84), compared to the unadjusted MLE estimate of 0.70 (0.62 to 0.79). Therefore although publication bias appears to be present, because a few large studies are dominating the intervention effect, it does not appear to be adversely affecting the meta-analysis results, i.e. the inputed studies are small and have little influence on the pooled estimate. The funnel plot with the three added studies is given in figure 4.11.

Figure 4.9: Funnel plot for lifestyle interventions

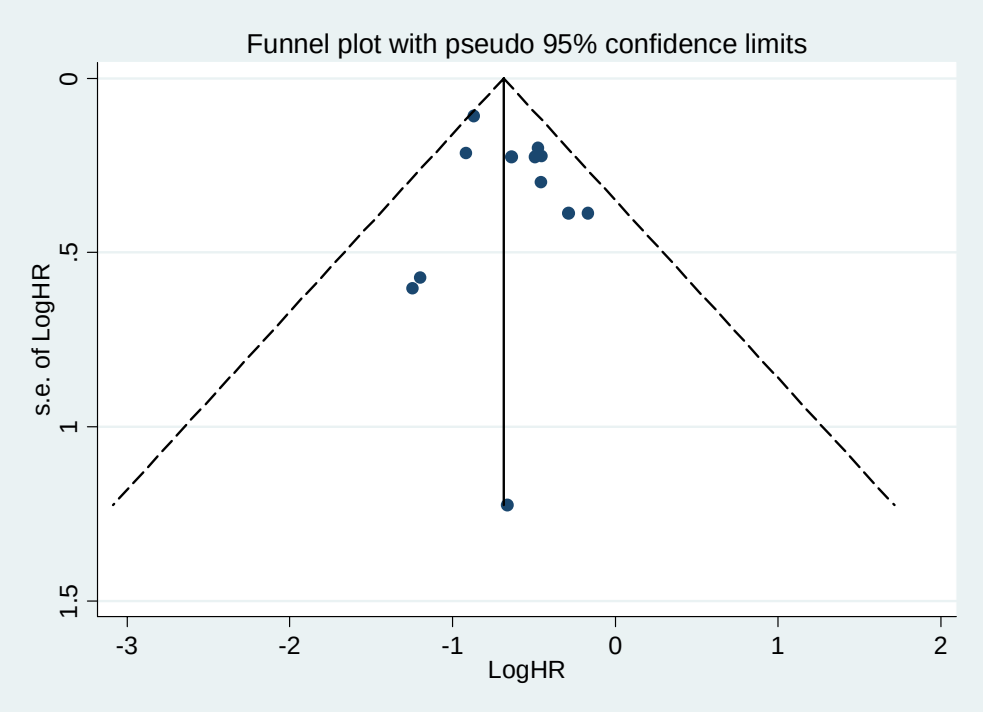


Figure 4.10: Funnel plot for anti-diabetic interventions

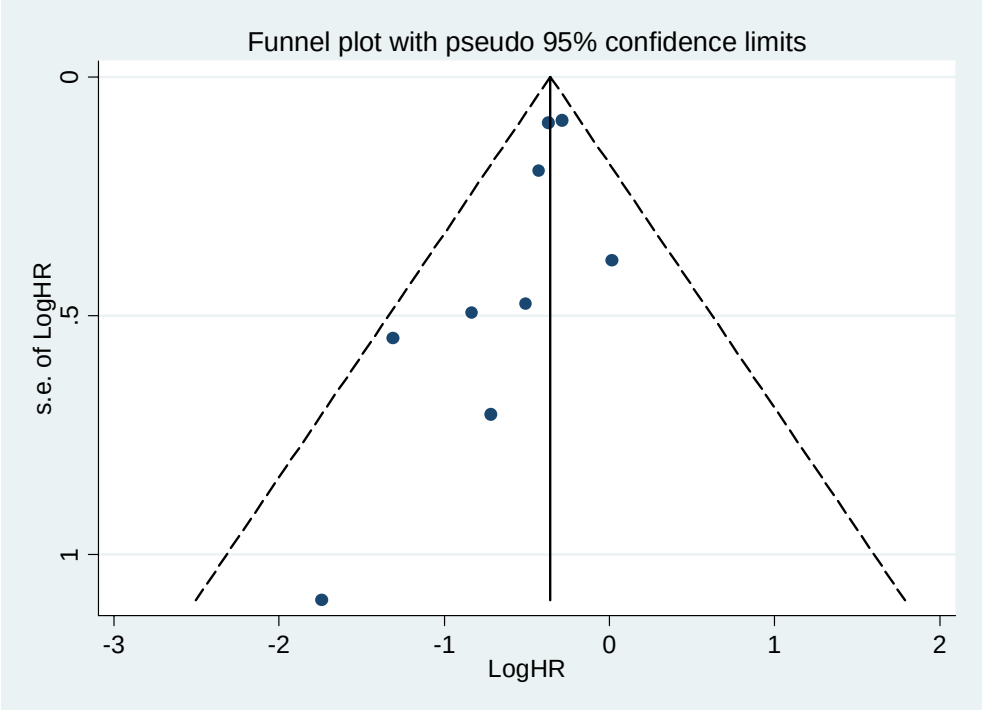
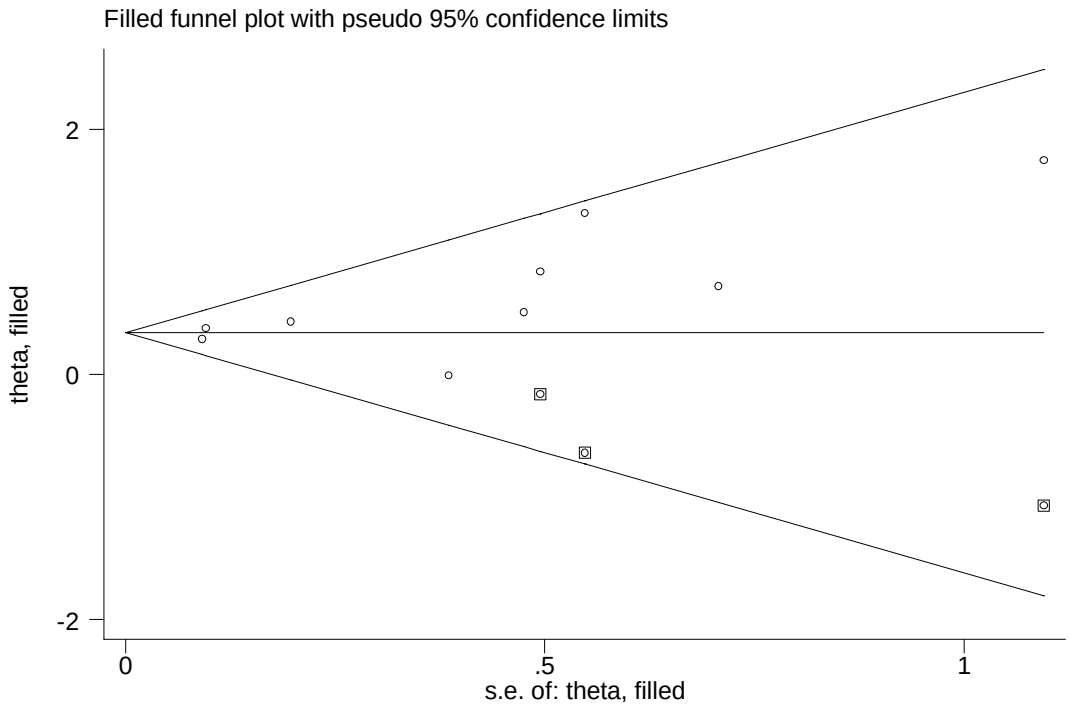


Figure 4.11: Trim and filled funnel plot for the anti-diabetic meta-analysis



* Added studies are indicated by squares

4.5.3 Adverse events

When considering the feasibility of possible interventions to delay/prevent T2DM, it is important to consider any possible side effects caused by the interventions. No adverse events were reported in the lifestyle intervention studies, those reported for pharmacological interventions are summarised in table 4.7. The majority of adverse events thought to be directly related to the intervention drugs were gastrointestinal or, in the case of troglitazone, a decline in liver functioning. Although the occurrence of adverse events varied widely between trials, all were higher in the intervention than the placebo groups.

Table 4.7: Reported adverse events possibly caused by the pharmacological interventions

Active Intervention	Trial	Event	Intervention Group	Placebo Group
Acarbose	Fang	Gastrointestinal side effects	8.0	0.0
	Pan	Gastrointestinal side effects	35.7	18.2
	STOP	Gastrointestinal side effects	13.0	2.5
Flumamine	Fang	Gastrointestinal side effects	6.3	0.0
Glipizide	Eriksson	Hypoglycaemic symptoms	41.0	32.0
Troglitazone	DPP	Liver function test >=3	4.3	3.6
		Liver function test >=10	1.2	0.2
Metformin	DPP	Gastrointestinal symptoms	77.8*	30.7*
	IDDP	Hypoglycaemia	8.4	0.0
		Gastrointestinal symptoms	1.9	0.0
	Li	Gastrointestinal side effects	4.4	0.0
Orlistat	XENDOS	At least one gastrointestinal event in the first year	91.0	65.0
		At least one gastrointestinal event in the fourth year	36.0	23.0
		Withdrawals due to adverse events	4.0	8.0

Figures are percentages except * which indicate number of events per 100 person years

4.6 Discussion

All of the meta-analyses show there is great potential for intervening to reduce the risk of T2DM in individuals with IGT, and lifestyle interventions appear to be at least as effective as pharmacological interventions. Both an increase in obesity and a decrease in physical activity in some westernised societies in recent years, are strongly linked with the increase in the prevalence and incidence of T2DM (Davies et al., 2004). Lifestyle interventions, which aim to reduce obesity and/or increase physical activity, help to directly address these risk factors. With T2DM affecting an estimated 171 million people worldwide in the year 2000 and a projected doubling of cases by 2030 (Wild et al., 2004), interventions to prevent T2DM will play an important role in future health policies. This analysis excluded studies that had assessed the effect of interventions on individuals with IFG. This was because the decision model developed here concentrates on the transition from IGT to T2DM, as more is known about progression between these two states. An assumption could have been made that

intervention effects would be the same in both IGT and IFG individuals, but as the trials containing IFG individuals were limited, minimal data was lost by their exclusion.

For the decision model the effect of interventions will be modelled over a number of years, therefore the long-term effectiveness of interventions and compliance to them is an important consideration. The DPP trial reported progression to T2DM after withdrawal from troglitazone and metformin (Knowler et al., 2005, Diabetes Prevention Program Research, 2003). Results showed the treatment effect was not sustained following discontinuation, therefore the effectiveness of pharmacological interventions is reliant on compliance and a long-term commitment to the treatment. It is essential therefore that individuals are comfortable on the intervention and even minor adverse effects, such as the gastrointestinal problems summarised here, take on greater importance if interventions have to be taken for life. Generally it would be fair to assume that lifestyle interventions would incur fewer and less serious side effects than pharmacological, but as with the pharmacological interventions, their effect may not be permanent and dietary and exercise advice may need to be reinforced on a regular basis. Additionally, although compliance was high in these trials where it was reported, it is still to be determined whether compliance could be maintained outside of a trial setting. When using the intervention effects in the decision model assumptions of both compliance and whether the intervention effects are likely to decrease over time need to be considered. The effect of reducing compliance to interventions was assessed in the comprehensive decision model, as described in Chapter 8.

Poor reporting of results in some of the intervention trials, made their inclusion in the meta-analyses either impossible or difficult. Most of the hazard ratios and incidence rate ratios included in the meta-analyses were unadjusted, except those used for Wein (Wein et al., 1999), STOP-NIDDM (Chiasson et al., 2002) and IDDP (Ramachandran et al., 2006). As trial arms were similar at baseline for unadjusted and adjusted characteristics, it is unlikely adjustment introduced any inconsistency into the meta-analyses.

An MTC analysis was carried out to enable trials comparing lifestyle interventions against controls, and/or pharmacological interventions against controls, to be combined, enabling a direct comparison of lifestyle vs. pharmacological interventions. The key assumptions made for a random effects MTC analysis are that although the effect sizes may differ between trials, they are from a common distribution, that is the same across trials (Caldwell et al., 2006) and that an additive scale of

measurement is being modelled, in that the relative effect of A vs. C can be predicted from A vs. B and B vs. C (Deeks, 2002), that is the indirect estimate is the same as would have been obtained from a head to head randomised control trial. Hazard ratios on the log scale were modelled, which is likely to be more appropriate for the second assumption. The first assumption is more difficult to check so sources of heterogeneity between study results were investigated as far as possible.

From the meta-regression results it appears lifestyle interventions may have a greater impact the higher the mean baseline body mass index of a group of individuals. It is acknowledged though that using study-level data can lead to problems of aggregation bias, where there appears to be a relationship when in fact one does not exist at an individual level (Sutton et al., 2000). To conduct a more conclusive assessment individual person data would be needed. Baseline risk of T2DM was not shown to affect the effectiveness of the interventions, showing that the intervention effects should be consistent across populations with different risks of developing T2DM.

The assessment of publication bias indicated that it might be an issue for the anti-diabetic agents meta-analysis, in that the funnel plot was asymmetric, with the smaller studies all showing large intervention effects. Re-estimating the effect size after using 'trim and fill' methods to correct for publication bias, only changed the pooled estimate minimally though.

There was great diversity in study quality, with the lifestyle trials generally scoring lower on the Jadad score, where blinding of treatment was not possible. The Heymsfield (Heymsfield et al., 2000) trial combined data from three randomised controlled trials, so technically was not a true single study. It was treated as such for the purposes of these analyses, as results were not available for each trial individually. All three trials were almost identical in their design, and the trials had not been individually powered to assess incidence of T2DM as an outcome.

Since the literature search was carried out for this review, a further relevant trial has been published, the Diabetes Reduction Assessment with Ramipril and Roglitazone Medication (DREAM) trial (DREAM trial investigators, 2006a, DREAM trial investigators, 2006b). Unfortunately a request for relevant data to enable this trial to be included in this review was unsuccessful. This trial only reported combined results for individuals with either IGT or impaired fasting glucose, so it is not directly comparable with our meta-analyses here, but in summary they found the ace-inhibitor

ramipril did not significantly reduce incidence of T2DM, hazard ratio 0.91 (95% confidence interval, 0.80 to 1.03) but roglitazone, an oral anti-diabetic agent, did 0.38 (0.33 to 0.44), although there has been some concern expressed over the rate of cardiovascular events in the roglitazone group (Heneghan et al., 2006).

4.7 Summary

A systematic review and meta-analysis has been described here, with the purpose of obtaining pooled effect estimates, and associated distributions, of both lifestyle and pharmacological interventions for the decision model. Whilst substantial evidence has been found to support the clinical effectiveness of both lifestyle and pharmacological interventions in significantly reducing the risk of developing T2DM, a number of issues remain. For pharmacological interventions adverse effects need to be fully understood to enable potential harms and benefits to be assessed. There is also the issue of whether what is fundamentally a lifestyle problem should really be treated with a lifelong course of medication. For lifestyle interventions, compliance is the key to their success, therefore strategies to assist compliance need to be carefully thought through and implemented. All these issues will affect the long-term effectiveness of interventions which is an important consideration for the decision model.

A great deal of effort has been used to ensure the meta-analyses of intervention trials are both comprehensive and systematic, with further analyses carried out to check for sources of heterogeneity and bias. Still the possibility of intervention effects being mis-specified has to be considered within the comprehensive decision model, particularly in terms of compliance and their long term effectiveness. Therefore an extension to the model, described in chapter 8, explores reduced compliance and hence reduced effectiveness of interventions, on the cost-effectiveness of the screening strategies where they are included.

Finally, the evidence meta-analysed here is on patients already identified as having IGT. The overall effectiveness and cost-effectiveness of a policy of T2DM prevention/delay must consider how different identification and screening strategies would impact on the overall evaluation of such policies. This will be done using the comprehensive decision model described in the next chapter.

5. MODELLING THE SCREENING AND INTERVENTION PATHWAY

5.1 Chapter overview

Before the comprehensive decision model was developed, decisions had to be made on how to structure the model. To do this a number of issues had to be considered, including clinical aspects of the IGT and T2DM disease pathway, and the likely approach that would be taken in the implementation of a screening strategy. The aim of this chapter is to explain the thought process that led to the structure of the final model. The actual data needed for the model will not be discussed here, but in detail in chapter 6. This chapter will firstly give a summary of how similar models have been constructed and their results, and then the model for this thesis is described along with an account of the decisions taken to determine the model structure.

5.2 Existing work

A literature search was carried out to identify previous decision models that had assessed screening for IGT or T2DM and models that had assessed interventions to delay T2DM in individuals at risk. The databases Medline, Embase and the Cochrane library were searched, as well as the ScHARR (School of Health and Related Research) website which has a section on reviewing modelling methods for the evaluation of screening programs. Models fell into three types, those that had considered screening tests, those that had assessed interventions for preventing T2DM, and those that had modelled the impact of screening and early diagnosis of T2DM.

5.2.1 Previous models that have considered screening tests for IGT and/or T2DM

Four studies were identified that had carried out economic evaluations comparing different screening tests for identifying individuals with IGT and/or T2DM (Zhang et al., 2003, Zhang et al., 2005, Shirasaya et al., 1999, Icks et al., 2004). Study characteristics are presented in table 5.1. Model structures were simple in that they were based solely on costs of screening tests and test sensitivities and specificities.

The study by Icks et al. (2007) considered four screening strategies for T2DM, using population-based data on subjects aged 55 to 74 years. They ran a decision analytic model for a time horizon of one year and compared the following screening tests: fasting glucose testing, fasting glucose followed by OGTT, OGTT only, and HbA_{1c} followed by OGTT. The main outcomes were costs, true-positive T2DM cases and incremental cost-effectiveness ratios. They found that a HbA_{1c} test followed by an OGTT in those with a HbA_{1c} value >5.6% identified the most cases of T2DM, but was also the most expensive strategy at €21.44 per patient. OGTT alone was the cheapest strategy at €4.90 per patient. They concluded that the decision regarding which is the most favourable approach to screening depends on whether the goal is to identify a high number of cases, or to incur lower costs at reasonable effectiveness.

Shirasaya et al. carried out an economic evaluation of three screening tests for IGT and T2DM that do not require the subjects to fast beforehand. 891 men between 26 and 80 years of age were each screened using three tests, which were 1,5-anhydroglucitol (1,5-AG), HbA_{1c}, and fructosamine (FRA). The primary health outcomes of the study were the number of IGT and T2DM diagnoses, the sensitivity and specificity of each test, and the area under the ROC curve for each test. Clinically no test gave an outstanding performance for detecting IGT and T2DM cases, with area under the ROC curves of 66.85%, 66.25% and 60.28% for each test respectively. Test efficiency was better when testing for T2DM alone, with area under the ROC curves of 90.83%, 88.72% and 78.49% respectively. The most cost-effective indicator was found to be FRA, although it was not clear how this was determined as the statistical analyses used to assess cost-effectiveness was not reported.

Zhang et al. (2003) looked at cost and efficiency of screening for pre-diabetes using five detection strategies; OGTT, FPG, HbA_{1c}, capillary blood glucose (CBG) and a risk assessment questionnaire. Main outcomes were proportion of cases identified, total costs and cost per case identified. They simulated a population based on the demographics of a U.S. population, aged between 45 to 74 years, who had visited a health care provider at least once in the last year. Data from a number of sources was utilised, including the Third U.S. National Health and Nutritional Examination Survey, the 2000 census, Medicare and published literature. They found that the cost per case of pre-diabetes or undiagnosed T2DM identified ranged from $176 to $236, with HbA_{1c} being the most expensive strategy and the risk assessment questionnaire, and CBG test being the cheapest. It was concluded that a trade-off must be made between clinical and cost-effectiveness when choosing an appropriate screening test strategy.

A similar economic evaluation was conducted by the same study group a few years later (Zhang et al., 2005), but in this study three screening tests were assessed; FPG, HbA_{1c} and CBG. Tests were evaluated for screening for T2DM alone, or both T2DM and pre-diabetes combined, and the cost-effectiveness of eight cut-off points for each test were evaluated. Again a U.S. population aged 45 to 74 years of age was simulated for the model. The number of cases identified by each test, at each cut-off was estimated using prevalence data, and test sensitivity data. The cost of opportunistic screening was estimated by multiplying the cost of screening one individual by the number of individuals who would be eligible for screening in the U.S. For all three screening tests cost per case identified first decreased and then increased as the cut-off value was increased. Using the most efficient cut-offs, FPG proved to be the cheapest test, followed by the CBG test, and the HbA_{1c} test was the most expensive.

In summary, the results of studies that have assessed the cost-effectiveness of different screening tests highlight the fact that it is still unclear which tests would be the most cost-effective for screening for IGT and T2DM. The more accurate tests such as OGTT are also the more expensive, but as all tests are relatively cheap in terms of health care costs, then clinical effectiveness, ease of use and acceptability of a test to a population, are likely to be the most important factors for determining which screening tests to adopt. Overall though it is difficult to assess the cost-effectiveness of screening tests, without considering the clinical and cost implications of subsequent treatments and interventions.

Table 5.1: Studies of economic evaluations of screening tests for IGT and/or T2DM

Study and Population	Evaluated	Analysis	Data	Conclusions
Icks (2004) 1353 participants 55-74 years Germany	Four screening strategies for T2DM; fasting glucose, fasting glucose and OGTT, HbA_{1c} and OGTT, and OGTT alone.	Calculated true positive cases identified, total costs and cost per patient	Screening and population data from the KORA 2000 study. Costs came from the German healthcare system.	HbA_{1c} and OGTT was the most effective but the most expensive at €21.44 per patient. OGTT alone was the cheapest at €4.90 per patient.
Shirasaya (1999) 891 men 26-80 years Japan	Three screening tests for IGT and T2DM: 1,5-anhydroglucitol, HbA_{1c} and fructosamine.	Economic evaluation based on sensitivity and specificity using optimal cut-offs.	Screening data collected as part of this study. Cost data was obtained from external sources.	FRA is the most cost-effective screening test for IGT and T2DM
Zhang (2003) 45-74 years U.S.A.	5 detection strategies for IGT and T2DM: OGTT, FPG, HbA_{1c}, capillary blood test (CBG) and a risk assessment questionnaire	Calculated cases identified, total costs and cost per case identified.	2000 census, Medicare, published literature and the U.S. national heath and nutrition examination survey	CBG and the risk assessment questionnaire are the cheapest for cost per case identified.
Zhang (2005) 45-74 years U.S.A.	Three screening tests for IGT and T2DM, FPG, HbA_{1c}, CBG, each evaluated over a range of cut-off.	Calculated cases identified, total costs and cost per case identified.	2000 census, Medicare, published literature and the U.S. national heath and nutrition examination survey	FPG is the cheapest test and HbA_{1c} the most expensive for cost per case identified.

5.2.2 Previous models that have considered interventions for T2DM prevention

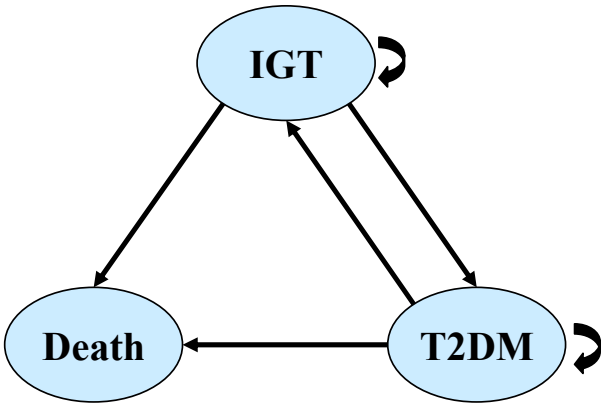
Eight studies were identified that had carried out economic evaluations of the clinical and cost-effectiveness of intervening in individuals with IGT to try and prevent T2DM (Avenell et al., 2004, Caro et al., 2004, Eddy et al., 2005, Herman et al., 2005, Icks et al., 2007, Jacobs-van der Bruggen et al., 2007, Palmer et al., 2004, Segal et al., 1998). These are summarised in table 5.2.

Seven of the models fitted were Markov models whereby a simulated population was moved between various states using estimated transition rates. One study used what they termed the Archimedes model (Eddy et al., 2005), which instead of modelling states, models what are termed 'objects'. Eddy defines 'objects' as being any relevant factors that affect model outcomes which in this example included fasting plasma glucose, basal hepatic glucose production, blood pressure and body mass index. Numerous differential equations are then used to model how all the objects change and interact continuously over time.

Only four of the studies gave diagrams of the model structure, allowing the configuration of the model to be fully understood. Figure 5.1 is reproduced from Avenell et al. (2004), and shows how their Markov model was based on three states. The model by Caro et al, (2004) followed a similar structure, although they included a fourth state of NGT. Both the papers by Icks and Palmer represented their models through fairly complex decision tree diagrams.

The simulated populations varied between models, often they were characterised to represent a national population, for example, Jacobs-van der Bruggen (2007) based their model on the population demographics of the Netherlands, whereas other studies used the sample characteristics of the clinical trials from which they were deriving the intervention effect, for example Caro (2004) used the baseline characteristics from the STOP-NIDDM trial.

Figure 5.1: Structure of the model as described and fitted by Avenell et al.



To briefly summarise some of the more important model assumptions made. The model by Avenell et al. was comprised of three Markov states of IGT, T2DM and death. It was assumed that individuals could not leave the model by reverting to NGT, and transition rates were used to move individuals between states on a yearly basis. The intervention assessed was lifestyle, which reduced the transition rate from IGT to T2DM when applied. The model by Caro et al was more complex and included an NGT state, in addition to IGT, T2DM and death, and included costs for identifying individuals with IGT. Their model cycled over 6 monthly periods and included a diabetes sub-model whereby the experience in the T2DM state was dependent upon the characteristics of an individual when they entered that state, such as HbA_{1c}, age and gender. The model by Icks limited the time horizon of their model to just three years as this represented the length of follow-up of the trial data being utilised (DPP study). This was unusual and most models extrapolated trial data for longer than the follow-up of the trial. The model by Palmer included the three states, IGT, T2DM and death. Probability of death was dichotomised and dependent on whether individuals were diagnosed at onset of T2DM, or after an average 8 years remaining undiagnosed. Both the models by Segal and Jacobs-van der Bruggen used age specific mortality rates, but in many of the models it was unclear how transitions were changed over time.

Data sources were varied with the models using a variety of sources from published trials, epidemiological studies and national statistics. The time span, or time horizon, over which the models were run, ranged from just three years post intervention, up to the expected lifetime of the model population. Costs attached to the models were often country specific and of the eight studies only three included costs of identifying individuals with impaired glucose tolerance (Caro et al., 2004, Herman et al., 2005, Segal et al., 1998). In general the data sources for the models were limited to a few sources. No attempt was made by any of the models to include all available data, so for example when modelling intervention effects using the results from just one trial was the norm.

All models compared a strategy of interventions against no interventions, rather than screening for impaired glucose tolerance followed by interventions, in comparison to no screening. All but one model simulated populations where all individuals had impaired glucose tolerance at the start of the model and the end state was development of diabetes, or death, hence only a limited section of the disease pathway was modelled. The models did not take into account that screening for IGT will at the same time allow individuals with undiagnosed T2DM to be identified, allowing for early treatment. Also individuals with IGT who receive interventions are

likely to be closely monitored, so that if they go on to develop T2DM, again appropriate treatment can begin early. Both situations may lead to reduced complication rates associated with T2DM. Hence, whilst these studies offer an assessment of the cost-effectiveness of interventions for diabetes prevention, none assess the impact of screening followed by interventions on the whole disease pathway.

Table 5.2: Decision models that assessed interventions to prevent or delay T2DM

Study and Population	**Evaluated**	**Model**	**Data**	**Conclusions**
Avenell (2004) IGT, 55 years, 33% male and BMI ≥ 30kg/m^2	Diet and exercise interventions	Markov model with 3 states: IGT, T2DM and death. 6 year horizon.	Transition rates taken from DPS. UK mortality rates from WHO. Costs based on UK figures.	Lifestyle interventions vs. no intervention cost £13,389 per QALY gained.
Caro (2004) Population characteristics of the STOP-NIDDM trial.	Acarbose, intensive lifestyle modification and metformin.	Markov model with 4 states, IGT, NGT, T2DM and death. 10 year horizon.	DPP, DPS and STOP-NIDDM trials. Mortality rates from Canadian life table data.	Lifestyle vs. no intervention cost $749 per LYG. Metformin and acarbose vs. no treatment cost -$999 and -$897 per LYG.
Eddy (2005) IGT individuals with BMI ≥ 25kg/m^2	4 strategies; no intervention, metformin or lifestyle interventions in individuals with IGT and lifestyle interventions once T2DM develops	Archimedes model, time horizons of 5-30 years.	Published epidemiological studies, DPP trial and Kaiser permanente administrative data	Compared to no intervention cost per QALY gained was $143,000 for lifestyle interventions in IGT, $24,500 for lifestyle interventions in T2DM and $35,400 for metformin.
Herman (2005) Individuals with IGT, > 25 years	Lifestyle and metformin interventions.	Markov model, lifetime horizon	Transition rates and costs from the DPP trial and published reports.	Compared to placebo lifestyle interventions cost $1124 and metformin $31,286 per QALY gained.
Icks (2007) Individuals with IGT, 60-74 years, BMI ≥	Lifestyle and metformin interventions.	Decision analytical model, horizon 3 years.	Population data from the KORA study. Intervention effects from DPP trial. German healthcare costs.	Lifestyle was more cost effective than metformin, £3127 and £12,731 per case of T2DM prevented respectively

25kg/m^2				
Jacobs-van der Bruggen (2007) A community population and a population of obese adults	A general community intervention and a healthcare lifestyle intervention (in the obese population).	Markov model, horizon 20 years.	Published literature and the Netherlands National Institute of Public Health.	Compared to no intervention, community intervention cost €3,100 to €3,900 and healthcare intervention €3,900 to €5,500 per QALY.
Palmer (2004) Overweight individuals with IGT	Lifestyle or metformin interventions.	Markov model with 3 states, IGT, T2DM or death. Lifetime horizon.	Costs of interventions and their effect from DPP trial. National mortality tables.	Lifestyle was more cost effective than metformin, €17,900 and €47,200 per LYG compared to controls.
Segal (1998) High risk groups, e.g. obese/previous gestational diabetes.	6 programmes; intensive diet, surgery, GP advice, media campaign and group behavioural modification.	Several Markov sub-models. 25 year horizon.	Transition matrices derived from the Swedish study (Eriksson, 1992). Mortality rates and costs from the Australian Bureau of Statistics.	Group programme, media campaign, behavioural programme and diet were all cost-effective (gross costs of A$500 to A$5900 per LYG). Surgery performed poorest, A$12,300 per LYG.

Due to variability in the models and the fact that costs from different countries were used, it is difficult to directly compare model results. All except the model by Eddy et al. concluded lifestyle interventions were likely to be cost effective at reducing cases of T2DM compared to no intervention. Eddy concluded that although lifestyle modification is likely to have important effects on the morbidity and mortality associated with T2DM, the intervention used by the DPP study may be too expensive for implementation as a national program. Of the four models that compared both pharmacological and lifestyle interventions, two found pharmacological interventions to be more cost-effective (Caro et al., 2004, Eddy et al., 2005), whilst two favoured intensive lifestyle interventions (Herman et al., 2005, Palmer et al., 2004). Icks (2007) who considered prevention of T2DM in a real-world setting, found cost-effectiveness was very sensitive to uptake of screening and compliance to the interventions by the population.

Time horizons varied between models which may affect their conclusions. It might be expected that models run for longer horizons would find the interventions to be more cost effective, due to the high initial costs of lifestyle interventions and the fact that it will take a few years for the negative costs of T2DM to gather pace.

Only two of the models assessed the cost-effectiveness of interventions in a U.K. setting. Avenell (2004) considered the effectiveness of lifestyle interventions in the obese, whilst Icks (2007) looked at lifestyle and metformin interventions in individuals with IGT between the ages of 60 to 74 years. Avenell concluded lifestyle interventions cost £13,389 per QALY gained, therefore appearing cost-effective, whilst Icks found interventions were fairly costly, at £3127 and £12,731 per case of T2DM prevented for lifestyle and metformin respectively.

5.2.3 Previous models that have considered screening for T2DM

Early diagnosis of T2DM allows for treatment to begin earlier and this is thought to lower the future risk of complications, although no randomised controlled trials have yet published results to support this hypothesis. Five simulation studies were found that considered the cost and clinical implications of screening for diabetes (CDC Diabetes Cost-Effectiveness Study Group, 1998, Chen et al., 2001, Glumer et al., 2006, Hoerger et al., 2004, Waugh et al., 2007). Most looked at the impact of early treatment on cardiovascular events, but some also included micro-vascular events such as retinopathy. The studies are summarised in table 5.3.

The model by CDC Study group consisted of a screening module, whereby test sensitivity and specificity affected number of cases identified, and a disease progression module whereby both diagnosed and undiagnosed diabetics were modelled in terms of complication rates for micorvascular complications including retinopathy and nephpropathy. They found the benefits of early detection and treatment accrued more from postponement of complications, and the resulting improvement in QALYs, than from additional life years gained.

The model by Chen et al. compared three strategies of no screening, 5 yearly screening and biennial screening. They utilised a Markov model to model the natural history of T2DM progression to complications and death. Complications were modelled by three states, retinopathy, nephropathy and neuropathy. The transition rates to complications increased as duration of T2DM increased. Incidence and mortality rates of cardiovascular disease were taken from the Framingham Heart Study and also included in the model. It was concluded that five-year screening appeared more cost-effective than biennial, and that screening younger individuals (30-39), was more cost-effective than screening the elderly.

Glumer et al. (2006) claimed their model was the only one developed so far to include uncertainty around model inputs, and by doing so produce confidence intervals around their results. The main model outcome was CHD risk, with the model utilising data from the Danish Inter 99 study and the UKPDS study. Compliance to screening was assumed to range between 30% to 75%. It was concluded that there was considerable uncertainty around the cost-effectiveness of screening for T2DM, with the most important, but still uncertain parameter, being the effect of early treatment of T2DM.

The model developed by Hoerger et al. utilised data from the UKPDS study and the Hypertension Optimal Treatment trial (HOT), to model progression of diabetics to nephropathy, neuropathy, retinopathy, coronary heart disease and stroke. It was assumed that individuals who were not screened would develop T2DM 10 years after its onset, and individuals screened would be diagnosed five years after onset. Earlier detection and treatment was modelled as lower HbA_{1c} and subsequently lower transition rates to complications. Conclusions from this model were that screening for T2DM is more cost-effective if targeted at hypertensives than compared to universal screening.

Table 5.3: Decision models that assessed screening for T2DM

Study and Population	Evaluated	Model	Data	Conclusions
CDC (1998) Demographic characteristics of U.S. population, ≥25 years.	1 time opportunistic screening ofT2DM.	Markov model with a lifetime horizon.	Clinical trials, epidemiological studies and population surveys.	Screening vs. no screening cost $56,649 per QALY gained.
Chen (2001) >30 years, Demographic characteristics of Taiwanese population.	No screening, biennial screening and 5 year screening.	Markov model with a 30 year horizon.	Clinical trials, epidemiological studies and demographic statistics of the Taiwanese population.	Biennial screening cost $17,833 and 5 year screening $17,113 per QALY gained compared to no screening.
Glümer (2006) Demographic characteristics of Danish population, 35-60 years.	No screening vs. screening and treatment of T2DM.	Markov model with uncertainty accounted for. 5 year horizon.	Danish inter99 population studies and clinical trials including UKPDS.	Screening costs £40,700 (95% CI: 23,300 to 82,000) per QALY gained compared to no screening.
Hoerger (2004) Demographic Characteristics of U.S. population	No screening, universal screening and targeted screening at those with hypertension.	Markov model with a lifetime horizon.	UKPDS, Hypertension Optimal Treatment trial and recent cost data.	At age 55 yrs, targeted screening cost $34,375 and universal screening cost $360,966 per QALY gained compared to no screening.
Waugh (2007) 40-70 years	No screening vs. a one-off universal screening. Sensitivity analyses looked at different age-groups and hypertensive and obese sub-groups	Decision tree and Markov model with a lifetime horizon.	Clinical trials including UKPDS. NHS cost data.	For all age groups the cost per QALY gained for screening vs. no screening was £2266. The cost was highest for the youngest age group, 40-49 years, £10,216. Cost per QALY was lower in the two risk groups, £1,505 for hypertensive and £1,046 for obese individuals.

The final model by Waugh et al. fitted a baseline scenario whereby the cost-effectiveness of a single screening of a population aged 40 to 70 years was assessed, and then ran a number of sensitivity analyse to assess different sub-groups including limiting the screening to different ten year age bands and targeting hypertensives or obese individuals. The model by Waugh was the only model that was hybrid in structure, as the model developed for this thesis is, in that it was comprised of both a decision tree and a Markov model. The model was fitted as one component of a *Health Technology Assessment* on screening for T2DM. In conclusion they stated that there was a strong case for screening for T2DM, although more clinical data was needed.

Overall most of the models produced favourable results for T2DM screening, but cost-effectiveness varied depending on age group screened and the population targeted for screening. Both the CDC Diabetes Cost-effectiveness Study group and Chen et al. (2001) found the incremental costs of screening increased with the age of the proposed screened population, although Waugh et al. (2007) found costs decreased with age. Both Hoerger et al (2004) and Waugh et al (2007) concluded targeted screening of individuals with hypertension or those who were obese, was more cost-effective than a universal screening program.

5.2.4 How the model developed for this thesis will add to current knowledge

As discussed a number of decision models already exist that have considered delaying T2DM through interventions or screening and early detection of T2DM to reduce risk of complications. Despite this, conclusive evidence is still not available for the cost-effectiveness of such strategies, as model results are conflicting. Only one of the models discussed here included uncertainty in their model inputs, data sources used was often limited, assumptions varied greatly and were often not clear and costs included were often not relevant for a U.K. health setting.

The model developed for this thesis aims to build on previous work by modelling the full pathway from screening for IGT, intervening to delay or prevent T2DM, right through to the development of T2DM, complications and death. It will therefore provide a much more comprehensive and useful overview than previous models, as it is difficult to assess the full impact of intervening to delay T2DM, unless the costs and clinical implications of screening for IGT are also considered. For example even extremely effective interventions for delaying

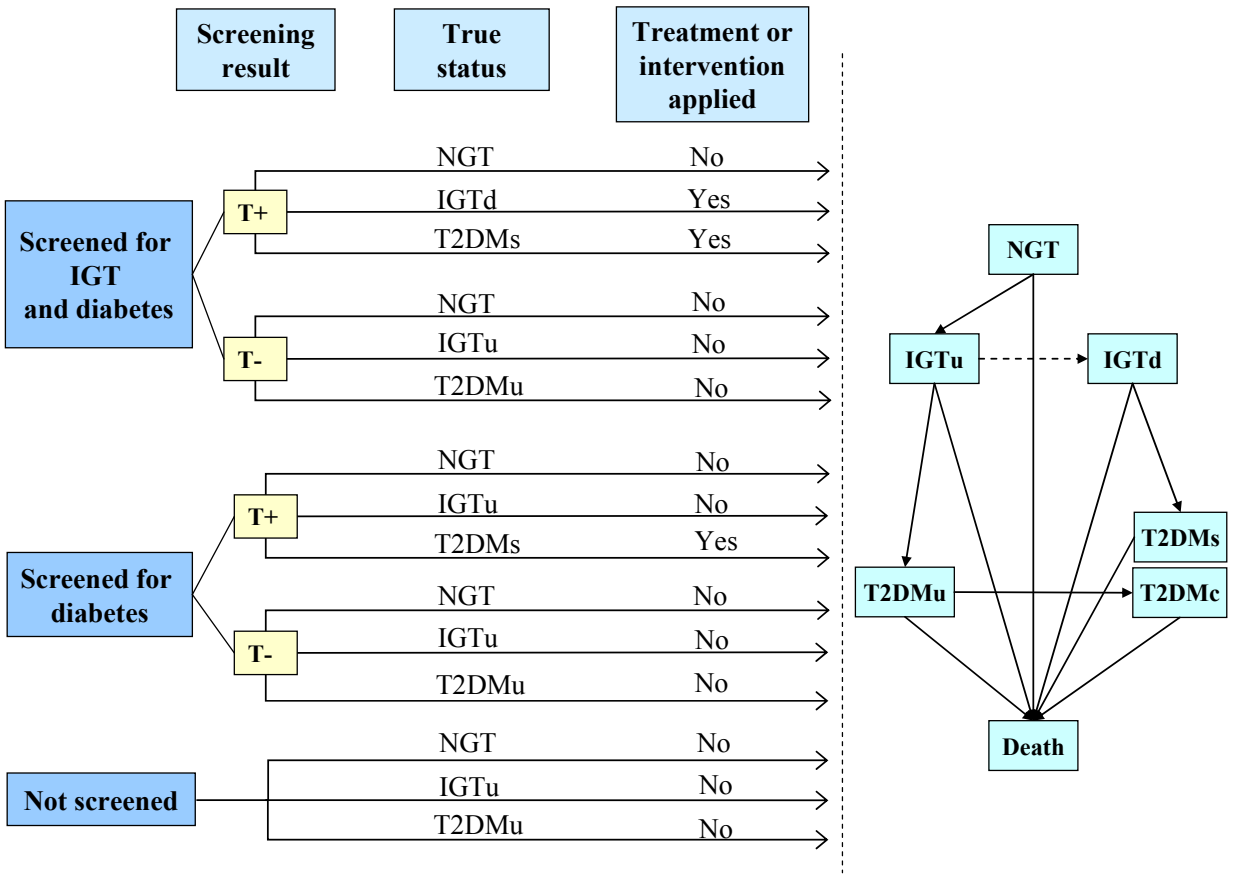
T2DM will have little impact if screening tests for identifying individuals at high risk of developing T2DM prove to have very low sensitivities and specificities, or to think of matters in different way, good screening tests are inconsequential if interventions or treatments for individuals identified are ineffective.

This is also the first model to consider both screening for IGT followed by appropriate intervention, and screening for T2DM, which will allow a comparison of the two scenarios. This is important if a decision on the most effective approach to screening for T2DM is to be reached. Current models are also improved upon by using as much of the available data relevant for the model as possible, where appropriate, rather than basing the model on results from a limited number of trials or sources. Further, by carrying out a number of sensitivities considering populations to be screened, accuracy of screening tests and model horizons, then the main drivers of the model can be fully understood and clearer conclusions concerning what an effective screening policy should encompass can be made. Additionally, carrying out a thorough series of sensitivity analyses will enable the identification of where further research is needed in the field.

5.3 The proposed model

The comprehensive decision model developed for this thesis is a hybrid model, which consists of a decision tree and a Markov model (Figure 5.2). The decision tree is structured to model the screening component of the health policy. Test sensitivities, specificities and prevalences of disease states are used to identify the number of individuals in a screen population who will be correctly identified as having IGT and T2DM. From the decision tree the numbers starting in each state of the Markov model can be identified. The Markov model models disease status of a screened population over a number of years. By attaching costs and utilities to each Markov state the cost-effectiveness of different screening strategies can be assessed. The implications of complications associated with T2DM on both costs and utilities, were modelled by incorporating UKPDS data within the model.

Figure 5.2: The proposed screening and intervention model



(NGT=Normal Glucose Tolerance, IGT=Impaired Glucose Tolerance, T2DM=Type 2 Diabetes Mellitus, u=undiagnosed, d=diagnosed, s=screen diagnosed and c=clinically diagnosed)

5.3.1 Structuring the decision tree

One of the research questions assessed for this thesis, focuses on whether screening for IGT and subsequently intervening to try and delay or prevent T2DM is a clinically and cost effective policy. To try and answer this, three possible approaches to screening are considered in this model. An option of no screening was included, as currently there is no systematic screening of the population by the NHS for IGT or T2DM and any new screening policy needs to be compared against current practice. To answer the research question an assessment of screening for IGT needs to be made. As glucose levels are measured on a continuous scale it is impossible to screen for IGT without picking up previously undiagnosed individuals with T2DM, consequently a strategy considering screening for both IGT and T2DM was included in the

model. The model could have been kept as a comparison of these two approaches, but if screening for IGT and T2DM was proved to be an effective strategy then policy makers would be interested in knowing if most of the benefit was coming from screening and intervening in the IGT patients, or in just from identifying undiagnosed T2DM cases. Hence the inclusion of the third screening strategy, which considers screening for undiagnosed T2DM only. This arm will have a higher threshold for the screening test than the IGT and T2DM combined screening test, and hence will pick up slightly fewer T2DM cases, due to lower sensitivity, although it will have a higher specificity for T2DM. Where T2DM alone is screened for, IGT patients may still be identified. As this screening programme is not aimed at identifying such individuals, they are treated as receiving standard advice and not a tailored intervention for prevention of T2DM.

These three screening scenarios are therefore what are compared in the model. The sensitivities and specificities of different screening tests are determined against the 'gold standard' oral glucose tolerance test (OGTT). The screening test to be used and the specific population to be screened were not set in the model structure but were changed between different versions of the model to allow for the assessment of different screening policies. Excluded from the screening programmes were any individuals who have already been diagnosed with T2DM.

5.3.2 Outline of the decision tree

As discussed above the decision tree is comprised of three main arms, representing three possible screening approaches which are no screening, screening for undiagnosed T2DM and screening for both IGT and undiagnosed T2DM. Individuals who have already been identified as having T2DM were excluded from the screening process. The same population can be entered into each of the three screening arms and the results then compared to assess the different screening strategies in terms of both costs and clinical effectiveness.

The decision tree uses prevalences of IGT and T2DM of a population of interest, for example a Southern Asian population or an older population have greater prevalences of IGT and T2DM compared to a Caucasian or younger population respectively. By utilising estimated prevalence, along with sensitivities and specificities of a proposed screening test, where the sensitivities are given in terms of NGT, IGT and T2DM, the model estimates the true status of screened individuals. This was done as specified in equation 5.1, where the positive predictive values (PPV) and negative predictive values (NPV) are calculated using sensitivities, $P(T+|i)$, and

specificities, *P(T-|i)*, with *i* representing either NGT, IGT or undiagnosed T2DM, and λ_i the prevalence of each glucose tolerance state. For the arm of the model where no screening is assessed, prevalences alone were used to predict the true status of individuals in the population. An example of how the decision tree works in practice is given in chapter 6, using data from the STAR study.

$$PPV_i = \frac{P(T+|i)\lambda_i}{\sum_{i=1}^{3} P(T+|i)\lambda_i} \qquad \text{[Equation 5.1]}$$

$$NPV_i = \frac{P(T-|i)\lambda_i}{\sum_{i=1}^{3} P(T-|i)\lambda_i}$$

In practice only those who test positive during the screening test will go forward to receive a 'gold standard' test to verify their condition. Therefore it is possible for individuals with IGT and T2DM to remain unidentified by the screening process. False positives at screening will receive an OGTT to confirm their true status. For each of the two strategies where screening takes place, screened individuals fall into six categories defined by their test result (negative or positive) and by their true glucose tolerance status (NGT, IGT or T2DM) as illustrated in figure 5.2.

The test result and their predicted true status will affect the starting state of an individual in the Markov model, as specified by the true status labels on each arm of the decision model in figure 5.2. All individuals with NGT start in the NGT state, the only consequence of incorrectly testing positive in the screening test is that they incur the extra costs of an OGTT test. Individuals with IGT start in the undiagnosed IGT state (IGTu) if they tested negative at screening, or if no screening took place. They start in the diagnosed IGT state (IGTd) if they tested positive during the screening tests. Individuals with T2DM at time of screening start in the screen detected T2DM state (T2DMs) if they have a positive screening test and undiagnosed T2DM (T2DMu) otherwise. No individuals start in the clinically detected T2DM state (T2DMc) or the death state, as these can only be entered as the model progresses.

The diagnosed and undiagnosed states of IGT and T2DM were used to model whether a diabetic receives appropriate treatment, or whether an individual with IGT receives an intervention to try

and delay T2DM. For example if an individual is in the one of the diagnosed states they will either receive an intervention to delay T2DM, or treatment for diabetes, whichever is appropriate. Alternatively if they are undiagnosed, they will not receive an intervention or treatment.

5.3.3 Structuring the Markov model

As will be described in section 5.3.4, there are seven possible states within the Markov model, and individuals were moved between states depending on expected transition rates. The decision was made not to allow individuals to move from NGT to any of the T2DM states without passing through IGT first as glucose tolerance can be thought of as a continuous scale, whereby IGT is passed through on the way to T2DM. As the model is run for yearly cycles it is clinically unlikely that an individual would move from NGT to T2DM within a year, therefore to have no direct transition between NGT and T2DM is clinically justifiable.

Individuals cannot move backwards in the model, so those defined as T2DM cannot move back into the IGT state. This is clinically correct, as once an individual is diagnosed with T2DM, even if their glucose tolerance improves, they are still clinically defined as diabetic. Additionally individuals are unable to move from IGT to NGT in the model. This was a difficult decision to make, but on balance it was decided that once an individual had suffered from impaired glucose tolerance, even if their glucose tolerance improved, their future risk of diabetes was probably more similar to individuals with IGT than individuals who had always been clinically NGT.

5.3.4 Outline of the Markov model

The Markov model consists of seven states into which an individual can be categorised. These are NGT, undiagnosed IGT (IGTu), diagnosed IGT (IGTd), undiagnosed T2DM (T2DMu), screen diagnosed T2DM (T2DMs), clinically diagnosed T2DM (T2DMc) and death. Those screen diagnosed represent individuals picked up by the initial screening programme, or those diagnosed with IGT through the initial screening programme who go on to receive interventions, as these individuals are monitored for their glucose tolerance status as part of the intervention programme. Clinically diagnosed individuals with T2DM are individuals diagnosed through standard clinical practice, as is currently the norm, and are therefore individuals who are diagnosed and move from the undiagnosed T2DM state.

The decision tree determines how many individuals from a given population start in each state. Four Markov models were run simultaneously, one for each of the screening strategies. Diagnosed IGT and T2DM individuals receive relevant treatments or interventions, whereas those undiagnosed do not receive anything and were modelled accordingly in terms of complications and death rates. No individuals start in the death state, or the clinically detected T2DM state.

The idea of the Markov model is that incidence rates were used to move individuals between states. For example if you know the yearly transition rates between states, you could run the model ten times to predict how many people you would expect to have T2DM in ten years time. For individuals with IGT, transition to T2DM will differ depending on if the individual is receiving an intervention or not. Transition rates may be age or time dependent, and therefore may vary over the time horizon of the model. Specific details of the transition rates used and how they were modelled, is given in Chapter 6.

Transition from T2DM to death was affected by whether an individual was undiagnosed, was diagnosed through screening, or was detected clinically. Undiagnosed individuals will have the highest age specific death rates, as these individuals will not be receiving appropriate treatment for their condition. Screen diagnosed individuals with T2DM, will have the lowest death rates of diabetics, as they would have been detected early, and therefore an assumption can be made that their blood glucose will be better controlled leading to fewer complications, than if they had been clinically diagnosed, as is currently the norm.

Output from the Markov model gives total person years spent in each of the seven states for a given population and time period, for each of the four screening strategies. Cost and utilities of each state can then be used, along with costs of screening, to assess the clinical and cost implications of each of the screening strategies, as shown in detail in Chapter 6. Complications associated with T2DM, which increase with duration of the disease, are modelled generally in terms of their implications on costs and QALYs, using published data from the UKPDS study.

5.4 Chapter summary

A number of cost-effectiveness analyses have been published that consider strategies for screening or prevention of T2DM, with inconclusive and often conflicting results. Data utilised has often been limited to a minimal number of sources, rather than trying to encompass all available data. Models assessing interventions have been limited in that they only considered a small part of the disease pathway between IGT and T2DM, so that the costs or effectiveness of screening for IGT was often not included in the decision model, or the effect of screening and interventions once T2DM had developed. Also no current model has directly compared the two active screening strategies of screening for T2DM followed by treatment, or screening for both IGT and T2DM followed by interventions and treatment.

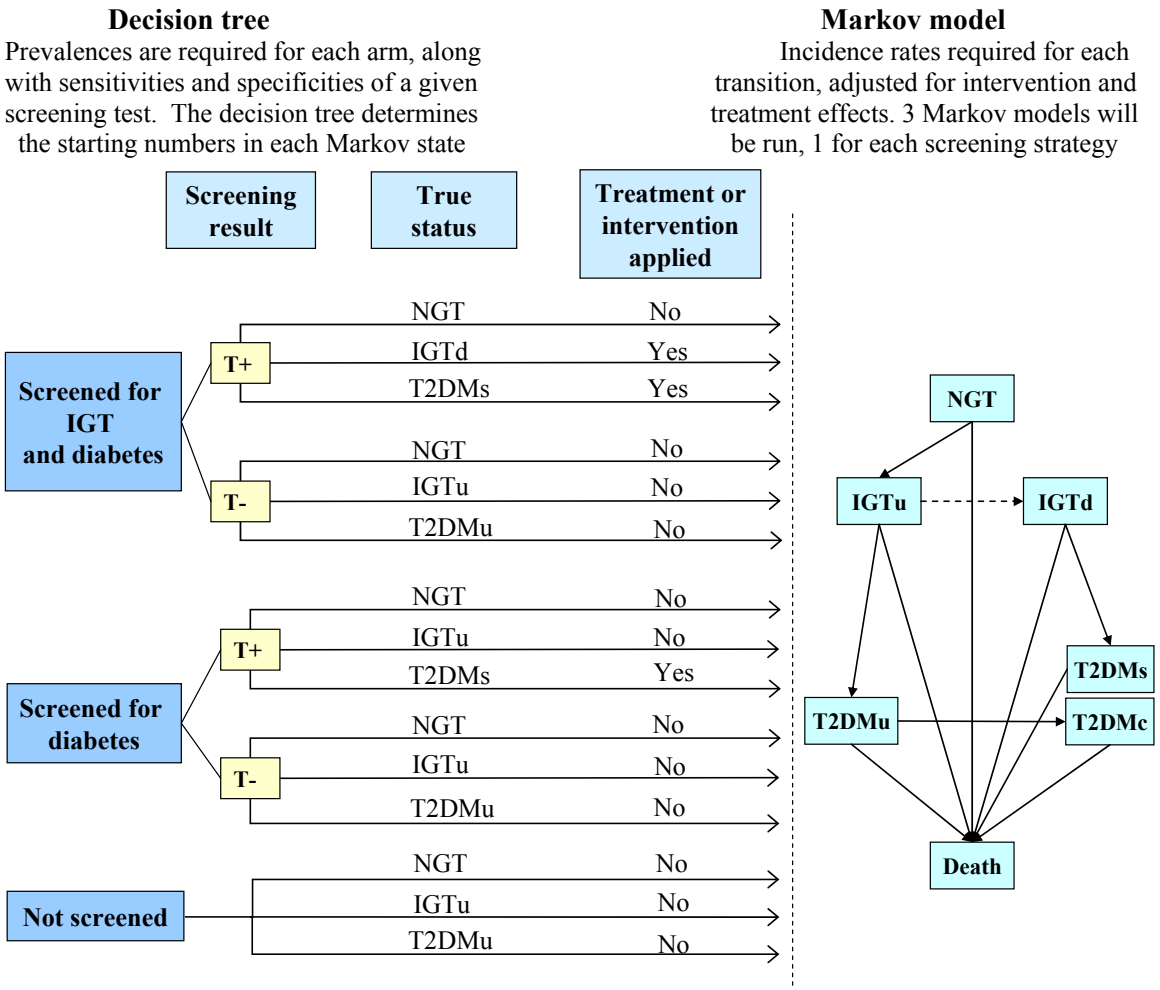
This study adds to current research by providing a full overview of the screening and intervention pathway, using all appropriate data sources where possible. The model provides the first direct comparison of no screening, screening for T2DM alone and screening for IGT and T2DM together. The structure of the model developed here was determined using clinical information on the IGT and T2DM disease pathway and was based on plausible future health policies for T2DM prevention. The meta-analyses carried out for Chapter 4 were incorporated within the full comprehensive decision model, to provide estimates of intervention effects. Details of the data utilised for the model, how the data was managed and combined, and how it was used to model different screening scenarios, is explained in full in Chapter 6.

6. DEVELOPMENT OF A PRIMARY MODEL

6.1 Chapter overview

This chapter describes the development of a base case, primary model. A detailed description of the data sources to be used is given, along with information on data extraction and transformations, and how these feed into the model. The assumptions made for the purpose of the primary model are described and discussed and the initial results from the primary model presented. The primary model is then expanded, improved upon and explored through further analyses in chapters 7 and 8. The primary model was initially run for a time horizon of 50 years, for a Caucasian population, who are 45 years of age at the start of the model. Figure 6.1 is a reminder of the proposed model structure.

Figure 6.1: The proposed screening and intervention model



(NGT=Normal Glucose Tolerance, IGT=Impaired Glucose Tolerance, T2DM=Type 2 Diabetes Mellitus, u=undiagnosed, d=diagnosed, s=screen diagnosed and c=clinically diagnosed)

6.2 Screening data and prevalences

For the decision tree component of the model data was needed on the accuracy of a chosen screening test and the prevalence of NGT, IGT and T2DM in the population to be modelled. This information will enable the probabilities of starting in each of the Markov states to be determined. For the primary model information on screening tests and prevalences were taken from IPD available from the STAR study. For this study individuals aged 40-75 years (Caucasians) or 25-75 years (Non-Caucasians), who had at least one recognised risk factor for T2DM, from 15 general practices in Leicestershire, were invited for screening. Risk factors included a known history of coronary heart disease, hypertension, dyslipidaemia, cerebrovascular disease, a first degree relative with T2DM and a body mass index greater than 25 kg/m^2. Therefore the screening data included in the primary model was from a population considered to be 'at risk' of T2DM.

The STAR study examined use of both the FPG test and the HbA_{1c} test as potential screening tools, with the true glucose tolerance status of the participant being confirmed using an OGTT. The receiver operating characteristic (ROC) curves for both tests, when testing for either T2DM alone or IGT and T2DM in combination, are given in figures 6.2 and 6.3. These curves are a plot of sensitivity against 1-specificity, at each possible cut-off of the screening test (Altman, 1991). If the 'cost' of a false negative result is considered to be the same as that of a false positive, then the optimum cut-off for the screening test can be identified as the cut-off that maximises the sum of the sensitivity and specificity, which is the point on the ROC curve nearest the top left hand corner (Altman, 1991).

From the ROC curves it can be seen that the FPG test was more accurate than the HbA_{1c} for both scenarios, and therefore the FPG test was assumed to be the screening test used for our initial model. Both tests performed better when testing for T2DM alone, rather than IGT and T2DM in combination. STAR examined a mixed population of both Asians and Caucasians. As our initial model was to assess a Caucasian 'at risk' population, just the data on Caucasians was used, although the data on Asians was used for a model extension investigating model results for different ethnic cohorts. For the two active screening strategies (screening for T2DM only, or screening for both IGT and T2DM) cut-offs for the FPG test were taken as optimal values, i.e. the point on the curve nearest the top left hand corner, as circled, on the respective ROC curve.

Figure 6.2: ROC curves for screening tests for T2DM

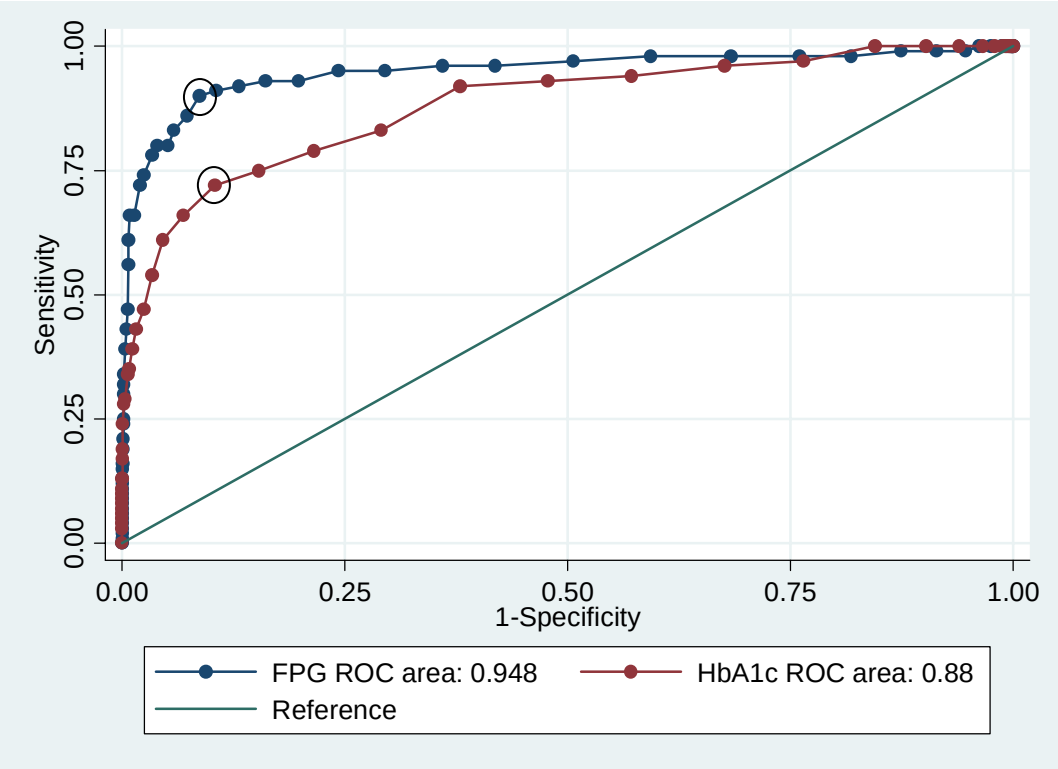


Figure 6.3: ROC curves for screening tests for IGT and T2DM combined

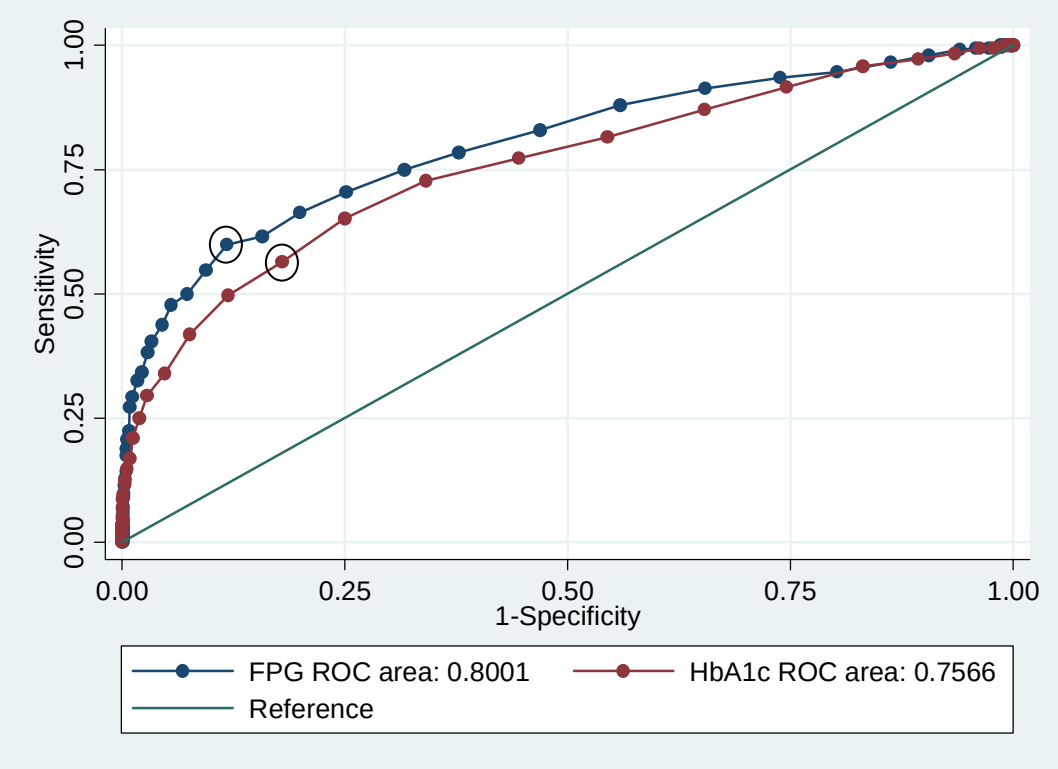


Table 6.1 displays the results of the STAR study for the FPG test in Caucasian adults over the age of 45 years. The area under the curve (AUC) are given, which represents how effective the test is at distinguishing between individuals with or without the disease. The area measures discrimination, that is, the ability of the test to correctly classify those with and without the disease. For example, if two individuals are randomly picked, one from a diseased group and one from a healthy group, and both are tested, the patient with the more abnormal test result should be the one from the diseased group. The area under the curve is the percentage of randomly drawn pairs for which this is true. A perfect test would have an AUC of 1, whilst a poor test has an AUC of 0.5. It can be seen that the use of traditional diabetes tests to screen for IGT is not ideal, with an area under the ROC curve of just 0.85 (95% CI: 0.82, 0.87) and low sensitivity at the optimum cut-off of just 59.4%. It is likely that a screening programme for IGT would have to include a questionnaire or risk score to improve sensitivity. This is explored further in Chapters 7 and 8, where sensitivity analyses are carried out on both test sensitivity and prevalence of IGT and T2DM.

Table 6.1: Performance of the FPG in the STAR study

	T2DM only	**IGT and T2DM**
Optimal cut-off	>=6.0 mmol/l	>=5.7 mmol/l
Sensitivity	90/105 (85.7%)	219/369 (59.4%)
Specificity	1957/2111 (92.7%)	1626/1847 (88.0%)
Area under ROC curve (95% CI)	0.95 (0.92, 0.98)	0.85 (0.82, 0.87)

Table 6.2: Results of the FPG test in the STAR study

	True status as confirmed by OGTT, N(%)			
FPG result (mmol/l)	**NGT**	**IGT**	**T2DM**	**Totals**
<5.7	1626 (88.0)	142 (53.8)	8 (7.6)	1776
>=5.7 and <6.0	138 (7.5)	51 (19.3)	7 (6.7)	196
>=6.0	83 (4.5)	71 (26.9)	90 (85.7)	244
Totals	1847 (83.3)	264 (11.9)	105 (4.7)	2216

Table 6.2 displays the STAR data in terms of participant numbers and FPG results. Individuals with a test result of FPG < 5.7 would test negative for both screening scenarios, that is they would test negative if they were screened for T2DM or both IGT and T2DM. Individuals with an FPG >=6.0mmol/l would test positive for both active screening scenarios, and those falling in between would test positive if they were being screened for both IGT and T2DM and negative if they were just being screened for T2DM only. The middle row is therefore combined with the top or bottom row depending on the screening scenario considered.

Table 6.2 was entered as data into the model with the three test outcomes (j) modelled under a multinomial distribution for each of the three true glucose tolerance status categories (i). A vague dirichlet prior distribution was placed on each of the three test probabilities, expressing the belief that each test outcome was equally likely within each glucose tolerance status

(equation 6.1, where r_{ij} represents the numbers for each of the three test outcomes (j) and for each of the three glucose tolerance states (i), π_{ij} represents the probabilities of each of the three tests results, and N_i represents the total number in each of the three states) (Briggs et al., 2003). The multinomial distribution is a generalization of the binominal distribution, and the dirichlet an extension of the beta distribution, both for when there are more than two outcomes (as described in more detail in section 2.3.3).

$$r_{ij} \sim multinomial(r_{ij}, N_i) \qquad \text{[Equation 6.1]}$$

$$\pi_{ij} \sim dirichlet(1,1,1)$$

$$and \quad \pi_{i1} + \pi_{i2} + \pi_{i3} = 1$$

From table 6.2 sensitivities and specificities for either screening scenario can be computed, e.g. P(test positive for T2DM|true status is T2DM). For the model though predictive values were needed, e.g. P(true status is T2DM|test positive for T2DM). Predictive values for each screening scenario were computed using equation 6.2, with i representing either NGT, IGT or undiagnosed T2DM, and λ_i the prevalence of each glucose tolerance state, where $\lambda_1 + \lambda_2 + \lambda_3 = 1$. For the primary model prevalences were taken from the Caucasians involved in the STAR study and were as follows; NGT = 1847/2216 = 83%, IGT = 264/2216 = 12%, T2DM = 105/2216 = 5%. The prevalences were modelled using a dirichlet distribution. The WinBUGS code for the screening decision tree is given in box 1.

$$PPV_i = \frac{P(T+|i)\lambda_i}{\sum_{i=1}^{3} P(T+|i)\lambda_i} \qquad \text{[Equation 6.2]}$$

$$NPV_i = \frac{P(T-|i)\lambda_i}{\sum_{i=1}^{3} P(T-|i)\lambda_i}$$

Table 6.3 displays the predictive values computed and utilised by the primary model

As the predictive values sum to 1 for each test result (T- or T+) an overall probability was calculated, which was the predictive probability multiplied by the probability of the test result (both calculated using data in table 6.2). The overall probabilities can then be used to determine

the probabilities of starting in each of the Markov states for each of the three screening scenarios (table 6.3 and 6.4). From table 6.3 it can be seen that when both IGT and T2DM were screened for, 50% of the positive tests were actually in individuals with NGT, which dropped to 34% when only T2DM was screened for. Of those who tested negative for the IGT and T2DM screen, 9% were misclassified as negative when they actually had IGT or T2DM, and for the T2DM screening of those who tested negative only 1% had T2DM.

Table 6.4 shows the probability of starting in each of the seven states, for each of the three screening possibilities; no screening, screening for T2DM only and screening for IGT and T2DM in combination. Where a strategy of no screening is modelled no individuals start in the diagnosed IGT or T2DM states (IGTd and T2DMd). Screening for IGT and T2DM leads to a greater number of T2DM individuals being identified, due to the lower cut-off of the screening test. Although individuals with IGT are identified when only T2DM is screened for, the model was structured so that they did not receive any interventions, as discussed in Chapter 5. This was because it was considered that a policy of screening for T2DM only would not lead to preventative interventions being presented to individuals identified with IGT. The probability of starting in the death state at time 1 was set to 0 for all three screening scenarios.

Table 6.3: Predictive values and overall probabilities

	Predictive value *P(i|T)*	Probability of test result	Overall probability for each screening strategy
Screening for both IGT and T2DM			
P(NGT|T+)	0.50	T+=0.20	0.100 (0.098, 0.102)
P(IGT|T+)	0.28		0.056 (0.049, 0.062)
P(T2DM|T+)	0.23		0.046 (0.036, 0.052)
P(NGT|T-)	0.91	T-=0.80	0.728 (0.720, 0.747)
P(IGT|T-)	0.08		0.064 (0.057, 0.072)
P(T2DM|T-)	0.01		0.004 (0.003, 0.004)
Screening for T2DM only			
P(NGT|T+)	0.34	T+=0.11	0.037 (0.036, 0.038)
P(IGT|T+)	0.29		0.032 (0.029, 0.036)
P(T2DM|T+)	0.39		0.043 (0.033, 0.049)
P(NGT|T-)	0.89	T-=0.89	0.793 (0.781, 0.812)
P(IGT|T-)	0.10		0.088 (0.077, 0.097)
P(T2DM|T-)	0.01		0.005 (0.006, 0.008)
No Screening			
Prev(NGT)	-	-	0.833 (0.818, 0.849)
Prev(IGT)	-		0.119 (0.106, 0.133)
Prev(T2DM)	-		0.047 (0.039, 0.057)

Table 6.4: Probabilities of starting in each Markov state

Markov State	Screening for IGT and T2DM	Screening for T2DM only	No screening
NGT	0.833 (0.818, 0.849)	0.833 (0.818, 0.849)	0.833 (0.818, 0.849)
IGTu	0.064 (0.057, 0.072)	0.087 (0.077, 0.097)	0.119 (0.106, 0.133)
IGTd	0.056 (0.049, 0.062)	0.032 (0.029, 0.036)	0
T2DMu	0.004 0.003, 0.004)	0.007 (0.006, 0.008)	0.047 (0.039, 0.057)
T2DMd	0.044 (0.036, 0.052)	0.041 (0.033, 0.049)	0
Death	0	0	0

For the primary model the number screened was set to 1, so that results could be interpreted as effects for a single individual. The cut function in WinBUGS was used between the decision tree and the Markov model to prevent information entered into the Markov model, influencing the results of the decision tree. The cut function forms a valve in the code, such that prior information is allowed to flow downwards through the cut, but likelihood information is prevented from flowing upwards. The WinBUGS code for the whole decision tree, used to model the screening part of the comprehensive decision model, is given in Box 1. Information from the three possible screening scenarios, provided starting numbers for four possible screening strategies; no screening, screening for T2DM alone and screening for IGT and T2DM in combination, followed by either lifestyle of pharmacological interventions.

Box 1: WinBUGS code for the decision tree

Status 1 NGT, 2 IGT, 3 T2DM, N[status] = N, T+ shows a positive result for both cut-offs, T- a negative result for both cut-offs and T0 a positive result for the high cut-off (T2DM only) and a negative result for the low cut-off (T2DM and IGT). Hence r[status,1] is pr(T- |status), r[status,2] is pr(T0|status), and r[status,3] is pr(T+|status)

```
for (status in 1:3) {
                    N[status] <- sum(x[status,1:3])
                    x[status,1:3] ~ dmulti(r[status,1:3],N[status])
                    r[status,1:3] ~ ddirch(prior_r[status,1:3])
                                                        }
        prev[1:3] ~ ddirch(preva[1:3])
```

Positive and negative predictive probabilities, neglow n(status | T-.low), poslow n(status | T+.low), neghigh n(status |T-.high), poshigh n(status | T+.high)

```
for (status in 1:3) {
neglow[status] <- (x[status,1]/N[status]*prev[status])*Nscreen
poslow[status] ((x[status,2]/N[status]*prev[status])+(x[status,3]/N[status]*prev[status]))*Nscreen
neghigh[status] ((x[status,1]/N[status]*prev[status])+(x[status,2]/N[status]*prev[status]))*Nscreen
poshigh[status] <- (x[status,3]/N[status]*prev[status])*Nscreen}
```

Numbers for each of three screening scenarios, T2DM only screening (high cut-off)[1], IGT and T2DM screening (low cut-off)[2], and no screening [3], in terms of whether a diagnosis will be made for each of the three glucose tolerance states (NGT, IGT and T2DM). D=diagnosed and U=undiagnosed.

```
NGT[1] <- poshigh[1]+neghigh[1]
IGTD[1] <- poshigh[2]
IGTU[1] <- neghigh[2]
T2DMD[1] <- poshigh[3]
T2DMU[1] <- neghigh[3]

NGT[2] <- poslow[1]+neglow[1]
IGTD[2] <- poslow[2]
IGTU[2] <- neglow[2]
T2DMD[2] <- poslow[3]
T2DMU[2] <- neglow[3]

NGT[3] <- prev[1]*Nscreen
IGTU[3] <- prev[2]*Nscreen
T2DMU[3] <- prev[3]*Nscreen
```

Box 1 continued.
Starting numbers in each state for each of the four strategies, at the start of the model (strategy,state,time=1). Where strategy 1=T2DM screening, 2=IGT & T2DM screening with lifestyle intervention, 3=IGT & T2DM screening with metformin intervention and 4=no screening. States, 1=NGT, 2=IGTu, 3=IGTd, 4=T2DMu (undiagnosed), 5=T2DMs (screen diagnosed), 6=T2DMc (clinically diagnosed) and 7=death

```
number[1,1,1] <- cut(NGT[1])
number[1,2,1] <- cut(IGTU[1])
number[1,3,1] <- cut(IGTD[1])
number[1,4,1] <- cut(T2DMU[1])
number[1,5,1] <- cut(T2DMD[1])
number[1,6,1] <- 0
number[1,7,1] <- 0
number[2,1,1] <- cut(NGT[2])
number[2,2,1] <- cut(IGTU[2])
number[2,3,1] <- cut(IGTD[2])
number[2,4,1] <- cut(T2DMU[2])
number[2,5,1] <- cut(T2DMD[2])
number[2,6,1] <- 0
number[2,7,1] <- 0
number[3,1,1] <- cut(NGT[2])
number[3,2,1] <- cut(IGTU[2])
number[3,3,1] <- cut(IGTD[2])
number[3,4,1] <- cut(T2DMU[2])
number[3,5,1] <- cut(T2DMD[2])
number[3,6,1] <- 0
number[3,7,1] <- 0
number[4,1,1] <- cut(NGT[3])
number[4,2,1] <- cut(IGTU[3])
number[4,3,1] <- 0
number[4,4,1] <- cut(T2DMU[3])
number[4,5,1] <- 0
number[4,6,1] <- 0
number[4,7,1] <- 0
```

6.3 Transition probabilities

Once the probability of starting in each state for each of the three screening strategies was established, a Markov model could be run for each strategy using annual transition rates to move individuals between states over a number of cycles. Information on each transition rate is described below. Transition rates were taken from studies that observed similar populations to those considered in the primary model. Although numbers starting in each state differ by screening strategy, transition rates were the same for each strategy, with the only exception being the transition from diagnosed IGT to screen diagnosed T2DM. For the two screening

strategies where no screening or T2DM only screening was implemented, the transition rate was the same as that from undiagnosed IGT to undiagnosed T2DM, and was the estimated population transition rate from IGT to T2DM. For the screening strategy where both IGT and T2DM were screened for, for individuals identified as having IGT, the transition rate was reduced from the population average by the estimated intervention effect of either lifestyle or pharmacological interventions. Therefore the transition rate was lower between diagnosed IGT and screen diagnosed T2DM for this screening strategy.

Transition rates were transformed, so that they were all in the same format of per 100 person years on the natural logarithm scale. The standard error (s) of the log transition rate per 100 person years was calculated as in equation 6.3 (Clayton and Hills, 1993), where d represents the number of events over 100 person years. Where there was just one estimate of the transition rate (λ), it was modelled using a normal distribution on a log scale (equation 6.4). If there was more than one estimate of a transition rate they were combined using a random effects meta-analysis, and the pooled mean rate (μ) was used in the model (equation 6.5) (Sutton et al., 2000). The rate/pooled rate was then exponentiated and divided by 100 to give a rate per single person year. As transition rates represent the instantaneous potential for change, whereas the yearly probabilities or risk of individuals moving between states were required for the model, rates were converted to probabilities (γ) using equation 6.6 (Miller and Homan, 1994), which assumes a constant rate within each cycle.

$$s = \frac{1}{\sqrt{d}} \qquad \text{[Equation 6.3]}$$

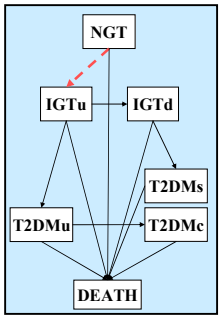
$$Ln(\lambda) \sim N(\mu, s^2) \qquad \text{[Equation 6.4]}$$

$$Ln(\lambda_i) \sim N(\theta_i, s_i^2) \qquad \text{[Equation 6.5]}$$

$$\theta_i \sim N(\mu, \tau^2)$$

$$\gamma = 1 - e^{-\mu} \qquad \text{[Equation 6.6]}$$

6.3.1 NGT to undiagnosed IGT



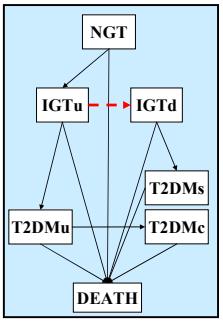
Two studies were identified that reported transition rates from NGT to IGT, these were the Baltimore longitudinal study of aging (Meigs et al., 2003) and the Isle of Ely study (Wareham et al., 1999). The results are reported in Table 6.5. As the transition rates appeared to differ between age groups, in that the mean transition was estimated as 12.06 cases ($e^{-2.49}$) per 100 person years for the over 65s and just 5.26 ($e^{-1.66}$) for the under 65s, only the Baltimore data was used in the primary model. This enabled different transition rates to be used when modelling individuals under 65 years of age or those 65 years and over. The transition rate was used to model the transition from NGT to IGTu, and not IGTd, as the transition would occur in the model post screening, therefore development of IGTu would remain undetected.

Table 6.5: Studies reporting transition rates from NGT to IGT

Study	Year	Location	N and Follow-up	Population	Log transition rate per 100pyrs (SE)
Baltimore study	1979-88	US	488 11 years	96% Caucasian 60.7% male Age 52.9 (16.6)	All: 1.86 (0.06) ≥65yrs: 2.49 (0.11) <65yrs 1.66 (0.08)
Isle of Ely study	1990-	UK	767 4.44 years	Caucasians 42% male Age 54.9	0.54 (0.13)

* Mean age in years reported, with SD given in parenthesis where available

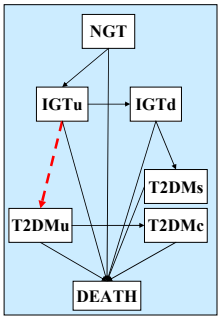
6.3.2 Undiagnosed IGT to diagnosed IGT



The transition rate from undiagnosed to diagnosed IGT will be zero unless a re-screening strategy is assumed, whereby the initial population is re-screened every set number of years allowing for further IGT patients to be identified. In the primary model it will be assumed that only one round of screening will take place and this will be at the start of the model and will allow the identification of

individuals starting in the diagnosed IGT state. As no further rounds of screening will be incorporated into the primary model the transition rate from undiagnosed to diagnosed IGT will be set to zero. A policy of multiple screening rounds is explored in chapter 8.

6.3.3 Undiagnosed IGT to undiagnosed T2DM



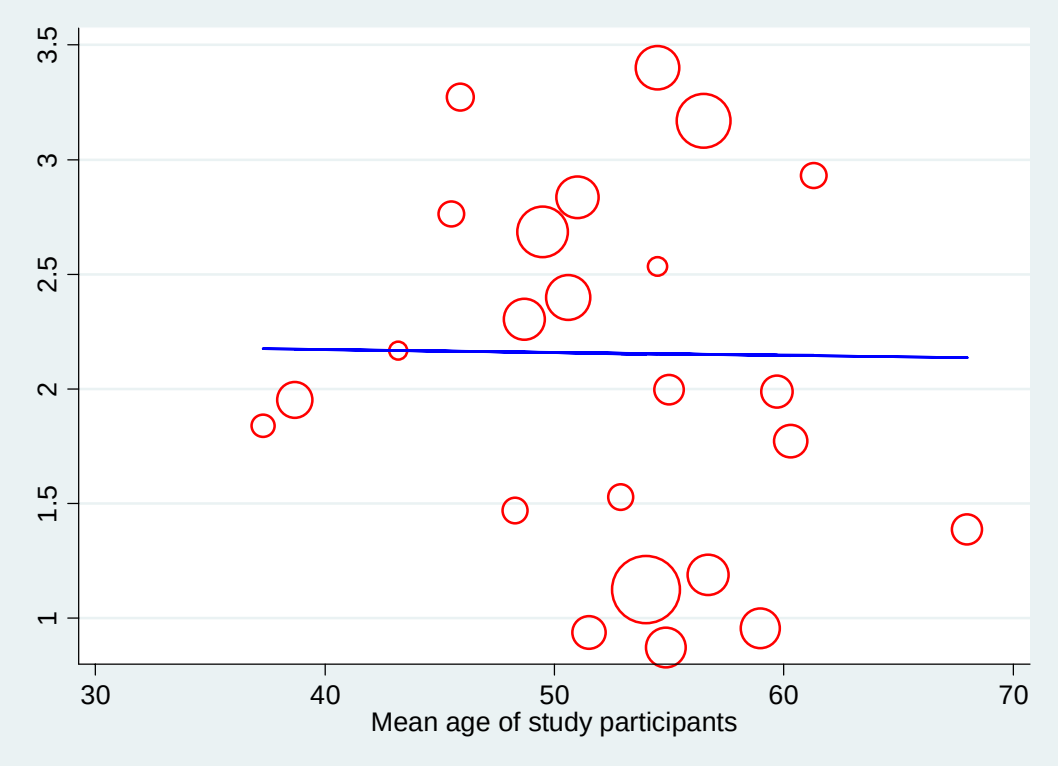
The transition rate from undiagnosed IGT to undiagnosed T2DM is the transition with the most data available. 24 studies were found that reported this transition, comprising of 10 epidemiological studies (Bonora et al., 2004, de Vegt et al., 2001, Edelstein et al., 1997a, Meigs et al., 2003, Rasmussen et al., 2006, Wareham et al., 1999) and the control groups of 14 trials investigating lifestyle and anti-diabetic drug interventions, as used in the meta-analyses in Chapter 4 (Chiasson et al., 2002, Eriksson et al., 2006, Fang et al., 2004, Jarrett et al., 1984, Knowler et al., 2002, Kosaka et al., 2005, Li et al., 1999, Liao et al., 2002, Lindstrom et al., 2003b, Pan et al., 2003, Pan et al., 1997, Ramachandran et al., 2006, Tao et al., 2004, Wein et al., 1999).

As control groups were being combined from both lifestyle and drug trials, where potentially the study populations may differ between the two types of trials, the control groups were compared to check they did not differ in terms of their risk of T2DM using the MTC model described in Chapter 4. The risk of T2DM did not differ significantly between the two types of controls, hazard ratio 1.032 (95% CI: 0.402 to 2.174).

The 24 studies were very diverse in terms of their study populations, therefore interactions between transition rate and mean age, ethnicity or gender were investigated by fitting meta-regression models in WinBUGS (both WinBUGS and meta-regression are described in Chapter 2, and the code is given in Appendix 3.4). Gender was entered as percentage of the study sample that were male, and age was entered as mean age of the study group in years. The studies were a mixture of seven broad ethnic categories, therefore categories were combined and ethnicity was dealt with by coding studies as of being an ethnic group at high risk of developing T2DM (Asian Indians, Mexican-Americans, Pima Indians, Hispanics, Micronesians, Japanese and Chinese), or at low risk (Caucasians), depending on the predominant ethnicity of the study.

There was no evidence to suggest an interaction between gender and the log transition rate from IGT to T2DM; gradient of the meta-regression line -0.002 (95% CrI: -0.001 to 0.006). Similarly with age, the meta-regression results showed a decrease in the log transition rate of -0.001 (95% CrI: -0.007 to 0.005) for an increase in the mean age of the study participants of 1 year, again indicating no association as the credible interval contains the null value of 0. For ethnicity, where studies of low risk ethnic groups were compared against studies of higher risk groups, the log transition rate was slightly higher in the high risk group, although the credible interval again contained the null value of 0, meta-regression results of 0.0003 (95% CrI: -0.095 to 0.095). However, as discussed in Chapter 2, meta-regression analyses have low power to detect any effect (Lambert et al., 2002), and ideally individual patient data is needed to fully investigate and understand the impact of participant characteristics on transition rates.

Figure 6.4: **Relationship between mean age of study participants and the log transition rate from IGT to T2DM**



For illustration of the meta-regressions figure 6.4 shows the meta-regression model fitted for age and the log transition rate. Many studies have reported increasing transition rates from NGT to T2DM as age increases (Meigs et al., 2003). Although evidence was found supporting an increase in the transition rate from NGT to IGT as age increases, no evidence was found supporting an increase in IGT to T2DM transition rates as age increases. Therefore it was decided to combine all relevant information into one mean transition rate and to use this for all ages in the model.

Although no evidence was found to support a difference in risk of developing diabetes between different ethnic populations, evidence is available supporting the fact that transition rates from IGT to T2DM differ between different ethnic groups. As the primary model was based on a Caucasian cohort, and as the data was available to enable transition rates to be modelled by ethnic group, only the 12 studies where Caucasians attributed to over 50% of the study sample (Table 6.6) were used to calculate the mean transition rate for the primary model, which was estimated as a mean log rate of 1.956 (0.252) per 100 person years.

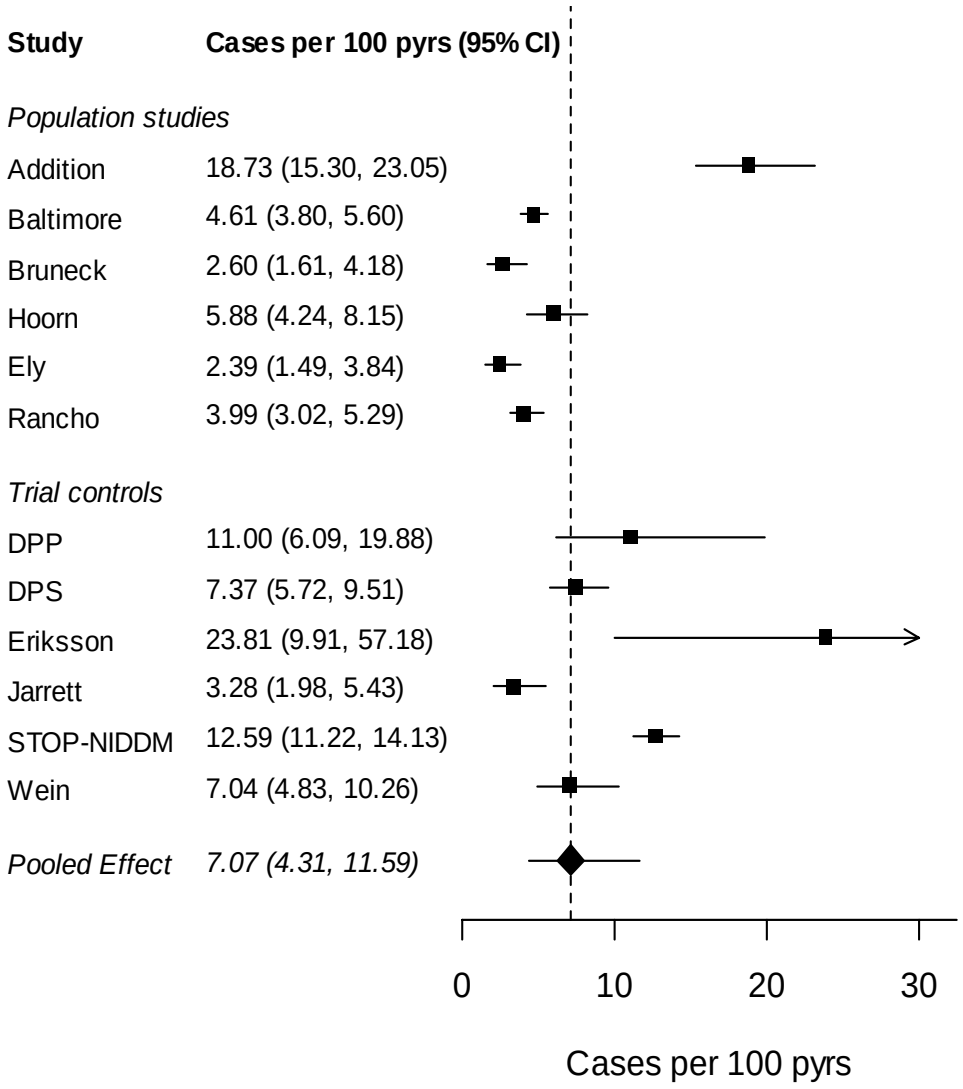
The forest plot detailing this meta-analysis of transition rates from IGT to T2DM is given in figure 6.5 and here study results are reported as cases per 100 years, to allow ease of interpretation and comparison between studies. The population studies tended to have narrower confidence intervals than the control arms on intervention trials. This is probably because the follow-up time was generally longer in the population studies, and therefore the number of events (cases of T2DM) tended to be greater. The estimated transition rate varied greatly between studies, with an estimated between study standard deviation of 0.76 (0.45, 1.33), although as already described meta-regression analyses failed to identify sources of heterogeneity. To fully understand why the transition rate estimates varied, IPD data would be required. The meta-analyses of transition rates from IGT to T2DM was carried out in WinBUGS, within the full comprehensive decision model, and therefore was an MCMC analysis. Details of convergence and autocorrelation tests, as well as sensitivity analyses for the prior distribution of the between study variance, are detailed in Chapter 7. Other ethnic groups and mixed ethnic populations were considered in extensions of the primary model in Chapter 8.

Table 6.6: Studies reporting transition rates from IGT to T2DM

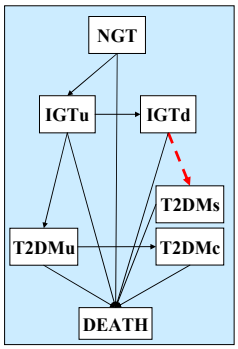
Study	Year	Location	N and Follow-up	Population*	Log transition rate per 100 pyrs (SE)
Population studies					
Addition	2001-2003	Denmark	503 12.5 months	Caucasian 44.9% male Age 61.3	2.93 (0.103)
Baltimore study	1979-1988	US	265 11 years	96% Caucasian 60.7% male Age 52.9 (16.6)	1.528 (0.099)
Bruneck study	1990-2000	Italy	837 10 years	Caucasians 51% male Age 59 (11)	0.955 (0.243)
Hoorn study	1989-1992	Netherlands	111 6.4 years	Caucasian 45% male Age 60.3 (6.9)	1.771 (0.167)
Isle of Ely study	1990	UK	170 4.44 years	Caucasian 42% male Age 54.85	0.872 (0.242)
Rancho Bernardo study	1984-1996	US	186 6.60 years	Caucasian 35.5% male Age 68.0	1.386 (0.143)
Control arms of intervention trials					
DPP	1996-2002	USA	335 2.80 years	54.7% Caucasian 32.3% male Age 50.6	2.398 (0.302)
DPS	1993-2004	Finland	257 3.20 years	Caucasian 33% male Age 55.0	1.998 (0.130)
Eriksson	2006	Finland	17 1.32 years	Caucasian 26% male Age 56.5	3.170 (0.447)
Jarrett	1968-1980	UK	53 4.36 years	Caucasian 100% male Age 56.7	1.187 (0.258)
STOP-NIDDM	1998-2001	Canada, Europe and Israel	715 3.30 years	97% Caucasian 49% male Age 54.5	2.533 (0.059)
Wein	1989-1991	Australia	100 4.25 years	Caucasian 0% male Age 38.7	1.952 (0.192)

* Mean age in years reported, with SD given in parenthesis where available

Figure 6.5: Meta-analyses of transition rates from IGT to T2DM



6.3.4 Diagnosed IGT to screen diagnosed T2DM



This is the one transition rate that varies between the three screening strategies. An assumption was made that all individuals identified as having IGT by the initial screening programme, would have their blood glucose monitored as part of the intervention programme, and therefore would be diagnosed with T2DM when it developed. There was therefore no movement between diagnosed IGT and undiagnosed T2DM, and individuals in the diagnosed IGT state moved straight to the state of screen diagnosed T2DM .

For the strategy where no screening takes place and for the screening strategy aimed at identifying individuals with T2DM only, then this transition rate will be the same as for from undiagnosed IGT to undiagnosed T2DM. For the screening strategy where both IGT and T2DM are screened for then it will be assumed that all the individuals identified as having IGT will receive an appropriate intervention, to try and prevent or delay T2DM. Therefore this transition rate will be the same as for from undiagnosed IGT to undiagnosed T2DM but with an intervention effect applied.

For the primary model both lifestyle interventions and anti-diabetic agents will be assessed, with both meta-analyses of intervention trials incorporated within the decision model. The pooled hazard ratio for lifestyle interventions, as described in chapter 4, is 0.53 (95% CrI: 0.44 to 0.65), and for anti-diabetic agents it is 0.66 (0.47 to 0.83). All available data was used for the pooled hazard ratios as interactions between intervention effect and baseline risk and age had been tested for and no strong evidence of effect was found, as detailed in Chapter 4. The hazard ratios, and there associated distributions, were applied to the transition rate before they were converted into probabilities. Figures 6.6 and 6.7 detail the two intervention meta-analyses run within the decision model. Details of convergence and autocorrelation tests for these two meta-analyses, as well as sensitivity analyses for the prior distribution of the between study variance, are described in Chapter 7.

Figure 6.6: Meta-analysis of lifestyle interventions run within the decision model

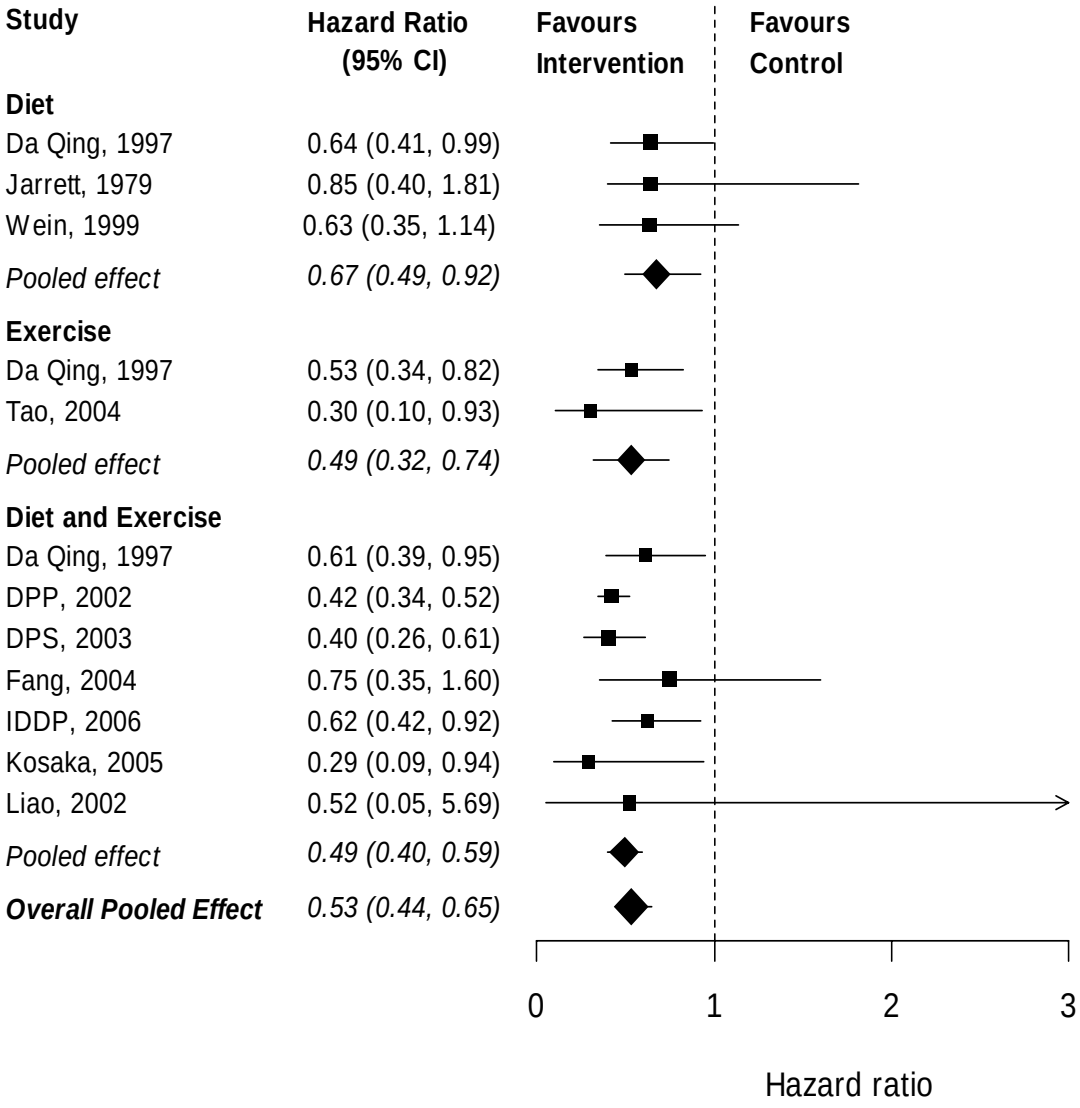
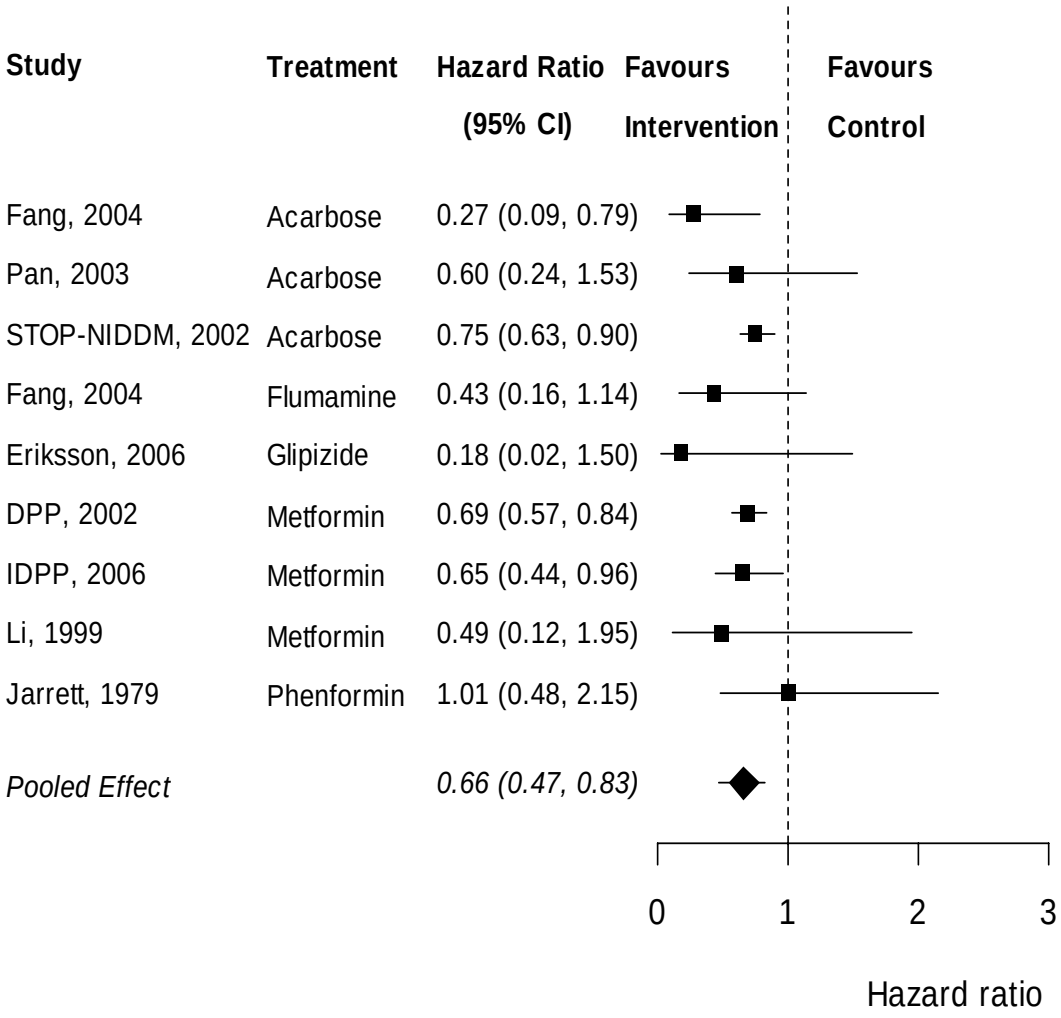


Figure 6.7: Meta-analysis of anti-diabetic agents run within the decision model



A further consideration was whether the intervention effect being entered into the model, was relevant to the estimated transition rate from IGT to T2DM. To determine this the results from the baseline risk models detailed in Chapter 4 were assessed. Figures 6.8 and 6.9 show the intervention effect (Log incidence rate ratio) plotted against the baseline risk (Log incidence rate in the control group) for each trial, for both anti-diabetic agents and lifestyle interventions. Each circle represents one trial and the size of the circle represents the weight given to the trial in the baseline risk model, with trials being weighted using the inverse of the SE of the incidence rate ratio. The plotted lines represent the effect of baseline risk on the intervention effect as estimated by the baseline risk models.

The pooled log transition rate from IGT to T2DM that will be entered into the overall model is 1.96 (95% CrI: 1.51 to 2.44), as described in section 6.3.3. Using the graph in figure 6.8, the effect of lifestyle interventions where the baseline risk of diabetes is a log incidence rate of 1.96, can be read from the graph as an incidence rate ratio on the log scale of -0.637. When exponentiated this is 0.529 and is close to the value used in the model for lifestyle intervention effect, which was 0.53 (95% CrI: 0.43 to 0.65). For anti-diabetic agents (figure 6.9), an incidence rate of 1.96 predicts an intervention effect on the log scale of -0.510, which when exponentiated gives an IRR of 0.60, as compared to the figure used in the decision model which was 0.66 (95% CrI: 0.47 to 0.83). Therefore both the baseline risk models predict a similar intervention effect when the baseline risk of the model population is accounted for, compared to those used within the comprehensive decision model, as estimated from the intervention trials. The impact of a reduction in intervention effects, on the decision model conclusions is explored in chapter 8, where models are run with reduced compliance to interventions.

Figure 6.8: Effect of baseline risk on the effectiveness of lifestyle interventions

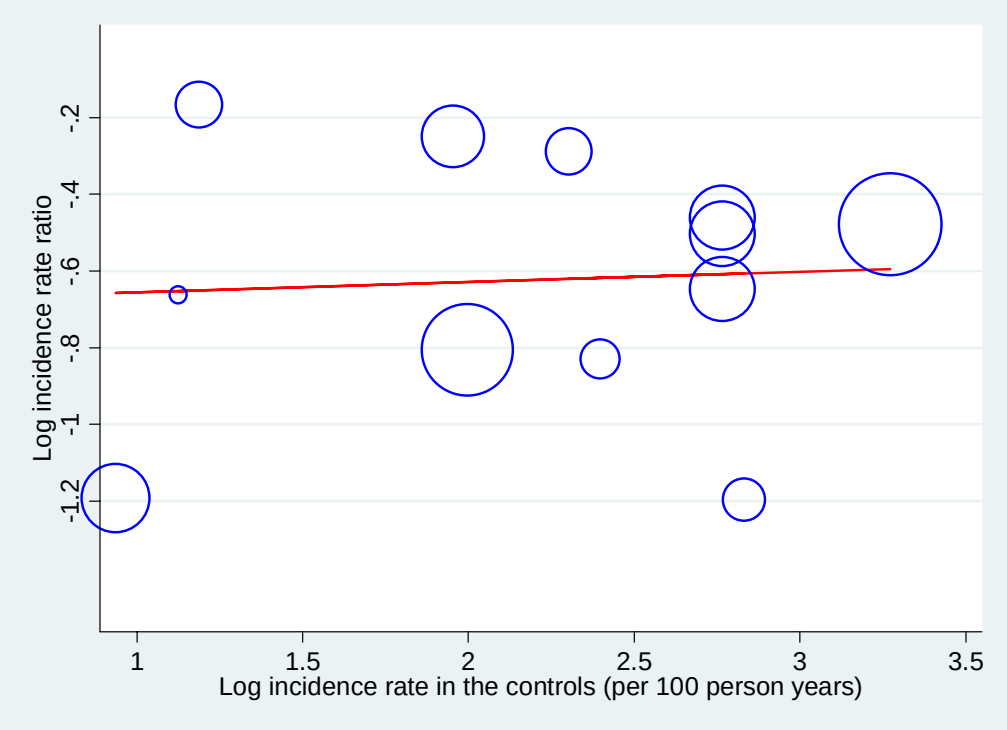
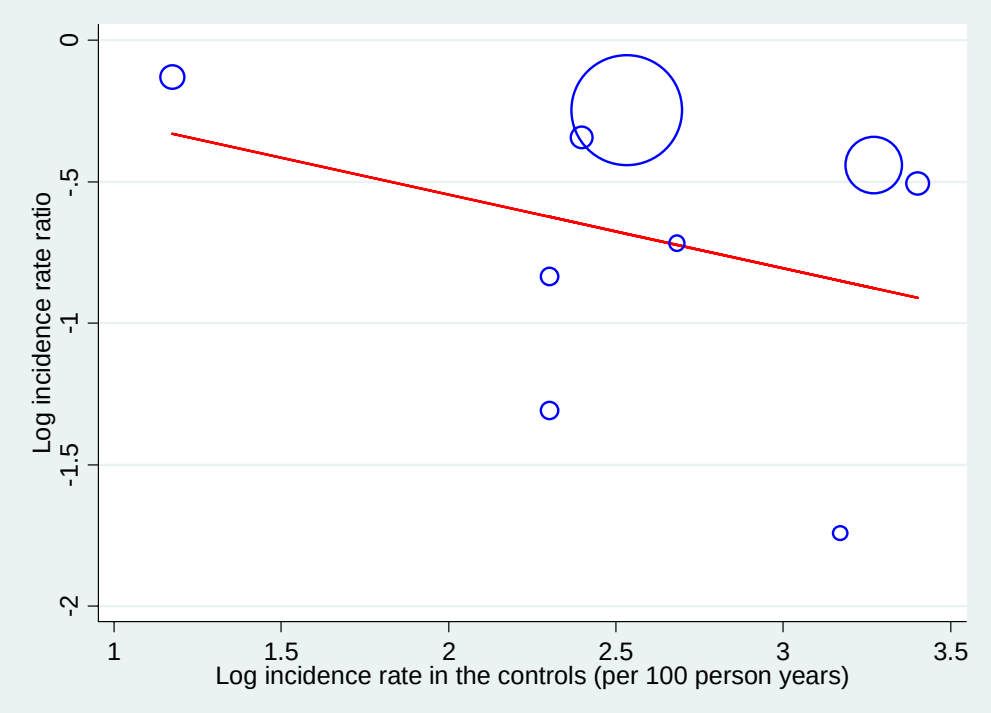
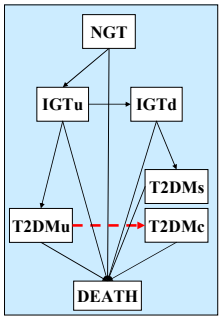


Figure 6.9: Effect of baseline risk on the effectiveness of anti-diabetic agents



6.3.5 Undiagnosed T2DM to clinically diagnosed T2DM

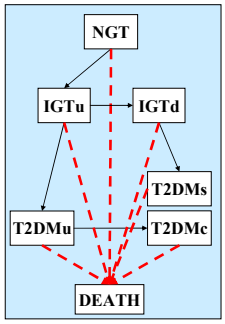


The transition rate from undiagnosed to clinically diagnosed T2DM was the most difficult transition to find information on. A few studies were found that reported the number of diabetes cases diagnosed each year, which would give the probability of being diagnosed from a study population, but understandably no studies reported the probability of being diagnosed given you already had undiagnosed T2DM. Therefore information was collated on the mean time a diabetic individual can expect to stay undiagnosed (t) and, assuming an exponential distribution, this was converted to a yearly transition rate (γ) (equation 6.7).

$$\gamma = \frac{1}{t} \qquad \text{[Equation 6.7]}$$

Only one relevant paper was identified (Harris et al., 1992). The paper described how retinopathy is usually the first observable vascular condition specific to diabetes to develop in diabetic patients. Prevalence of retinopathy increases linearly with longer duration of diabetes and by extrapolating this linear relationship, time when retinopathy prevalence was estimated to be zero can be approximated. Onset of retinopathy was estimated to have occurred 4 to 7 years before diagnosis of diabetes in two populations. Specifically, for a Wisconsin study group 6.5 years (95% CI: 4.1 to 9.9) and for an Australian study group 4.2 years (95% CI: 2.1 to 7.4). Although other data indicates diabetes may be present up to 5 years before onset of retinopathy, this was the best estimate of delay to diagnosis that could be located. These two estimates were pooled on the log scale within WinBUGS, to distribution for the transition rate which had a mean of 0.22 (SE 0.218).

6.3.6 Mortality rates



Mortality rates were taken from Department of Health statistics for England and Wales for the year 2000. They were taken as known and no uncertainty was placed around them in the model. As the model was run for 50 years, starting with one 45 year old, different mortality rates were used for different cycles, depending on the predicted age of the individual for that Markov cycle. Mortality rates per 100 person years

by age were 45 to 54 years of age 0.32, 55 to 64 years 0.84, 65 to 74 years 2.36, 75 to 84 years 6.09, 85+ years 15.68.

Mortality rates were assumed to be the same as population mortality rate from NGT, and both the IGT states. From the three T2DM states the mortality rates for each age group were increased using a hazard ratio of increased mortality of type 2 diabetics as reported by the DECODE study (The DECODE study group, 1999), hazard ratio 2.13 (95% CI: 1.79 to 2.52), showing a greater than doubling of the mortality rate in diabetics compared to non-diabetics. HbA_{1c} levels represent a measure of blood glucose control, and mortality rates in diabetics have been shown to be linked with HbA_{1c} levels, in that a 1% increase in HbA_{1c} is associated with an increase risk in the hazard of mortality, hazard ratio 1.11 (95% CI: 1.03 to 1.20) (Rossing et al., 1996). Although the DECODE study did not directly report the HbA_{1c} of their study population, it was assumed the sample most directly represented the individuals with clinically detected diabetes in this model. Therefore the death rates for screen diagnosed T2DMs and undiagnosed T2DM were adjusted further depending on their predicted difference in HbA_{1c} compared to the clinically diagnosed group.

HbA_{1c} levels for individuals with T2DM within the decision model were predicted from results taken from the UKPDS study (UK Prospective Diabetes Study (UKPDS) Group, 1998a). The UKPDS study is composed of clinically detected individuals where one group received standard therapy, whilst a second group received intensive therapy for both glucose and blood pressure control. For the clinically detected group, an average value over time was needed, for a group of conventionally detected and treated individuals with T2DM. Therefore the HbA_{1c} levels of the group receiving standard therapy in the UKPDS study was used, 7.9 (Inter-quartile range (IQR): 6.9 to 8.8) (UK Prospective Diabetes Study (UKPDS) Group, 1998a).

For the model state of undiagnosed T2DM, it would be expected that HbA_{1c} would be higher, as individuals in this state would not be receiving any medication to control their blood glucose levels. For this state the HbA_{1c} levels of individuals clinically detected at entry to the UKPDS study, before T2DM treatment commenced, was used, HbA_{1c} 9% (IQR: 7.5 to 10.5)(UK Prospective Diabetes Study (UKPDS) Group, 1991). No long term statistics of average HbA_{1c} levels in individuals detected through a screening programme were available. So for the decision model an assumption was made that a screen detected population would have a long term HbA_{1c} level similar to the arm of the UKPDS trial who were receiving intensive therapy,

which was reported as an average of 7.0 (IQR: 6.2 to 8.2), (UK Prospective Diabetes Study (UKPDS) Group, 1998a).

As HbA_{1c} levels were reported using medians and IQRs, for the purposes of the decision model normality of the data was assumed, with the median taken as being equivalent to the mean and the standard deviation calculated, using the standard normal distribution, as the IQR divided by 1.349.

Using the above data, the probabilities of dying (*M)* for all the Markov states were estimated for each of the five age groups (45-54, 55-64, 65-74, 75-84 and 85+) as follows:

NGT and IGT states	$M = 1 - e^{-\lambda_j}$	[Equation 6.8]
Clinically diagnosed T2DM	$M = 1 - e^{(-\lambda_j . \gamma)}$	
Screen diagnosed T2DM	$M = 1 - e^{(-\lambda_j . \gamma . \kappa^{\delta})}$	
Undiagnosed T2DM	$M = 1 - e^{(-\lambda_j . \gamma . \kappa^{\delta})}$	

Where λ_j is the mortality rate in age group j, γ is the increased mortality risk of an individual with T2DM, κ is the increased mortality rate for a one unit increase in HbA_{1c} and δ is the difference in HbA_{1c} between predicted HbA_{1c} levels of screen diagnosed or undiagnosed individuals with T2DM, and those clinically diagnosed.

6.3.7 Calculating the number in each state at each cycle

For each of the three Markov models (one for each screening strategy), the number in each of the seven states, for each yearly cycle of the model, was calculated using the *inprod* command in WinBUGS, along with transition rates between states, and the numbers estimated to be in each state at the previous time/cycle. The code to run this in WinBUGS is specified in box 2.

Box 2: WinBUGs code for calculating the numbers in each state at each time point, for each strategy:

```
for (strategy in 1:4){
        for (state in 1:7){
        for (time in 2:10) {
number[strategy,state,time] <- inprod(number[strategy,1:7, time-1], trans[strategy,1,1:7,state])
        }
        for (time in 11:20) {
number[strategy,state,time] <- inprod(number[strategy,1:7, time-1], trans[strategy,2,1:7,state])      }
        for (time in 21:30) {
number[strategy,state,time] <- inprod(number[strategy,1:7, time-1], trans[strategy,3,1:7,state])
        }
        for (time in 31:40) {
number[strategy,state,time] <- inprod(number[strategy,1:7, time-1], trans[strategy,4,1:7,state])
        }
        for (time in 41:50) {
number[strategy,state,time] <- inprod(number[strategy,1:7, time-1], trans[strategy,5,1:7,state])
        }
        total[strategy,state] <- sum(number[strategy,state,1:horizon])
```

The code works by estimating the number of individuals, for each of the four screening strategies that could be predicted to be in each of the Markov states for each cycle (time) of the model horizon. This is done by taking the number of individuals in each of the seven Markov (n_M) states at the previous time point (t), multiplying these by the probabilities they will move to a different state (p, where p_{abt} is the probability of moving from a to b at time t) and then summing results to estimate the number in each state at the new time point ($t+1$), as specified below.

$$\begin{pmatrix} n_{1t} & n_{2t} & n_{3t} & n_{4t} & n_{5t} & n_{6t} & n_{7t} \end{pmatrix} \times \begin{pmatrix} p_{11t} & p_{12t} & p_{13t} & p_{14t} & p_{15t} & p_{16t} & p_{17t} \\ p_{21t} & p_{22t} & p_{23t} & p_{24t} & p_{25t} & p_{26t} & p_{27t} \\ p_{31t} & p_{32t} & p_{33t} & p_{34t} & p_{35t} & p_{36t} & p_{37t} \\ p_{41t} & p_{42t} & p_{43t} & p_{44t} & p_{45t} & p_{46t} & p_{47t} \\ p_{51t} & p_{52t} & p_{53t} & p_{54t} & p_{55t} & p_{56t} & p_{57t} \\ p_{61t} & p_{62t} & p_{63t} & p_{64t} & p_{65t} & p_{66t} & p_{67t} \\ p_{71t} & p_{72t} & p_{73t} & p_{74t} & p_{75t} & p_{76t} & p_{77t} \end{pmatrix} = \begin{pmatrix} n_{1t+1} \\ n_{2t+1} \\ n_{3t+1} \\ n_{4t+1} \\ n_{5t+1} \\ n_{6t+1} \\ n_{7t+1} \end{pmatrix}$$

For example at year two the number in the death state would be the sum of:

number in state NGT at year 1 x probability of moving from NGT to Death +

number in state IGTu at year 1 x probability of moving from IGTu to Death +

number in state IGTd at year 1 x probability of moving from IGTd to Death +

number in state T2DMu at year 1 x probability of moving from T2DMu to Death +

number in state T2DMd at year 1 x probability of moving from T2DMd to Death

The transition rates were varied where necessary between strategies, as discussed above. Transition rates were also changed depending on the year cycle of the model, to mimic the effect of the screened population aging, from 45 years of age at the start of the model to 90 years of age by the 50th cycle. This is why the code is split into 10 year time bands, so that the transition rates could be varied.

6.4 Costs

To allow for an assessment of cost effectiveness between the screening strategies, costs were added into the model for screening, interventions and clinical status of the screened population. As described in chapter 2, costs were obtained by estimating resource use and then applying a unit cost.

6.4.1 Cost of screening

When screening for IGT the opportunity is often taken to screen for a number of additional clinical problems, for example high cholesterol. For the initial model it was decided just to cost the necessary tests for screening and confirming glucose tolerance status. Therefore the costs added to the model were 40p for every FPG test carried out and £1.30 for every OGTT test (costs provided by University Hospitals of Leicester NHS Trust, 2006).

The nurse's time to used to carry out the tests was also costed, based on an estimated cost for a nurse in a GP practice of £26 per hour of patient contact (Curtis and Netten, 2006) and an estimated time for the nurse to carry out an FPG test of 5 minutes and an OGTT of 30 minutes. Uncertainty was placed around the time taken to complete each test, such that the time to complete an FPG test was estimated as 5 minutes, with a variance of 2 minutes, and the OGTT as 30 minutes with a variance of five minutes. They were modelled using a gamma distribution to restrict simulations to positive values as specified in Equation 6.9. The parameters of these gamma distributions were estimated using method of moments techniques, which use the properties of the gamma distribution as detailed in chapter 2.

$$T_{FPG} \sim gamma(12.5,2.5)$$
$$T_{OGTT} \sim gamma(150,5.5) \qquad \text{[Equation 6.9]}$$

6.4.2 Cost of interventions

The costs of pharmacological interventions were taken from costs published by the NHS Northern & Yorkshire regional drug and therapeutics centre (2006). Metformin at a dose of 250mg three times a day will have an annual cost of £16.10 per patient. Acarbose at a dose of 50mg 3 times a day will cost £80.30 annually. The doses used for costings were the most commonly given dosages in the intervention trials.

Lifestyle interventions were much more problematic to cost, as the intensity and form of a lifestyle intervention can vary widely and none of the published trials that had considered lifestyle interventions for T2DM prevention had used U.K. costs. A study was found that had assessed the effects of lifestyle interventions for reducing obesity in a U.K. setting, where the lifestyle intervention had been modelled on the intervention used in the DPS trial (Avenell et al., 2004). The costs were for 2001 and included dietician costs (7 visits in the first year and 4 visits annually thereafter) and supervised group exercise sessions twice weekly. Assuming a class size of 20 participants and a 60% attendance rate, total costs will be £324 per person in the first year of the intervention and £228 in subsequent years. As costs of lifestyle interventions were difficult to predict, sensitivity analyses will be carried out in Chapter 8 to assess the impact of different estimated costs on lifestyle interventions on model outcomes.

6.4.3 Costs of undiagnosed T2DM

Although individuals undiagnosed with T2DM are not incurring any specific treatment costs related to T2DM, it has been found that five years before a diabetes diagnosis is made

individuals start to consult their general practitioner (GP) more frequently and receive more prescription items, compared to individuals that do not go on to receive a T2DM diagnosis (Gulliford et al., 2005). Using average U.K. costs of a GP of £21 per 10 minute surgery consultation and average prescription costs per consultation of £34.60 (Curtis and Netten, 2006) and using predicted increased rates of GP visits of 3.05 the year before diabetes is diagnosed and an average rate of 1.39 for the four years previous to that (Gulliford et al., 2005), then additional costs incurred by undiagnosed T2DM can be calculated as:

Year before diagnosis: 2.05 x (£21 + £34.60) = £113.98

Years >1 before diagnosis: 0.39 x (£21 + £34.60) = £21.68

6.4.4 Costs of diagnosed T2DM

The additional costs that a diabetic patient compared to a non-diabetic patient will incur to the NHS, was derived from the UKPDS study (Clarke et al., 2004b). The UKPDS study reported costs of T2DM treatment, implementation (e.g. retinopathy screening and GP visits), and complications (e.g. amputations and blindness) (table 6.7), for each of the two treatment arms within the trial, which consisted of those on conventional treatment, and those on intensive treatment for blood glucose control and hypertension.

Costs attached to the model were varied for the diagnosed T2DM state, in that those clinically detected were modelled to incur costs of complications as reported for the UKPDS participants on conventional treatment. For those screen detected in the diagnosed T2DM state, for whom treatment begins earlier thereby reducing future complications, the costs quoted for intensive treatment of UKPDS participants was used. These individuals had better controlled blood glucose and fewer complications which is more representative of a screen detected population. The UKPDS also quoted costs of hypertensive treatment but as this is not a cost solely attributable to diabetics, it was not included in this decision model.

Table 6.7 Costs incurred by T2DM over the course of the UKPDS trial

	T2DM clinically detected (£)	T2DM screen detected (£)
Anti-diabetic treatment	640 (997)	640 (997)
Cost of implementation	674 (193)	674 (193)
Cost of complications	27,350 (46,842)	24,585 (36,609)
Total cost	28,644 (SE 656.2)	25,899 (SE 554.3)
Average annual cost[+]	2754 (SE 63.1)	2490 (SE 53.3)

* Figures reported are mean (SD) unless otherwise stated

+ As the UKPDS trial lasted 10.4 years, annual costs are total cost divided by 10.4.

6.4.5 Standardising costs to 2006

As the costs collated for the model were from different years they needed to be standardised to the same year, as described in section 2.4.7, to make sure the results from the model were accurate and interpretable. Standardisation can be done by using hospital and community health service pay and price inflation indices (Curtis and Netten, 2006) to either inflate the data from an earlier year to the chosen year, or to deflate the data from a later year. Costs for this model were standardised to 2006, as the majority of the costs were already relevant for this year (table 6.8).

Table 6.8: Standardisation of costs to 2006 using price inflation indices

Item	Cost	Year	2006 Cost
Screening tests			
FPG test	40p	2006	40p
OGTT test	£1.30		£1.30
Nurse cost per hour	£26	2006	£26
Pharmacological interventions			
Acarbose (per annum)	£80.30	2006	£80.30
Metformin (per annum)	£16.10		£16.10
Lifestyle intervention			
Year 1	£324	2001	£398
Subsequent years	£228		£280
Undiagnosed T2DM			
Year before diagnosis	£113.98	2006	£113.98
Years 2 to 5 before diagnosis	£21.68		£21.68
Diagnosed T2DM			
Screen detected (per annum)	£2490	2004	£2672
Clinically detected (per annum)	£2754		£2945

Taking costs of T2DM as an example, to inflate these costs from 2004 to 2006, the pay and prices indices from these years are taken from the report by Curtis and Netten, and the 2006 value is divided by that of 2004 (241.3/224.8 =1.073). The costs from 2004 are then multiplied by 1.073 to estimate the 2006 costs (£2490 x 1.073 = £2672). Chapter 7 details sensitivity analyses on all cost data that was estimated with uncertainty, to assess the impact of changing estimated costs on model outcomes.

6.5 Utilities

For the states of NGT, undiagnosed IGT and diagnosed IGT, the utility was assumed to be that of full health and was set at 1. There has been some research to show that IGT may carry a reduced health utility but reported estimates are varied, so for the purposes of keeping the primary model simple, the assumption of complete health was made.

Utilities for individuals with undiagnosed T2DM and screen detected T2DM were calculated using individual patient data made available by the Leicester arm of the ADDITION study (Srinivasan et al., 2007). The ADDITION study collected quality of life data, using EQ-5D questionnaires, on a screen detected population. The EQ-5D utility score was calculated for 140 individuals who were identified as having T2DM when screened at the start of the ADDITION study, using the MVH-A1 algorithm (Dolan, 1997). This algorithm was developed using a study in which direct valuations were elicited for 42 EuroQol health states (using the time trade-off method) from a representative sample of the UK population. A "tariff" of EuroQol values was then developed from this, to represent each of the 243 possible health states the EuroQol index generates. After applying the MVH-A1 algorithm to the ADDITION data, the mean utility was calculated as 0.788 (SE 0.020). For those undiagnosed the utility was kept constant for all years spent in this state, as it was assumed that to remain undiagnosed the presence of complications or ill health would not be present. The screen and clinically detected utility was decreased each year in line with predicted complication rates and their estimated effect on health utility, as will be described in more detail later.

For people with clinically diagnosed T2DM utilities were taken from those reported by the UKPDS study, as entry to this study was at time of clinical diagnosis. The utility of the trial arm receiving conventional therapy was used, which was a reported value, when complications were

adjusted for, of 0.725 (SE 0.035) (Clarke et al., 2002). This is lower than the value used for undiagnosed and screen diagnosed T2DM, which is realistic as it could be expected that symptoms of T2DM would be more severe.

For both the screen diagnosed and clinically diagnosed T2DM states in the decision model, utility was decreased for each year spent in the state, to account for increasing complications and a decline in health due to duration of T2DM. This was done by combining data from two UKPDS papers, one of which reported complication rates (Clarke et al., 2004a) and another which reported the effects of complications on utility values (Clarke et al., 2002). At development of T2DM, the utility value assuming no complications, was different depending on whether it was undiagnosed, screen diagnosed or clinically diagnosed.

Although there is currently no trial evidence on the long-term effects of screening and early detection of T2DM, it is generally believed that early detection would lead to better long-term glucose control and general health. Therefore the complication rates for the conventionally treated group in the UKPDS study were used for modelling the utility value in the clinically diagnosed T2DM group and the complication rates for the UKPDS arm receiving intensive treatment was utilised for the model state of screen diagnosed T2DM. (Clarke et al., 2002, Clarke et al., 2004a). Therefore the utility value decreased more rapidly in individuals clinically diagnosed, compared to those who were screen detected. Utility values were decreased by duration of T2DM using figures reported by the UKPDS study. One published paper reported results of a model forecasting the occurrence of diabetes-related complications, using data on 3642 patients who had been followed up for over 10 years during the course of the study (Clarke et al., 2004a). They fitted a series of Weibull proportional hazards regression models, which estimated the occurrence of several complications associated with T2DM, these included ischaemic heart disease (IHD), myocardial infarction (MI), coronary heart disease (CHD), stroke, amputation and blindness in one eye due to diabetes retinopathy.

The model fitted to each of these outcomes is given in equation 6.10, where t is time to event, j is complication, λ is the scale parameter, γ is the shape parameter, and β represents the effect of HbA_{1c}, and the parameters of each of these six models, as published by Clarke, 2004a), are given in table 6.9. HbA_{1c} appears to be strongly linked with complication rates, in that a lower HbA_{1c} value, indicating better glucose level control, is associated with lower complication rates. As the decision model predicts different HbA_{1c} levels for the two diagnosed T2DM states (screen and clinically diagnosed), complication rates and their effect on utilities were modelled

accordingly. Consequently higher complication rates were predicted in clinically detected individuals compared to those screen detected. As discussed previously the utility rate was not decreased in the undiagnosed T2DM state as it was assumed that a diagnosis would be made as soon as any complications developed. This assumption is a reasonable one, particularly for severe complications.

$$S_j(t) = \exp(-\lambda_j \exp(\beta_j x) t^{\gamma_j}) \qquad \text{[Equation 6.10]}$$

Using the estimated survival for each complication, the cumulative probability of developing the complication at time t_k, can be estimated $C_j(t_k)$, and hence the probability for each year of duration of T2DM can also be specified ($D_j(t_k)$ (equation 6.11).

$$C_j(t_k) = 1 - S_j(t_k) \qquad \text{[Equation 6.11]}$$

$$D_j(t_k) = C_j(t_k) - C_j(t_k - 1)$$

The cumulative probabilities for each of the six complications, over 20 years of duration with T2DM, are displayed in figure 6.10. It can be seen that MI has the highest probability, reaching 25% at 20 years of T2DM duration.

To consider an example, using the parameters in table 6.9, the 5-year survival for IHD, cumulative probability of IHD after 5-years duration of T2DM, and the probability of IHD in year five post-development of T2DM, can be calculated as:

$$S_{IHD}(5) = \exp(-\exp(-5.310) \times \exp(0.125 \times HbA1_c) \times 5^{1.150} = 0.965$$

$$C_{IHD}(5) = 1 - S_{IHD}(5) = 0.035$$

$$D_{IHD,5} = C_{IHD}(5) - C_{IHD}(4) = 0.035 - 0.027 = 0.008$$

Therefore the cumulative probability of IHD up to five years duration of T2DM is 3.5% and the probability of having IHD in the fifth year of duration with T2DM is 0.8%.

6.9: Data from the UKPDS study on complication rates and associated utility decrements

Complication	Scale parameter (log λ)	Shape parameter (γ)	HbA1c effect (β)	Utility decrement	
				Previous yr (*Y*)	Prior to previous yr (*X*)
IHD	-5.310 (0.174)	1.150 (0.067)	0.125 (0.035)	-0.141 (0.060)*	-0.079 (0.020)*
MI	-4.977 (0.160)	1.257 (0.060)	0.118 (0.025)	-0.081 (0.052)	-0.044 (0.021)*
CHF	-8.018 (0.408)	1.711 (0.158)	0.157 (0.057)	-0.058 (0.066)	-0.134 (0.038)*
Stroke	-7.163 (0.342)	1.497 (0.126)	0.128 (0.042)	-0.131 (0.073)	-0.199 (0.035)*
Amputation	-8.718 (0.613)	1.451 (0.232)	0.435 (0.066)	-0.451 (0.131)*	-0.335 (0.068)*
Blindness	-6.464 (0.326)	1.154 (0.121)	0.221 (0.075)	-0.074 (0.070)*	-0.080 (0.029)*

* statistically significant decrements in utility

The models reported above, included other covariates which were centred, and therefore the model is for a population with a BMI of 27.77kg/m^2, $HbA1_c$ 7.09%, systolic blood pressure of 135.09 mm Hg, and total HDL cholesterol 5.23.mmol/l. The influence of $HbA1_c$ was the only value included in the decision model, to adjust complication rates between the three T2DM states. A second paper from the UKPDS study reported the estimated utility decrements that were associated with each complication (table 6.9), as predicted from EQ-5D data collected on the same group of patients (Clarke et al., 2002). The decrement in utility differed, depending on if the complication had occurred in the previous year, or prior to the previous year. Therefore the total decrement in the utility value ($U_j(t_k)$), for each year of duration (t_k), for each complication (j), can be calculated as in equation 6.12, where the annual probability of a complication ($D_j(t_k)$) is multiplied by the utility decrement for the complication occurring in the past year (Y_j), and is added to the probability of the complication occurring prior to the previous year

($C_j(t_{k-1})$) multiplied by the utility decrement if the complication occurred prior to the previous year (X_j). The IHD example is extended further to show how the figures in table 6.9 are used in the calculation.

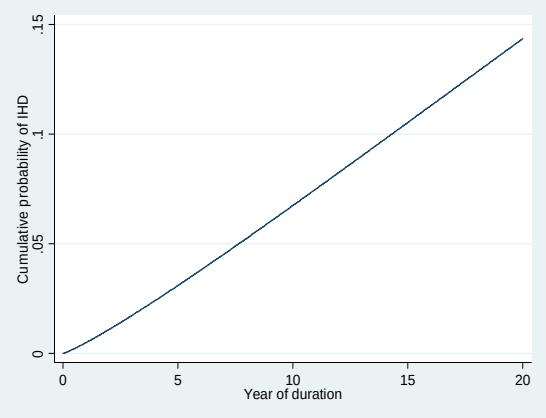
$$U_{jt} = \left(D_{jt} \times Y_j\right) + \left(C_{jt-1} \times X_j\right) \qquad \text{[Equation 6.12]}$$

$$U_{IHD,5} = (0.008 \times -0.141) + (0.027 \times -0.079) = -0.002461$$

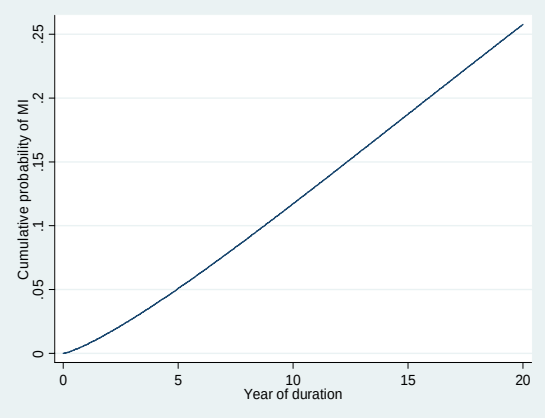
Therefore the total decrement in utility, in year 5 of duration of T2DM, associated with the probability of IHD is -0.002461. Hence, by combining data from the two papers it was possible to estimate the total utility decrement that could be associated with each year of duration of T2DM (t_k) due to all complications ($\sum_{1}^{J} U_j(t_k)$), and therefore the expected overall utility value. All the parameters in the model included uncertainty, as specified in table 6.9. The QALY for each year of duration with T2DM was then calculated by subtracting the expected decrement due to all six complications, from the baseline utility of T2DM that was estimated for diabetics with no complications. The WinBUGS code used to calculate the utilities is given in box 3. As the utility decrements are calculated for each year of duration of T2DM, transitions in the model had to be tracked, so that the probability of moving into a T2DM state at each cycle could be monitored, and then followed through future cycles to allow duration of T2DM to be estimated.

Figure 6.10: Cumulative probability of six complications over 20 years of duration with T2DM

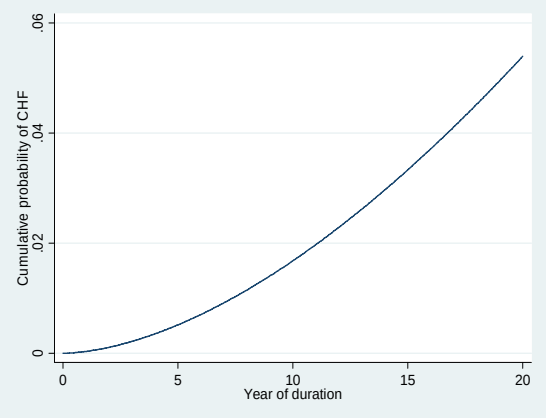
a) CHD



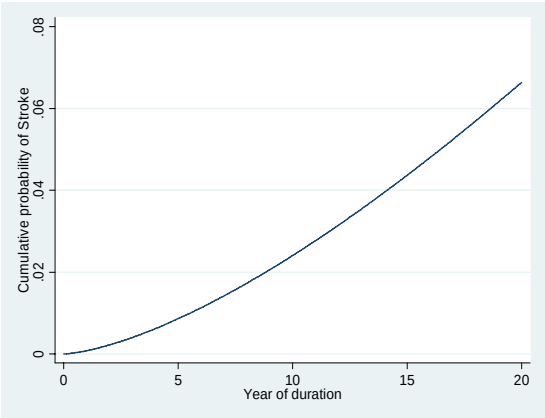
b) MI



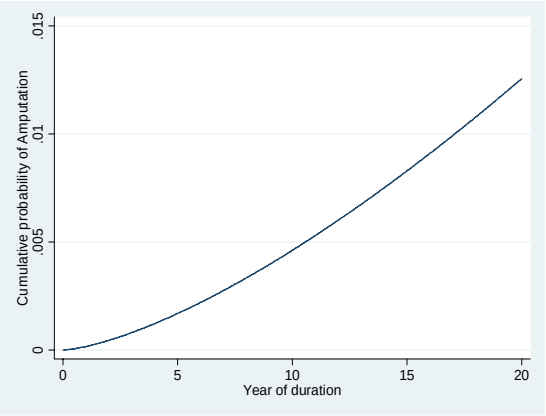
c) CHF



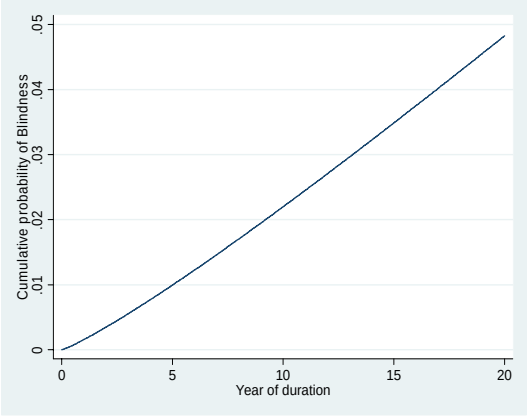
d) Stroke



e) Amputation



f) Blindness



Box 3: WinBUGs code for modelling complication rates and their effect on utility

```
for (comp in 1:6) {
        for (det in 1:2) {
        s[det,comp,1] <- 1                  #All start complication-free
        c[det,comp,1] <- 0
}

        lambda[comp] <- exp(beta0[comp])        #Distribution on scale parameter
        beta0.p[comp] <- 1/pow(beta0.se[comp],2)
        beta0[comp] ~ dnorm(beta0.m[comp],beta0.p[comp])

        rho[comp] <- exp(gamma[comp])   #Distribution on shape parameter
        gamma.m[comp] <- log(rho.m[comp])
        gamma.p[comp] <- pow(rho.m[comp],2)/pow(rho.se[comp],2)
        gamma[comp] ~ dnorm(gamma.m[comp],gamma.p[comp])
        uy.p[comp] <- 1/pow(uy.se[comp],2)              #Distributions on utility decrements
        uy[comp] ~ dnorm(uy.m[comp],uy.p[comp])
        upy.p[comp] <- 1/pow(upy.se[comp],2)
        upy[comp] ~ dnorm(upy.m[comp],upy.p[comp])

        gender.p[comp] <- 1/pow(gender.se[comp],2)
        gender[comp] ~ dnorm(gender.m[comp],gender.p[comp])
        hba1c.p[comp] <- 1/pow(hba1c.se[comp],2)
        hba1c[comp] ~ dnorm(hba1c.m[comp],hba1c.p[comp])
}
ufemale ~ dnorm(ufemale.m,ufemale.p)
                        #Utility from UKPDS expressed as female baseline and male effect
ufemale.p <- 1/pow(ufemale.se,2)
umale ~ dnorm(umale.m,umale.p)
umale.p <- 1/pow(umale.se,2)
clin.ubase <- ufemale + (0.5*umale)
screen.ubase ~ dnorm(uscreen.m,uscreen.p)
uscreen.p <- 1/pow(uscreen.se,2)

for (dur in 2:horizon) {
    for (comp in 1:6) {
        s[1,comp,dur] <- exp(-
lambda[comp]*(exp(0.5*gender[comp]*adjust[comp])*pow(dur,rho[comp])))
                                        #Weibull survival model for screen detected

s[2,comp,dur] <- exp(-lambda[comp]*(exp((0.5*gender[comp]*adjust[comp]) +
(0.9*hba1c[comp]))*pow(dur,rho[comp])))
                                    #Weibull survival model for clinically detected (higher hba1c)

for (det in 1:2) {
        c[det,comp,dur] <- 1 - s[det,comp,dur]      #Cumulative probability of complication
        d[det,comp,dur] <- c[det,comp,dur] - c[det,comp,dur-1]
                                        #Probability of complication in previous yr
        py[det,comp,dur] <- c[det,comp,dur] - d[det,comp,dur]
                                    Probability of complications  in yrs prior to previous yr
        ud[det,comp,dur] <- d[det,comp,dur]*uy[comp] + py[det,comp,dur]*upy[comp]
                }}
        tud[1,dur] <- sum(ud[1,1:6,dur])
        tud[2,dur] <- sum(ud[2,1:6,dur])            #Total utility decrement for all complications
        clin.qaly[dur] <- clin.ubase + tud[2,dur]   #Total QALY in each year duration in T2DMc
        screen.qaly[dur] <- screen.ubase + tud[2,dur]
                                        #Total QALY for each year duration in T2DMs
```

6.6 Model outcomes

Additional code was incorporated within the model to compute outcomes concerning both clinical and cost estimates, which would enable an assessment of the cost-effectiveness of each of the four screening strategies. Total life years were calculated for each of the four screening strategies, by summing the time spent in each of the Markov states apart from death. QALYs were calculated by adjusting time spent in each of the Markov states by the appropriate utility decrement. Costs of T2DM and interventions were applied to each state as appropriate and then summed with screening costs, to allow the total cost for each strategy over the time horizon of the model to be computed.

Additional clinical outcomes of interest included the total number of clinically diagnosed T2DM cases and screen diagnosed cases, which was calculated by tracking the number of individuals who entered this state at each cycle, and then summing this over the whole time horizon. The total number of T2DM cases was calculated by tracking individuals as they moved into the undiagnosed T2DM state and screen diagnosed state and summing these together. Those clinically diagnosed did not have to be incorporated into this calculation, as these are individuals who already have T2DM but have moved from the undiagnosed state.

Finally diabetes-free life years was calculated by summing years spent with NGT or IGT. All clinical and costs outcomes were calculated for each of the four screening strategy, and the differences in outcomes between each of the three active screening strategies compared to the strategy of no screening were also computed. As well as undiscounted results, both costs and QALYs were calculating incorporating an annual discount of 3.5%. The WinBUGS code for calculating model outcomes is given in box 4.

Box 4: WinBUGS code for model outcomes

```
#Clinical Outcomes
totalIGT[strategy] <- sum(total[strategy,2:3])
totalT2DM[strategy] <- sum(total[strategy,4:6])
totallife[strategy] <- sum(total[strategy,1:6])
totaldiabfree[strategy] <- sum(total[strategy,1:3])
totallifedis[strategy] <- sum(lifedis[strategy,1:50])
totaldiabfreedis[strategy] <- sum(diabfreedis[strategy,1:50])

for (time in 1:50) {
        lifedis[strategy,time] <- sum(number[strategy,1:6,time])*pow(0.965,time-1)
        diabfreedis[strategy,time] <- sum(number[strategy,1:3,time])*pow(0.965,time-1) }

difflife[strategy] <- totallife[strategy] - totallife[4]
diffdiabfree[strategy] <- totaldiabfree[strategy] - totaldiabfree[4]
diffcases[strategy] <- totalcases[strategy] - totalcases[4]
difflifedis[strategy] <- totallifedis[strategy] - totallifedis[4]
diffdiabfreedis[strategy] <- totaldiabfreedis[strategy] - totaldiabfreedis[4]


#Diabetes cases
casesT2DM[strategy,1] <- number[strategy,4,1] + number[strategy,5,1] + number[strategy,6,1]

for (time in 2:horizon) {
        allcases[strategy,time] <- (number[strategy,2, time-1]*trans[strategy,1,2,4]) +
        (number[strategy,3,time-1]*trans[strategy,1,3,5])

        clincases[strategy,time] <- number[strategy,4, time-1]*trans[strategy,1,4,6]
                screencases[strategy,time] <- number[strategy,3, time-1]*trans[strategy,1,3,5]
                                }
        allcases[strategy,1] <- number[strategy,4,1] + number[strategy,5,1]
        clincases[strategy,1] <- number[strategy,6,1]
        creencases[strategy,1] <- number[strategy,5,1]
        totalcases[strategy] <- sum(allcases[strategy,1:horizon])
        totalclincases[strategy] <- sum(clincases[strategy,1:horizon])
        totalscreencases[strategy] <- sum(screencases[strategy,1:horizon])
        }
```

To assess the cost-effectiveness of each of the three active screening strategies compared to no screening, the incremental cost-effectiveness ratios (ICERs), as well as cost-effectiveness acceptability curves were calculated. The ICER is simple the difference in costs between two strategies, divided by the difference in QALYs, as detailed in Chapter 2. The CEACs were calculated by estimating the probability of a strategy being cost-effective compared to another, at a number of different willingness-to-pay thresholds (these were £100, £1000, £2500, £5000, £7500, £10000, £20000, £25000, £30000 and £50000). The methods for this were described in more detail in Chapter 2, sections 2.4.4 and 2.4.6.

Box 5: WinBUGS code for calculating the ICERs for each active strategy minus no screening, and the CEACs

```
        diffcost[1] <- (totalcost[1] -  totalcost[4])
        diffcost[2] <- (totalcost[2] -  totalcost[4])
        diffcost[3] <- (totalcost[3] -  totalcost[4])
        diffqaly[1] <- (qaly[1] - qaly[4])
        diffqaly[2] <- (qaly[2] - qaly[4])
        diffqaly[3] <- (qaly[3] - qaly[4])
        inccost[1] <- diffcost[1]  / diffqaly[1]
        inccost[2] <- diffcost[2]  / diffqaly[2]
        inccost[3] <- diffcost[3]  / diffqaly[3]

for(k in 1:NK) {
        Q[1,k] <- step(1- ( (totalcost[1] -  totalcost[4]) - K[k] * (qaly[1] - qaly[4]) ))
        Q[2,k] <- step(1- ( (totalcost[2] -  totalcost[4]) - K[k] * (qaly[2] - qaly[4]) ))
        Q[3,k] <- step(1- ( (totalcost[3] -  totalcost[4]) - K[k] * (qaly[3] - qaly[4]) ))

Where k = c(100,1000,2500,5000,7500,10000,20000,25000,30000,50000)
```

6.7 Running the model

A summary of all the model inputs are given in tables 6.10 and 6.11. Four Markov models were run simultaneously, one for each of the screening/intervention strategies of no screening, screening for T2DM only, screening for IGT and T2DM, with lifestyle interventions administered to individuals with IGT to attempt to prevent T2DM, and lastly as above but with pharmacological interventions. Each model cycle represents one year and the model for the base case scenario was run for a time horizon of 50 years. Model results include both clinical and cost-effectiveness outcomes, with cost per quality adjusted life year being the primary outcome. The model was implemented within WinBUGS using a Bayesian comprehensive decision modelling approach (Cooper et al., 2004). The full WinBUGS code written to implement the model, comprising of over ten pages with additional code to what has been presented in this chapter, is given in Appendix 3.6. As described vague prior distributions were assumed for all model parameters where they were required, that is for the diagnostic data for the decision tree and the four meta-analyses of transition rates and intervention effects within the decision model. The influence of these 'vague' priors are explored further in Chapter 7. Model parameters were estimated by using Markov Chain Monte Carlo simulation methods (Spiegelhalter et al.). Results are based on a sample of 20,000 simulations, following a 'burn-in' of 10,000 simulations. Convergence of the Markov chain was assessed, as discussed further in Chapter 7.

Table 6.10: Summary of clinical model inputs

Parameter	Distribution	Value (SE)	Source(s)
Data for the decision tree			
Prevalences	Dirichlet	NGT= 1847/2216= 83% IGT= 264/2216= 12% T2DM= 105/2216= 5%	STAR study
Screening test efficiency	Multi-nominal	Entered as 3 x 3 table, which represents: *For T2DM:* Sensitivity 89.5% Specificity 91.3% *For IGT and T2DM:* Sensitivity 59.4% Specificity 88.0%	STAR study
Transition rates (per 100 person years)			
NGT to IGT <65yrs ≥65yrs	 Log normal Log normal	 1.66 (0.08) 2.49 (0.11)	Baltimore study
IGT to T2DM	Log normal	1.956 (0.252)	12 studies
Time spent with undetected diabetes (years)	Log normal	1.647 (0.181)	Harris
Mortality rates (per 100 person years)			
45-54 years 55-64 years 65-74 years 75-84 years 85+ years	- - - - -	0.32 0.84 2.36 6.09 15.68	Department of Health statistics (2000)
Increased risk of death if have diabetes (hazard ratio)	Log normal	0.756 (0.087)	DECODE
Increased risk of death for 1% increase in HbA_{1c} (hazard ratio)	Log normal	0.104 (0.039)	Rossing
Intervention effects on risk of developing type 2 diabetes (hazard ratio)			
Lifestyle vs. standard treatment	Log normal	-0.646 (0.099)	12 studies
Anti-diabetic drugs vs. placebo	Log normal	-0.425 (0.141)	9 studies
HbA_{1c}			
Undiagnosed T2DM	Normal	9.0% (0.056)	UKPDS
Screen detected T2DM	Normal	7.0% (0.028)	UKPDS
Clinically detected T2DM	Normal	7.9% (0.042)	UKPDS
Utilities			
Undetected T2DM	Normal	0.788 (0.020)†	ADDITION
Screen detected T2DM	Normal	0.788 (0.020)‡	ADDITION
Clinically detected T2DM	Normal	0.771 (0.035)‡	UKPDS

† Utility kept constant for duration undiagnosed

‡ Baseline value only and Utility decreased with every year of duration of T2DM

Table 6.11: Summary of model costs

Parameter	Distribution	Value (SE)	Source(s)
Costs*			
Screening tests FPG test OGTT test	 - -	 40p per person £1.30 per person	 NHS (2006)
Nurse cost	-	£26 per hour	Curtis
Metformin intervention	-	£16.10 per annum	NHS (2006)
Lifestyle intervention Year 1 Subsequent years	 - -	 £398 per annum £280 per annum	 Avenell
Undiagnosed diabetes Year before diagnosis Years 2 to 5 before diagnosis	 - -	 £114 per annum £22 per annum	 Gulliford Curtis
Diagnosed diabetes Screen detected Clinically detected	 Normal Normal	 £2490 (53.3) per annum £2754 (63.1) per annum	 UKPDS

6.8 Results

Tables 6.12 and 6.13 present a summary of the results of the decision model for the base case scenario, where the effects of different screening and intervention strategies are assessed for 45 year old, Caucasian adults, for a time horizon of 50 years. At the time of screening 83% have NGT, 12% IGT and 5% T2DM. Although in reality a population would be screened, results are presented as costs and clinical outcomes of one individual, to enable ease of interpretation.

From table 6.12 it can be seen the average cost over 50 years, per person, for each of the four screening/intervention strategies varies between £17,290 for no screening to £18,040 for screening for T2DM only. The majority of costs are incurred in the diagnosed T2DM states, due to costs of treatment, screening and complications, hence the strategy of no screening has a lower cost than that of screening for T2DM, due to a longer period spent in the undiagnosed state. The total costs of the two intervention strategies are lower than the screening for T2DM only strategy, due to the delay in T2DM progression, and the fact that the intervention costs are relatively low.

The estimated probability of developing T2DM for the model population, which was an 'at risk' population, was approximately 68% over the 50 year time horizon. The majority of cases would

be clinically rather than screen diagnosed. This is probably due to the model incorporating only one screening programme at the start of the model, rather than a strategy of re-screening a population at regular intervals. The impact of implementing more than one screening programme is investigated as a model extension in Chapter 8.

The clinical effects of each strategy appear to vary very little. The total life years is slightly increased in the active screening/intervention strategies, no screening 30.34 years (95% CrI: 27.75 to 32.86), screening for T2DM only 30.40 years (27.82 to 32.93) and screening for both IGT and T2DM with lifestyle interventions 30.49 years (27.90 to 33.01) and pharmacological interventions 30.46 years (27.88 to 32.99). In terms of all the clinical outcomes the strategy of no screening performed the worse, followed by screening for T2DM only. Both the screening strategies which involved interventions to delay T2DM performed the best, with lifestyle interventions slightly outperforming pharmacological interventions, due to the fact that the transition rate from IGT to T2DM was decreased a fraction more by lifestyle rather than pharmacological interventions.

Table 6.13 reports the absolute differences in terms of cost and clinical outcomes between the strategy of no screening and the three other screening/intervention strategies, and also reports the relative cost-effectiveness outcomes. The estimated cost for each QALY gained compared to a strategy of no screening was £8,681 for screening for T2DM only, £2,863 for screening and lifestyle interventions and £3,429 for screening and pharmacological interventions.

Both the IGT and T2DM screening/intervention strategies had a high probability that they were cost-effective compared to no screening at the willingness-to-pay threshold of £20,000, with the strategy incorporating lifestyle interventions having a probability of 0.986, and that using pharmacological interventions of 0.947. Screening for T2DM compared to no screening only had a probability of being cost-effective of 0.681 at £20,000 and 0.765 at the £30,000 per QALY willingness-to-pay threshold, indicating that more uncertainty surrounded the comparison of the two strategies, compared to the intervention strategies compared to no screening.

As well as comparing the three active screening strategies to the current practice of no screening, the two strategies that screened for both IGT and T2DM were also compared directly to the strategy of screening for T2DM alone. As both the intervention strategies had a lower overall cost and better clinical outcomes in terms of QALYs, the incremental costs were -£794 per QALY gained for the strategy involving lifestyle interventions compared to T2DM only screening, and -£1485 per QALY gained for the strategy involving pharmacological interventions compared to no screening. The probabilities of cost-effectiveness also highly favoured the two intervention strategies, with probabilities of 0.99 for both at the £20,000 willingness-to-pay threshold.

Figure 6.11 shows the cost-effectiveness planes, as described in section 2.4.5, for each of the three active screening strategies compared to no screening. All three plots show the majority of estimates lie in the top-right quadrant, which represents that a trade-off between costs and benefits needs to be assessed. For the cost-effectiveness plane of screening for T2DM only compared to no screening, the estimates are close to the axis, showing both the difference in clinical effectiveness and the difference in costs between screening for T2DM and no screening are small. The cost-effectiveness acceptability curves (Figure 6.12) plot the probability of cost-effectiveness against the willingness-to-pay threshold for each of the three active strategies compared to no screening, and the low probabilities achieved by T2DM screening indicates uncertainty around the cost-effectiveness of this strategy. Both the screening/intervention strategies achieve a high-probability of their cost-effectiveness at a fairly low willingness-to-pay threshold, supporting the cost-effectiveness of such strategies.

Table 6.12: Outcomes of the decision model by screening strategy

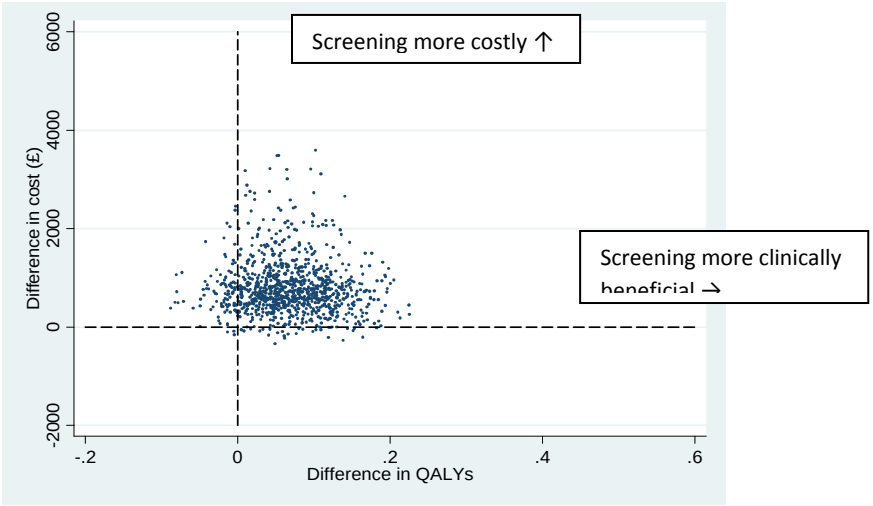
	No screening	Screening for T2DM only	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
Years spent in different model states				
Total life years	30.34 (27.75, 32.86)	30.40 (27.82, 32.93)	30.49 (27.90, 33.01)	30.46 (27.88, 32.99)
QALYs	28.06 (23.49, 32.01)	28.12 (23.58, 32.08)	28.26 (23.74, 32.23)	28.22 (23.69, 32.18)
Years T2DM free	20.85 (10.36, 29.45)	20.85 (10.36, 29.45)	21.17 (10.66, 29.79)	21.07 (10.54, 29.66)
Years IGT	9.06 (1.95, 14.87)	9.06 (1.95, 14.87)	9.40 (2.28, 15.21)	9.27 (2.15, 15.10)
Years T2DM	9.49 (3.38, 18.01)	9.55 (3.45, 18.08)	9.31 (3.19, 17.85)	9.49 (3.38, 18.01)
Probability of developing T2DM (%)				
T2DM	68.40 (18.02, 91.83)	68.40 (18.02, 91.83)	67.46 (17.09, 91.09)	67.89 (17.41, 91.42)
Clinically detected	49.24 (13.60, 82.42)	43.19 (7.94, 75.77)	41.23 (6.21, 73.55)	41.23 (6.21, 73.55)
Screen detected	0	6.80 (5.96, 7.70)	8.13 (6.83, 9.41)	8.53 (7.21, 9.77)
Costs (GBP)				
Total cost	17,290 (5,746, 39,580)	18,040 (7,083, 39,970)	17,910 (7,124, 39,740)	17,900 (7,061, 39,710)
Screening	0	3.95 (2.96, 5.38)	5.12 (4.06, 6.56)	5.12 (4.06, 6.56)
Intervention	0	0	295.40 (212.20, 377.60)	14.57 (9.93,19.81)
Undiagnosed T2DM	121 (35, 228)	106 (21.00, 207)	101 (16, 201)	101 (16, 201)
Screen diagnosed T2DM	0	4,147 (3,504, 4,868)	4,655 (3,851 5,568)	4,900 (4,046, 5,842)
Clinically diagnosed T2DM	17,170 (5,653, 39,440)	13,700 (3,046, 35,460)	12,810 (2,289, 34,330)	12,810 (2,289, 34,330)
All T2DM states	17,290 (5,746, 39,580)	18, 030 (7,080, 39,970)	17,630 (6,793, 39,470)	17,880 (7,037, 39,700)

Table 6.13: A comparison of cost and clinical outcomes between screening strategies

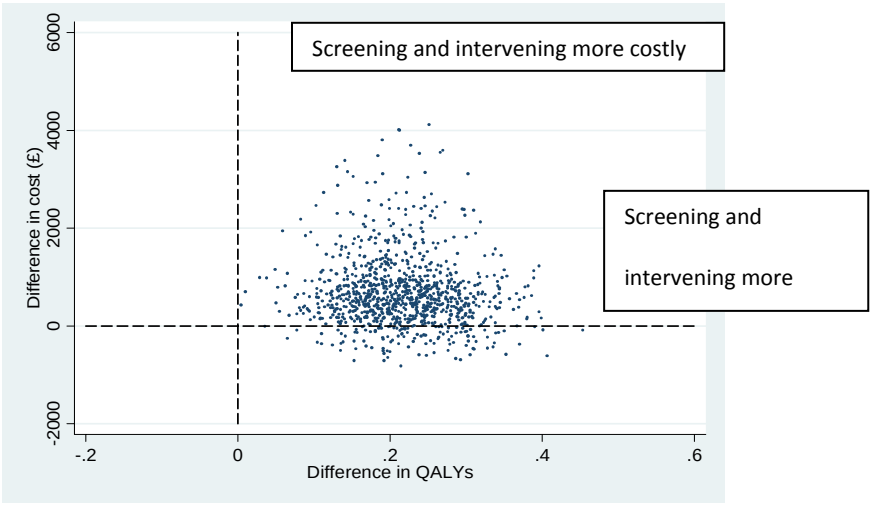
	No screening	**Screening for T2DM only**	**Screening for T2DM and IGT, lifestyle interventions**	**Screening for T2DM and IGT, pharmacological interventions**
	Mean (95% CrI)	Mean difference from, or in comparison to, no screening (95% CrI)		
Total life years	30.34 (27.75, 32.86)	0.06 (0.02, 0.12)	0.15 (0.08, 0.22)	0.13 (0.06, 0.20)
QALYs	28.06 (23.49, 32.01)	0.07 (-0.03, 0.18)	0.22 (0.08, 0.36)	0.17 (0.03, 0.32)
Years T2DM free	20.85 (10.36, 29.45)	-	0.33 (0.21, 0.43)	0.20 (0.10, 0.37)
Lifetime risk of T2DM (%)	64.55 (18.02, 91.83)	-	-0.98 (-0.50, -1.42)	-0.54 (-0.21, -1.17)
Total cost	17,290 (5,746, 39,580)	730 (-9, 2,341)	610 (-373, 2,693)	579 (-428, 2,658)
Cost per life year gained	-	11,460	4,179	4,768
Cost per QALY gained	-	8,681	2,863	3,429
Cost per case prevented	-	-	62,810	105,000
Probability of cost- effectiveness at willingness-to-pay per QALY: £20,000 £30,000	-	 0.681 0.765	 0.986 0.996	 0.947 0.973

Figure 6.11: Cost-effectiveness plots

Screening for T2DM only compared to no screening



Screening for IGT and T2DM with lifestyle interventions compared to no screening



Screening for IGT and T2DM with pharmacological interventions compared to no screening

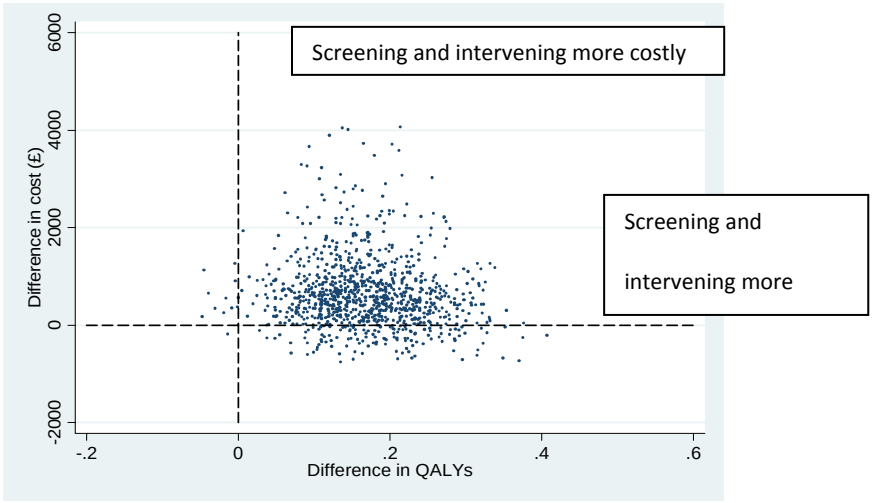
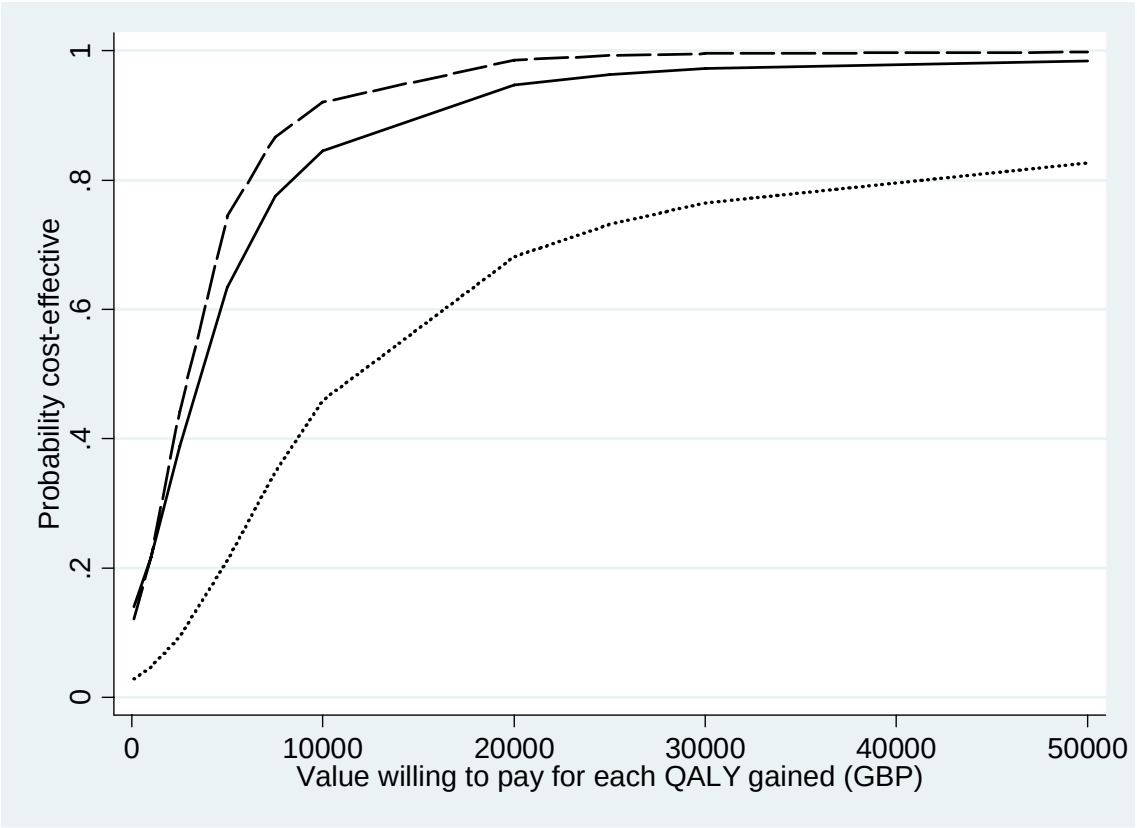


Figure 6.12: Cost-effectiveness acceptability curves



6.9 Discussion

As discussed in detail in chapter 5 previous studies have compared the cost and clinical effectiveness of intervening in IGT individuals to delay T2DM (Avenell et al., 2004, Caro et al., 2004, Eddy et al., 2005, Herman et al., 2005, Icks et al., 2007, Jacobs-van der Bruggen et al., 2007, Palmer et al., 2004, Segal et al., 1998). Results of the models were all favourable in terms of cost and clinical effectiveness, but as the models were designed to assess the effectiveness of interventions rather than screening and intervening, the models only considered a population of IGT individuals, only part of the disease pathway was modelled, and it was assumed that management of T2DM started as soon as the disease developed. The model developed here models the whole screening and intervention pathway from screening to death, and offers the first comparison of different approaches to T2DM screening and prevention at the population level, and is totally probabilistic, allowing for all sources of uncertainty.

From the results presented here it appears that screening for IGT, in addition to T2DM, and intervening with either lifestyle or pharmacological interventions is a cost-effective health policy. Although screening for T2DM alone gave a relatively low predicted incremental cost per QALY of £8,681, due to uncertainty in the model the probability of this strategy being cost effective was only 77% at the £30,000 willingness-to-pay threshold. Therefore uncertainty remains as to whether screening for T2DM alone would be a viable policy.

Differences in clinical outcomes between the no screening strategy and the three active screening strategies were small. This is partly because results were reported as an average for a screened population of mixed glucose tolerance, where only 12% had IGT and 5% T2DM at the time of screening. Also the sensitivity of the test for IGT was fairly low. Therefore screening a population with higher prevalences of glucose intolerance, implementing a screening strategy whereby the population is re-screened at regular intervals, or improving the efficiency of the screening test will probably lead to an improvement in the clinical outcomes. This will be explored further in Chapters 7 and 8 through a series of sensitivity analyses and model extensions. Also worth noting is that microvascular and macrovascular outcomes were not measured individually in this model, and as these may show benefits from the early detection or delay of T2DM, additional clinical benefits in terms of these outcomes are plausible.

The model described here makes a number of assumptions, as detailed in section 5.3. No transition was allowed from NGT to T2DM without first passing through IGT. This is because it is clinically unlikely that an individual would change from NGT to T2DM within a year, which is one model cycle. No transition was allowed from T2DM back to IGT or from IGT to NGT. This is clinically accurate, as once an individual is diagnosed with T2DM, even if their glucose tolerance improves, they are still clinically defined as having diabetes. Also once an individual has suffered from IGT, even if their glucose tolerance improves, their future risk of T2DM is probably more similar to individuals with IGT rather than those who have always had NGT.

Another assumption was that the HbA_{1c} of those with diabetes who were clinically diagnosed would be similar to the 10 year average of an intensively treated group of people with T2DM from the UKPDS study (UK Prospective Diabetes Study (UKPDS) Group, 1998a). This assumption was made in the absence of long-term clinical data on individuals whose diabetes was screen detected. Although 10 year averages of HbA_{1c} levels were used for diagnosed diabetics, where our model was run for longer time horizons the HbA_{1c} levels were potentially underestimated, which means complication rates and their effects on utilities and mortality rates may also be moderately underestimated. Further data is needed on how HbA_{1c} could be expected to increase over time to allow for more accurate modelling.

Complications associated with T2DM, in terms of both costs and their effect on quality of life, were modelled using data from the UKPDS study. They had fitted and reported Weibull models for six complications, aswell as, in a separate paper, the effects of complications on utility values. The use of both this cost and utility data was assuming that the intensive treatment arm of the UKPDS trial was representative of a screen detected T2DM population. It was also assumed that the Weibull models fitted by UKPDS, were appropriate for the data, and that both model fit and model assumptions had been checked. The use of WinBUGS and a Bayesian framework, for the comprehensive decision model, provided a flexible framework that enabled the model of complication rates and their effects on utilities to be incorporated within the full model, which enabled the incorporation of all the uncertainty around complication rates and their effects to be included also. A limitation of this decision model is that the model parameters were considered independent, when ideally the potential correlation between different parts of the model needs to be accounted for.

As the model was run for a time horizon of 50 years, the screened population (aged 45 years at the start) aged with each cycle of the model, thus, where possible, time dependent model parameters were incorporated. However, for some model parameters, such as the treatment intervention effects, it was assumed that the effect was constant over time. The effects on model outcomes of compliance to treatment and screening, will be explored further in Chapter 8.

Most of the data utilised for the model came from studies where the majority of the study population were Caucasian adults. The meta-analyses of intervention effects was the most diverse in terms of the ethnicities of study populations, but as a previous analysis discussed in Chapter 4 showed no interaction between ethnicity and intervention effect, no bias should have been introduced to the model. The results are therefore most relevant for Caucasian populations.

6.10 Chapter summary

In conclusion, a policy of a one-off screening for both T2DM and IGT, with appropriate intervention for those identified with IGT, appears to be cost-effective. However, given the uncertainty in the results presented here, particularly for the assessment of T2DM screening, further research is needed on the long term clinical effects of early diagnosis of T2DM. Furthermore, to model more accurately the two strategies that involved interventions, additional information on long-term compliance to interventions and their potential harms and benefits, is required.

This model builds on previous work by utilising a comprehensive decision modelling framework that included as much of the relevant, available evidence as was feasible, was probabilistic in terms of model inputs and all uncertainty was included where relevant. The model was complex, incorporating four meta-analyses, and a series of Weibull survival models, for modelling complication rates and their associated utilities. It was also the first model to directly compare screening for T2DM alone, to screening for IGT and T2DM in combination, and it was the first decision model to contain the full screening/treatment/disease pathway, from NGT through to death.

7: MODEL CHECKING AND SENSITIVITY ANALYSES

7.1 Chapter overview

Once the primary, base case model had been developed, as described in the previous chapter, it was essential to carry out a thorough check of the model and its assumptions. This involved checking convergence of the model parameters, considering the effect of distributions given to data and parameters within the model, and verifying that the vague prior distributions utilised were not influencing the results. In addition a number of sensitivity analyses were carried out on several model parameters including prevalence of NGT, IGT and T2DM, test efficiency and the time horizon of the model. This was to enable a better understanding of the model, particularly to identify which model parameters had the greatest effect on the cost-effectiveness of different screening strategies. By identifying the model parameters that have the greatest influence on the model results, the most important factors that need to be considered when implementing a screening strategy can be ascertained. In addition sensitivity analyses are useful in that they can help to identify clinical areas where further research is needed. This chapter will describe and present results for all the model checks and the sensitivity analyses that were carried out on the primary model.

7.2 Convergence of model parameters

7.2.1 Graphics utilised

When assessing models that have been fitted using Markov Chain Monte Carlo (MCMC) methodology, as in this example, a number of issues need to be considered when investigating the validity of the model (Brooks and Gelman, 2007, Cowles and Carlin, 1996). MCMC methods aim to successively sample values from a convergent Markov chain and they are not without their problems, in particular there are three primary concerns that need to be assessed.

Firstly, as the chain requires starting values, the choice of initial values will influence the early part of the chain. The initial part of the chain, before the model settles, is called the burn-in and is discarded, with the results of the model being taken from later iterations. Therefore the first check to make is whether the burn-in was long enough, and whether chains with different initial values will culminate in the chains converging before the end of the burn-in. This can be assessed visually by plotting a number of chains, with a range of different initial values, against

model simulation. Good convergence will produce a plot whereby the chains all converge early on in the simulation run.

Secondly an assessment of whether the chain has been run for long enough to allow for convergence of model parameters, needs to be made. Whilst some models converge quickly, others, usually when there is high autocorrelation, will require 100,000s of simulations before convergence is adequately achieved. Convergence can be assessed using trace plots. Trace plots also allow for an assessment of how well the chain is mixing across all possible values. Trace plots are produced by plotting the value of the model parameter for each model simulation, thereby enabling the parameter to be monitored across simulations and any problems to be identified. A problem with convergence may be identified if the trace plot drifts rather than remaining stable, and problems of mixing can be identified if the chain appears to 'stick' at or near one value for several iterations, rather than showing a rapid movement across all possible values.

Density plots, as well as assessing if the distributions of model parameters are as expected, can also be used to assess convergence. By dividing a chain into two halves and plotting each half, as well as the whole chain, onto one axis, convergence will be shown if all three densities are similar to each other. A measure of convergence (D) was calculated from the density plots, whereby D represents the maximum distance between two densities as a percentage of the maximum height of the combined density (Thompson et al., 2006). A value of D less than 20 represents reasonable agreement and a value less than ten shows good agreement and convergence.

The third issue that needs to be assessed is correlation between successive values in a chain. High autocorrelation between successive iterations means that values in the chain cannot be treated as a random sample from the posterior, a problem that can be overcome by running the model for an increased number of iterations to obtain an adequately random sample. Problems of correlation within chains can be assessed using autocorrelation plots. Autocorrelation plots comprise of the correlation between two points in a chain plotted against their distance from each other within the chain (lag).

To assess the primary model, plots of overlaid chains with different initial values were drawn for all parameters within the model that required initial values to be specified. These were the

three prevalences of NGT, IGT and T2DM, the values of the three by three table that determined the sensitivity and specificity of the screening tests and the means and standard deviations of the four meta-analyses run within the model. From the four meta-analyses *mu1* and *sd1* represent the mean and between study standard deviation of the meta-analysis for the IGT to T2DM transition, *mu2* and *sd2* the meta-analysis performed for the pharmacological interventions, *mu3* and *sd3* the lifestyle interventions and *mu4* and *sd4* the transition from undiagnosed to diagnosed T2DM. Autocorrelation and trace plots were also compiled for the above parameters and in addition, to assess parameters computed within the model, both the QALYs and the total costs for each of the four strategies were plotted. Lastly density plots on the aforementioned parameters were constructed to monitor both converge and the distributions of these model parameters, to allow for the assessment of any anomalies.

7.2.2 Results of convergence checks

The diagnostic graphs are presented in figures 7.1 to 7.4, with additional graphs given in Appendix 4. The base case model, and all subsequent sensitivity analyses and model extensions, were run with a burn-in of 10,000 iterations, with results being taken from a subsequent 20,000 iterations. From figure 7.3, where four chains comprising of different initial values chosen to represent the full plausible range of the parameter, were run, it can be seen that convergence of chains occurs within the first few iterations of the model. This indicates that the initial values were not influencing the end results computed by the model, and that the length of burn-in was more than adequate.

All of the trace plots showed no problems with any of the parameters drifting over the course of the simulations (figure 7.1). Both the autocorrelation plots and the trace plots for the means and standard deviations of the four meta-analyses run within the decision model, highlighted that there was an issue with correlation between values in concurrent simulations, particularly for *mu2* and *mu3* which were the mean values from the lifestyle interventions and pharmacological interventions meta-analyses. Auto-correlation is not a problem if enough iterations are run to enable results to be drawn from a random sample. The WinBUGs code was run for 20,000 iterations for all models, running the base case model for a further 20,000 iterations did not change the estimated means and standard deviations from the four meta-analyses, or their estimated uncertainty, therefore it was concluded that all models had been run for sufficient iterations for correlation between simulations not to influence the results.

All the density plots, given in Figure 7.4 along with the corresponding D value, also indicated adequate convergence of model parameters. Splitting the chains into two and plotting both halves showed good agreement between the two halves and the D values assessing agreement were all below twenty, representing either reasonable or good agreement.

When assessing the density plots it was noted that the plots for the QALYs for each of the four strategies, followed an unusual distribution (Figure 7.4). To investigate this further, as the QALYs were composed of time spent in each of the Markov model states, individual density plots were drawn for each of the seven states. The plots for the seven states for the T2DM screening strategy are displayed in Figure 7.5, while the plots for the other three strategies are given in Appendix 4. All four strategies had similar plots for each of the seven states. From the plots it can be seen that the plots varied widely between states, due to different transition rates moving to and from states. For example for the NGT state there was only movement to undiagnosed IGT or death, with the death rates being relatively low and the movement to undiagnosed IGT being stratified by two age groups (below or equal to and above 65 years), although constant within age group. Due to the transition to undiagnosed IGT dominating the distribution the density curve effectively represents two exponential survival distributions, one for each age range, whereby the distributions are exponential due to risk or 'hazard' of moving to undiagnosed IGT being constant. The diversity in the density plots for each of the six 'alive' Markov states, leads to the unusual distribution when they are used to form QALYs for each strategy.

Figure 7.1: Trace plots of the means and standard deviations from the four meta-analyses

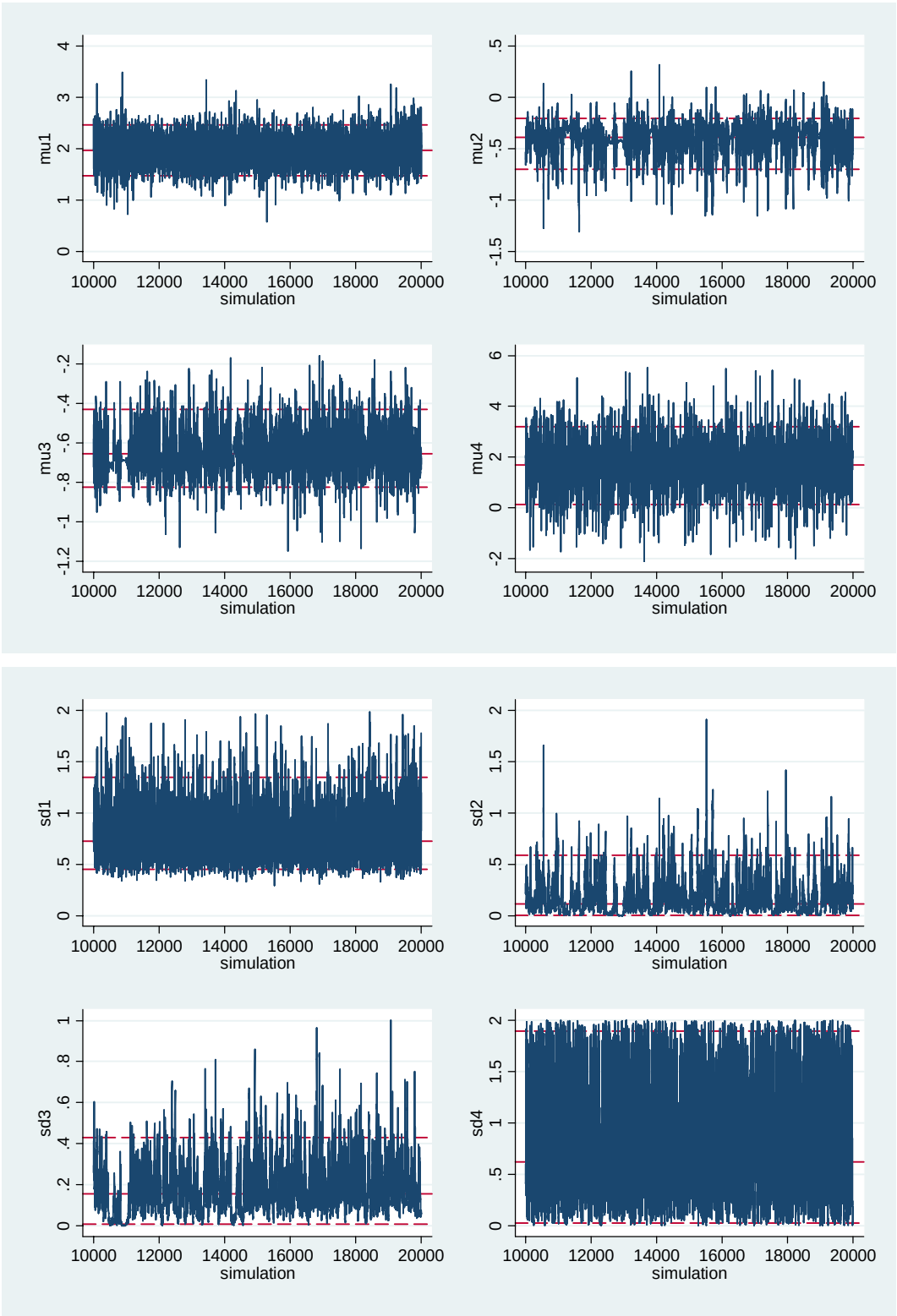


Figure 7.2: Auto-correlation plots of the means and standard deviations from the four meta-analyses

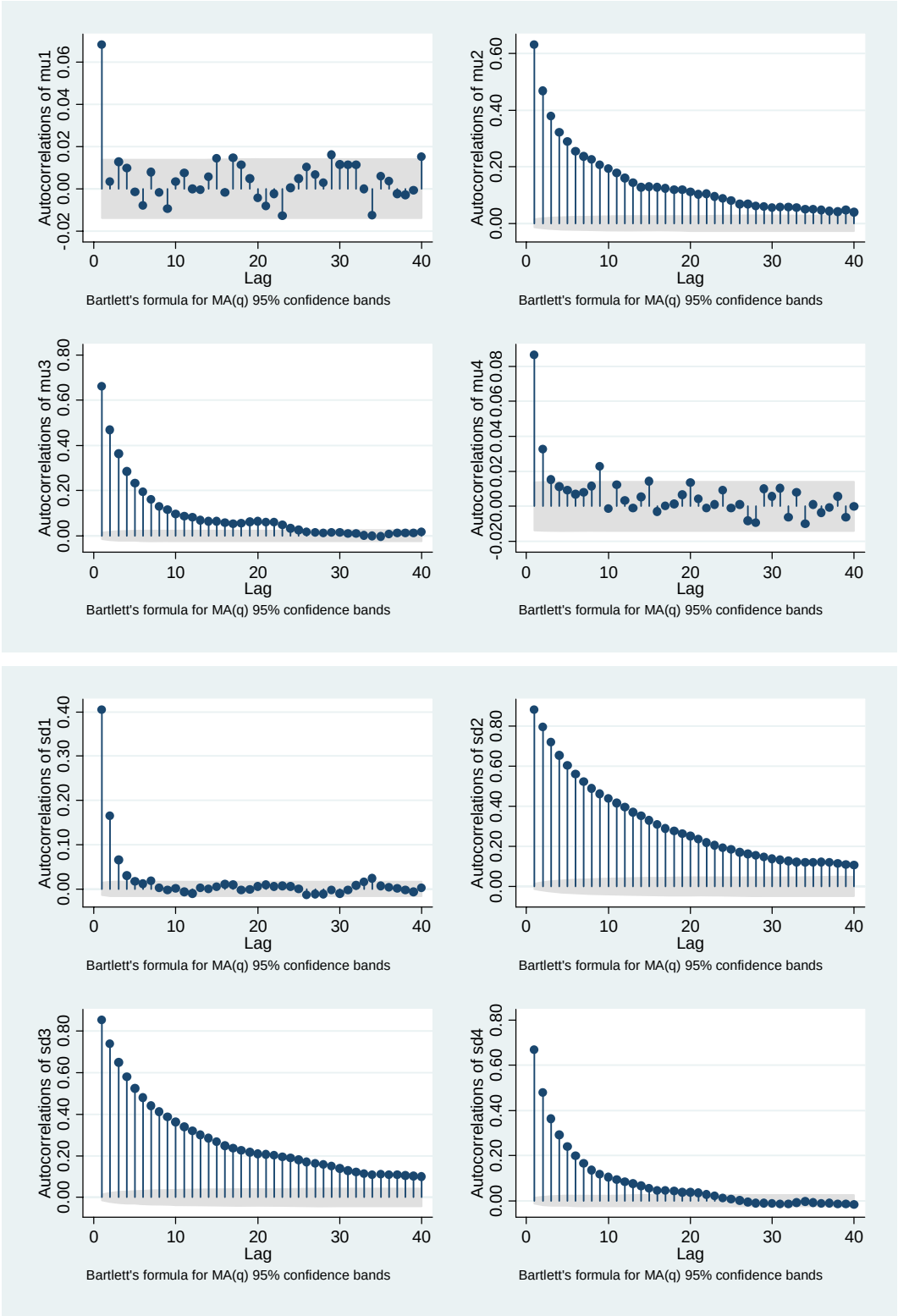


Figure 7.3: Overlaid chains, each with a different initial value, plotted for the mean values from the four meta-analyses

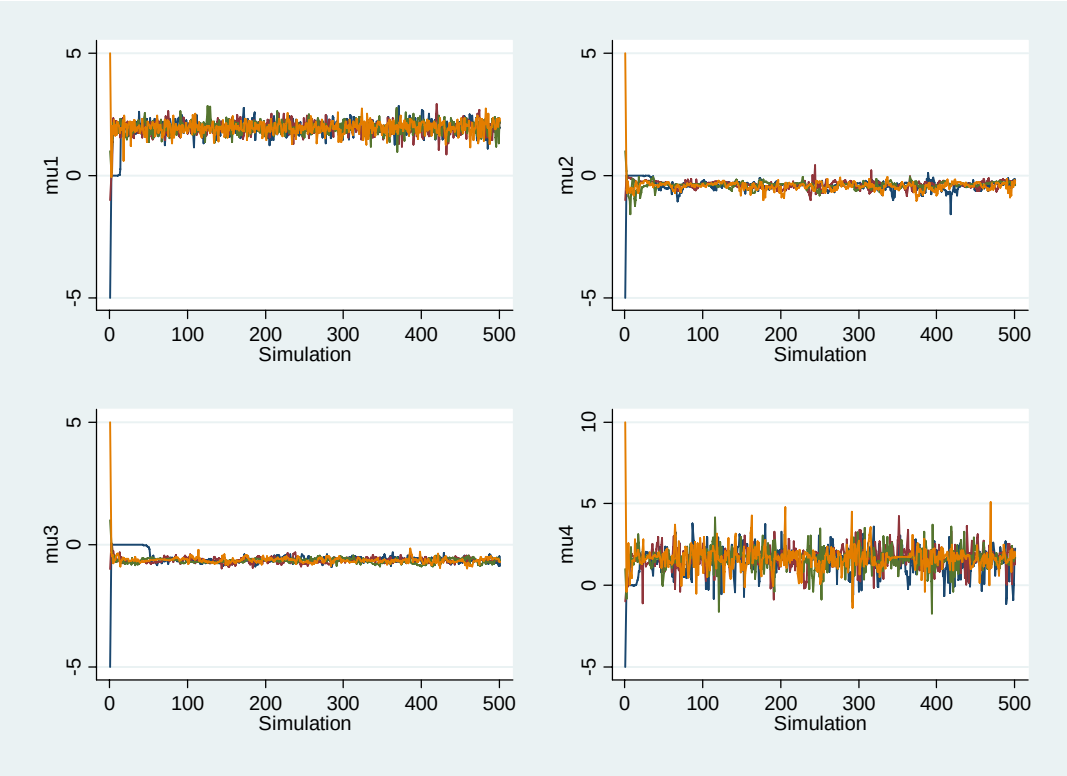


Figure 7.4: Density plots of the mean values from the four meta-analyses and QALY estimates

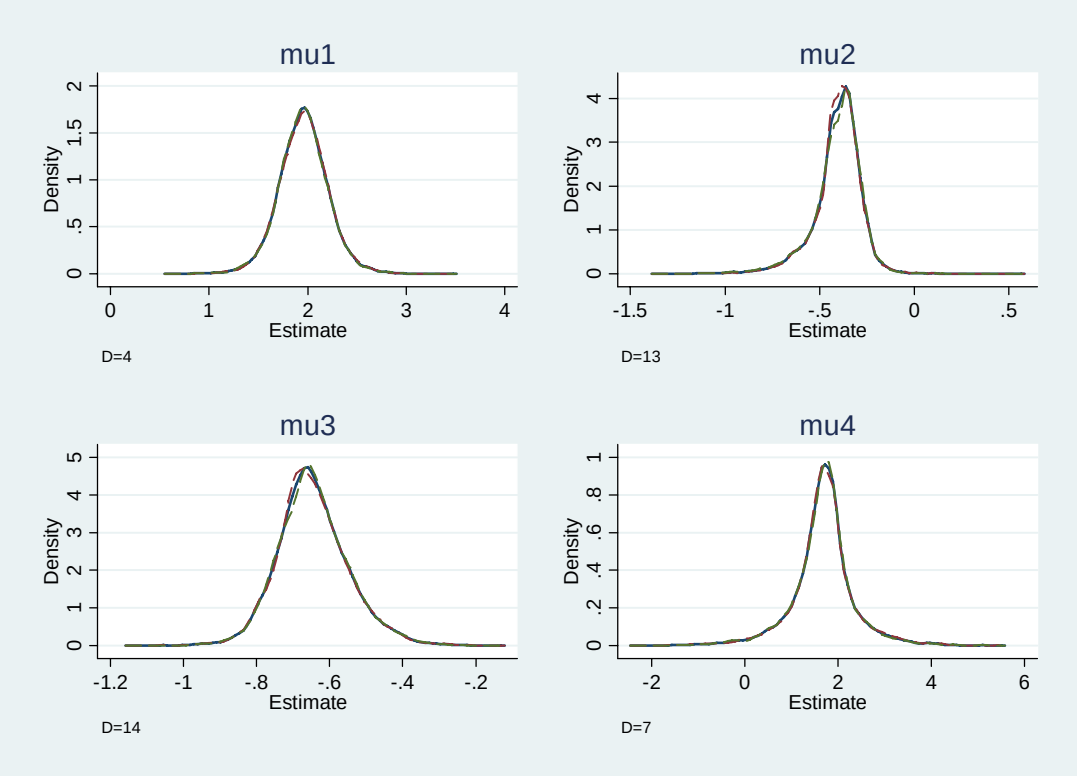
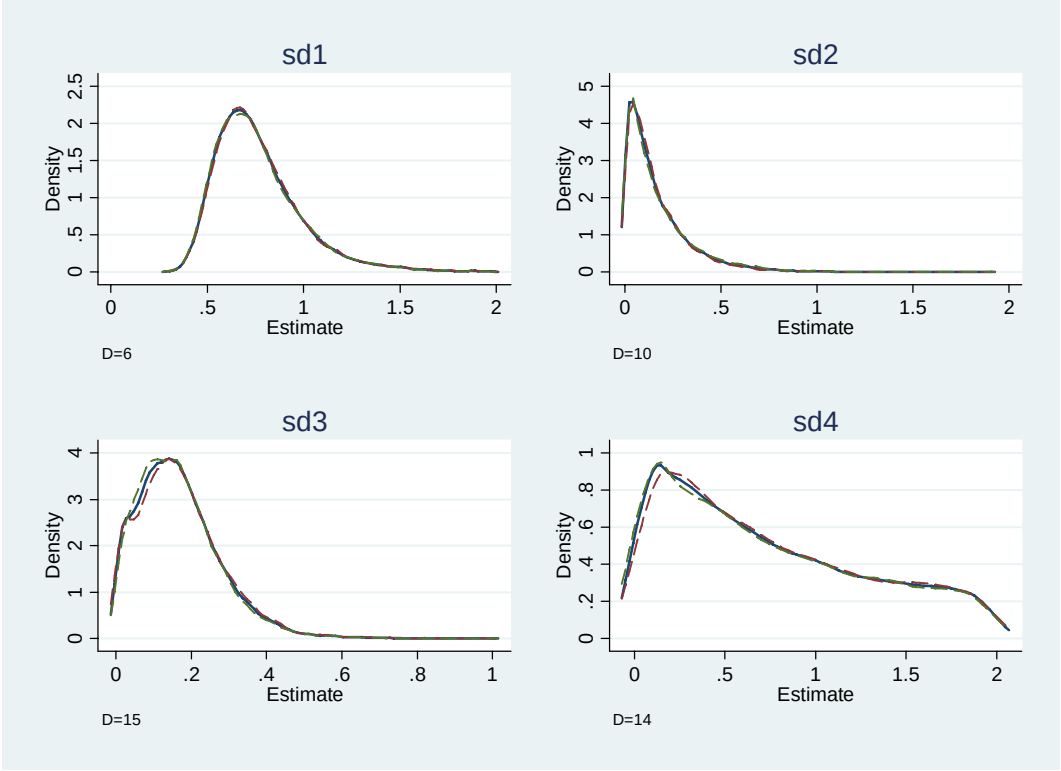


Figure 7.4 continued:



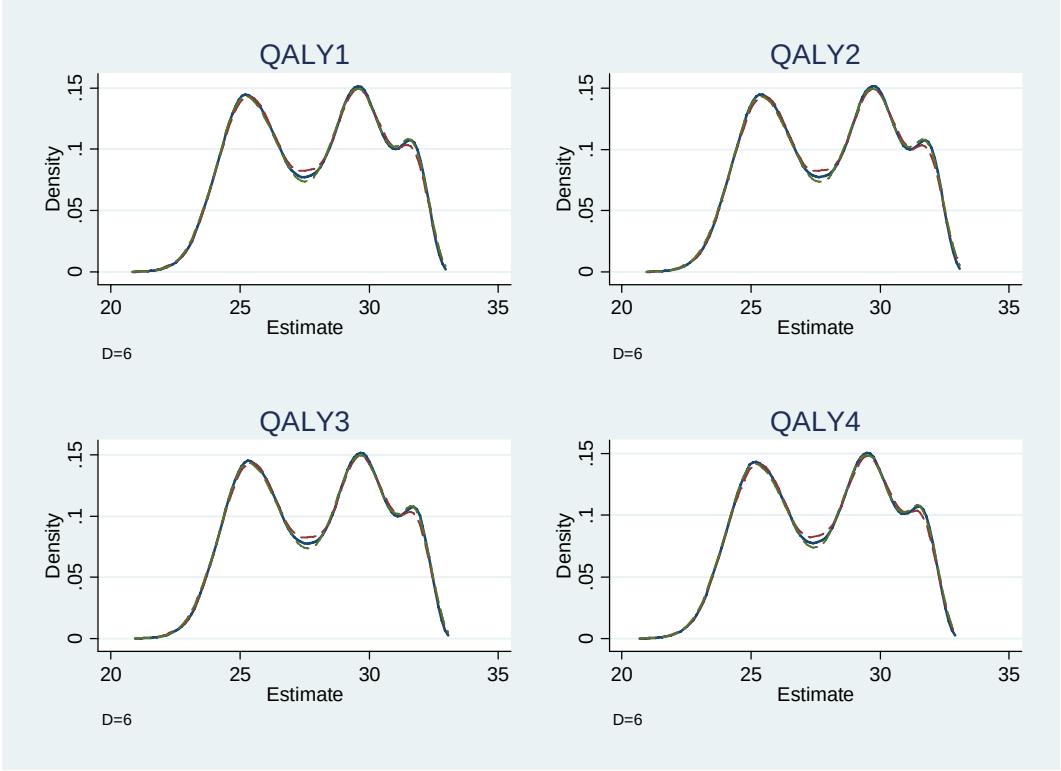
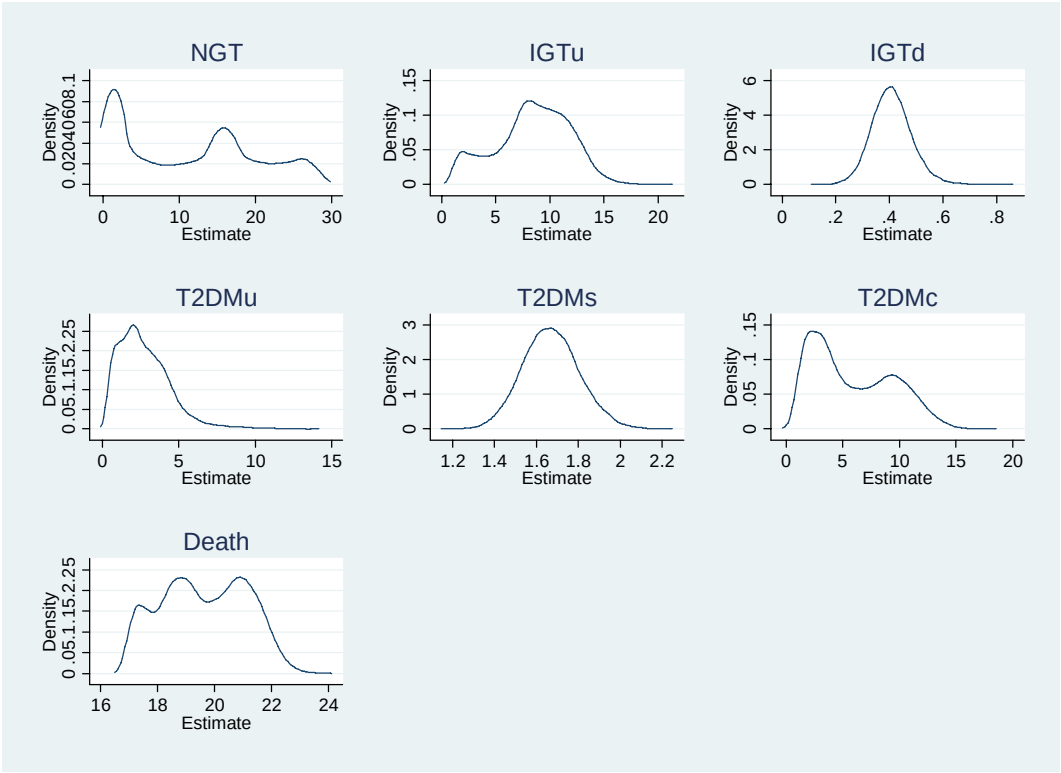


Figure 7.5: Density plots for each state for the T2DM screening strategy



7.3 Distributions of model data

A number of the parameters entered into the decision model had distributions placed around them to allow for any uncertainty present in their estimation. These are described in detail in Chapter 6. For this decision model both costs and utilities were modelled with normal distributions, although it is common practice to model costs using a gamma or log normal distribution, to avoid the range of possible values including negative values, and to model utilities on a beta distribution to ensure the range of possible values have an upper limit of 1. It was attempted to model costs with a gamma distribution and utilities with a beta distribution but due to the small standard errors of these data inputs, method of moments estimation could not provide sensible parameter estimates that could be used to specify the distributions. As all the costs and utilities do have small standard errors though, modelling them normally should not have posed a problem as the plausible range will still fall within an appropriate range for these parameters.

7.4 Prior distributions

For the decision model only vague rather than informative priors were used, to allow the data to dominate model results. When incorporating vague priors it is important to check that the choice of prior is not affecting or influencing the results in any way (Lambert et al., 2005). Therefore several models were run with different prior distributions used for the between study standard deviation (τ) of the four meta-analyses which comprised part of the decision model. Although some checks on prior distributions for the two intervention meta-analyses were carried out and reported in chapter 4, they are expanded here and the effect on the model results, in terms of the incremental costs per QALY, rather than just the meta-analysis were considered.

In the base case model the distribution applied to all the four τ values, within the meta-analyses, were uniform(0,2). Each of the four prior distributions were altered one at a time to either a half normal or a gamma distribution and the effects on the model are reported in table 7.1. The distributions used are described in detail in Chapter 2,

Table 7.1: The effects of changing the prior distributions

Meta-analysis	Prior distribution for the between study standard deviation (τ)	Mean (95% CrI)	Cost per QALY gained compared to the strategy of no screening (GBP)		
			Screening for T2DM only	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
Pharmacological interven	*Uniform(0,2)*	-0.39 (-0.21, -0.71)	8,681	2,863	3,429
	Gamma(0.1,0.1)	-0.38 (-0.27, -0.49)	8,822	2,829	3,613
	Normal(0,1)I(0,)	-0.39 (-0.20, -0.61)	8,709	2,848	3,453
Lifestyle interventions	*Uniform(0,2)*	-0.65 (-0.43, -0.82)	8,681	2,863	3,429
	Gamma(0.1,0.1)	-0.67 (-0.51,-0.79)	8,828	2,797	3,376
	Normal(0,1)I(0,)	-0.65 (-0.43, -0.86)	8,800	2,810	3,452
Transition from IGT to T2DM	*Uniform(0,2)*	1.96 (1.46, 2.46)	8,681	2,863	3,429
	Gamma(0.1,0.1)	1.95 (1.49, 2.42)	8,694	2,776	3,362
	Normal(0,1)I(0,)	1.95 (1.48, 2.43)	8,735	2,802	3,355
Transition from T2DMu to T2DMc	*Uniform(0,2)*	1.70 (0.15, 3.15)	8,681	2,863	3,429
	Gamma(0.1,0.1)	1.73 (0.22, 2.98)	8,759	2,903	3,582
	Normal(0,1)I(0,)	1.70 (0.49, 2.84)	8,715	2,881	3,486

From the results it can be seen that although changing the prior distributions of the between study standard deviation had some effect on both the mean values of the meta-analyses and the estimated incremental costs of each of the three active screening strategies compare to no screening, the effects were minimal and did not change the overall conclusions of the decision model.

7.5 Economic sensitivities

Once a thorough check of the model and its assumptions had been carried out, extensive sensitivity analyses were run on the model. By carrying out numerous sensitivity analyses and comparing model results, a fuller understanding of the importance of model parameters, and the clinical implications of different approaches to a screening/intervention programme, can be fully understood. A number of sensitivity analyses were carried out on both economic and clinical model inputs and assumptions. The economic sensitivity analyses will be described and interpreted here with the clinical sensitivity analyses discussed in detail in section 7.6.

7.5.1 Discounting costs and benefits

The first sensitivity analyses assessed the impact of discounting both costs and benefits, on the conclusions of the model. Economic evaluation studies often report results from discounted models as it can be argued that more emphasis should be given to more immediate costs and benefits, rather than long-term implications, as described in more detail in Chapter 2. The National Institute of Clinical Excellence (NICE) currently recommend that cost-effectiveness analyses should discount both costs and benefits at a rate of 3.5% per year (National Institute for Clinical Excellence, 2004), therefore these were the values used to discount the primary model. There has been dispute over these discounts rates though, with some advocating a return to differential discounting, whereby effects are discounted at a lower annual rate than costs, so that future health effects are not discounted so heavily (Brouwer et al., 2005). All other parameters remained unchanged from the base case scenario.

The results from a model with both costs and benefits discounted at 3.5% per annum are presented in table 7.2. When both costs and benefits were discounted, as expected both the total cost of each of the four screening/intervention strategies, as well as the estimated QALYs, were reduced. Total costs were reduced dramatically by over half, which is unsurprising, as because

the majority of costs are incurred once T2DM develops, the bulk of the costs are accrued in the later years of the model.

The more interesting results to compare between discounted and undiscounted models are the incremental cost-effectiveness ratios, that is the estimated cost for each QALY gained, and the probabilities of a strategy being cost-effective at a certain willingness-to-pay threshold. From table 7.2 it can be seen that for each of the three active screening/intervention strategies compared to no screening, the incremental costs increased when costs and benefits were discounted. The results for the T2DM screening strategy increased from £8,681 to £14,150 per QALY, screening followed by lifestyle interventions increased from £2,863 to £6,242 per QALY, and screening with pharmacological interventions from £3,429 to £7,023 per QALY. This is because the discounted analysis gives more weight to the years immediately after screening, when the clinical effect is small. Although the incremental costs were higher, they were still relatively low and both the screening and intervention strategies still retained a high probability of being cost-effective at the £20,000 willingness-to-pay threshold, with screening and lifestyle interventions having a probability of 0.93 and screening with pharmacological interventions a probability of 0.85. The strategy of screening for T2DM only though, when costs and benefits were discounted, had a probability of cost-effectiveness of 0.49 at the £30,000 threshold and 0.61 at the £20,000 threshold, making this strategy borderline as to whether it would be an acceptable health policy for implementation.

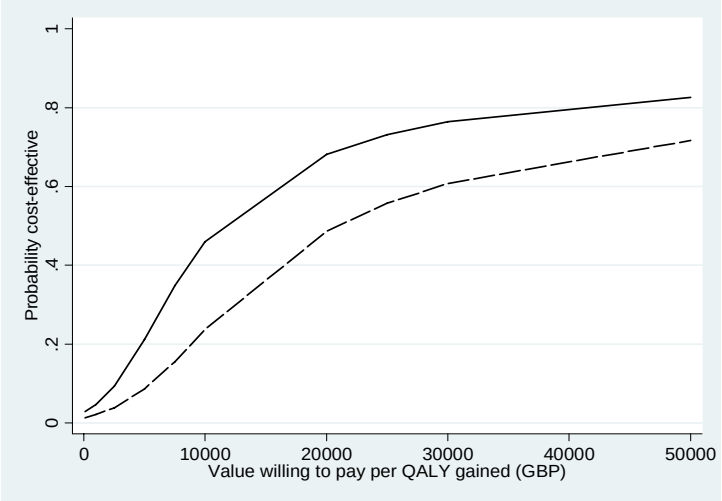
Figure 7.6 presents the cost-effectiveness acceptability curves for each of the three active screening/intervention policies compared to no screening, for both discounted and undiscounted results. It can be seen that discounting both costs and benefits reduces the probability of each strategy being cost-effective, over a range of willingness-to-pay thresholds.

Table 7.2: A comparison of results from an undiscounted model and a model discounted at 3.5% per annum for both costs and benefits

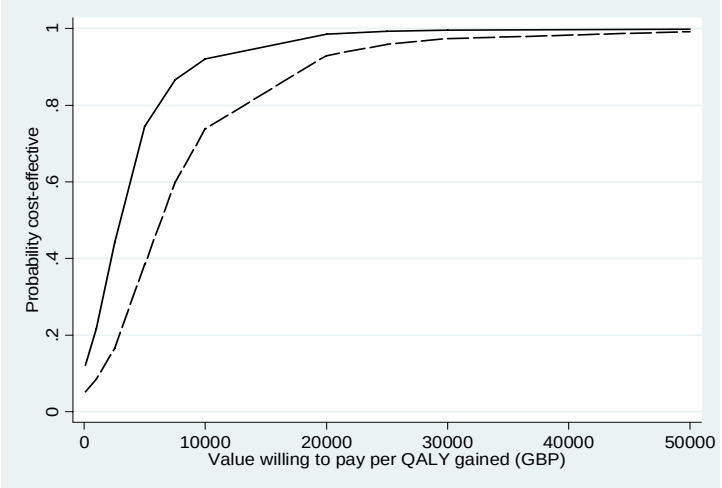
	No screening	Screening for T2DM only	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
Undiscounted (base case results)				
Total life years	30.34 (27.75, 32.86)	30.40 (27.82, 32.93)	30.49 (27.90, 33.01)	30.46 (27.88, 32.99)
QALYs	28.06 (23.49, 32.01)	28.12 (23.58, 32.08)	28.26 (23.74, 32.23)	28.22 (23.69, 32.18)
Years T2DM free	20.85 (10.36, 29.45)	20.85 (10.36, 29.45)	21.17 (10.66, 29.79)	21.07 (10.54, 29.66)
Total cost (GBP)	17,290 (5,746, 39,580)	18,040 (7,083, 39,970)	17,910 (7,124, 39,740)	17,900 (7,061, 39,710)
Cost per life year gained (GBP)	-	11,460	4,179	4,768
Cost per QALY gained (GBP)	-	8,681	2,863	3,429
Probability of cost- effectiveness at willingness-to-pay per QALY:	-			
£20,000		0.681	0.986	0.947
£30,000		0.765	0.996	0.973
Discounted at 3.5% per year for both costs and benefits				
Total life years	18.19 (17.25, 18.98)	18.22 (17.28, 18.96)	18.25 (17.31, 18.98)	18.19 (17.25, 18.93)
QALYs	17.13 (15.02, 18.49)	17.16 (15.07, 18.51)	17.22 (15.14, 18.58)	17.20 (15.12, 18.56)
Years T2DM free	13.69 (7.99, 17.08)	13.69 (7.99, 17.08)	13.86 (8.16, 17.26)	21.07 (10.54, 29.66)
Total cost (GBP)	7,636 (2,636, 19,370)	8,244 (3,702, 19,690)	8,260 (3,789, 19,590)	8,199 (3,710, 19,570)
Cost per life year gained (GBP)	-	23,710	10,900	11,690
Cost per QALY gained (GBP)	-	14,150	6,242	7,023
Probability of cost- effectiveness at willingness-to-pay per QALY:	-			
£20,000		0.486	0.930	0.850
£30,000		0.608	0.974	0.916

Figure 7.6: Cost-effectiveness acceptability curves for discounted and undiscounted results

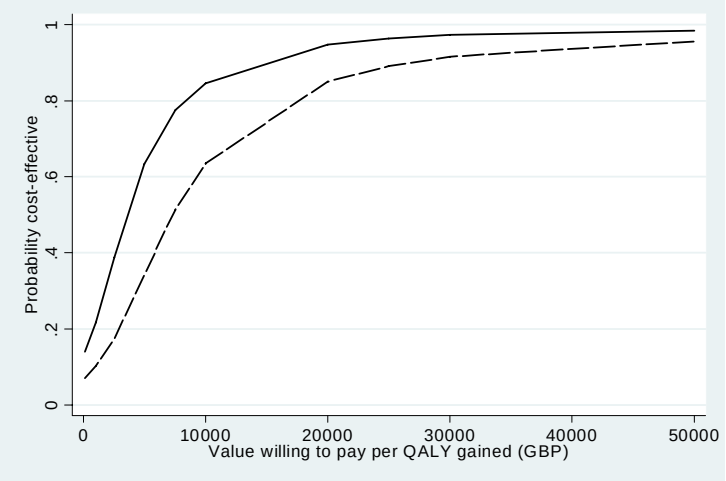
Diabetes screening

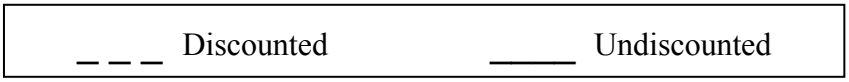


Screening with lifestyle interventions



Screening with pharmacological interventions





7.5.2 Sensitivity analyses for costs

In the decision model costs were attached to screening tests, interventions applied to individuals with IGT, and estimated costs incurred by T2DM which differed for undiagnosed, clinically diagnosed and screen diagnosed. As the costs of screening tests were relatively low (40p for a FPG test and £1.30 for an OGTT), and the accuracy of these costs relatively assured, they were not adjusted in sensitivity analyses. Sensitivity analyses were carried out though on both intervention costs and the costs incurred by T2DM, through monitoring, treatment and complications of the disease.

For the primary, base case model intervention costs were estimated as, for lifestyle interventions £398 in the first year and £280 in subsequent years, and for pharmacological interventions £16.10 per annum. In particular the lifestyle interventions were very difficult to cost, as described in Chapter 6, and costs could vary depending on the intensity of the intervention adopted. To assess the effect on model conclusions if these estimates had been incorrectly specified, particularly as costs were estimated from clinical trial settings, and may be higher in practice, models were run with intervention costs at two, five and ten times the base case rate. The results of these models are given in table 7.3. As the strategy incorporating screening only for T2DM involved no interventions, the results of this strategy remained unchanged from the base case scenario, although they are still included in table 7.3 to enable comparisons with the two intervention strategies.

Total costs for the screening and lifestyle intervention strategy increased from £17,910 (95% CrI: 7,124 to 39,740) for the base case scenario to £20,580 (95% CrI: 9,895 to 42,010) if intervention costs were increased ten fold. Correspondingly the strategy involving pharmacological interventions increased from £17,900 (95% CrI: 7,061, 39,710) to £18,030 (95% CrI: 7,206 to 39,580). Therefore in terms of overall costs of each strategy, increasing intervention costs had relatively little impact. This is because the total cost of each strategy was dominated by the costs of T2DM.

In terms of cost-effectiveness the incremental costs increased from £2,863 to £15,410 per QALY, for the strategy of screening with lifestyle interventions when lifestyle costs were increased ten fold, and the probability of cost-effectiveness at the £30,000 willingness-to-pay threshold reduced from 1.00 to 0.97. For the strategy involving pharmacological interventions

cost per QALY increased from £3,429 to £4,231 and the probability of cost-effectiveness at the £30,000 threshold remained unchanged at 0.97. The impact of increasing costs of lifestyle interventions was greater than that of increasing pharmacological, as the lifestyle costs were much greater per annum in the base case model. Increasing intervention costs though had little effect on the overall cost-effectiveness of the two strategies involving interventions, meaning their results are robust to any potential inaccuracies in the estimated costs of interventions, or for future increases.

As well as carrying out sensitivity analyses for intervention costs, the costs attached to T2DM were also assessed. Base case costs for the primary model were estimated as £133.98 for undiagnosed T2DM for the year prior to diagnosis and £21.68 for previous years, £2490 per annum for screen detected T2DM and £2756 per annum for clinically detected T2DM. Models were run with the costs attached to each of the three T2DM states increased by two, five and ten fold, and the results are presented in table 7.4.

As the majority of the total costs of each screening strategy are attributable to T2DM, increasing the costs of T2DM increased the total costs of each strategy by a similar factor. As the costs of all four strategies were increased, impacts on the comparisons of the three active strategies compared to no screening were less dramatic. Doubling T2DM costs resulted in an increase in the costs per QALY to £17,310 for T2DM only screening, £4,281 for screening and lifestyle interventions and £6,696 for screening and pharmacological interventions and increasing T2DM costs ten-fold resulted in costs per QALY of £86,240, £15,640 and £32,869 for each active strategy respectively. After doubling the cost of T2DM, both the intervention strategies still had a high probability of cost-effectiveness at the £20,000 willingness-to-pay per QALY threshold, whilst the cost-effectiveness of the T2DM only screening strategy looked uncertain at these costs. Increasing the costs five-fold introduced uncertainty into the cost-effectiveness of all three screening strategies, although the likelihood of costs reaching such levels, that is an estimated average annual cost of £13,780 for a clinically detected and £12,450 for screen detected diabetics, in clinical practice seems unlikely.

A series of sensitivity analyses were also carried out to assess the cost-effectiveness of each screening strategy over a number of time horizons, the results of which are in table 7.5. The results show that the three active strategies appear to become more cost-effective the longer

they are monitored. So at five years the cost per QALY gained was £112,400 for T2DM screening, £90,490 for IGT and T2DM screening followed by lifestyle interventions, and £93,320 for IGT and T2DM screening followed by pharmacological interventions. These costs were reduced steadily as the time horizon increased, until at a horizon of 50 years the costs were £8,681, £2,863 and £3,429 per QALY, for each of the three active screening strategies respectively. This is because the clinical impact of screening and/or intervening occurs in later years of the model, when T2DM is either delayed, or complications of T2DM are reduced through early diagnosis and treatment.

The benefits of any screening strategy are therefore likely to be more apparent in the long rather than the short term, which emphasises the need to monitor the long-term effects when assessing the benefits of any new screening strategy, particularly as it is the long-term effects for which there is currently the least evidence. Figure 7.7 shows the effect of increasing the time horizon of the model on the incremental cost-effectiveness ratios for the three active screening strategies compared to a strategy of no screening, and how the three active screening strategies appear more cost-effective the longer the follow-up.

Table 7.3: Effect of changing intervention costs on model outcomes

		No screening	Screening for T2DM only	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
Total cost (GBP)	Base case	17,290 (5,746, 39,580)	18,040 (7,083, 39,970)	17,910 (7,124, 39,740)	17,900 (7,061, 39,710)
	Costs x 2	17,290 (5,746, 39,580)	18,040 (7,083, 39,970)	18,210 (7,443, 40,000)	17,920 (7,077, 39,720)
	Costs x 5	17,290 (5,746, 39,580)	18,040 (7,083, 39,970)	19,100 (8,368, 40,750)	17,960 (7,128, 39,760)
	Costs x 10	17,290 (5,746, 39,580)	18,040 (7,083, 39,970)	20,580 (9,895, 42,010)	18,030 (7,206, 39,580)
Cost per case prevented (GBP)	Base case	-	-	62,810	105,000
	Costs x 2	-	-	93,310	107,800
	Costs x 5	-	-	184,800	115,900
	Costs x 10	-	-	337,400	129,300
Cost per QALY gained (GBP)	Base case	-	8,681	2,863	3,429
	Costs x 2	-	8,681	4,254	3,516
	Costs x 5	-	8,681	8,487	3,785
	Costs x 10	-	8,681	15,410	4,231
Probability cost effective at willingness-to-pay of £20,000 / £30,000 per QALY	Base case	-	0.68 / 0.76	0.99 / 1.00	0.95 / 0.97
	Costs x 2	-	0.68 / 0.76	0.98 / 0.99	0.95 / 0.97
	Costs x 5	-	0.68 / 0.76	0.94 / 0.98	0.94 / 0.97
	Costs x 10	-	0.68 / 0.76	0.73 / 0.92	0.94 / 0.97

***** Base case costs for the primary model were, for lifestyle £398 in the first year and £280 in subsequent years, and for pharmacological interventions £16.10 per annum

Table 7.4: Effect of changing T2DM costs on model outcomes

		No screening	Screening for T2DM only	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
Total cost (GBP)	Base case	17,290 (5,746, 39,580)	18,040 (7,083, 39,970)	17,910 (7,124, 39,740)	17,900 (7,061, 39,710)
	Costs x 2	34,530 (11,520, 79,110)	36,050 (14,170, 79,890)	35,540 (13,920, 79,180)	35,770 (14,100, 79,300)
	Costs x 5	86,310 (28,810, 197,900)	90,070 (35,430, 199,500)	88,440 (34,260, 197,300)	89,410 (35,200, 198,300)
	Costs x 10	172,600 (57,620, 395,500)	180,200 (70,860, 398,900)	176,700 (68,230, 394,600)	178,700 (70,350, 396,400)
Cost per case prevented (GBP)	Base case	-	-	62,810	105,000
	Costs x 2	-	-	94,520	206,900
	Costs x 5	-	-	145,300	512,300
	Costs x 10	-	-	345,900	1,020,000
Cost per QALY gained (GBP)	Base case	-	8,681	2,863	3,429
	Costs x 2	-	17,310	4,281	6,696
	Costs x 5	-	43,140	8,537	16,480
	Costs x 10	-	86,240	15,640	32,860
Probability cost effective at willingness-to-pay of £20,000 / £30,000 per QALY	Base case	-	0.68 / 0.76	0.99 / 1.00	0.95 / 0.97
	Costs x 2	-	0.46 / 0.60	0.93 / 0.97	0.85 / 0.92
	Costs x 5	-	0.15 / 0.27	0.77 / 0.87	0.56 / 0.71
	Costs x 10	-	0.07 / 0.11	0.57 / 0.70	0.33 / 0.46

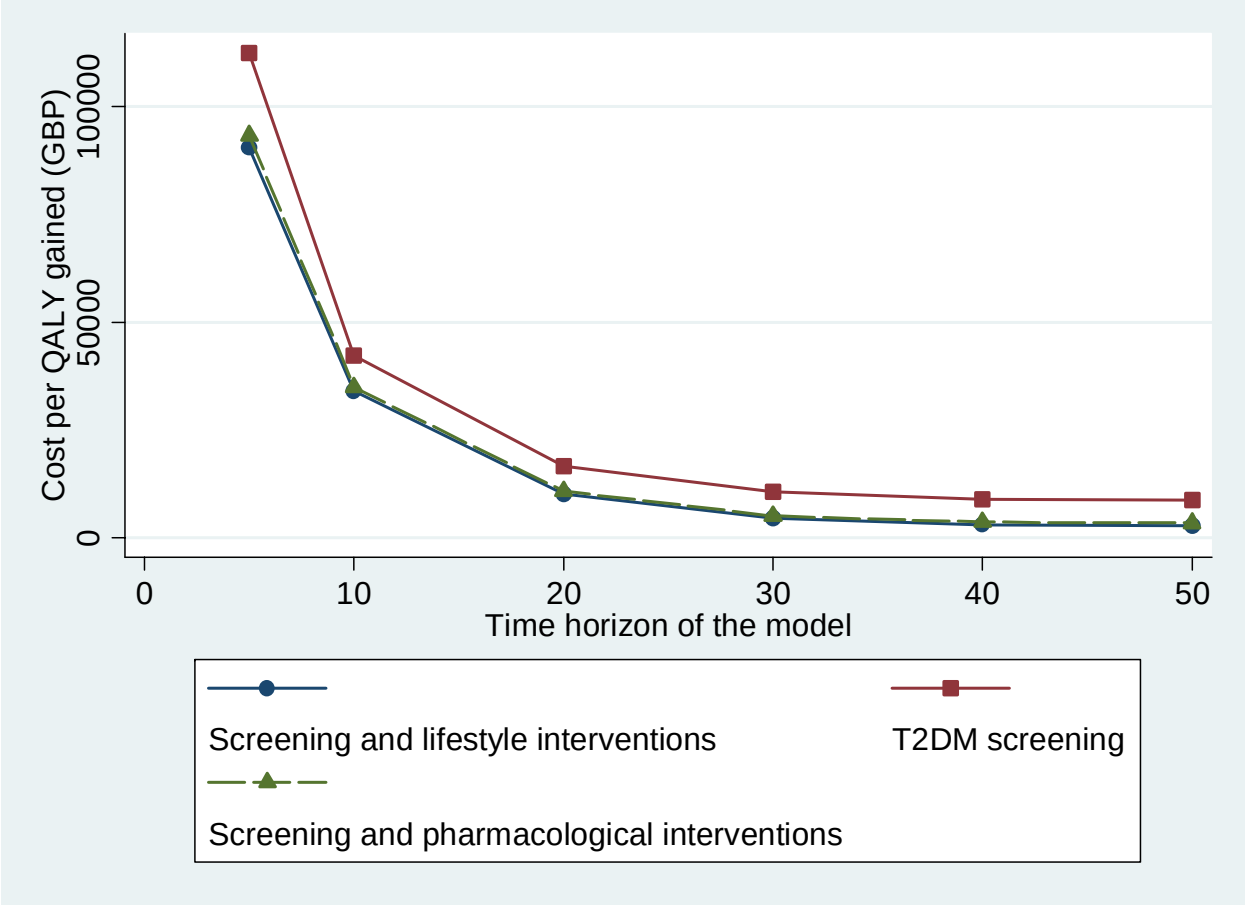
***** Base case costs for the primary model were for undiagnosed T2DM £133.98 for the year prior to diagnosis and £21.68 for previous years, and £2490 for screen detected T2DM and £2756 for clinically detected T2DM per annum

Table 7.5: Effect of changing the time horizon on model outcomes

	Horizon (years)	No screening	Screening for T2DM only	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
QALY	5	4.89 (4.80,4.91)	4.89 (4.81, 4.91)	4.89 (4.81, 4.92)	4.89 (4.80,4.92)
	10	9.61 (9.16, 9.71)	9.62 (9.17, 9.71)	9.63 (9.18, 9.72)	9.63 (9.18, 9.72)
	20	18.15 (16.46, 18.70)	18.17 (16.49, 18.71)	18.21 (16.55, 18.75)	18.20 (16.53, 18.74)
	30	24.48 (21.32, 26.04)	24.53 (21.39, 26.07)	24.62 (21.49, 26.16)	24.60 (21.46, 26.14)
	40	27.50 (23.24, 30.56)	27.56 (23.34, 30.61)	27.69 (23.48, 30.74)	27.65 (23.44, 30.70)
	50	28.06 (23.49, 32.01)	28.12 (23.58, 32.08)	28.26 (23.74, 32.23)	28.22 (23.69, 32.18)
Total cost (GBP)	5	249 (76, 895)	635 (514, 869)	730 (609, 952)	669 (545, 895)
	10	1,244 (405, 3,886)	1,768 (1,258, 4,198)	1868 (1,403, 4,264)	1,791 (1,309, 4,193)
	20	5,392 (1,869, 17,410)	6,091 (3,227, 17,760)	6,117 (3,367, 17,670)	6,053 (3,267, 17,630)
	30	11,730 (3,922, 31,510)	12,520 (5,404, 31,900)	12,450 (5,491, 31,670)	12,410 (5,411, 31,660)
	40	16,370 (5,483, 38,600)	17,140 (6,823, 38,880)	17,040 (6,864, 38,610)	17,040 (6,792, 38,640)
	50	17,290 (5,746, 39,580)	18,040 (7,083, 39,970)	17,910 (7,124, 39,740)	17,900 (7,061, 39,710)
Cost per QALY gained (GBP)	5	-	112,400	90,490	93,320
	10	-	42,190	34,050	34,980
	20	-	16,610	10,220	10,820
	30	-	10,610	4,467	5,043
	40	-	8,903	3,055	3,631
	50	-	8,681	2,863	3,429
Probability cost effective at willingness-to-pay of £20,000 / £30,000 per QALY	5	-	0.01 / 0.03	0.02 / 0.04	0.02 / 0.05
	10	-	0.08 / 0.15	0.20 / 0.40	0.20 / 0.36
	20	-	0.33 / 0.45	0.76 / 0.86	0.67 / 0.77
	30	-	0.54 / 0.63	0.93 / 0.97	0.86 / 0.91
	40	-	0.65 / 0.74	0.98 / 0.99	0.93 / 0.96
	50	-	0.68 / 0.76	0.99 / 1.00	0.95 / 0.97

* the base case model was run for a time horizon of 50 years

Figure 7.7: The effect of increasing the time horizon of the model on the incremental cost-effectiveness ratios for the three active screening strategies compared to a strategy of no screening



7.6 Clinical sensitivities

A number of sensitivity analyses were carried out on the clinical parameters within the decision model. Firstly the test sensitivities of the screening tests were changed for both T2DM and the test for IGT and T2DM combined. Test specificities were not assessed by sensitivity analyses as poor test specificity would result in a greater number of confirmatory OGTT tests being carried out, but as this test is relatively cheap (£1.30 per test) the impact on the total costs of each strategy would be minimal. Secondly the prevalence of both IGT and T2DM were increased in unison to assess the predicted results of each screening strategy if populations with different risks of glucose intolerance were targeted.

Lastly two models were run with either a population who all suffered from T2DM, and then a population who all had IGT. The models were run with screening tests with 100% sensitivity, and therefore the results of the model were actually showing the predicted effects of either identifying T2DM early, compared to the strategy of no screening whereby they would just be identified clinically, or the effect of the interventions compared to no interventions in individuals with IGT. As with the economic sensitivity analyses, sensitivity analyses were carried out with all other model parameters remaining constant and at the same levels as the base case scenario.

7.6.1 Test sensitivity

The test sensitivities utilised in the base case model were taken from the STAR study. They were for the fasting plasma glucose test (FPG) which was shown to have a sensitivity of 86% when testing for T2DM alone, and 59% when testing for both IGT and T2DM in combination. Due to the poor sensitivity when testing for both IGT and T2DM it was important to understand the effect of the present unavailability of an efficient test, and the impact any improvement in the screening test for IGT and T2DM would have. The sensitivities of both test, T2DM alone and IGT and T2DM in combination, were increased in unison encompassing a range from just 20% to 100%. Test specificities were kept constant, as sensitivity was changed.

Results of the sensitivity analyses are presented in table 7.6. Comparing two scenarios with either just 20% test sensitivities to that of 100% test sensitivities, all the estimated costs per QALY were higher when the test sensitivity was lower. So for T2DM only screening the cost

per QALY was £7,784 at a test sensitivity of 20% and £7,449 at a test sensitivity of 100%. Changing test sensitivity therefore had little impact on the comparison of T2DM screening, compared to no screening, probably because only 5% of the screened population had T2DM at the time of screening. Also, as test sensitivity was changed, the change in the positive predictive values of a test, which were used in the decision tree, were less dramatic. For example for T2DM screening, the PPV was 19% at a screening test sensitivity of 20% and 44% at 100% test sensitivity. For IGT and T2DM screening with lifestyle interventions the costs were £3,498 at 20% and £1,935 at 100%, and for IGT and T2DM screening with lifestyle interventions the costs were £4,120 at 20% and £2,351 at 100%. It can be seen by looking at the QALYs for these strategies, that reducing the test sensitivity decreases the numbers with IGT that receive interventions, and leads to a reduction in QALYs. Costs show less of a pattern as whilst intervention costs decrease as test sensitivity decreases, costs associated with T2DM will increase. The overall conclusions of the cost-effectiveness of each of the three active screening strategies compared to no screening, were not altered when test sensitivity was changed, with the probabilities of cost-effectiveness remaining relatively constant. Test sensitivity would have more of an impact though, in populations where the prevalence of IGT and T2DM were higher than in this base case scenario, or when repeat screenings of a population were undertaken.

7.6.2 Prevalence of glucose tolerance status

The base case model was run for prevalences of 83% for NGT, 12% for IGT and 5% for T2DM. These prevalences were taken from the STAR study, which was an analysis of individuals with at least one risk factor for T2DM, and therefore represents prevalences of a population with a slightly increased risk of T2DM. To allow for a comparison between the screening for T2DM only strategy and the strategies of screening for both IGT and T2DM followed by either lifestyle and pharmacological interventions, it made sense to increase both the prevalences of IGT and T2DM in unison, keeping the proportion between the two states roughly constant. Models were therefore run with the base case prevalences of 83%, 12% and 5% for NGT, IGT and T2DM respectively and then 70%, 20% and 10%; 40%, 40% and 20%; and finally 10%, 60% and 30%. Increasing the prevalences of IGT and T2DM, represent the affect of targeting populations for screening with, a greater the risk of developing T2DM. Results of the models are presented in table 7.7.

As would be expected lifetime risk of T2DM increased as prevalence of both IGT and T2DM increased. The total costs of each of the four strategies increased fairly dramatically as

prevalence of IGT and T2DM increased, this is because the majority of costs incurred by each strategy are due to the costs of treating and monitoring patients with T2DM. QALYs, on the other hand, decreased as prevalence of IGT and T2DM increased, which is unsurprising as the T2DM state was modelled with a decreased utility value and an increased risk of death.

As all four strategies increased in costs and decreased in QALYs, when comparing the three active strategies with no screening, increasing the prevalence of glucose intolerance in the screened population had little impact, and both the estimated costs per QALY gained, and the probability of cost-effectiveness, remained fairly constant. For example for T2DM screening the cost per QALY varied between £8,451 and £8,681 as prevalences were changed, for screening followed by lifestyle interventions cost varied between £2,863 and £3,201, and for screening followed by pharmacological interventions costs varied between £3,429 and £3,809 per QALY. Therefore although targeting populations with different risks of glucose intolerance will affect the overall costs of different screening strategies, the decision as to which strategy is the best approach for screening and/or intervening will not alter.

Table 7.6: Results of models run for different screening test sensitivities

		No screening	Screening for T2DM only	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
QALY	100%	28.05 (23.47, 32.01)	28.10 (23.54, 32.07)	28.47 (23.93, 32.43)	28.36 (23.81, 32.34)
	80%	28.09 (23.47, 32.02)	28.13 (23.53, 32.07)	28.42 (23.86, 32.36)	28.34 (23.77, 32.28)
	60%	28.09 (23.47, 32.02)	28.13 (23.51, 32.05)	28.36 (23.75, 32.27)	28.29 (23.70, 32.21)
	40%	28.00 (23.47, 32.04)	28.02 (23.50, 32.06)	28.18 (23.68, 32.21)	28.13 (23.56, 32.12)
	20%	28.04 (23.46, 32.04)	28.05 (23.47, 32.04)	28.15 (23.58, 32.14)	28.00 (23.47, 32.04)
Total cost (GBP)	100%	17,580 (5,730, 39,470)	18,120 (6,753, 39,620)	18,520 (7,567, 39,400)	18,450 (7,453, 39,420)
	80%	17,130 (5,551, 39,650)	17,550 (6,475, 39,820)	17,900 (7,356, 39,680)	17,890 (7,235, 39,770)
	60%	17,120 (5,585, 39,530)	17,450 (6,282, 39,600)	17,760 (7,139, 39,600)	17,740 (7,045,39,660)
	40%	17,460 (5,670, 39,420)	17,690 (6,123, 39,490)	18,000 (6,871, 39,489)	17,980 (6,805, 39,510)
	20%	17,290 (5,685, 39,520)	17,390 (5,893, 39,950)	17,690 (6,558, 39,610)	17,660 (6,510, 39,640)
Cost per QALY gained (GBP)	100%	-	7,449	1,935	2,351
	80%	-	7,582	2,119	2,624
	60%	-	7,611	2,302	2,781
	40%	-	7,488	2,617	3,132
	20%	-	7,784	3,498	4,120
Probability cost effective at willingness- to-pay of £20,000 / £30,000 per QALY	100%	-	0.69 / 0.76	1.00 / 1.00	0.98 / 0.99
	80%	-	0.68/ 0.75	1.00 / 1.00	0.97 / 0/99
	60%	-	0.68/ 0.76	0.99 / 1.00	0.97 / 0.99
	40%	-	0.68 / 0.75	0.99 / 1.00	0.95 / 0.98
	20%	-	0.68 / 0.76	0.97 / 0.98	0.92 / 0.95

* the base case model had screening test sensitivities of 86% for T2DM and 59% for IGT and T2DM combined

Table 7.7: Results of models run for different prevalences of glucose tolerance status

	Prevalence (NGT/IGT/ T2DM)	No screening	Screening for T2DM only	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
Lifetime risk of T2DM (%)	83/12/5	68.4 (18.0, 91.8)	68.4 (18.0, 91.8)	67.5 (17.1, 91.1)	67.9 (17.4, 91.4)
	70/20/10	71.9 (29.5, 92.4)	71.9 (29.5, 92.4)	70.3 (27.8, 91.2)	70.9 (28.5, 91.8)
	40/40/20	80.1 (55.2, 93.6)	80.1 (55.2, 93.6)	76.8 (51.5, 91.2)	78.2 (53.1, 92.2)
	10/60/30	87.8 (77.6, 94.9)	87.8 (77.6, 94.9)	82.8 (71.5, 91.7)	84.9 (73.6, 93.3)
QALY	83/12/5	28.06 (23.49, 32.01)	28.12 (23.58, 32.08)	28.26 (23.74, 32.23)	28.22 (23.69, 32.18)
	70/20/10	28.26 (24.72, 31.18)	28.26 (24.79, 31.14)	28.47 (25.02, 31.34)	28.41 (24.96, 31.29)
	40/40/20	25.44 (22.58, 27.86)	25.71 (22.95, 28.10)	26.21 (23.46, 28.57)	26.05 (23.31, 28.43)
	10/60/30	23.75 (21.82, 25.58)	24.16 (22.40, 25.85)	24.91 (23.15, 26.55)	24.67 (22.89, 26.35)
Total cost (GBP)	83/12/5	17,290 (5,746, 39,580)	18,040 (7,083, 39,970)	17,910 (7,124, 39,740)	17,900 (7,061, 39,710)
	70/20/10	21,320 (9,132, 41,270)	22,780 (12,470, 41,840)	22,620 (12,650, 41,370)	22,560 (12,540, 41,420)
	40/40/20	30,200 (15,030, 44,890)	32,930 (23,140, 46,060)	32,530 (23,840, 45,190)	32,460 (23,540, 45,210)
	10/60/30	38,440 (19,740, 49,690)	42,580 (32,660, 51,190)	41,980 (33,990, 49,980)	41,830 (33,530, 50,090)
Cost per QALY gained (GBP)	83/12/5	-	8,681	2,863	3,429
	70/20/10	-	8,617	3,203	3,809
	40/40/20	-	8,451	3,161	3,791
	10/60/30	-	8,464	3,148	3,781
Probability cost effective at willingness-to-pay of £20,000 / £30,000 per QALY	83/12/5	-	0.68 / 0.76	0.99 / 1.00	0.95 / 0.97
	70/20/10	-	0.68 / 0.76	0.98 / 0.99	0.93 / 0.96
	40/40/20	-	0.69 / 0.77	0.98 / 0.99	0.93 / 0.96
	10/60/30	-	0.68 / 0.76	0.98 / 0.99	0.93 / 0.96

* the base case model had prevalences of 83% for NGT, 12% for IGT and 5% for T2DM

7.6.3 Multi-way sensitivity analyses

The sensitivity analyses discussed so far have assessed the impact of changing model parameters individually, on model results. It is possible however for multi-way sensitivity analyses to be carried out, whereby a number of model inputs are varied simultaneously. To demonstrate this, a sensitivity analysis was performed whereby costs of T2DM, that is the costs attributed to the undiagnosed, clinically diagnosed and screen diagnosed T2DM states, were doubled, and prevalence of glucose intolerance was raised (NGT=40%, IGT=40% and T2DM=20%). The results are displayed in Table 7.8.

From the results it can be seen that raising the prevalence of glucose intolerance impacts on the clinical outcomes of lifetime risk of T2DM and QALYs, and unsurprisingly raising the costs attributable to T2DM does not. Both increasing prevalence and increasing costs of T2DM, affect the total costs of each strategy. Increasing these two parameters in unison, leads to a greater difference in costs between the three active screening strategies and no screening. This is because, as the prevalence of T2DM is increased, the numbers undiagnosed with T2DM in the no screening strategy increases. Undiagnosed T2DM costs either £133.98 or £21.68 per annum in the base case scenario, whilst diagnosed T2DM costs £2490 or £2756 per annum. Increasing these costs, whilst increasing the prevalence of T2DM, increases the differences in total costs between strategies, by adding more benefit to T2DM remaining undiagnosed.

Due to an increased difference in total costs between the three active screening strategies and no screening, the incremental costs are raised and the probabilities of cost-effectiveness are reduced dramatically. This example highlights the importance of multi-way sensitivity analyses, as whilst model parameters may have minimal impact when changed individually, changing them in unison may have a multiplicative rather than additive effect. Although multi-way sensitivity analyses are an important consideration, they have to be well planned. For example if sensitivity analyses are carried out on five parameters, each with three possible values, the total number of different possible scenarios for multi-way sensitivity analyses is 3^5=243.

Table 7.8: Results of models run for both high prevalences of glucose tolerance status, and double T2DM costs

	Prevalence (NGT/IGT/ T2DM)	No screening	Screening for T2DM only	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
Lifetime risk of T2DM (%)	Base case	68.4 (18.0, 91.8)	68.4 (18.0, 91.8)	67.5 (17.1, 91.1)	67.9 (17.4, 91.4)
	Prev (40,40,20)	80.1 (55.2, 93.6)	80.1 (55.2, 93.6)	76.8 (51.5, 91.2)	78.2 (53.1, 92.2)
	T2DM costs x 2	68.4 (18.0, 91.8)	68.4 (18.0, 91.8)	67.5 (17.1, 91.1)	67.9 (17.4, 91.4)
	Prev (40,40,20) and costs x 2	80.1 (55.2, 93.6)	80.1 (55.2, 93.6)	76.8 (51.5, 91.2)	78.2 (53.1, 92.2)
QALY	Base case	28.06 (23.49, 32.01)	28.12 (23.58, 32.08)	28.26 (23.74, 32.23)	28.22 (23.69, 32.18)
	Prev (40,40,20)	25.44 (22.58, 27.86)	25.71 (22.95, 28.10)	26.21 (23.46, 28.57)	26.05 (23.31, 28.43)
	T2DM costs x 2	28.06 (23.49, 32.01)	28.12 (23.58, 32.08)	28.26 (23.74, 32.23)	28.22 (23.69, 32.18)
	Prev (40,40,20) and costs x 2	25.44 (22.58, 27.86)	25.71 (22.95, 28.10)	26.21 (23.46, 28.57)	26.05 (23.31, 28.43)
Total cost (GBP)	Base case	17,290 (5,746, 39,580)	18,040 (7,083, 39,970)	17,910 (7,124, 39,740)	17,900 (7,061, 39,710)
	Prev (40,40,20)	30,200 (15,030, 44,890)	32,930 (23,140, 46,060)	32,530 (23,840, 45,190)	32,460 (23,540, 45,210)
	T2DM costs x 2	34,530 (11,520, 79,110)	36,050 (14,170, 79,890)	35,540 (13,920, 79,180)	35,770 (14,100, 79,300)
	Prev (40,40,20) and costs x 2	43,600 (24,360, 62,310)	54,000 (40,830, 71,190)	54,461 (42,390,71,550)	55,400 (42,840, 72,480)
Cost per QALY gained (GBP)	Base case	-	8,681	2,863	3,429
	Prev (40,40,20)	-	8,451	3,161	3,791
	T2DM costs x 2	-	17,310	4,281	6,696
	Prev (40,40,20) and costs x 2	-	38,600	19,100	25,790
Probability cost effective at willingness-to- pay of £20,000 / £30,000	Base case	-	0.68 / 0.76	0.99 / 1.00	0.95 / 0.97
	Prev (40,40,20)	-	0.69 / 0.77	0.98 / 0.99	0.93 / 0.96
	T2DM costs x 2	-	0.46 / 0.60	0.93 / 0.97	0.85 / 0.92
	Prev (40,40,20) and costs x 2		0.12 / 0.24	0.44 / 0.77	0.24 / 0.55

7.6.4 Models with populations of either 100% T2DM or 100% IGT

To remove any influence of screening test efficiency or population prevalences, and to assess solely the model estimated effects for either the interventions or the effect of identifying individuals with T2DM early, two further models were run. To assess early identification of T2DM a model was run with a population comprised solely of individuals with T2DM, and the screening test sensitivity was set at 100%, so that for the strategy of screening for T2DM only all individuals were diagnosed in the first year and entered the T2DM screened diagnosed state, followed by death as the model iterations progressed. This strategy was compared to no screening, whereby all individuals started in the undiagnosed T2DM state and gradually moved to the clinically diagnosed state or death over the course of the model. The results of this model are presented in table 7.9. It can be seen from the results that the model predicts a gain in over one QALY if individuals are screened and diagnosed with T2DM early. The estimated costs per QALY is £7,333 and the probability of cost-effectiveness is 0.78 at the £30,000 willingness-to-pay threshold. The screening strategy therefore appears more cost-effective in this model compared to the base case, probably because clinical effects are not averaged across a whole population, some of which have a healthy glucose tolerance, and no individuals are screened unnecessarily.

The results of the model whereby a population with 100% IGT was simulated are presented in table 7.10. It can be seen that the model predicts an increase in QALYs if interventions are implemented, so for the strategy where no screening, and hence no interventions were administered the QALYs were estimated as 24.25 (95% CrI: 22.64 to 26.92), for lifestyle interventions 27.89 (95% CrI: 25.68 to 29.80) and for pharmacological interventions 27.01 (95% CrI: 24.77 to 29.16). Lifetime risk of T2DM and years spent free from diabetes were also improved in both the intervention strategies, and the cost per QALY was just £900 for lifestyle interventions and £964 for pharmacological interventions, as although pharmacological interventions are cheaper their clinical effect was smaller than lifestyle. As with the previous model the active screening strategies showed a much greater impact on clinical outcomes than in the base case model, primarily because the results are average effects for a population with IGT, rather than averaged across a population where just 12% had IGT and the majority were healthy.

Table 7.9: Predicted results if a population with a 100% prevalence of T2DM is targeted

	No screening	Screening for T2DM only
Lifetime risk of T2DM (%)	100	100
Years diabetes free	0	0
QALY	20.02 (18.09, 21.94)	21.19 (19.55, 22.93)
Total cost (GBP)	58,990 (31,290, 70,970)	69,510 (64,660, 74,560)
Cost per QALY gained (GBP)	-	7,333
Probability cost effective at willingness-to-pay of £20,000 / £30,000 per QALY	-	0.71 / 0.78

Table 7.10: Predicted results if a population with a 100% prevalence of IGT is targeted

	No screening/ interventions	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
Lifetime risk of T2DM (%)	86.4 (73.0, 94.3)	68.5 (50.6, 84.4)	76.2 (57.5, 89.5)
Years diabetes free	12.67 (8.52, 17.42)	18.74 (13.52, 23.40)	16.42 (11.26, 21,71)
QALY	24.75 (22.64, 26.91)	27.89 (25.68, 29.80)	27.01 (24.77, 29.16)
Total cost (GBP)	31,960 (14,890, 44,400)	34,820 (26,790, 44,500)	34,350 (23,680, 45,310)
Cost per QALY gained (GBP)	-	900	964
Probability cost effective at willingness-to-pay of £20,000 / £30,000 per QALY	-	0.999 / 1.000	0.995 / 0.998

7.7 Chapter summary

The decision model developed for this thesis was thoroughly assessed using a range of diagnostic checks, to ensure the model had both sound assumptions and that convergence of model parameters had occurred. Results all indicated that the model was converging well, and although correlation between simulations may be an issue for a few of the model parameters, a large number of iterations were run before collating results, so that results should effectively be drawn from a random sample of simulations.

A number of both economic and clinical sensitivity analyses were carried out to enable a better understanding of the model. It was found that costs of either interventions or T2DM, prevalence of glucose intolerance or screening test sensitivity did not effect model conclusions. Both the time horizon of the model and discounting costs and benefits, did effect the cost-effectiveness of different screening strategies and therefore would impact on the choice of the most cost-effective approach to screening. Considering the impact of screening over a longer term (i.e. a longer time horizon) provided more evidence in support of a screening programme, than if only short-term benefits were considered. Unfortunately the long-term effects of screening programmes are still unclear and more trial evidence is needed to strengthen the model results, particularly for the assessment of screening for T2DM alone. As the model results were fairly robust to most changes in data inputs, this indicates that the results should be robust to any inaccuracies in model estimates. The multi-way sensitivity analyses carried out for prevalence and costs of T2DM, did show a greater impact on model results. This highlights the importance of multi-way sensitivity analyses, and more would have been carried out if time had allowed. For this model further multi-way analyses that would be of particular interest, would be the effect of changing either test sensitivities or costs of interventions, with prevalence of glucose intolerance, as it would be interesting to assess the effect of changing important model parameters for different ‘at risk’ populations that may be targeted by a screening programme. When considering a possible screening strategy, it would be useful to carry out more sensitivity analyses on the population of interest, to investigate the estimated effect of model parameters for that particular population.

8. MODEL EXTENSIONS

8.1 Chapter Overview

This chapter describes a number of additional extensions to the model that were carried out to demonstrate the flexibility of the framework, as well as to assess further options that may need to be considered when implementing a screening/treatment strategy. Further extensions included utilising the mixed treatment comparison analysis of interventions, rather than the meta-analyses as used in the primary model, assessing the effect of screening and intervention strategies on an Asian and an ethnically mixed population, considering the effects of compliance to both screening and interventions, assessing the effects of multiple screenings rather than just a one-off screening, the inclusion of additional information on screening tests, and assessing the value of obtaining perfect information for the model. The methods and results of each of the extensions to the primary model will now be described.

8.2 Utilising the mixed treatment comparison analysis in the decision model

In Chapter 4 a systematic review and meta-analysis was carried out incorporating all relevant trials that had investigated the effectiveness of interventions for the prevention or delay of T2DM. The trials had been analysed using conventional Bayesian meta-analysis methods for lifestyle and pharmacological interventions separately, and using a mixed treatment comparison analysis framework, whereby data from all trials was included within a single model rather than a series of pair-wise comparisons. As described in Chapter 4 the data from factorial trials had to be treated differently by the two methods, in that in the mixed treatment comparison an additional group who received both lifestyle and pharmacological interventions was created, which allowed for the assessment of any interaction between pharmacological and lifestyle treatments.

In table 8.1 the results of both the individual meta-analyses and the mixed treatment comparison analyses are presented. Although the estimated effects of lifestyle and pharmacological interventions are similar between the two methods, they vary slightly as some of the data incorporated in the individual analyses was utilised in the separate category of both lifestyle and pharmacological interventions in the mixed treatment comparison. As no interaction effect was found between lifestyle and pharmacological interventions as discussed in Chapter 4, utilising

the data from the separate meta-analyses appears to make best use of the available data, so this was the approach taken for the primary model. To demonstrate the flexibility of the decision model framework the first extension to the model undertaken was to incorporate the mixed treatment comparison analyses within the decision model, and to take the estimated pooled effect sizes of both lifestyle and pharmacological interventions from this analysis to predict the effectiveness of each of the screening and intervention strategies.

Table 8.1: Intervention effects estimated by different methods

	Individual meta-analysis	Mixed treatment comparison analysis
Lifestyle Interventions	0.53 (0.43, 0.64)	0.63 (0.48, 0.85)
Anti-diabetic interventions	0.66 (0.47, .83)	0.67 (0.47, 0.89)
Both interventions	-	0.57 (0.31, 0.99)

Table 8.2: Results of the decision model using the different intervention analyses

	Model of intervention effects	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
QALY	Meta-analysis	28.26 (23.74, 32.23)	28.22 (23.69, 32.18)
	MTC	28.23 (23.68, 32.21)	28.22 (23.67, 32.20)
Total cost (GBP)	Meta-analysis	17,910 (7,124, 39,740)	17,900 (7,061, 39,710)
	MTC	18,190 (7,265, 39,760)	18,010 (7,039, 39,630)
Cost per QALY gained (GBP)	Meta-analysis	2,863	3,429
	MTC	4,385	3,303
Probability cost effective at willingness-to-pay per QALY of £20,000/£30,000	Meta-analysis	0.99 / 1.00	0.95 / 0.97
	MTC	0.94 / 0.97	0.94 / 0.97

The assumptions of the mixed treatment comparison model were discussed in more detail in Chapter 4. Due to the added assumptions, it is better in practice, to only use MTC methodologies where there is insufficient direct evidence (National Institute for Clinical Excellence, 2004). The MTC analysis differs from carrying out individual meta-analyses, in that a common τ is assumed across all studies, whereas each individual meta-analysis would have a separate τ. For the individual meta-analyses, τ was estimated as 0.20 (95% CrI:0.01 to 0.66) for anti-diabetic interventions, and 0.16 (0.02 to 0.44) for lifestyle interventions. For the MTC analysis the estimate of τ was slightly larger; 0.32 (0.11 to 0.62), but as both lifestyle and pharmacological trials were combined in this analysis, then the between study standard deviation could probably be expected to be greater. As expected the results of the decision model are similar as to when the individual meta-analyses were utilised within the model,

although cost-effectiveness is slightly reduced due to a slight reduction in the estimated intervention effects.

8.3 Modelling different ethnic cohorts

8.3.1 Adapting the model for different ethnicities

For the base case model prevalences of NGT, IGT and T2DM were taken from the Caucasian participants in the STAR study, and were as follows NGT = 83.3%, IGT = 11.9% and T2DM = 4.7%. As it is suspected that rate of transition from the IGT state to T2DM varies by ethnicity, with some ethnic groups having a higher risk of T2DM, the transition rate between IGT to T2DM was estimated using data derived from studies on populations where the majority were Caucasian, for the primary model. The purpose of the decision model though was to assess the cost-effectiveness of IGT and T2DM screening and intervention policies on a U.K. population, it was therefore important to consider other ethnic groups where data was available. The STAR study had assessed both Caucasians and Southern Asians, as it was based in Leicestershire where the population has a large component that are of South Asian decent. From the STAR data the prevalence of each glucose tolerance state could be estimated for an Asian population, for use in the decision model, these were estimated as NGT = 77.9%, IGT = 15.6% and T2DM = 6.5%, which show a higher risk of T2DM compared to Caucasians, where the prevalences were 83.3%, 11.9% and 4.7% respectively. Only one study reported a transition rate from IGT toT2DM for an Asian population and this was the Indian Diabetes Prevention Program (IDDP). From the control group monitored in this intervention trial, the estimated transition rate from IGT to T2DM was 26.36 (95% CI: 21.01 to 33.06) cases per 100 person years, which is higher than the estimated rate for Caucasians used in the primary model of 7.07 (95% CI: 4.31 to 11.59). It would have been more appropriate to use U.K. based estimates of intervention effects on an Asian population, as lifestyles here may be very different to those in India, but unfortunately this was not available.

Using the STAR data, diagnostic tables for an FPG screening test could be constructed for both an Asian and a Caucasian population (table 8.3), and this in turn could be used to predict the efficiency of FPG as a screening test for the two ethnic groups. From table 8.4 it can be seen that the FPG test performed better in Caucasians than Southern Asians, with the test sensitivity, specificity and area under the ROC curve all being higher in the Caucasian population. Using all the STAR data on both Caucasians and Asians a diagnostic table could also be constructed for a

population of mixed ethnic origin, which in the case of the STAR study was 29% Southern Asian and 71% Caucasian. This is reflective of the City of Leicester population as a whole, where in the 2001 census 29.92% of people classed themselves as belonging to Indian, Pakistani or Bangladeshi ethnic groups.

Table 8.3: Results of the FPG test in the STAR study by ethnic group

	True status as confirmed by OGTT			
Caucasians				
FPG result (mmol/l)	NGT	IGT	T2DM	Totals
<5.7	1626	142	8	1776
>=5.7 and <6.0	138	51	7	196
>=6.0	83	71	90	244
Totals	1847	264	105	2216
South Asians				
FPG result (mmol/l)	NGT	IGT	T2DM	Totals
<5.7	596	71	8	675
>=5.7 and <6.0	62	25	3	90
>=6.0	35	43	47	125
Totals	693	139	58	890
Mixed				
FPG result (mmol/l)	NGT	IGT	T2DM	Totals
<5.7	2222	213	16	1744
>=5.7 and <6.0	200	76	10	206
>=6.0	118	114	137	266
Totals	2540	403	163	3106

Assessing a population of mixed ethnic origin, whereby 29% of the population are assumed Southern Asian and 71% Caucasian, could be approached in two ways using the available data. Firstly a model could be run using the 3x3 diagnostic table of all the Caucasian and Southern Asian participants of the STAR study, and using a transition probability from IGT to T2DM that is a combination of the two transition rates for the two ethnic groups, weighted by their proportions. Therefore as the transition rate for Caucasians is 7.3 (95% CrI: 4.3 to 11.6) cases per 100 person years and for Asians 26.4 (21.0 to 32.9), the transition rate for a mixed population, weighted to comprise of 29% Southern Asians and 71% Caucasians, is 12.8 (10.2 to 16.3). Using the diagnostic data of the FPG test, STAR data for prevalences and the averaged transition rate, the model could then be run for a mixed population, whereby the mixing of data takes place at the start of the model.

Table 8.4: Efficiency of the FPG screening test by ethnic group

	T2DM only (>=6.0 mmol/l)	IGT and T2DM (>=5.7 mmol/l)
Sensitivity		
Caucasians	90/105 (85.7%)	219/369 (59.4%)
Asians	47/58 (81.0%)	118/197 (59.8%)
Mixed	137/163 (84.0%)	337/566 (59.5%)
Specificity		
Caucasians	1957/2111 (92.7%)	1626/1847 (88.0%)
Asians	754/832 (90.6%)	596/693 (86.0%)
Mixed	2711/2943 (92.1%)	2222/2540 (87.5%)
Area under ROC curve (95% CI)		
Caucasians	0.95 (0.92, 0.98)	0.85 (0.82, 0.87)
Asians	0.93 (0.89, 0.97)	0.80 (0.77, 0.84)
Mixed	0.94 (0.92, 0.97)	0.80 (0.78, 0.82)

Alternatively the data could be combined at the end of the model. To achieve this the model was run for both a Caucasian population and an Asian population individually and the results from 20,000 simulations of the model were monitored and saved for several outcomes of interest. The data from these simulations was then transferred to Stata, whereby the results where merged for every simulation, to represent a mixed population for a number of outcomes of interest, which included lifetime risk of T2DM, years T2DM free, QALYs and total cost for each strategy. They were merged using equation 8.1, where M_i represents the mixed, C_i Caucasian and A_i Southern Asian estimate for every simulation (i).

$$M_i = (C_i \text{ x } 0.71) + (A_i \text{ x } 0.29) \qquad \text{[Equation 8.1]}$$

Using the merged variables representing total cost of each strategy and total QALYs the incremental cost effectiveness ratios and the cost-effectiveness acceptability curves could then be computed.

8.3.2 Results for different ethnic cohorts

The results from the four ethnic models, which were 100% Caucasian, 100% Southern Asian, mixed with data combined at the start of the model and mixed where data is combined at the end of the model simulations, are displayed in table 8.5. Firstly if Caucasian and Asian results are compared, it can be seen that estimated lifetime risk and costs were higher in Asians, whilst

QALYs and years T2DM free were lower. This is unsurprising as the prevalence of IGT and T2DM was slightly higher in the Asian population, as was the transition rate from IGT to T2DM.

When the data was mixed the estimated outcomes fell in-between the 100% Caucasian and 100% Asian cohorts. Whether the data for the two ethnicities were combined at the start or the end of the model simulations did affect the results. Where data were mixed at the start of the model the lifetime risk of T2DM was higher, which in turn affected the outcomes of years T2DM free, QALYs and costs. When investigated further it was found that combining the data at the end of the model simulations predicted an average time spent in the undiagnosed IGT state of 5.84 years, but in the model where data was combined at the start estimated time spent with undiagnosed IGT was 5.76 years.

It therefore appears that combining data at the start leads to slight underestimation of time spent with IGT and overestimation of time spent with T2DM. This could be due to the effect of applying the weighted combined transition rate from IGT to T2DM, to a population with mixed risk of T2DM. Individuals of Asian ethnicity may be expected to leave the IGT state earlier than Caucasians due to their increased risk, meaning that in later model simulations the proportion of Caucasians in the IGT state has increased, and applying the combined transition rate overestimates the transition. A similar phenomenon can be induced in survival models, whereby populations with varying 'frailties' need to be considered carefully when deciding on the appropriate model (Hougaard, 1995).

Table 8.5: Clinical and cost-effectiveness of each screening strategy for different ethnic cohorts

	Ethnicity	No screening	Screening for T2DM only	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
Lifetime risk of T2DM (%)	Caucasian	68.4 (18.0, 91.8)	68.4 (18.0, 91.8)	67.5 (17.1, 91.1)	67.9 (17.4, 91.4)
	S. Asian	86.8 (26.4, 98.4)	86.8 (26.4, 98.4)	86.6 (26.1, 98.3)	86.7 (26.2, 98.4)
	Mixed- *start*	78.8 (21.4, 95.9)	78.8 (21.4, 95.9)	78.3 (20.8, 95.4)	78.6 (21.0, 95.7)
	Mixed- *end*	71.6 (28.9, 90.6)	71.6 (28.9, 90.6)	70.8 (27.4, 90.6)	71.2 (27.7, 90.2)
Years diabetes free	Caucasian	20.85 (10.36, 29.45)	20.85 (10.36, 29.45)	21.17 (10.66, 29.79)	21.07 (10.54, 29.66)
	S. Asian	14.01 (4.36, 26.47)	14.01 (4.36, 26.47)	14.24 (4,58, 26.72)	14.15 (4.48, 26.61)
	Mixed- *start*	17.68 (7.52, 28.27)	17.68 (7.52, 28.27)	18.01 (7.82, 28.59)	17.88 (7.68, 28.46)
	Mixed- *end*	18.41 (10.55, 25.75)	18.41 (10.55, 90.58)	18.71 (10.85, 26.04)	18.59 (10.73, 25.92)
QALY	Caucasian	28.06 (23.49, 32.01)	28.12 (23.58, 32.08)	28.26 (23.74, 32.23)	28.22 (23.69, 32.18)
	S. Asian	25.24 (20.65, 30.79)	25.35 (20.83, 30.91)	25.47 (20.96, 31.02)	25.43 (20.92, 30.98)
	Mixed- *start*	26.70 (20.04, 31.53)	26.78 (22.16, 31.60)	26.92 (22.32, 31.74)	26.88 (22.27, 31.70)
	Mixed- *end*	27.10 (23.79, 30.31)	27.18 (23.88, 30.39)	27.32 (24.02, 30.53)	27.27 (23.99, 30.53)
Total cost (GBP)	Caucasian	17,290 (5,746, 39,580)	18,040 (7,083, 39,970)	17,910 (7,124, 39,740)	17,900 (7,061, 39,710)
	S. Asian	28,250 (10,170, 55,120)	29,390 (12,270, 55,490)	29,420 (12,500, 55,220)	29,480 (12,550, 55,270)
	Mixed- *start*	22,180 (7,359, 47,310)	23,080 (9,147, 47,570)	22,940 (9,176, 47,210)	22,990 (9,196, 47,300)
	Mixed- *end*	22,145 (8,345, 41,657)	23,051 (9,820, 42,131)	22,973 (9,809, 41,962)	22,976 (11,885, 42,006)
Cost per QALY gained (GBP)	Caucasian	-	8,681	2,863	3,429
	S. Asian	-	8,168	4,657	5,643
	Mixed- *start*	-	8,452	3,093	4,095
	Mixed- *end*	-	8,523	3,555	4,497
Probability cost effective at willingness- to-pay of £20,000 / £30,000 per QALY	Caucasian	-	0.68 / 0.76	0.99 / 1.00	0.95 / 0.97
	S. Asian	-	0.68 / 0.75	0.89 / 0.94	0.83 / 0.88
	Mixed- *start*	-	0.69 / 0.76	0.97 / 0.99	0.92 / 0.95
	Mixed- *end*	-	0.69 / 0.77	0.98 / 0.99	0.96 / 0.98

This example highlights the issue that combining data or taking averages for use in decision models needs to be done with extreme care. Although the interpretation of both the mixed models carried out here would probably lead to the same conclusions, the estimated outcomes did vary by each approach. The effects of heterogeneity in ethnicity could be explored in this example, due to the availability of data by ethnic sub-groups, but not all sources of heterogeneity could be explored. To fully assess the impact of using averages across sub-groups IPD would be required.

8.4 Compliance to screening

8.4.1 Adapting the model to assess compliance to screening

For the base case scenario it had been assumed that, for the three strategies that incorporated a screening component, the whole population of interest had been screened. In a real life situation though, it is likely that if individuals are offered a screening test, or invited for screening, a percentage will refuse or fail to take-up the opportunity. It was therefore important to assess the potential impact of different compliance rates to screening, in particular whether a low uptake to screening would reduce the cost-effectiveness of the three active screening strategies compared to no screening.

To assess compliance to screening, the 3x3 diagnostic table entered at the start of the decision model was modified, so that as compliance to screening was reduced the proportion of individuals identified by the screening test was reduced accordingly, and the number of individuals with IGT and T2DM remaining unidentified increased. Also the initial screening costs entered into the model were reduced to correspond with the compliance rates.

For the purposes of this model it was assumed that individuals who refused screening were the same in terms of their risk of IGT and T2DM as those who agreed to be screened. This may not have been the case, for example individuals leading unhealthy lifestyles with a greater risk of T2DM may refuse screening for fear of what they may find out, or conversely attendees may comprise individuals with a greater risk of T2DM, such as those who have relatives with the disease, as they may be more aware of the benefits of early diagnosis. If data had been available on individuals who had refused T2DM screening, particularly in terms of their risk of IGT and T2DM, the model could have been modified to include this. As no data was available though

the assumption of no differences between individuals who agreed or refused screening was made. The results from the models where compliance to screening was changed are reported in table 8.6.

Table 8.6: The effect of compliance to screening on model outcomes

	Compliance to screening (%)	Screening for T2DM only	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
QALY	100	28.12 (23.58, 32.08)	28.26 (23.74, 32.23)	28.22 (23.69, 32.18)
	90	28.09 (23.54, 32.08)	28.23 (23.69, 32.22)	28.19 (23.63, 32.18)
	80	28.18 (23.54, 32.07)	28.30 (23.67, 32.19)	28.26 (23.63, 32.16)
	70	28.07 (23.52, 32.05)	28.17 (23.64, 32.16)	28.14 (23.60, 32.13)
	60	28.09 (23.52, 32.06)	28.19 (23.63, 32.15)	28.16 (23.60, 32.13)
	50	28.04 (23.51, 32.04)	28.13 (23.61, 32.13)	28.10 (23.59, 32.11)
Total cost (GBP)	100	18,040 (7,083, 39,970)	17,910 (7,124, 39,740)	17,900 (7,061, 39,710)
	90	18,010 (6,894, 39,840)	17,990 (6,984, 39,640)	17,930 (6,943, 39,630)
	80	17,740 (6,818, 39,820)	17,730 (6,984, 39,650)	17,700 (6,926, 39,650)
	70	18,070 (6,777, 39,800)	18,080 (6,957, 39,620)	18,070 (6,907, 39,710)
	60	17,790 (6,580, 39,570)	17,830 (6,851, 39,630)	17,800 (6,810, 39,630)
	50	17,870 (6,409, 39,750)	17,930 (6,705, 39,680)	17,910 (6,671, 39,690)
Cost per QALY gained (GBP)	100	8,681	2,863	3,429
	90	8,836	2,942	3,465
	80	8,739	3,014	3,583
	70	8,732	3,112	3,800
	60	8,632	3,283	3,876
	50	8,743	3,515	4,192
Probability cost effective at willingness-to-pay per QALY of £20,000/£30,000	100	0.68 / 0.76	0.99 / 1.00	0.95 / 0.97
	90	0.68 / 0.76	0.98 / 1.00	0.94 / 0.97
	80	0.68 / 0.76	0.98 / 0.99	0.94 / 0.97
	70	0.69 / 0.77	0.98 / 0.99	0.93 / 0.96
	60	0.68 / 0.77	0.97 / 0.99	0.92 / 0.96
	50	0.68 / 0.77	0.97 / 0.98	0.92 / 0.95

8.4.2 Results for varying screening compliance rates

From table 8.6 it can be seen that varying screening compliance rates from 100% to just 50% did not significantly affect the cost-effectiveness of each of the three strategies that involved a screening component, compared to no screening. The probability of cost-effectiveness only varied marginally, and the estimated incremental costs per QALY gained were always within £1000 of those estimated assuming 100% compliance. From these results it appears that compliance to screening is not an important issue that needs to be considered to achieve a cost-effective screening strategy. This may not be the case if populations with greater prevalence of T2DM were considered though, or if a strategy of multiple screens was implemented, and this could be investigated through multi-way sensitivity analyses, as demonstrated in chapter 7.

8.5 Compliance to interventions

8.5.1 Adapting the model to assess compliance to interventions

As discussed in Chapter 4, compliance to interventions was described as high where it was reported, although many of the intervention trials did not formally assess or even discuss compliance when they reported trial results. Compliance to interventions under trial conditions may be very different though, to levels of compliance that will be achieved in a real-life setting, therefore it was important to assess the cost-effectiveness of the two intervention strategies assuming different levels of compliance. To do this, the estimated intervention effects from the trial data was assumed to show the effectiveness of interventions at 100% compliance. The effectiveness of the interventions on delaying the transition rate from diagnosed IGT to T2DM was then reduced depending on the compliance rate being assumed (equation 8.2, where $LogHR_R$ is the reduced log hazard ratio accounting for compliance, $LogHR_F$ is the full log hazard ratio and C is the compliance rate.

$$LogHR_R = (LogHR_F \times C) + (1 \times (1 - C) \qquad \text{[Equation 8.2]}$$

Costs remained constant, as although costs may be reduced in an individual with poor compliance, this may not always be the case and it was impossible to predict any changes in costs. For example the individual may still receive the pharmacological intervention but fail to take it, or lifestyle interventions, such as group counselling sessions and exercise classes, may

still take place but with fewer attendees. As with screening compliance, rates of between 50% and 100% were assumed, and the results are displayed in table 8.7.

8.5.2 Results for varying intervention compliance rates

Results show that reducing compliance to interventions reduces the predicted QALYs for the two strategies that involved interventions, and increases costs. Consequently the costs per QALY gained also increase, from £2,863 where 100% compliance is assumed for lifestyle interventions to £5,775 if compliance is assumed to be just 50%, and from £3,429 to £6,243 for pharmacological interventions. From the probabilities of cost-effectiveness though it can be seen that both the intervention strategies are still cost-effective, even when compliance is reduced to just 50%.

Table 8.7: The effect of compliance to interventions on model outcomes

	Compliance to intervention (%)	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
Years T2DM free	100	21.17 (10.66, 29.79)	21.07 (10.54, 29.66)
	90	21.13 (10.61, 29.75)	21.04 (10.52, 29.64)
	80	21.09 (10.57, 29.71)	21.02 (10.49, 29.61)
	70	21.05 (10.54, 29.67)	20.99 (10.48, 29.59)
	60	21.01 (10.51, 29.63)	20.97 (10.46, 29.56)
	50	20.98 (10.47, 29.59)	20.95 (10.44, 29.54)
QALY	100	28.26 (23.74, 32.23)	28.22 (23.69, 32.18)
	90	28.25 (23.72, 32.21)	28.21 (23.68, 32.17)
	80	28.23 (23.70, 32.19)	28.22 (23.67, 32.16)
	70	28.22 (23.69, 32.18)	28.19 (23.66, 32.15)
	60	28.20 (23.67, 32.16)	28.18 (23.65, 32.14)
	50	28.19 (23.66, 32.15)	28.17 (23.64, 32.13)
Total cost (GBP)	100	17,910 (7,124, 39,740)	17,900 (7,061, 39,710)
	90	17,990 (7,203, 39,820)	17,950 (7,109, 39,770)
	80	18,070 (7,272, 39,900)	17,990 (7,163, 39,820)
	70	18,140 (7,343, 39,950)	18,040 (7,209, 39,880)
	60	18,210 (7,404, 39,980)	18,080 (7,253, 39,920)
	50	18,261 (7,455, 40,050)	18,120 (7,302, 39,960)
Cost per QALY gained (GBP)	100	2,863	3,429
	90	3,493	3,950
	80	4,189	4,470
	70	4,947	5,039
	60	5,775	5,637
	50	6,634	6,243
Probability cost effective at willingness-to- pay per QALY of £20,000/ £30,000	100	0.99 / 1.00	0.95 / 0.97
	90	0.98 / 0.99	0.93 / 0.96
	80	0.96 / 0.98	0.91 / 0.95
	70	0.94 / 0.97	0.89 / 0.94
	60	0.91 / 0.95	0.87 / 0.92
	50	0.88 / 0.93	0.84 / 0.90

8.6 Re-screening

8.6.1 Adapting the model to assess the effects of more than one screening

The primary model that has been developed considers the effects of a one-off screening for individuals at risk of T2DM, at 45 years of age. In practice a screening health policy may involve a series of screenings taking place at a number of ages, for example 45, 55 and 65 years of age. The model was therefore adapted to assess the impact of having more than one screening of a population, on the cost-effectiveness of the screening and intervention strategies. Adapting the model was complex as the numbers in each state at each time had been calculated in the primary model by using the inprod command which utilised matrices of the transition rates and number in each state at time-1 to calculate the number in each state at time.

The model was adapted by firstly changing the scenario so that instead of a one-off screening, the sample population were re-screened 5 years later, at the age of 50. Therefore at cycle six the numbers in the undiagnosed T2DN and undiagnosed IGT states were reduced and the numbers in the diagnosed IGT and screen diagnosed T2DM increased, to reflect the amount that would be detected by a second screening. This was calculated by using the relevant sensitivities (the probabilities of testing positive given you have the disease, P(T+| D+)) from the 3 x 3 diagnostic table entered at the start of the model. It was assumed that everyone in the NGT, undiagnosed IGT and undiagnosed T2DM states would receive the FPG test, and everyone testing positive would receive and OGTT, and the costs of these tests were added into the model.

To assess the impact of screening on more than one occasion the results from the base case model, where screening at age 45 years was modelled, were compared to two scenarios whereby two screenings took place at ages 45 and 50 years, or three screenings took place at ages 45, 50 and 55 years of age. The results of these model extensions are in table 8.8.

Table 8.8: The effect of multiple screenings of the model population on predicted outcomes

	Number of screenings*	No screening	Screening for T2DM only	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
Lifetime risk of T2DM (%)	1	68.4 (18.0, 91.8)	68.4 (18.0, 91.8)	67.5 (17.1, 91.1)	67.9 (17.4, 91.4)
	2	68.4 (18.0, 91.8)	68.4 (18.0, 91.8)	65.3 (16.6, 87.2)	66.6 (17.1, 89.3)
	3	68.4 (18.0, 91.8)	68.4 (18.0, 91.8)	63.7 (16.5, 85.7)	65.5 (17.0, 88.5)
Years diabetes free	1	20.85 (10.36, 29.45)	20.85 (10.36, 29.45)	21.17 (10.66, 29.79)	21.07 (10.54, 29.66)
	2	20.85 (10.36, 29.45)	20.85 (10.36, 29.45)	21.69 (12.04, 29.91)	21.40 (11.35, 29.74)
	3	20.85 (10.36, 29.45)	20.85 (10.36, 29.45)	22.05 (12.42, 29.96)	21.64 (11.58, 29.77)
QALY	1	28.06 (23.49, 32.01)	28.12 (23.58, 32.08)	28.26 (23.74, 32.23)	28.22 (23.69, 32.18)
	2	28.06 (23.49, 32.01)	28.13 (23.74, 32.06)	28.56 (24.74, 32.30)	28.44 (24.45, 32.24)
	3	28.06 (23.49, 32.01)	28.15 (23.86, 32.16)	28.80 (25.04, 32.32)	28.62 (24.70, 32.26)
Total cost (GBP)	1	17,290 (5,746, 39,580)	18,040 (7,083, 39,970)	17,910 (7,124, 39,740)	17,900 (7,061, 39,710)
	2	17,290 (5,746, 39,580)	18,850 (7,491, 40,980)	19,300 (7,570, 41,160)	19,150 (7,468, 41,150)
	3	17,290 (5,746, 39,580)	19,670 (7,735, 42,110)	20,220 (7,740, 42,210)	19,860 (7,621, 42,210)
Cost per QALY gained (GBP)	1	-	8,681	2,863	3,429
	2	-	9,544	2,777	3,317
	3	-	10,360	2,966	3,517
Probability cost effective at willingness- to-pay of £20,000 / £30,000 per QALY	1	-	0.68 / 0.76	0.99 / 1.00	0.95 / 0.97
	2	-	0.57 / 0.66	0.99 / 1.00	0.96 / 0.98
	3	-	0.54 / 0.64	0.99 / 1.00	0.97 / 0.99

* If one screening was made it was done at age 45years, 2 screenings 45 and 50 years, and 3 screenings 45, 50 and 55 years of age.

8.6.2 Results for additional screenings

Figures 8.1 and 8.2 show how adding additional screenings at cycles 6 and 11, increase the probabilities of having screen detected diabetes or diagnosed IGT, at each of these cycles. It can be seen that for IGT, the additional screening at 50 years of age would identify more cases, than the additional screening at 55 years of age, due to there being more individuals with undiagnosed IGT at this time point. The probability of having IGT reduces over time as individuals go on to develop T2DM or die.

From table 8.8 it can be seen that increasing the number of screens of the population leads to an increase in the estimated QALYs for the strategy of T2DM screening. The years T2DM free and lifetime risk of T2DM remain constant, as the increased screens only results in more cases being diagnosed and less time being spent undiagnosed. As would broadly be expected, for the two screening and intervention strategies, QALYs, years diabetes free and estimated lifetime risk of T2DM, all improved as the number of screenings increased. For the T2DM only screening strategy the estimated cost per QALY gained compared to no screening, increased as number of screenings increased, and the probability of cost-effectiveness decreased. For the two screening and intervention strategies the probability of their cost-effectiveness changed very little. It appears that clinical outcomes rather than cost-effectiveness may therefore be the deciding factor as to how many screenings are adopted for a UK health policy.

Figure 8.1: The probability of having screen diagnosed T2DM at each cycle, for programmes of either one, two or three screenings

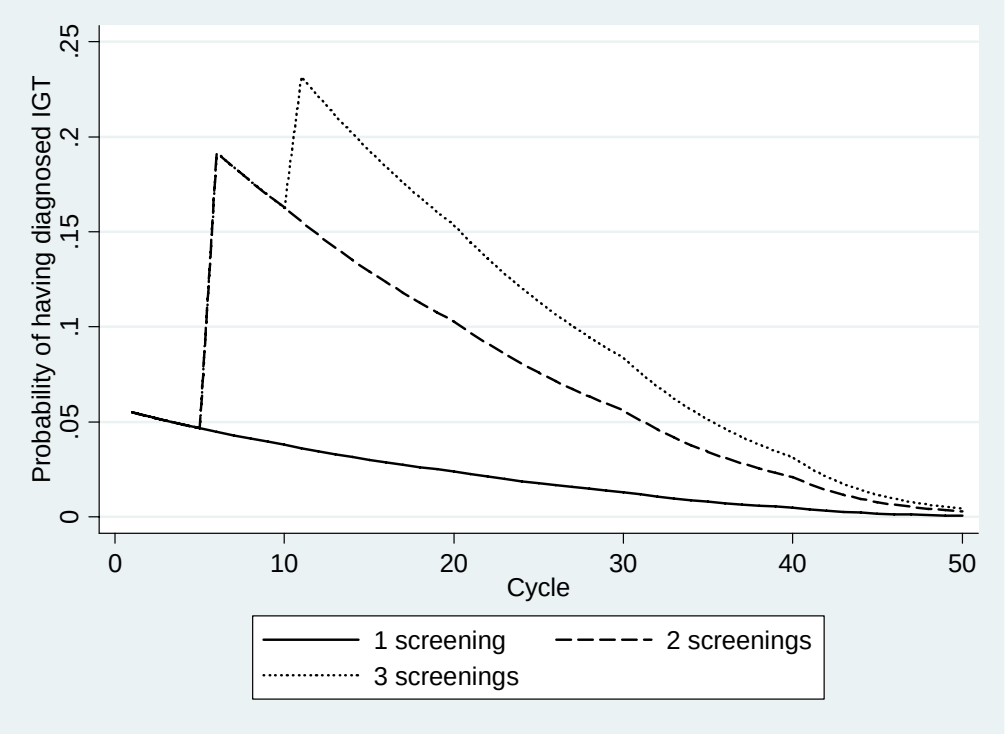
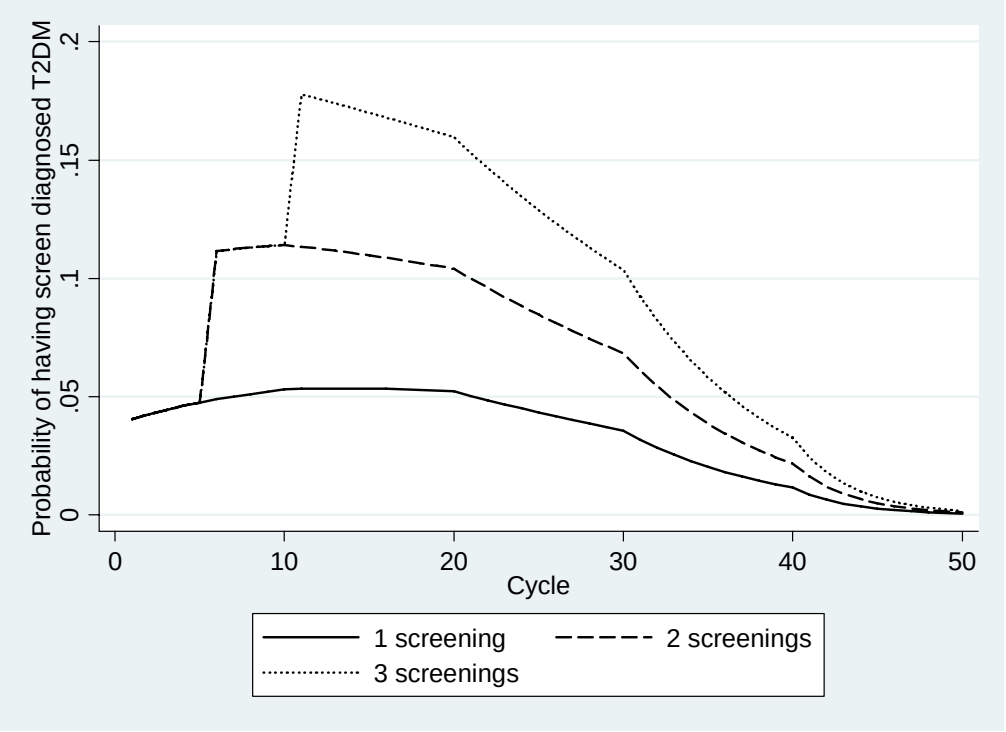


Figure 8.2: The probability of having screen diagnosed T2DM at each cycle, for programmes of either one, two or three screenings



8.7 Alternative sources of diagnostic data

For the primary model the FPG screening test was modelled using individual patient data available from the STAR study. Additional information on screening tests had been sourced by carrying out a search of both Medline and Embase databases as detailed in Appendix 1.4. Due to how the results were reported in these studies, including the additional data in the model was not straight forward. Therefore the primary model was based solely on STAR data, and how the inclusion of additional data could be achieved will instead be discussed briefly here.

The table of results needed for the decision model is given in table 8.9, whereby for each of the three glucose tolerance states (NGT, IGT and T2DM), and for each of two possible cut-offs (a lower cut-off for when both IGT and T2DM are screened for and a higher cut-off for when T2DM alone is screened for), the probabilities of testing negative for both cut-offs (T-), testing positive for both cut-offs (T+), or falling in between the two cut-offs (T0), are needed. An example of how data is typically reported in the literature is a study by Shirasaya (Shirasaya et al., 1999) which assessed the efficiency of three potential screening tests for IGT and T2DM, which were 1,5-anhydroglucitol (1,5-AG), glycosolated haemoglobin (HbA_{1c}) and fructosamine (FRA). 1,5-AG proved the most accurate of the three screening tests, so the results of this test will be considered here.

This study reported sensitivity and specificity of the test for two cut-offs, a lower cut-off whereby both IGT and T2DM were screened for, and a higher cut-off, whereby just T2DM was screened for. From these results two diagnostic tables could be defined (tables 8.10 and 8.11). From this data alone the 3x3 diagnostic table needed for the model could not be fully compiled, with the only data known given in table 8.12. This pattern was repeated in other screening studies, whereby the sensitivity and specificity of a test was reported for one or two cut-offs and the full 3x3 diagnostic table could not be specified. What can be done though is that both the collapsed tables given in table 8.10 and 8.11, can have each cell probability written in terms of the probabilities needed for the 3x3 table (i.e. in terms of a to i). These assumptions can then be used to estimate the missing cell frequencies.

Table 8.9: A break-down of the 3x3 diagnostic table

	NGT	IGT	T2DM	
T-	a	b	c	R1
T0	d	e	f	R2
T+	g	h	i	R3
	C1	C2	C3	N

Table 8.10: Results for the lower cut-off of the 1,5-AG screening test

	NGT (n)	IGT or T2DM (n)	Totals
Tested negative	421 (a)	75 (b+c)	496 (a+b+c)
Tested positive	269 (d+g)	126 (e+f+h+i)	395 (d+e+f)
	690 (a+d+g)	201 (b+e+h+c+f+i)	891 (g+h+i)

Table 8.11: Results for the higher cut-off of the 1,5-AG screening test

	NGT or IGT (n)	T2DM (n)	Totals
Tested negative	722 (a+b+d+e)	6 (b+c+e+f)	728 (a+b+c)
Tested positive	132 (g+h)	31 (h+i)	163 (d+e+f)
	854 (a+b+d+e+g+h)	37 (b+c+e+f+h+i)	891 (g+h+i)

Table 8.12: Shirasaya data available for the model

	NGT	IGT	T2DM	Totals
Tested negative	421	-	-	496
Test result between the two cut-offs	-	-	-	232
Tested positive	-	-	31	163
	690	164	37	891

A model was derived in WinBUGS to estimate the missing data (code given in Appendix 2.3). This was done by using the assumptions that can be drawn from the two 2x2 tables, as specified in tables 8.10 and 8.11. For each of the three groups (j), NGT, IGT and T2DM, the following definitions, representing the possible testing scenarios, were modelled, whereby T+ represents testing positive for the high cut-off, T- testing negative for the lower cut-off and T0 test result is between the two cut-offs;

$$probability(\ T+ \mid j) = p[1,j]$$

$$probability(\ T0 \mid T0\ or\ T-) = p[2,j]$$

$$probability(\ T0 \mid j) = p[3,j] = p[2,j](1-p[1,j])$$

$$probability(\ T- \mid j) = p[4,j] = 1 - p[1,j] - p[3,j]$$

probability(*T+ or T0* | *j*) = p[5,*j*] = p[1,*j*] + p[3,*j*]

probability(*T0 or T-* | *j*) = p[6,*j*] = p[3,*j*] + p[4,*j*]

where p[1,j] and p[2,j] are basic parameters, and the rest are functional, in that they can be written in terms of the two basic parameters. The known data from tables 8.10 and 8.11 were then entered into the model as follows:

$$\pi_1 = ((690 \times p[1,1]) + (164 \times p[1,2])) / 854$$

$$r_1 \sim Bin(\pi_1, N_1) \qquad \text{where} \qquad r_1 = 132 \;\&\; N_1 = 854$$

$$\pi_2 = ((164 \times p[4,2]) + (37 \times p[4,3])) / 201$$

$$r_2 \sim Bin(\pi_2, N_2) \qquad \text{where} \qquad r_2 = 75 \;\&\; N_2 = 201$$

where $\pi_1 = p[T+ \mid NGTorIGT]$ and $\pi_2 = p[T- \mid IGTorT2DM]$ which represent two of the cells that combined groups in the Shirasaya data (tables 8.10 and 8.11), that is cells *g*+*h* and *b*+*c* respectively. From the Shirasaya data, two cells in the 3x3 table are known (table 8.12), and these can be modelled directly as follows:

$$r_3 \sim dbin(p[4,1], N_3)$$
$$r_3 = 421 \;\&\; N_3 = 690$$
$$r_4 \sim dbin(p[1,3], N_4)$$
$$r_4 = 31 \;\&\; N_4 = 37$$

where *p*[4,1] is the p[T-|NGT] and *p*[1,3] is the p[T+|T2DM]. Distributions were then placed on the probabilities to ensure the three probabilities that comprised testing positive summed to 1, as did the three that comprised testing negative or having a test result between the two cut-offs. Also assumptions were made on the probabilities in terms of their relative magnitude. So, for example, p[1,1] was forced to be between 0 and p[1,2], which is assuming that the probability of T+ if you have NGT, will be smaller or equal to the probability of testing positive if you have

IGT. This can be done by putting constraints on the prior distributions of the following probabilities:

$$p[1,1] \sim Unif(0, p[1,2])$$
$$p[2,1] \sim Unif(0, p[2,2])$$
$$p[1,2] \sim Unif(p[1,1], p[1,3])$$
$$p[2,2] \sim Unif(p[2,1], p[2,3])$$
$$p[1,3] \sim Unif(p[1,2],1)$$
$$p[2,3] \sim Unif(p[2,2],1)$$

The remaining probabilities, which are functions of p[1,j] and p[2,j], were the same for each glucose tolerance state (j) and were specified as:

$$p[3, j] = p[2, j] \times (1 - p[1, j])$$
$$p[4, j] = 1 - p[1, j] - p[3, j]$$
$$p[5, j] = p[1, j] + p[3, j]$$
$$p[6, j] = p[3, j] + p[4, j]$$

By placing limits on the probabilities, as well as using the assumptions listed above and specifying the two known cell values for the 3x3 table, as well as the row and column totals, estimates of the missing data could be made (table 8.13). Using the same programme to estimate the 3x3 table for the STAR data, if it assumed that only the values available from the reported Shirasaya paper were known, shows the method closely replicates the actual results (tables 8.14 and 8.15).

Table 8.13: Estimated values for the Shirasaya data

	NGT	IGT	T2DM	Totals
Tested negative	422.2	71.5	2.0	495.7
Test result between the two cut-offs	179.1	48.2	4.6	231.9
Tested positive	88.7	44.3	30.3	163.3
	690	164	36.9	890.9

Table 8.14: Actual STAR results

	NGT	IGT	T2DM	Totals
Tested negative	2014	177	12	2203
Test result between the two cut-offs	384	101	8	493
Tested positive	142	125	143	410
	2540	403	163	3106

Table 8.15: STAR results as predicted by the WinBUGS program

	NGT	IGT	T2DM	Totals
Tested negative	2014.0	181.6	6.8	2202.4
Test result between the two cut-offs	378.6	101.0	13.9	493.5
Tested positive	147.5	120.4	142.2	410.1
	2540.1	403	162.9	3106

Using the estimated Shirasaya test data in the comprehensive decision model gave similar results as the primary model when the STAR data was used. The estimated QALYs were 28.08 (95% CrI: 23.5 to 32.02) for no screening, 28.15 (23.6 to 32.08) for T2DM only screening, 28.33 (23.8 to 32.27) for screening and lifestyle interventions and 28.27 (23.73 to 32.21) for screening and pharmacological interventions. The incremental cost per QALY gained were £8817 for T2DM screening, £2575 for screening and lifestyle interventions and £3188 for screening and pharmacological interventions, as compared to no screening.

The methods described in this section would enable the use of data that would otherwise be excluded. If a number of studies described the same test, then data could be meta-analysed to obtain pooled estimates. As diagnostic studies may vary in terms of cut-offs used and populations studied though, this could be difficult to achieve in practice (Hellmich et al., 1999).

8.8 Value of information

8.8.1 Methods for assessing value of information

A cost-effectiveness analysis, as has been carried out in this thesis, utilises evidence on numerous parameters relating to treatment effects, utilities and costs. Ideal data will never exist on all of these parameters, and this has been incorporated into the model here by including all appropriate uncertainty within the decision model. Having imperfect data though raises the question as to whether further research to improve data sources would be a useful utilisation of resources. This key question can be answered by carrying out a value of information analyses on the decision model (Claxton and Sculpher, 2006, Sculpher and Claxton, 2005).

Value of information analyses centre around the calculation of the expected value of perfect information (EVPI), that is the difference between the expected net benefit of having perfect information compared to the current information (net benefit was described in section 2.4.5). For example if there are two alternative interventions *(j=1,2)*, with unknown parameters θ, then the optimal decision is the intervention that generates the maximum expected net benefit, $max_j \{E_\theta NB(j,\theta)]\}$. This is the maximum net benefit over all the iterations from the Monte Carlo simulations, with each iteration representing a possible future realisation of existing uncertainty of θ. With perfect information the value of θ would be known, and the expected value of a decision taken with perfect information can be found by averaging the maximum net benefits over the distribution of θ, i.e. $E_\theta [max_j \{NB(j,\theta)\}]$. The EVPI for a single patient is then simply the difference between the expected value of the decision model with perfect information about the uncertain parameter θ, and the decision made on the basis of existing evidence:

$$EVPI = E_\theta [max_j \{NB(j,\theta)\}] - max_j \{E_\theta [NB(j,\theta)]\} \qquad \text{[Equation 8.2]}$$

The EVPI effectively gives a ceiling to the maximum amount that should be spent on further research. The individual patient EVPI can be extended to represent a population, and if this "population" EVPI exceeds the expected cost of additional research, then it is potentially cost-effective to conduct further research. Alternatively if the cost of further research exceeds the EVPI, then further work is not cost-effective. Obviously an assessment of current resource availability and the importance of conflicting demands for resources will need to be made, before a final decision on the feasibility of further research can be reached.

This idea of EVPI can be extended to assess the cost-effectiveness of reducing the uncertainty of a group of parameters within the model, rather than uncertainty around all model parameters. For example, it may be of interest to assess the EVPI for all the economic parameters in the model, or for just the utilities. This is termed the expected value of perfect partial information (EVPPI) (Claxton and Sculpher, 2006).

For the decision model developed in this thesis the EVPI was calculated for a number of willingness-to-pay thresholds, As the EVPI is calculated using a comparison of a new versus standard treatment, three EVPIs were calculated, one for each of the three active screening strategies compared to no screening. Additionally the population EVPI was also calculated for each of the three active screening strategies. Equation 8.3 shows how this was done, adjusting the individual EVPI, as defined in equation 8.2, to represent the current and future population of England and Wales. (Spiegelhalter et al., 2004).

$$EVPI_{pop} = EVPI. \sum_{t=1}^{T} \frac{I_t}{(1+\delta)^{t-1}} \qquad \text{[Equation 8.3]}$$

T represents the time horizon of a healthcare intervention, which for this example was taken as 50 years, I_t represents the annual incidence for year t, and δ is the annual discount rate, which for this example was taken as 3.5%, as recommended by NICE (National Institute for Clinical Excellence, 2004). For the analyses described here the annual incidence was the number of individuals turning 45 years of age in a given year, and therefore becoming eligible for screening and able to enter the model. The annual incidence was taken from national statistics for England and Wales, whereby it was estimated that 797,400 individuals turned 45 years of age in 2007. The incidence rate was assumed constant over the time horizon of the model. The WinBUGS code used to estimate the EVPI is given in Appendix 3.8.

8.8.2 Results of value of information analyses

The EVPI is intended for use in deciding whether to pursue a research programme, how to design it, and when to stop (Spiegelhalter et al., 2004). Using the methods as detailed above, the individual patient EVPIs and population EVPIs are presented in table 8.16.

Table 8.16: Expected value of information estimates

	Screening for T2DM only	Screening for T2DM and IGT, lifestyle interventions	Screening for T2DM and IGT, pharmacological interventions
Individual EVPI at different willingness-to-pay thresholds (GBP)			
£5000	814.6	738.1	710.0
£10,000	814.5	737.7	709.7
£20,000	814.4	737.5	709.5
£30,000	814.3	737.1	709.2
Population EVPI at different willingness-to-pay thresholds (results in million GBP)			
£5000	18,880	17,110	16,460
£10,000	18,880	17,100	16,450
£20,000	18,880	17,100	16,440
£30,000	18,880	17,090	16,440

From table 8.16 it can be seen that EVPI when considering an individual patient vary minimally by willingness-to-pay threshold, and range between £709 and £815. The assessment of T2DM only screening had the greatest EVPI value, which is unsurprising as it has already been seen that a lot of uncertainty surrounded the assessment of this screening strategy. Evaluating the EVPI for a whole population is probably a more useful exercise than just considering the individual EVPI, and these results are also presented in table 8.16. The costs are reported in millions (GBP). At the £20,000 willingness-to-pay threshold per QALY, the three active screening strategies had EVPI of £18,880,000,000 for T2DM screening, £17,100,000,000 for IGT and T2DM screening followed by lifestyle interventions and £16,440,000,000 for IGT and T2DM screening followed by pharmacological interventions. Although these are considerably larger, they are still comparable to other EVPI estimates. A number of EVPIs calculated for a range of health care programmes (Claxton and Sculpher, 2006), estimated an EVPI of £865 million for stroke prevention interventions and £710 million for screening to prevent myocardial infarction. These are smaller than our estimates, partly because the populations they would impact on are smaller.

The EVPIs calculated here indicate that a lot of uncertainty surrounds the parameters within the model. The EVPI for T2DM only screening was the highest as this was the strategy for which there was most uncertainty, as shown by the results presented in Chapter 6, where the probability of the strategy being cost-effective at the £20,000 willingness-to-pay threshold, was just 68% . As the EVPIs are high, continuing a research programme that will provide further data on the effects of screening and early treatment, interventions for the delay of T2DM, transition rates between different glucose tolerance states and the costs and utilities associated with T2DM, should be encouraged. A considerable amount of research will need to be conducted before the EVPI will be close to being exceeded by the cost of conducting an additional study.

The population EVPIs estimate here are inflated by the fact that the annual incidence of turning 45 years of age was used. Strictly speaking the model was constructed to model the effects of screening in individuals 45 years of age, who are ‘at risk’ of developing T2DM. Unfortunately the proportion ‘at risk’ could not easily be estimated. For example though, if just 25% of 45 year olds could be considered at risk and therefore eligible for screening, the EVPIs are for T2DM screening £4,720,000,000, for IGT and T2DM screening followed by lifestyle

interventions £4,274,000,000 and for IGT and T2DM screening followed by pharmacological interventions £4,112,000,000.

The population EVPIs calculated here are extremely high, probably due to the fact that the model has been constructed in such a way at to include all predicted uncertainty around parameters where they are not known for certain. Comprehensive decision models that do not correctly model all parameter uncertainty may under estimate the EVPI, which could result in further research being stopped, when in fact it would be a cost-effective exercise.

8.9 Discussion and chapter summary

This chapter explored the versatility of the modelling framework developed for this thesis by carrying out a number of extensions to the model. All were fairly straight forward to implement, indicating that once a decision model has been set up using methodologies as described in chapter 6, it is unproblematic to manipulate the model to consider different scenarios and outputs.

A number of model extensions were considered. These included utilising different data sources, such as described for incorporating the mixed treatment comparison of interventions, and the additional data on diagnostic tests. The model was also run for different ethnic groups, and this extension highlighted the very important issue of when data should be combined when compiling a comprehensive decision model. Using averages to model transition rates and other model parameters may lead to misleading results, so the construction of decision models needs to be carefully thought through, and sensitivity analyses carried out where possible. Importantly data may not be available on different groups and therefore the effect of using an average value in a model may be very difficult to assess. IPD data would allow the impact of the use of averages to be explored fully.

The model was adapted to assess both compliance to screening and interventions. These are important considerations as in a real-life setting 100% compliance will never be achieved. Results from the model suggest that compliance to either screening or interventions did not change qualitatively the conclusions when considering the cost-effectiveness of the three active screening strategies.

A further extension to the model considered a policy of re-screening the population of interest at ages additional to the one-off screening at 45 years of age that was modelled in the primary model. The results predict that screening on more than one occasion will not increase the cost-effectiveness of the three active screening strategies, in fact their cost-effectiveness decreased slightly if additional screenings were introduced to the model, probably due to the fact that costs of T2DM were higher once a diagnosis was made.

The final model extension assessed the expected value of perfect information, and it was identified that it would definitely be cost-effective to carry out further research to reduce uncertainty in the decision model parameters. The analysis could have been taken one step further and EVPPI for different groups of parameters could have been calculated. This may have helped identify where further research was most needed, that is the parameter groups with the highest EVPPI values. This was not carried out due to the complexity of the analysis and time limitations.

Overall the model extensions were all very informative in terms of both clinical considerations for the implementation of a screening strategy, and for developing a deeper understanding of drivers within the decision model. When a decision model has been developed it therefore seems prudent to consider further extensions, to enable the most information and benefit to be gained from the model.

9. APPLYING THE METHODOLOGY TO OTHER CLINICAL EXAMPLES

9.1 Chapter overview

The methodologies utilised and developed in previous chapters can be expanded from screening and prevention of T2DM to other examples. This chapter outlines two additional clinical situations and discusses how the methods from the diabetes model could be applied, and also how distinctive features specific to each clinical example would affect the modelling. Firstly the example of screening to identify individuals at high risk of coronary heart disease (CHD) and intervening to reduce risk of coronary events will be considered and secondly an example assessing the impact of screening for precursors for colon cancer will be described.

9.2 Screening cholesterol levels and intervening to reduce risk of CHD

9.2.1 CHD

CHD results from the accumulation of fatty deposits of cholesterol and waste substances which form plaque and clog up the arteries (NHSDirect, 2007). This makes them narrower and restricts blood flow, increasing blood pressure and putting a strain on the heart. Most individuals with CHD show no evidence of the disease for decades, and often a sudden heart attack is the first indication of the problem. There are approximately 270,000 heart attacks every year in the UK (NHSDirect, 2007).

9.2.2 Identifying individuals at high risk

Risk factors for CHD include modifiable ones such as high cholesterol (specifically serum low density lipoprotein (LDL) concentrations), smoking, hypertension, hyperglycaemia, obesity and physical inactivity. Unalterable risk factors include a family history of heart disease, age, gender and ethnicity (Ward et al., 2007). Much attention has been focused on screening people for CHD risk by measuring blood cholesterol concentrations, but although blood cholesterol is an important risk factor, by itself it is a relatively poor predictor of who will go on to have a CHD event (Ebrahim et al., 1998). Identifying those at risk by considering a number of risk factors together provides a much more reliable prediction of CHD risk. One of the most widely used tools for assessing risk is the Framingham risk score, which scores an individual depending on their age, gender, systolic blood pressure, ratio of total cholesterol to high density lipoprotein (HDL) cholesterol, presence of diabetes, and smoking habits (Ramachandran et al., 2000). It

gives a predicted ten year risk of CHD and cardio-vascular disease (CVD). The Framingham risk score was developed using a Caucasian population and therefore predicts CHD risk most accurately in Caucasians. Recently though it has been recalibrated so that the score can now be additionally used to assess seven British black and minority ethnic groups (Brindle et al., 2006).

9.2.3 Intervening to prevent CHD

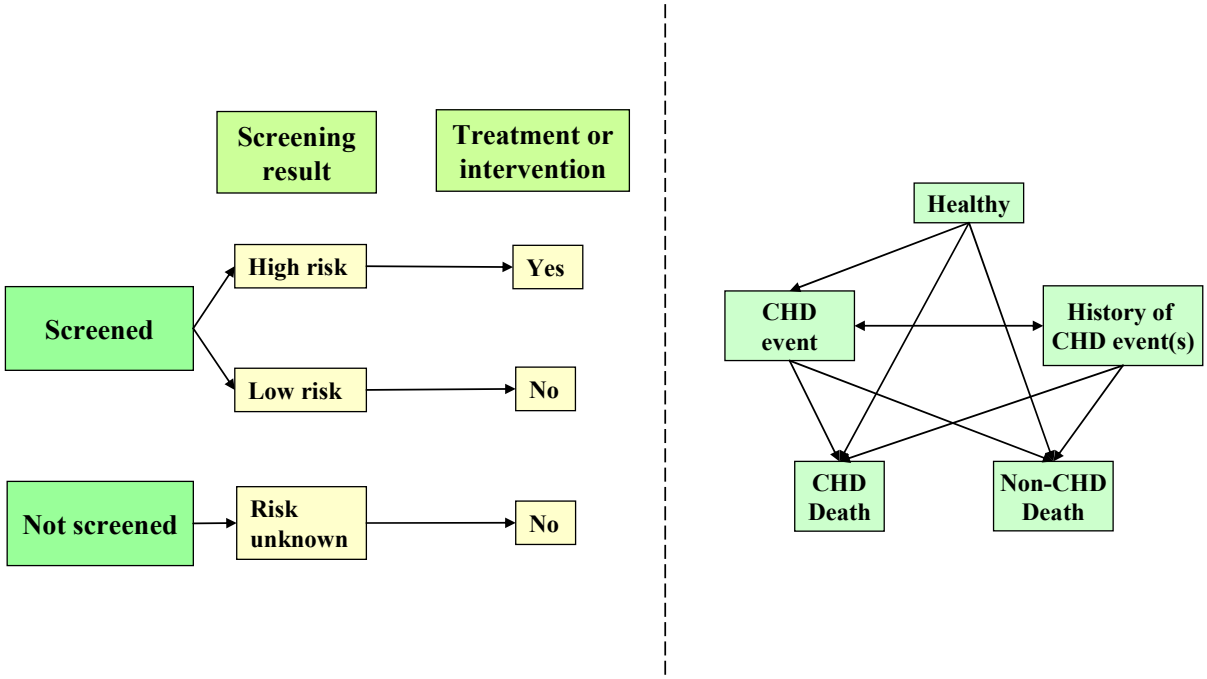
Individuals identified as being at high risk of CHD can be treated in a number of different ways. Risk can be significantly reduced by lifestyle changes such as increasing exercise, eating a healthier diet and stopping smoking. Pharmacological treatments can also play an important role, especially in those at highest risk, and include statins to lower LDL cholesterol, blood pressure lowering medications such as beta-blockers and angiotensin converting enzyme (ACE) inhibitors, and low dose aspirin to help prevent the blood from clotting and reduce the risk of a heart attack and angina (Ebrahim et al., 1998). Lowering cholesterol levels through either diet or pharmacological interventions, can lead to a significant reduction in CHD risk. It has been estimated from clinical trials of statins that CHD risk is reduced by 15% for every 10% reduction in plasma LDL cholesterol (Gould et al., 1998).

9.2.4 Modelling the information

An outline of how screening and intervening to reduce CHD risk could be modelled is given in figure 9.1. The model compares no screening with screening, where any screening test for which data is available could be considered. Firstly, a decision tree determines how many individuals would be detected through screening and have an intervention applied. A Markov model is then used to model yearly transitions from a healthy state to one representing occurrence of a CHD event, with two Markov models being run simultaneously, one for a screened population and one for an unscreened population. Transition rates could then be reduced where high risk individuals have been identified and interventions have been applied to reduce CHD risk. As with screening a number of different interventions could be considered, providing the data is available quantifying their effect on CHD risk.

Figure 9.1: Model outline of screening for CHD risk and intervening to reduce risk

Decision tree ***Markov model***



Once a CHD event has occurred, in the next cycle the Markov model moves an individual to a state that represents those who have a history of CHD events. Transition rates to a second CHD event or to death are higher from this state than from the healthy state. Both CHD death and non-CHD death are absorbing states. Additional event states could be included in the model, such as stroke or diabetes, but this would be dependent on data availability. As with the diabetes example, costs and utilities could be applied to each state, to allow for the assessment of both clinical and cost-effectiveness of screening for CHD risk.

9.3 Screening and prevention of colorectal cancer

9.3.1 Colorectal cancer

Colorectal cancer (CRC) or colon cancer is amongst the most common malignancies and remains a leading cause of cancer related morbidity and mortality (Gill and Sinicrope, 2005). Overall about one half of patients diagnosed with CRC ultimately die of the disease and this poor prognosis is primarily due to late detection of the disease by which time available treatments are limited and less effective (Sangha et al., 2004). Colorectal cancer is a multi-step process characterized by molecular and cellular alterations, that begins when normal mucosa changes to an adenomatous polyp. This adenoma is an easily identifiable precursor lesion that if untreated will eventually progress to carcinoma (Gill and Sinicrope, 2005). Due to the high prevalence of CRC, it's long asymptomatic phase, and the presence of a treatable precancerous lesion, CRC meets the criteria for screening (Sonnenberg et al., 2000).

9.3.2 Screening for precursors of colon cancer

Most colorectal cancer develops from precursor lesions which can be used as a target for early detection and therapy (Ramsoekh et al., 2007). Unlike other cancers there are several options for screening for CRC. The most extensively reviewed method is faecal occult blood testing (FOBT) and this has been shown, if offered biennially, to reduce mortality from CRC by as much as 20% (Atkin, 2003). Screening alternatives include flexible sigmoidoscopy and colonoscopy. Colonoscopy has the advantage over the other two techniques in that it allows the whole of the colon to be examined, but unfortunately it is also the most resource intensive and expensive method of screening (Ramsoekh et al., 2007).

A number of studies have been carried out to assess the cost-effectiveness of different screening approaches. One study used a Markov decision model to compare FOBT alone, FOBT combined with flexible sigmoidoscopy, flexible sigmoidoscopy alone and colonoscopy (Vijan et al., 2001). All screening strategies were found to be cost-effective when compared to no screening and twice-lifetime colonoscopy proved to be the most cost-effective screening strategy overall. The assessment of all screening strategies was very sensitive to assumed compliance levels.

9.3.3 Interventions

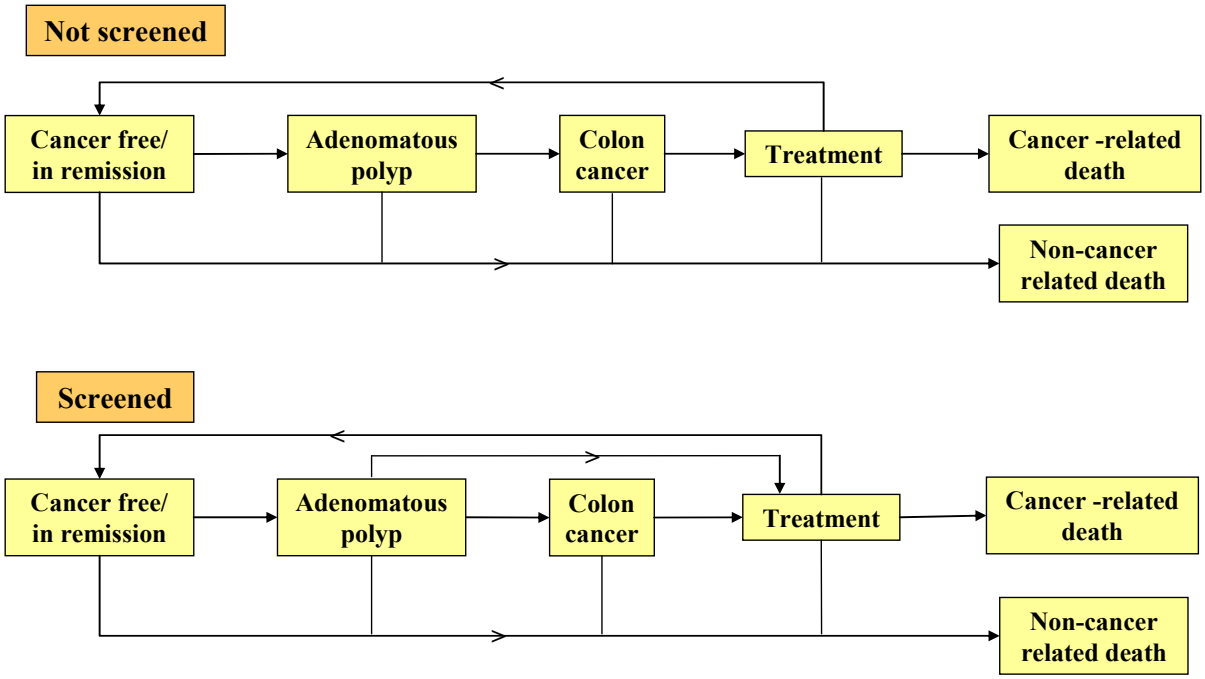
Once an adenomatous polyp has been identified the most effective intervention to prevent cancer is a polypectomy, which is a surgical removal of the abnormal growth (Sangha et al., 2004). Alternatively, or in addition to resection, various agents can be prescribed for chemoprevention of CRC. These include non-steroidal anti-inflammatory drugs (NSAIDs) such as aspirin, which inhibit tumorigenesis, and other agents such as folate, calcium, vitamins, anti-oxidants and fibre. The effectiveness of some of these agents is still speculative and research is very much on-going. Strong evidence already exists to support the therapeutic benefit of aspirin though, which has been found to be associated with significant reductions in both colorectal adenomas and carcinoma incidence (Gill and Sinicrope, 2005).

9.3.4 Modelling the information

A possible model for assessing the benefits of screening for precursors to colon cancer is given in figure 9.2. The model compares no screening vs. screening for pre-cancerous lesions, using two Markov models whereby yearly transition rates are used to model a theoretical population between states.

Where no screening occurs it is assumed that individuals with adenomas will not be identified and therefore will not receive a surgical resection or any interventions. The problem will only be identified once cancer has developed. Treatment of the cancer may be successful, returning the individual to a cancer free state or death due to cancer may occur. Where a screening programme is assumed the model allows for identification of individuals at the adenoma stage. Identification rates depend on uptake rates of screening, accuracy of the screening test and frequency of screening. These can all be altered within the model to assess their impact. Individuals identified with adenomas receive treatment and may move back to the cancer free state, or if treatment fails they may go on to develop cancer.

Figure 9.2: Model outline of screening for precursors to colon cancer



9.4 Comparison of the three models

In many ways the three decision models of diabetes prevention, CHD risk reduction and colon cancer prevention are very similar. They all rely on the availability of relevant data on screening tests and effectiveness of interventions before constructing a decision model can even be considered, and often what can be modelled will be restricted by data availability. All three models had a choice of possible screening tests and interventions that could be modelled. To assess the viability of different screening tests several models can be run, one for each screening test, and then the results compared.

Comparing interventions can be a little more complicated, as already shown with the diabetes example. To directly compare interventions where they may not have been compared directly in clinical trials, a mixed treatment comparison is needed. As it is unclear from the literature which interventions are the most effective in the two examples described in this chapter, there is great potential for using mixed treatment comparisons to try and identify the best treatments. Mixed treatment comparisons were used to directly compare pharmacological and lifestyle interventions using both direct and indirect evidence. The assumptions made for this type of analysis are discussed in chapter 4.

Many of the issues identified during the compilation of the diabetes model would also be important factors for the two models described in this chapter. The assessment of data quality is important for all three models, as the accuracy of model results is dependent on inputs entered into the model. For the diabetes model clinical trials were assessed for quality using the Jadad score and the same would be recommended for the trials utilised for these two models. As with the diabetes model, correlations, where one trial may provide data for more than one part of the model, need to be considered and modelled appropriately, and data has to be combined carefully if it has been derived from different populations. The availability of IPD for all three examples, would provide greater opportunities for modelling sub-groups within the decision model. For all decision models, sensitivity analyses on distributions and assumptions are very important, to enable a full understanding of the model, including the identification of the key parameters that are driving and influencing the results.

The three examples also have pronounced differences that effect how they are modelled. Screening for both CHD and IGT can be done simply by a blood test or by using a risk score. The most common screening tests for CRC though are faecal occult blood testing, colonoscopy or flexible sigmoidoscopy. These are more invasive and disagreeable to the individuals involved and therefore uptake rates of screening are likely to have a greater impact for this example, and therefore need to be taken account of and modelled realistically.

Unlike interventions for diabetes and CHD, the intervention of adenoma removal for CRC prevention is not a continuous treatment, and therefore it can be a state that is passed through in the model, instead of being modelled just as an effect on transition rates. Also, because intervention may not be a continuous process, then re-screening may be a more important issue in the cancer model, interventions in individuals at risk of diabetes and CHD may permanently reduce the risk if compliance is high, but once an adenoma has been removed an individual may still be at higher risk of future CRC than an individual who has had no precancerous lesions.

9.5 Chapter Summary

By considering two further examples it has become clear that many of the issues that have been considered for the diabetes model will be important issues for the majority of decision models. Synthesising data from a number of sources has the potential for many pitfalls and has to be carefully thought through if unbiased, accurate and interpretable results are to be produced.

10: DISCUSSION AND CONCLUSIONS

10.1 Summary

This thesis reviews the methods utilised for evidence synthesis, developing them further where appropriate, through the development of a comprehensive decision model to assess different health policies in respect to screening for T2DM. Four strategies were compared which were, no screening (current policy), screening for T2DM alone, screening for IGT and T2DM, with lifestyle interventions applied to those with IGT in an attempt to delay the onset to T2DM, and finally screening for both IGT and T2DM with pharmacological interventions applied to individuals with IGT. The model was structured to monitor the impact of carrying out a one-off screening of individuals aged 45 years of age and the impact this would have over a 50 year time horizon. The model was thoroughly checked for problems associated with MCMC analyses, and numerous sensitivity analyses and model extensions were considered, to explore a number of screening scenarios. The work carried out for this thesis needs to be interpreted in terms of both its methodological and clinical implications.

10.1.1 Methodological summary

The comprehensive decision model constructed here is innovative when compared to current published models. Firstly the entire model, including all evidence synthesis, was encompassed within a single flexible framework. This has many advantages including, all uncertainty can easily be incorporated within model parameters and assumptions on posterior distributions of summary estimates do not need to be made if they are being combined within the model (Cooper et al., 2004). Secondly a Bayesian approach was taken to the analysis, and the model was developed within the statistical package WinBUGS. This has the advantage over many traditional software packages in that it is relatively simple to specify and run computationally intensive and complex models, whilst accounting for all uncertainty and sources of correlation between model inputs. Also, although not utilised in this example, it allows subjective prior beliefs to be included within the model. Running the model in WinBUGS, which simulates parameters over thousands of iterations, has the additional advantage that direct statements can be obtained, such as the probability of one parameter being higher than another.

The model developed here sought to improve on current published decision models by incorporating as much published data as was relevant to the model as was feasible. For example

to estimate the effectiveness of interventions, a full systematic review was carried out. This involved the assessment of over ten thousand published abstracts, including foreign language papers, in an attempt to locate all randomised controlled trials, which had assessed interventions aimed at delaying or preventing T2DM. To allow a direct comparison of lifestyle and pharmacological interventions, an MTC analysis was carried out. This systematic review, meta-analysis and MTC were described in detail in Chapter 4, and were published in the British Medical Journal (Gillies et al., 2007).

To endeavour to make the model as realistic as possible, parameters were changed over time where relevant within the comprehensive decision model, e.g. transition rates and utilities were modified over the model horizon. Individual patient data (IPD) as well as summary data was utilised for the model. IPD was particularly useful for constructing the screening test decision tree, due to difficulty in extracting data from published reports as discussed in Chapter 8. Methods to extract data on screening tests, where only collapsed tables were reported, i.e. when groups are combined over rows or columns, was investigated and discussed in Chapter 8. To model the effects of complications associated with T2DM on quality of life, complication rates were estimated using a series of Weibull survival models, which were run within the decision modelling framework. Once the model had been compiled a full assessment of parameter convergence, and a number of model checks were carried out (Cowles and Carlin, 1996), as described in detail in Chapter 7. Therefore not only was the model the most comprehensive in comparison to previous models developed in the area of T2DM screening, it was also the most complex in its structure.

10.1.2 Clinical summary

The comprehensive decision model developed here was the first to assess a strategy of screening and intervening in individuals with IGT, and the first to compare screening for T2DM alone with a strategy of screening for both IGT and T2DM in combination. From the model results it appears that there is evidence to support strategies where both IGT and T2DM are screened for, followed by an intervention comprising of either a lifestyle programme or a course of prescription drugs for individuals diagnosed with IGT, and early treatment for individuals diagnosed with T2DM. Both the two strategies that included interventions had a high probability of being cost-effective at fairly low willingness-to-pay thresholds. Screening for T2DM alone compared to no screening was more difficult to assess in terms of its cost-effectiveness. A lot of uncertainty surrounded the model results and therefore although the

predicted cost per QALY gained for this strategy was fairly low, the probability of it being cost-effective, only exceeded 80% at a willingness-to-pay threshold of £40,000.

As both the strategies concerning lifestyle and pharmacological interventions gave similar results, this raises the question as to which is best. Lifestyle had a higher cost per annum than pharmacological, but the intervention effect on reduction to progression to T2DM was greater in these trials. Both have further issues aside from cost-effectiveness. With lifestyle interventions it may be difficult to maintain compliance outside of a trial setting, which will reduce their effectiveness. Alternatively lifestyle interventions which lead to an increase in exercise and an improved diet may have benefits additional to the reduction of T2DM risk, which would not be the case for pharmacological interventions. Pharmacological interventions have disadvantages in that their use will mean the medicalisation of what is fundamentally a lifestyle problem rather than an illness, and it may encourage individuals to continue with unhealthy lifestyles, rather than directly address the cause of their ill-health. As discussed in chapter 4, pharmacological interventions have issues of adherence if side effects become problematic, especially as they have been shown be effective for only as long as they are taken, with the benefit stopping when the medication is withdrawn. Also, as the RCTs of interventions averaged only 4 to 5 years of follow-up, the long-term effectiveness of both lifestyle and pharmacological interventions in unknown.

Overall the model results were very robust to changes in model inputs, as explored through extensive sensitivity analyses and model extensions, as described in chapters 7 and 8. In the base case model compliance to both screening and interventions was assumed to be 100%. This is unrealistic, and would not be the case in a real-life setting. Therefore sensitivity analyses were carried out to assess the effects of reducing compliance to both screening and interventions, and although reducing compliance did reduce the cost-effectiveness of the three active screening strategies, the overall conclusions of the decision model were unchanged.

One-way sensitivity analyses on most of the model inputs only minimally affected the model outcomes and did not effect the model conclusions. For example, changing the prevalence of each glucose tolerant state, or screening test sensitivity, had little impact on the cost-effectiveness of each screening strategy. Therefore decisions on how a screening strategy in a U.K. setting should be structured can be based on the most viable and publicly acceptable options. Although as models with higher prevalences of IGT and undiagnosed T2DM (that is to

say populations with higher risk of T2DM) had better clinical outcomes, then possibly a screening programme should be aimed at targeting specific groups. Previous models that have assessed screening or intervention strategies often advocated the targeting of high risk groups, such as the obese, or individuals diagnosed with hypertension, as discussed in Chapter 5. The multi-way sensitivity analysis, carried out on costs of T2DM and prevalence of glucose tolerance status, had more of an impact on model outcomes, resulting in the three active screening strategies being less cost-effective when both costs and prevalence were high. This highlights the importance of multi-way sensitivity analyses, whereby model parameters may not impact on results during one-way sensitivity analyses, but may impact when changed simultaneously.

10.2 Discussion and limitations of this work

10.2.1 Methodological discussion

The model fitted here improved on current decision models in a number of ways. The model was comprehensive in that the complete model, including any meta-analyses carried out to compute pooled estimates, was contained within one framework. This has a number of advantages as already discussed. An additional advantage not fully explored by this thesis, is that it enables correlations to be included within the model, if one data source or trial contributes data to more than one part of the model. In this model, this issue only occurred when estimating the effectiveness of interventions, and the transition rate between IGT and T2DM, where the control groups of some of the randomised trials were utilised. As the same trials were used to inform both estimates, induced correlations could have been incorporated within the modelling framework. This was not investigated due to time constraints.

To model the effects of screening over a time horizon of fifty years, some assumptions were made. Model inputs were varied where data was available, so for example transition rates to death were increased over the model cycles, to simulate the effects of an aging population. Other model inputs were more difficult to adjust, for example the effectiveness of interventions. In the base case model the intervention effects remained constant over the full time horizon. As most of the data was taken from trials that had followed up participants for an average of just four or five years, the estimated intervention effects were effectively extrapolated over a much

longer time frame, but because the average time spent with diagnosed IGT was approximately seven years, the assumption should not have resulted in inaccurate results.

The decision model utilised meta-analyses methods to estimate pooled estimates for both lifestyle and pharmacological interventions. The use of meta-analyses, and meta-regression methodologies, raise a number of issues, including publication bias, the quality of included data, the prior distribution placed on the between study heterogeneity, and the use of summary data (Sutton et al., 2000). These were explored where appropriate and discussed in detail in Chapters 4 and 7. Sources of bias in the intervention meta-analyses were assessed, in terms of both publication and study quality. Also the possibility that one study may be greatly influencing the results was checked, by removing each study individually from the meta-analyses and comparing pooled estimates. Further model-data consistency checks could have been carried out, if time had allowed, whereby individual studies are removed from each meta-analyses, and then the meta-analyses used to predict the result of the missing study (Ades, 2002).

A problem highlighted by the extension considering different ethnic groups, is the use of averages within decision models and the problems this can cause. Combining data on different ethnic groups at the start of the model, and using average transition rates within the Markov model, resulted in inaccurate results. This is because where a population with different risks of an event are modelled using an average risk, the effects of these '*frailties*' can give misleading results as discussed in more detail in Chapter 8 (Hougaard, 1995). Although the problem with the ethnic data was overcome by combining results at the end of the model, it raises concerns for other averages utilised within the model. Where enough data was available checks were made, so for example the effect of baseline risk of T2DM and intervention effects were assessed for any interaction between the two. An alternate model structure, which would eliminate the risks involved with using averages, is to model a population by simulating each individual separately within the model and then combine the data at the end. Obviously this method is much more computational intensive and time-consuming than the more conventional decision model.

A limitation of the work carried out here, is that all model uncertainty was not considered. Model uncertainty may be structural, e.g. what branches to include in a decision model or states in a Markov Model, or statistical, e.g. the choice of statistical model to use to obtain input distributions for probabilities, hazard ratios, costs and utilities. Model uncertainty can be

investigated using sensitivity analyses or Bayesian model averaging methodologies (Hoeting et al., 1999).

Finally, the estimation of hazard ratios entered into the model was modelled using a posterior distribution. It has been suggested that, due to heterogeneity in patient groups, the mean treatment effect from a random effects meta-analysis will seldom be an appropriate representation of the efficacy expected in a future implementation (Ades et al., 2005, Welton et al., 2007). It has therefore been suggested that a more appropriate approach to modelling treatment effects would be to use a predictive distribution for a future treatment effect. The predictive distribution would be centred on the same value as a posterior, but would be wider; therefore its use would result in more model uncertainty.

10.2.2 Clinical discussion

Results show that screening for IGT and T2DM followed by interventions, and possible screening for T2DM alone, are potential cost-effective health policies. This raises the question as to whether such screening programmes would be viable in practice. In Chapter 7 the requirements necessary for a condition to be appropriate for screening were discussed, for example acceptable screening tests should be available. T2DM was found to meet most of the necessary requirements, although probably the most problematic issue is that benefits of early detection and treatment of undiagnosed T2DM are yet to be proved (Wareham and Griffin, 2007). Information to rectify this short fall in knowledge is currently being collated by on-going studies. Disadvantages and implications of screening are also important and should be considered before a health policy is implemented. Also the best use of limited NHS resources needs to be considered as perhaps it could be argued that the clinical management of people with established T2DM needs to be optimised before a screening programme is contemplated.

The model described here improves considerably on previous models assessing either interventions or screening for T2DM. The models were described in detail in Chapter 5, but briefly the major problems were the fact that often only a limited part of the treatment/disease pathway was modelled, and data sources were very limited, with the model usually based on data drawn from as few sources as possible. The model here considers the complete disease pathway, from NGT through to death, and by doing so is able to model the long-term outcomes of any screening strategy. It also allows a full assessment of an intervention strategy, as just assessing the effectiveness of interventions, without the costs of identifying the individuals. More importantly this is the first decision model that has directly compared a strategy of screening for T2DM alone or screening for IGT and T2DM in unison, which surely would be a key question that would need to be addressed if a screening policy was considered.

An evaluation of EVPI indicated there is still a need for further research in the area of T2DM screening and prevention. A great deal of research is already in progress, as T2DM is a growing health problem in the U.K., due to westernised lifestyles, and therefore is a health priority of national importance.

10.3 Further work

A number of extensions have already been carried out to extend the primary model, as described in Chapter 8. These included incorporating a mixed treatment comparison within the model,

extending the model to investigate the effects of implementing multiple screens rather than just a one-off screening, extracting further diagnostic data from the literature, and running the model for both an Asian population and a population of both Caucasian and Southern Asian origin.

Further work could include the extension of the model to incorporate more indirect evidence sources. The primary model utilised information on the transition rate from NGT to IGT, and IGT to T2DM, but additional, indirect, evidence is available on the transition from NGT to T2DM. This could be incorporated into the model to inform the NGT to IGT and IGT to T2DM transitions. Additionally, this model was based on the states of IGT and T2DM, but it would be interesting to also include the state of IFG. This was excluded in the primary model due to limited information on the disease pathway from IFG to T2DM, but the model could be extended when more data is available.

One of the extensions considered was an assessment of the EVPI. This could be extended in further work to assess the expected cost of perfect partial evidence (EVPPI). This takes EVPI one step further and investigates the costs of obtaining perfect information on specific groups of parameters. So for example the EVPPI for economic inputs to the model could be assessed, or the EVPPI for the transition rates. Obtaining this additional information would help to target future research, for example if the transition rates had a low EVPPI then it may not be worth investing further money in studies aimed at assessing incidence rates for IGT or T2DM.

The most difficult part of the decision model to develop was that of the effects of early diagnosis and treatment of T2DM through screening. This is because there are currently no large scale clinical trials that have assessed the impact of early diagnosis of T2DM, on factors such as complication rates and quality of life. For this model assumptions were made in that trial data from the UKPDS study was used and it was assumed individuals identified clinically, but who were placed on intensive treatment, would have similar complication rates and utility values to individuals diagnosed early. This was a strong assumption but made the most of the data available. The ADDITION study is currently assessing the impact of screening and early diagnosis of T2DM, and it would be interesting to update the model with these results when they are published.

The model developed for this thesis could be extended further by including more states, and increasing its complexity. For example, at the moment complication rates are only modelled in terms of their impact on costs and utilities, within the state of T2DM. The model could be extended to include states for each complication, including both microvascular and macrovascular outcomes, to fully assess the impact of screening on individual complication rates, and to fully model the inter-relationships between complications. To do this further clinical information on the disease pathway post-screening is needed. The model could also be improved by carrying out full systematic reviews for all model inputs, this was only done for some of the model inputs due to time limitations. Additionally, if IPD was available, a full assessment of the use of averages across sub-groups, or the impact of different screening strategies for different sub-groups, could be fully explored.

Other further work could involve applying the methodologies developed here to further clinical examples. This was discussed theoretically in Chapter 9, where the examples of screening for coronary heart disease, or for colon cancer were considered. As T2DM is associated with other components of the metabolic syndrome, such as hypertension, cholesterol and coronary heart disease, it would be interesting to extend the model to include screening and interventions for several components of the metabolic syndrome in unison. This would be very complex as not only would the model have to incorporate several screening tests, the interventions post screening would be likely to impact on several outcomes. Although it would be interesting to assess further examples, as different clinical conditions would raise different issues, as the literature searching, data extraction and model compilation developed here took three years to carry out thoroughly, applying the methodologies to another example in the same detail would obviously be a very time consuming endeavour.

10.4 Conclusions

In conclusion the comprehensive decision model developed for this thesis improved on both published models in the clinical area of type 2 diabetes, and published decision models in general. The results of the model provide evidence that a screening strategy for IGT and T2DM, allowing for preventative interventions to be given to individuals with IGT and early treatment to be provided for individuals with T2DM, would be a cost-effective screening strategy. Uncertainty still surrounds the cost-effectiveness of screening for T2DM alone and further research is required. Running decision models within a Bayesian, comprehensive framework, allows for model flexibility and has advantages over more conventional modelling techniques.

APPENDIX 1: SEARCH STRATEGIES

1.1 The search strategy for reviewing all previous meta-analyses and systematic reviews in the area of impaired glucose tolerance

This search strategy combines the systematic reviews and meta-analysis search filter from CRD report number 4, (appendix 1) with a search for IGT and related terms (from Cochrane protocol 'Lifestyle interventions for preventing type 2 diabetes mellitus'**)**

Medline (1966 to 20/12/04) and Embase (1980 to 20/12/04), Cochrane database- for protocols and systematic reviews (including DARE/HTA)

1. (systematic$ adj5 review$).tw.
2. (data adj synthesis).tw.
3. (published adj studies).ab.
4. (data adj extract$).ab.
5. metaanalysis/
6. meta analysis/
7. (meta analys$ or metaanalys$).tw.
8. comment.pt.
9. letter.pt.
10. editorial.pt.
11. animal/
12. human/
13. 11 not (11 and 12)
14. "prediabetic-state"/
15. ((prediabet$ or pre diabet$) adj5 state).ti,ab.
16. glucose intolerance/
17. (impaired glucose tolerance or glucose intoleran$ or insulin$ resist$).ti,ab.
18. impaired fasting glucose.ti,ab.
19. (IGT or IFG).tw.
20. (metabolic syndrome or syndrome x).mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
21. "hyperinsulinemia"/
22. (hyperinsulin$ or hyper insulin$).ti,ab.
23. glucose tolerance test.tw.
24. impaired fasting blood glucose.tw.
25. (impaired fasting glycaemia or impaired fasting glycemia).tw.
26. (impaired glucose stat$ or impaired glucose respons$ or impaired glucose control$).tw.
27. (impaired glucose regul$ or impaired glucose metab$).tw.
28. (impaired glucose homeost$ or reduced glucose metab$).tw.
29. (reduced glucose toleran$ or glucose intolerant$).tw.
30. (prediabet$ or praediabet$).tw.
31. (borderline diabet$ or mild diabet$).tw.
32. (impaired insulins secret$ or reduced insulin secret$).tw.
33. or/14-32
34. 33 not (8 or 9 or 10 or 13)
35. or/1-7
36. 34 and 35

1.2 Search strategy for the systematic review of interventions to prevent or delay type 2 diabetes in individuals with IGT

Search strategy

Both Medline (1966 to July, week 4, 2005) and Embase (1980 to week 32, 2005) databases were searched using the following search terms:

RCT filter

1. randomized controlled trial.pt.
2. controlled clinical trial.pt.
3. Randomized Controlled Trials/
4. random allocation/
5. double blind method/
6. single-Blind Method/
7. clincial trial.pt.
8. clinical trials/
9. clinical trial.tw.
10. ((singl$ or doubl$ or trebl$ or tripl$) and (mask$ or blind$)).tw.
11. PLACEBOS/
12. placebo$.tw.
13. random$.tw.
14. (clin$ adj5 trial$).ti,ab.
15. or/1-14
16. (animals not human).sh.
17. 15 not 16

Type II diabetes

18. diabetes-mellitus,-non-insulin-dependent/
19. insulin-resistance/
20. obesity-in-diabetes.mp. or Obesity in Diabetes/
21. (MODY or DM2 or NIDDM or IIDM).ti,ab.
22. (non insulin$ depend$ or noninsulin$ depend$).ti,ab.
23. (("typ$ 2" or typ$ II) adj10 (diabet$ or DM)).ti,ab.
24. (insulin$ defic$ adj5 relativ$).ti,ab.
25. (adult$ onset or matur$ onset or late$ onset).mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
26. 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25

Prevention combined with type II diabetes (using and)

27. "PREVENTIVE MEDICINE"/
28. "PREVENTIVE-HEALTH-SERVICES"/
29. (PREVENT$ or PROPHYLA$ or AVOID$ or DELAY$).ti,ab.
30. 26 and (27 or 28 or 29)

Exclusions
31. "dermatomyositis".mp. or DERMATOMYOSITIS/
32. "myotonic-dystrophy"/
33. exp Diabetes Insipidus/
34. mellitus.ti,ab.
35. 33 not (18 or 34)
36. (diabet$ adj5 (insipidus not mellitus)).ti,ab.
37. ((keto$ resist$ or nonketo$ or non keto$ or slow onset or stabl$) adj5 (diabet$ or DM or DM2)).mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
38. (fragil$ X or X linked).mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
39. (plurimetabolic$ syndrom$ or pluri metabolic$ syndrom$).mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
40. "PREGNANCY-IN-DIABETES".ti,ab.
41. (pregnan$ adj5 diabet$).ti,ab.
42. 31 or 32 or 35 or 36 or 37 or 38 or 39 or 40 or 41
43. 30 not 42

IGT and similar conditions
44. "prediabetic-state"/
45. ((prediabet$ or pre diabet$) adj5 state).ti,ab.
46. "glucose-intolerance"/
47. (impaired glucose tolerance or glucose intoleran$ or insulin$ resist$).ti,ab.
48. impaired fasting glucose.ti,ab.
49. (IGT or IFG).tw.
50. (metabolic syndrome or syndrome x).mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
51. "hyperinsulinemia"/
52. (hyperinsulin$ or hyper insulin$).ti,ab.
53. glucose tolerance test.tw.
54. impaired fasting blood glucose.tw.
55. (impaired fasting glycaemia or impaired fasting glycemia).tw.
56. (impaired glucose stat$ or impaired glucose respons$ or impaired glucose control$).tw.
57. (impaired glucose regul$ or impaired glucose metab$).tw.
58. (impaired glucose homeost$ or reduced glucose metab$).tw.
59. (reduced glucose toleran$ or glucose intolerant$).tw.
60. (prediabet$ or praediabet$).tw.
61. (borderline diabet$ or mild diabet$).tw.
62. (impaired insulin secret$ or reduced insulin secret$).tw.
63. 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62

IGT + ((diabetes and prevention)-exclusions)
64. 63 or 43
Above combined with RCT (using and)
65. 64 and 17

1.3 Search strategy for published economic evaluation studies of screening for impaired glucose tolerance and/or intervening to prevent type 2 diabetes

Search strategy for Medline and Embase:

1. (IGT or impaired glucose tolerance).mp.
2. ((IGT or impaired glucose tolerance) and screen$).mp.
3. (prevent$ and (non-insulin dependent diabetes or NIDDM or type II diabetes mellitus or type 2 diabetes mellitus or T2DM)).mp.
4. 1 or 2 or 3
5. (economic evaluation or Markov or cost-effective$).tw.
6. 4 and 5

MEDLINE (1966 to February Week 2 2005)

EMBASE (1980 to week 7 2005)

Search strategy for Cochrane:

((IGT or impaired glucose tolerance) or ((IGT or impaired glucose tolerance) and screening) or (prevention and (non-insulin dependent diabetes or NIDDM or type II diabetes))) and (economic evaluation or Markov or cost-effectiveness)

in All Fields, from 1800 to 2005 in The Cochrane Database of Systematic Reviews"

The **ScHARR** website 'reviewing modelling methods for the evaluation of screening programmes' was also searched.

1.4 Search strategy for studies that had evaluated screening tests for type 2 diabetes and/ or impaired glucose tolerance

Search Strategy for Medline and Embase:

1. sensitivity.mp. or exp "Sensitivity and Specificity"/ (237893)
2. DIAGNOSIS/ (990)
3. specificity.mp. (217633)
4. exp DIAGNOSIS/ (1227511)
5. 1 or 3 or 4 (1436847)
6. Glucose Intolerance/ or impaired glucose tolerance.mp. (3086)
7. IGT.mp. (894)
8. 6 or 7 (3161)
9. 5 and 8 (2271)
10. screen$.mp. (124704)
11. 9 and 10 (266)
12. limit 11 to (english language and abstracts) (217)

MEDLINE (1966 to January Week 4 2005)

EMBASE (1980 to week 4 2005)

APPENDIX 2: FORM USED TO CALCULATE JADAD QUALITY SCORE

Study name and paper used for assessment of quality (study design paper):

	Yes=1, No=0
1. Was the study described as randomised? (this includes words such as randomly, random and randomisation)	
Give 1 additional point if the method to generate the sequence of randomisation was described and it was appropriate (e.g. random number tables, computer generated).	
Deduct 1 point if the method of randomisation was described but was inappropriate (e.g. alternate allocation, date of birth, hospital number).	
2. Was the study described as double blind?	
Give 1 additional point if the method of double blinding was described and it was appropriate (identical placebo, active placebo, dummy etc.)	
Deduct 1 point if the study was described as double blind but the method of blinding was inappropriate (e.g. comparison of tablet vs. injection with no double dummy).	
3. Was there a description of withdrawals and dropouts?	
Total	

Guidelines for Assessment

1. Randomisation
A method to generate the sequence of randomisation will be regarded as appropriate if it allowed each study participant to have the same chance of receiving each intervention and the investigators could not predict which treatment was next. Methods of allocation using date of birth, date of admission, hospital numbers, or alternation should not be regarded as appropriate.

2. Double blinding
A study must be regarded as double blind if the words 'double blind' are used. The method will be regarded as appropriate if it is stated that neither the person doing the assessments or the study participant could identify the intervention being assessed., or if in the absence of such a statement the use of active placebos, identical placebos, or dummies is mentioned.

3. Withdrawals and dropouts
Participants who were included in the study but did not complete the observation period or who were not included in the analysis must be described. The number and the reasons for withdrawal in each group must be stated. If there were no withdrawals, it should be stated in the article. If there is no statement on withdrawals, this item must be given no points.

Additional Question: Was allocation concealed? **Yes/No/Unclear**

APPENDIX 3: WinBUGS CODE

3.1 Code for the three intervention meta-analyses of lifestyle, anti-diabetic agents and anti-obesity agents

Lifestyle

Model{
for(i in 1:n)
{
prec[i] <- 1/(SElogHR[i]*SElogHR[i])
LogHR[i] ~ dnorm(theta[i],prec[i])
theta[i] ~ dnorm(mu,tau)
HR[i] <- exp(theta[i])
}
mu ~ dnorm(0,0.001)
tau <- 1/(sd*sd)
sd ~ dunif(0,2)
pooledHR <- exp(mu)
}
Data
(Da Qing (diet), Wein, Da Qing (exercise), DPP (lifestyle), DPS, Da Qing (exercise + diet), Kosaka, Jarrett, Liao (diet),Fang (d&e), IDDP(d&e), Tao(d&e))

list(LogHR=c(-0.451,-0.4557,-0.638,-0.8675,-0.916,-0.492,-1.244,-0.1669,-0.6619,-0.288,-0.4732,-1.1980)
SElogHR=c(0.223,0.298,0.226,0.1084,0.216,0.226,0.603,0.3873,1.225,0.387,0.2005,0.5718),n=12)

Inits
list(mu=0,sd=0.5,theta=c(0,0,0,0,0,0,0,0,0,0,0,0))

Oral anti-diabetic agents

Model
{
for(i in 1:n)
{
prec[i] <- 1/(SElogHR[i]*SElogHR[i])
LogHR[i] ~ dnorm(theta[i],prec[i])
theta[i] ~ dnorm(mu,tau)
HR[i] <- exp(theta[i])
}
mu ~ dnorm(0,0.001)
tau <- 1/(sd*sd)
sd ~ dunif(0,2)

pooledHR <- exp(mu)
}
Data

(DPP (metformin), Li (metformin), Pan (acarbose), STOP (acarbose), Jarrett (phenformin), Fang(acarbose), Fang(flumamine), Eriksson(glipizide), IDDP (metformin))

list(LogHR=c(-0.3711,-0.7172,-0.506,-0.2877,0.0128,-1.3106,-0.8353,-1.7430,-0.4292), SElogHR=c(0.0959,0.7071,0.4754,0.091,0.3852,0.5477,0.4944,1.0954,0.1969),n=9)

Inits

list(mu=0,sd=0.5,theta=c(0,0,0,0,0,0,0,0,0))

Anti-obesity agents

Model

```
{
for(i in 1:n)
{
prec[i] <- 1/(SElogHR[i]*SElogHR[i])
LogHR[i] ~ dnorm(theta[i],prec[i])
theta[i] ~ dnorm(mu,tau)
HR[i] <- exp(theta[i])
}
mu ~ dnorm(0,0.001)
tau <- 1/(sd*sd)
sd ~ dunif(0,10)
pooledHR <- exp(mu)
}
```

Data

(Xendos, Heymsfield)

list(LogHR=c(-0.7298,-0.9447), SElogHR=c(0.3066,0.3536),n=2)

Inits

list(mu=0,sd=0.5,theta=c(0,0))

3.2 Code for calculating the absolute difference between treatments and the NNT

Model

```
{
for(i in 1:17)
{
prec[i] <- 1/(var[i])
logH[i] <- log(H[i]/100)
logH[i] ~ dnorm(theta[i],prec[i])
theta[i] ~ dnorm(mu1,tau1)
}

logdelta ~ dnorm(mu2,tau2)
mu1 ~ dnorm(0,0.001)
tau1 ~ dgamma(0.01, 0.01)

baseH<- exp(mu1)

CI.B <- 1- exp(-baseH*5)
CI.I  <- 1- exp(-baseH*exp(logdelta)*5)
diff <- CI.B - CI.I
NNT <- 1/diff
}
```

Data

#17trials:DaQing,Jarrett,DPS,Pan,Li,Heymsfield,STOP,Xendos,Wein,Fan,Tao,Kosako,Fang,Liao,DPP,Eriksson, IDDP

Anti-diabetic agents: mu2=-0.358, tau2=564
Anti-obesity agents: mu2=-0.821, tau2=90.1
Lifestyle: mu2=-0.666, tau2=674
Herbal: mu2=-1.143, tau2=1.7

List

(mu2=as detailed above,tau2=as detailed above,
H=c(15.7,2.61,7.38,30.03,7.14,4.77,12.63,5.91,7.05,13.33,16.99,2.55,10,3.08,11,23.8,18.3),
var=c(0.011,0.1667,0.017,0.08,0.1667,0.25,0.0035,0.02,0.037,0.33,0.077,0.03
,0.0667,0.5,0.003,0.2,0.0133))

Inits

list(mu1=0,tau1=1,logdelta=0.1,theta=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0))

3.3 Code for the mixed treatment comparison analysis

Notes

16 trials *(including four that considered both lifestyle and pharmacological interventions),*
39 data points
Treatments:

1=control (placebo and/or standard diet and exercise advice)
2=lifestyle
3=anti-diabetic
4=both lifestyle and antidiabetic
5=anti-obesity
6=pharmacological where there are 2 pharm trts in 1 trial

model{

```
# Model for log-hazards of diabetes
        for(i in 1:16){
                        w[i,1] <- 0
                        delta[i,t[i,1]] <- 0
                        mu[i] ~ dnorm(0,.0001)
                #vague priors for baseline

                        for (k in 1:na[i]){
                                LH[i,t[i,k]] ~ dnorm(theta[i,t[i,k]],prec[i,t[i,k]])
                                prec[i,t[i,k]] <- 1/(se[i,t[i,k]]*se[i,t[i,k]])
                                theta[i,t[i,k]]<-mu[i] + delta[i,t[i,k]]
                                }

                        for (k in 2:na[i]){
                                delta[i,t[i,k]] ~ dnorm(mu2[i,t[i,k]],taud[i,t[i,k]])
                                mu2[i,t[i,k]] <- d[t[i,k]] - d[t[i,1]] + sw[i,k]
                                taud[i,t[i,k]] <- tau*2*(k-1)/k
                                w[i,k] <- (delta[i,t[i,k]] - d[t[i,k]]  + d[t[i,1]])/(k-1)
                                sw[i,k] <- sum(w[i,1:k-1])/(k-1)
                        }
        }
        d[1] <- 0
        for (k in 2:5){
                        d[k] ~ dnorm (0,.0001)
                                }
        d[6] <- d[3]
        tau ~ dgamma(0.001, 0.001)


# interaction between lifestyle and pharmacological in factorial trials
int <- d[4] - (d[3] + d[2])
expint <- exp(int)

# All pairwise log hazard ratios and hazard ratios
        for (c in 1:4){
```

```
                for (k in (c+1):5){
                        LogHR[c,k] <- d[k] - d[c]
                        HR[c,k] <- exp(LogHR[c,k])

                        }
                }
prob <- step(LogHR[2,3])
```

3.4 Meta-regression models fitted to explore sources of heterogeneity in the transition rate from IGT to T2DM

Model #(3 models adjusted for age/gender/ethnicity)

```
for(i in 1:24){
        prec[i] <- 1/(SEID[i]*SEID[i])
        LogIRID[i] ~ dnorm(theta[i],prec[i])
        #theta[i] <- theta.star + beta*(age[i]-age.bar)
        #theta[i] <- theta.star + beta*(male[i]-male.bar)
        theta[i] <- theta.star + beta*(ethnic.risk[i])
                                }
        beta ~ dnorm(0,0.0001)
        theta.star ~ dnorm(0,0.0001)
        #age.bar <- mean(age[])
        #male.bar <- mean(male[])
        IGTtoT2DM <- exp(theta.star)    #pooled rate per 100 pyrs at the mean age/percent male or if
ethnic.risk=1
}
```

(Study order, n=24: Baltimore, Rancho, SanAntonio, Nauru, SanLuis, Pima, Hoorn, Ely, DaQing, DPP, DPS, Kosako, Jarrett, Liao, Wein, Tao, Fang, IDDP, Li, Pan, STOP, Erikkson,Addition,Bruneck)

Data-.age

```
list(LogIRID=c(1.528,1.386,1.468,1.837,1.987,2.167,1.77,0.872, 2.76, 2.398,1.998,
0.936,1.187,1.124,1.952, 2.833, 2.303, 3.272, 2.683, 3.402, 2.533, 3.17,2.93,0.955),
SEID=c(0.099, 0.143, 0.097, 0.084, 0.156, 0.048, 0.167, 0.242, 0.106, 0.302, 0.130, 0.174, 0.258, 0.707,
0.192, 0.277, 0.258, 0.116, 0.408, 0.289, 0.059, 0.447,0.103,0.243),
age=c(52.9, 68, 48.3, 37.3, 59.7, 43.2, 60.3, 54.9, 45.5, 50.4, 55, 51.5, 56.7, 54, 38.7, 51, 48.7, 45.9, 49.5,
54.5, 54.5, 56.5,59.8,59))
```

Data-.male

```
list(LogIRID=c(1.528,1.386,1.468,1.837,1.987,2.167,1.77,0.872, 2.76, 2.398,1.998,
0.936,1.187,1.124,1.952, 2.833, 2.303, 3.272, 2.683, 3.402, 2.533, 3.17,2.93,0.955),
SEID=c(0.099, 0.143, 0.097, 0.084, 0.156, 0.048, 0.167, 0.242, 0.106, 0.302, 0.130, 0.174, 0.258, 0.707,
0.192, 0.277, 0.258, 0.116, 0.408, 0.289, 0.059, 0.447,0.103,0.243),
male=c(61,36, 36, 46, 40, 37, 45, 42, 53, 32, 33, 100, 100, 45, 0, 57, 55, 79, 71, 40, 49, 26, 45,51))
```

Data.ethnicity

```
list(LogIRID=c(1.528,1.386,1.468,1.837,1.987,2.167,1.77,0.872, 2.76, 2.398,1.998,
0.936,1.187,1.124,1.952, 2.833, 2.303, 3.272, 2.683, 3.402, 2.533, 3.17,2.93,0.955),
SEID=c(0.099, 0.143, 0.097, 0.084, 0.156, 0.048, 0.167, 0.242, 0.106, 0.302, 0.130, 0.174, 0.258, 0.707,
0.192, 0.277, 0.258, 0.116, 0.408, 0.289, 0.059, 0.447,0.103,0.243),
ethnic.risk=c(1, 1, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1,1,1))

#ethnic.risk, 1=low risk, 2=high risk
#ethnic=c(1, 1, 2, 3, 4, 5, 1, 1, 6, 1, 1, 7, 1, 7, 1, 6, 6, 8, 6, 6, 1, 1,1,1)
#(mostly 1=caucasian, 2=mexican American, 3=micronesian, 4=hispanic, 5=pima indian, 6=chinese,
7=Japanese, 8=Asian Indians)
```

Inits list(beta=0,theta.star=0)

3.5 Code for the assessment of baseline risk on the intervention effect

Model

```
{
            for(i in 1 : 9)
          {
                    d1[i] ~ dpois(mu1[i])
                      d2[i] ~ dpois(mu2[i])
                    log(mu1[i]) <- log(Py1[i]) + alpha[i]
                    log(mu2[i]) <- log(Py2[i]) + alpha[i]  + delta[i]                    #alpha=log(mu) -
log(py)=log rate in control grp
   delta[i] <- delta.star[i] + beta*(alpha[i]-alpha.bar)              # delta.star= logIRR at mean alpha
   delta.star[i] ~ dnorm(gamma,prec)
                    alpha[i] ~ dnorm(0.0,0.0001)
}
   gamma ~ dnorm(0.0,0.0001)
   beta ~ dnorm(0.0,0.0001)
   prec <- 1/(tau*tau)
   tau ~ dunif(0,10)
   IRR <- exp(gamma)
   alpha.bar <- mean(alpha[])
          }
```

Data for anti-diabetic agents

d1[]	d2[]	Py1[]	Py2[]
11	7.8	1.00	1.00
14	13	4.325	4.575
6	3	0.41	0.42
12	7	0.3996	0.3866
285	221	22.638	22.506
7.5	9	0.75	2.075
7.5	6	0.75	2.225
74.8	53.9	2.838	3.1815
5	1	0.21	0.24

Data for lifestyle interventions

d1[]	d2[]	Py1[]	Py2[]
30	62	1.8897	6.20
30	58	1.8897	6.988
30	58	1.8897	6.042
11	4.8	1.00	1.00
59	27	8.00	8.192
33	3	12.94	3.88
15	12	4.575	4.325
2	1	0.65	0.63
27	26	3.832	4.7368
13	4	0.765	0.78
15	12	1.50	1.60
74.8	52.3	2.838	3.2055

Initial Values

```
list(alpha=c(0,0,0,0,0,0,0,0,0),
```

```
delta.star=c(0,0,0,0,0,0,0,0,0),gamma=0,tau=0.5,beta=0)
```

3.6 Code for the full decision model

```
model{
#Screening decision tree
# status 1 NGT, 2 IGT, 3 T2DM, N[status] = N, T+ shows a positive result for both cut-offs, T- a negative
result for both cut-offs and T0 a positive result for the high cut-off (T2DM only) and a negative result for
the low cut-off  (T2DM and IGT). r[status,1] pr(T- |status),   r[status,2] pr(T0|status),   r[status,3]
pr(T+|status)

for (status in 1:3) {
                    N[status] <- sum(x[status,1:3])
                    x[status,1:3] ~ dmulti(r[status,1:3],N[status])
                    r[status,1:3] ~ ddirch(prior_r[status,1:3])
                    prev[1:3] ~ ddirch(preva[1:3])

#positive and negative predictive probabilities, neglow  n(status | T-.low), poslow  n(status | T+.low),
neghigh n(status |T-.high ), poshigh  n(status | T+.high)

for (status in 1:3) {
        neglow[status] <- (x[status,1]/N[status]*prev[status])*Nscreen
        poslow[status] <- ((x[status,2]/N[status]*prev[status])+(x[status,3]/
                N[status]*prev[status]))*Nscreen
        neghigh[status] <((x[status,1]/N[status]*prev[status])+(x[status,2]/
                N[status]*prev[status])) *Nscreen
        poshigh[status] <- (x[status,3]/N[status]*prev[status])*Nscreen
                                                                }

# Starting numbers for Markov model for each screening strategies, high cut-off [1], low cut-off &
intervention [2], none [3]. D=detected, U=undetected
        NGT[1] <- poshigh[1]+neghigh[1]
        IGTD[1] <- poshigh[2]
        IGTU[1] <- neghigh[2]
        T2DMD[1] <- poshigh[3]
        T2DMU[1] <- neghigh[3]

        NGT[2] <- poslow[1]+neglow[1]
        IGTD[2] <- poslow[2]
        IGTU[2] <- neglow[2]
        T2DMD[2] <- poslow[3]
        T2DMU[2] <- neglow[3]

        NGT[3] <- prev[1]*Nscreen
        IGTU[3] <- prev[2]*Nscreen
        T2DMU[3] <- prev[3]*Nscreen

#Starting numbers in each state (strategy,state,time=1):
#strategy 1=DM screening, 2=IGT & DM screening, lifestyle intervention, 3=IGT & DM screening,
metformin intervention, 4=no screening
#states, 1=NGT, 2=IGTu, 3=IGTd, 4=T2DMu, 5=T2DMscreen.d, 6=T2DMclin.d  7=death
#agegrp, 1=45-54, 2=55-64, 3=65-74, 4=75-84, 5=85+

                    number[1,1,1] <- cut(NGT[1])
```

```
number[1,2,1] <- cut(IGTU[1])
number[1,3,1] <- cut(IGTD[1])
number[1,4,1] <- cut(T2DMU[1])
number[1,5,1] <- cut(T2DMD[1])
number[1,6,1] <- 0
number[1,7,1] <- 0
number[2,1,1] <- cut(NGT[2])
number[2,2,1] <- cut(IGTU[2])
number[2,3,1] <- cut(IGTD[2])
number[2,4,1] <- cut(T2DMU[2])
number[2,5,1] <- cut(T2DMD[2])
number[2,6,1] <- 0
number[2,7,1] <- 0
number[3,1,1] <- cut(NGT[2])
number[3,2,1] <- cut(IGTU[2])
number[3,3,1] <- cut(IGTD[2])
number[3,4,1] <- cut(T2DMU[2])
number[3,5,1] <- cut(T2DMD[2])
number[3,6,1] <- 0
number[3,7,1] <- 0
number[4,1,1] <- cut(NGT[3])
number[4,2,1] <- cut(IGTU[3])
number[4,3,1] <- 0
number[4,4,1] <- cut(T2DMU[3])
number[4,5,1] <- 0
number[4,6,1] <- 0
number[4,7,1] <- 0
```

 # Markov model

```
for (strategy in 1:4){                          #transition probabilities
        for (agegrp in 1:5){
                trans[strategy,agegrp,1,1] <- 1 - (trans[strategy,agegrp,1,2] +
                                                   trans[strategy,agegrp,1,7])
#transition from 1 (NGT) to 2 (IGTu) different by age, see below
        trans[strategy,agegrp,1,3] <- 0
        trans[strategy,agegrp,1,4] <- 0
        trans[strategy,agegrp,1,5] <- 0
        trans[strategy,agegrp,1,6] <- 0
        trans[strategy,agegrp,2,1] <- 0
        trans[strategy,agegrp,2,2] <- 1 - (trans[strategy,agegrp,2,4] + trans[strategy,agegrp,2,7])
        trans[strategy,agegrp,2,3] <- 0
        trans[strategy,agegrp,2,4] <- 1 - exp(-IGTtoT2DM)
        trans[strategy,agegrp,2,5] <- 0
        trans[strategy,agegrp,2,6] <- 0
        trans[strategy,agegrp,3,1] <- 0
        trans[strategy,agegrp,3,2] <- 0
        trans[strategy,agegrp,3,3] <- 1 - (trans[strategy,agegrp,3,5] + trans[strategy,agegrp,3,7])
        trans[strategy,agegrp,3,4] <- 0
#transition from 3 (IGTd) to 5 (T2DMd) different for each strategy, see below
        trans[strategy,agegrp,3,6] <- 0
        trans[strategy,agegrp,4,1] <- 0
```

```
trans[strategy,agegrp,4,2] <- 0
trans[strategy,agegrp,4,3] <- 0
trans[strategy,agegrp,4,4] <- 1 - (trans[strategy,agegrp,4,6] + trans[strategy,agegrp,4,7])
trans[strategy,agegrp,4,5] <- 0
trans[strategy,agegrp,4,6] <- 1 - exp(-T2ud)
trans[strategy,agegrp,5,1] <- 0
trans[strategy,agegrp,5,2] <- 0
trans[strategy,agegrp,5,3] <- 0
trans[strategy,agegrp,5,4] <- 0
trans[strategy,agegrp,5,5] <- 1 - trans[strategy,agegrp,5,7]
trans[strategy,agegrp,5,6] <- 0
trans[strategy,agegrp,6,1] <- 0
trans[strategy,agegrp,6,2] <- 0
trans[strategy,agegrp,6,3] <- 0
trans[strategy,agegrp,6,4] <- 0
trans[strategy,agegrp,6,5] <- 0
trans[strategy,agegrp,6,6] <- 1 - trans[strategy,agegrp,6,7]
trans[strategy,agegrp,7,1] <- 0
trans[strategy,agegrp,7,2] <- 0
trans[strategy,agegrp,7,3] <- 0
trans[strategy,agegrp,7,4] <- 0
trans[strategy,agegrp,7,5] <- 0
trans[strategy,agegrp,7,6] <- 0
trans[strategy,agegrp,7,7] <- 1  }


trans[strategy,1,1,2] <- 1 - exp(-NGTtoIGT1)          #NGT to IGT increase with age (<65s)
trans[strategy,2,1,2] <- 1 - exp(-NGTtoIGT1)
trans[strategy,3,1,2] <- 1 - exp(-NGTtoIGT2)                                    #(>65s)
trans[strategy,4,1,2] <- 1 - exp(-NGTtoIGT2)
trans[strategy,5,1,2] <- 1 - exp(-NGTtoIGT2)
trans[strategy,1,1,7] <- 1- exp(-0.0032)
                                        #mortality probabilities- increase as age increases
trans[strategy,1,2,7] <- 1- exp(-0.0032)
trans[strategy,1,3,7] <- 1- exp(-0.0032)
trans[strategy,1,4,7] <- 1- exp(-0.0032*HRdeathdiab*pow(HRhba1c,(hba1cu-hba1cc)))
trans[strategy,1,5,7] <- 1- exp(-0.0032*HRdeathdiab*pow(HRhba1c,(hba1cs-hba1cc)))
trans[strategy,1,6,7] <- 1- exp(-0.0032*HRdeathdiab)
trans[strategy,2,1,7] <- 1- exp(-0.0084)
trans[strategy,2,2,7] <- 1- exp(-0.0084)
trans[strategy,2,3,7] <- 1- exp(-0.0084)
trans[strategy,2,4,7] <- 1- exp(-0.0084*HRdeathdiab*pow(HRhba1c,(hba1cu-hba1cc)))
trans[strategy,2,5,7] <- 1- exp(-0.0084*HRdeathdiab*pow(HRhba1c,(hba1cs-hba1cc)))
trans[strategy,2,6,7] <- 1- exp(-0.0084*HRdeathdiab)
trans[strategy,3,1,7] <- 1-exp(-0.0236)
trans[strategy,3,2,7] <- 1-exp(-0.0236)
trans[strategy,3,3,7] <- 1-exp(-0.0236)
trans[strategy,3,4,7] <- 1-exp(-0.0236*HRdeathdiab*pow(HRhba1c,(hba1cu-hba1cc)))
trans[strategy,3,5,7] <- 1-exp(-0.0236*HRdeathdiab*pow(HRhba1c,(hba1cs-hba1cc)))
trans[strategy,3,6,7] <- 1-exp(-0.0236*HRdeathdiab)
trans[strategy,4,1,7] <- 1-exp(-0.0609)
trans[strategy,4,2,7] <- 1-exp(-0.0609)
trans[strategy,4,3,7] <- 1-exp(-0.0609)
```

```
        trans[strategy,4,4,7] <- 1-exp(-0.0609*HRdeathdiab*pow(HRhba1c,(hba1cu-hba1cc)))
        trans[strategy,4,5,7] <- 1-exp(-0.0609*HRdeathdiab*pow(HRhba1c,(hba1cs-hba1cc)))
        trans[strategy,4,6,7] <- 1-exp(-0.0609*HRdeathdiab)
        trans[strategy,5,1,7] <- 1-exp(-0.1568)
        trans[strategy,5,2,7] <- 1-exp(-0.1568)
        trans[strategy,5,3,7] <- 1-exp(-0.1568)
        trans[strategy,5,4,7] <- 1-exp(-0.1568*HRdeathdiab*pow(HRhba1c,(hba1cu-hba1cc)))
        trans[strategy,5,5,7] <- 1-exp(-0.1568*HRdeathdiab*pow(HRhba1c,(hba1cs-hba1cc)))
        trans[strategy,5,6,7] <- 1-exp(-0.1568*HRdeathdiab)
                                                                    }
for (agegrp in 1:5){
        trans[1,agegrp,3,5] <- 1 - exp(-IGTtoT2DM)
        trans[2,agegrp,3,5] <- 1 - exp(-IGTtoT2DM*LifeHR)
        trans[3,agegrp,3,5] <- 1 - exp(-IGTtoT2DM*PharmHR)
        trans[4,agegrp,3,5] <- 1 - exp(-IGTtoT2DM)
                        }

for (strategy in 1:4){
for (state in 1:7){
for (time in 2:10) {
 number[strategy,state,time] <- inprod(number[strategy,1:7, time-1], trans[strategy,1,1:7,state])}
for (time in 11:20) {
 number[strategy,state,time] <- inprod(number[strategy,1:7, time-1], trans[strategy,2,1:7,state])}
for (time in 21:30) {
 number[strategy,state,time] <- inprod(number[strategy,1:7, time-1], trans[strategy,3,1:7,state])}
for (time in 31:40) {
number[strategy,state,time] <- inprod(number[strategy,1:7, time-1], trans[strategy,4,1:7,state])}
for (time in 41:50) {
number[strategy,state,time] <- inprod(number[strategy,1:7, time-1], trans[strategy,5,1:7,state])}

total[strategy,state] <- sum(number[strategy,state,1:horizon]) }
                                        #sum up the total person years spent in each state
```

#Clinical Outcomes

```
totalIGT[strategy] <- sum(total[strategy,2:3])
totalT2DM[strategy] <- sum(total[strategy,4:6])
totallife[strategy] <- sum(total[strategy,1:6])
totaldiabfree[strategy] <- sum(total[strategy,1:3])
totallifedis[strategy] <- sum(lifedis[strategy,1:50])
totaldiabfreedis[strategy] <- sum(diabfreedis[strategy,1:50])
        for (time in 1:50) {
                lifedis[strategy,time] <- sum(number[strategy,1:6,time])*pow(0.965,time-1)
                diabfreedis[strategy,time] <- sum(number[strategy,1:3,time])*pow(0.965,time-1) }

difflife[strategy] <- totallife[strategy] - totallife[4]
diffdiabfree[strategy] <- totaldiabfree[strategy] - totaldiabfree[4]
diffcases[strategy] <- totalcases[strategy] - totalcases[4]
difflifedis[strategy] <- totallifedis[strategy] - totallifedis[4]
diffdiabfreedis[strategy] <- totaldiabfreedis[strategy] - totaldiabfreedis[4]
```

```
#Diabetes cases
casesT2DM[strategy,1] <- number[strategy,4,1] + number[strategy,5,1] + number[strategy,6,1]
for (time in 2:horizon) {
        allcases[strategy,time] <- (number[strategy,2, time-1]*trans[strategy,1,2,4]) +
                        (number[strategy,3,time-1]*trans[strategy,1,3,5])
        clincases[strategy,time] <- number[strategy,4, time-1]*trans[strategy,1,4,6]
        screencases[strategy,time] <- number[strategy,3, time-1]*trans[strategy,1,3,5] }
        allcases[strategy,1] <- number[strategy,4,1] + number[strategy,5,1]
        clincases[strategy,1] <- number[strategy,6,1]
        screencases[strategy,1] <- number[strategy,5,1]
        totalcases[strategy] <- sum(allcases[strategy,1:horizon])
        totalclincases[strategy] <- sum(clincases[strategy,1:horizon])
        totalscreencases[strategy] <- sum(screencases[strategy,1:horizon])      }
```

```
#Costs
# coded as costs[strategy,i] where i=1 for screening costs, 2 for intervention costs, 3 for undiagnosed
diabetes costs and 4 for diagnosed diabetes costs
# screening tests costs (cost[strategy,1]): FPG=40p, OGTT=£1.30, everyone who tests positive with the
fpg test will have an OGTT. Nurse costs £26 per hour. Costs calculated for the each screening strategy
(high and low cut-offs)

cost[1,1] <- (sum(poshigh[1:3])*(1.70+(26*(timefpg+timeogtt)/60)))
                + (sum(neghigh[1:3])*(0.40+(26*timefpg/60)))
cost[2,1] <- (sum(poslow[1:3])*(1.70+(26*(timefpg+timeogtt)/60)))
        +(sum(neglow[1:3])*(0.40+(26*timefpg/60)))
cost[3,1] <- (sum(poslow[1:3])*(1.70+(26*(timefpg+timeogtt)/60)))
        +(sum(neglow[1:3])*(0.40+(26*timefpg/60)))
cost[4,1] <- 0

timefpg ~ dgamma(12.5,2.5)          #E(timefpg)=r/m=5 mins and Var(timefpg)=r/m^2= 2
timeogtt ~ dgamma(150,5.5)          #E(timeogtt)=r/m=30 mins and Var(timeogtt)=r/m^2= 5


#intervention costs (cost[strategy,2])  lifestyle £398 in first yr, £280 thereafter , £16.10 metformin, per
person per yr (this will need to be updated if a rescreening strategy is assessed so that more people move
to IGTd)
cost[1,2] <- 0
cost[2,2] <- (398*number[2,3,1]) + ((total[2,3] - number[2,3,1])*280)
cost[3,2] <- total[3,3]*16.1
cost[4,2] <- 0

# costs of undiagnosed diabetes (£133.98 in last year, £21.68 previous years)
for (strategy in 1:4){
  for (time in 1:horizon-1){
        costT2DMu[strategy,time] <- ((number[strategy,4,time] -
                clincases[strategy,time+1])*21.68) + (clincases[strategy,time+1]*133.98)}
        costT2DMu[strategy,horizon] <- number[strategy,4,horizon]*21.68
                cost[strategy,3] <- sum(costT2DMu[strategy,1:horizon])      }

# costs of detected diabetes, screen diagnosed and clinically diagnosed
```

```
        for (strategy in 1:4){
                for (time in 1:horizon){
                        costT2DMs[strategy,time] <- number[strategy,5,time]*diabcosts
                        costT2DMc[strategy,time] <- number[strategy,6,time]*diabcostc   }
cost[strategy,4] <- sum(costT2DMs[strategy,1:horizon]) + sum(costT2DMc[strategy,1:horizon])
        }
        diabcosts ~ dnorm(2490,0.0004)
        diabcostc ~ dnorm(2756,0.0003)

for (strategy in 1:4)
        { totalcost[strategy] <- sum(cost[strategy,1:4]) }        # total costs for each strategy


# discounted costs (3.5% per year)
#screening costs happen at start so don't need to be discounted
#discount intervention costs (for strategies 2 & 3) and diagnosed diabetes costs (for all 4 strategies)
discost[2,2,1] <- (398*number[2,3,1])
for (time in 2:horizon) {
        discost[2,2,time] <- number[2,3,time]*(280*pow(0.965,time-1))  }
for (time in 1:horizon){
        discost[3,2,time] <- number[3,3,time]*(16.1*pow(0.965,time-1))
    for (strategy in 1:4) {
        discostu[strategy,time] <- costT2DMu[strategy,time]*pow(0.965,time-1)
        discosts[strategy,time] <- costT2DMs[strategy,time]*pow(0.965,time-1)
        discostc[strategy,time] <- costT2DMc[strategy,time]*pow(0.965,time-1)
        discost[strategy,3,time] <- discostu[strategy,time] + discosts[strategy,time] +
                                        discostc[strategy,time] }}
sumdiscost[1,2] <- 0
sumdiscost[2,2] <- sum(discost[2,2,1:horizon])
sumdiscost[3,2] <- sum(discost[3,2,1:horizon])
sumdiscost[4,2] <- 0
sumdiscost[1,3] <- sum(discost[1,3,1:horizon])
sumdiscost[2,3] <- sum(discost[2,3,1:horizon])
sumdiscost[3,3] <- sum(discost[3,3,1:horizon])
sumdiscost[4,3] <- sum(discost[4,3,1:horizon])

for (strategy in 1:4) {
    totdiscost[strategy]<-cost[strategy,1] +sumdiscost[strategy,2] +sumdiscost[strategy,3] }
```

#Utilities

```
#calculating yearly utilities from UKPDS complications model
# Ignores renal as no utility decrement reported in UKPDS 62
# Assumes utility decrements associated  with complications are Normal

for (comp in 1:6) {
 for (det in 1:2) {
        s[det,comp,1] <- 1          # All start complication-free
        c[det,comp,1] <- 0  }
        lambda[comp] <- exp(beta0[comp])          # Distribution on scale parameter
        beta0.p[comp] <- 1/pow(beta0.se[comp],2)
        beta0[comp] ~ dnorm(beta0.m[comp],beta0.p[comp])
```

```
        rho[comp] <- exp(gamma[comp])          # Distribution on shape parameter
        gamma.m[comp] <- log(rho.m[comp])
        gamma.p[comp] <- pow(rho.m[comp],2)/pow(rho.se[comp],2)
        gamma[comp] ~ dnorm(gamma.m[comp],gamma.p[comp])

        uy.p[comp] <- 1/pow(uy.se[comp],2)
        uy[comp] ~ dnorm(uy.m[comp],uy.p[comp])
        upy.p[comp] <- 1/pow(upy.se[comp],2)
        upy[comp] ~ dnorm(upy.m[comp],upy.p[comp])

        gender.p[comp] <- 1/pow(gender.se[comp],2)
        gender[comp] ~ dnorm(gender.m[comp],gender.p[comp])
        hba1c.p[comp] <- 1/pow(hba1c.se[comp],2)
        hba1c[comp] ~ dnorm(hba1c.m[comp],hba1c.p[comp])
}
ufemale ~ dnorm(ufemale.m,ufemale.p)
                        #utility from UKPDS expressed as female baseline and male effect
ufemale.p <- 1/pow(ufemale.se,2)
umale ~ dnorm(umale.m,umale.p)
umale.p <- 1/pow(umale.se,2)
clin.ubase <- ufemale + (0.5*umale)
screen.ubase ~ dnorm(uscreen.m,uscreen.p)
uscreen.p <- 1/pow(uscreen.se,2)


for (dur in 2:horizon) {
  for (comp in 1:6) {
        s[1,comp,dur] <- exp(-
lambda[comp]*(exp(0.5*gender[comp]*adjust[comp])*pow(dur,rho[comp])))
                                        # Weibull survival model for screen detected
        s[2,comp,dur] <- exp(-lambda[comp]*(exp((0.5*gender[comp]*adjust[comp]) +
                                        (0.9*hba1c[comp]))*pow(dur,rho[comp])))
                                # Weibull survival model for clinically detected (higher hba1c)
for (det in 1:2) {                                                      c[det,comp,dur]
<- 1 - s[det,comp,dur]                  # Cumulative prob of complication
        d[det,comp,dur] <- c[det,comp,dur] - c[det,comp,dur-1]
                                        # prob. of complication in previous yr
        py[det,comp,dur] <- c[det,comp,dur] - d[det,comp,dur]
                                        # prob. of comp. in yrs prior to previous yr
        ud[det,comp,dur] <- d[det,comp,dur]*uy[comp] + py[det,comp,dur]*upy[comp]  }}
        tud[1,dur] <- sum(ud[1,1:6,dur])
        tud[2,dur] <- sum(ud[2,1:6,dur])
                                        # total utility decrement for all complications
        clin.qaly[dur] <- clin.ubase + tud[2,dur]
                                    # total qaly in each yr duration in clinically detected
        screen.qaly[dur] <- screen.ubase + tud[1,dur-4]
                                    # total qaly for each yr duration in screened detected
}
clin.qaly[1] <- clin.ubase          # set utilities for 1st year duration of T2DM
screen.qaly[1] <- screen.ubase
```

```
for (strategy in 1:4){
 for (time in 1:10){
  for (dur in 1:11-time){
          qs[strategy,time,dur] <- screencases[strategy,time]*pow(1-trans[strategy,1,5,7],
          dur-1)*screen.qaly[dur]
          qc[strategy,time,dur] <- clincases[strategy,time]*pow(1-trans[strategy,1,6,7],
          dur-1)*clin.qaly[dur] }
for (dur in 12-time:21-time){
          qs[strategy,time,dur] <- screencases[strategy,time]*pow(1-trans[strategy,1,5,7],10-time)*pow(1-
          trans[strategy,2,5,7],dur-(11-time))*screen.qaly[dur]
          qc[strategy,time,dur] <- clincases[strategy,time]*pow(1-trans[strategy,1,6,7],10-time)*pow(1-
          trans[strategy,2,6,7],dur-(11-time))*clin.qaly[dur]   }
for (dur in 22-time:31-time){
          qs[strategy,time,dur] <- screencases[strategy,time]*pow(1-trans[strategy,1,5,7],10-time)*pow(1-
          trans[strategy,2,5,7],10)*pow(1-trans[strategy,3,5,7],dur-(21-time))*screen.qaly[dur]
          qc[strategy,time,dur] <- clincases[strategy,time]*pow(1-trans[strategy,1,6,7],10-time)*pow(1-
          trans[strategy,2,6,7],10)*pow(1-trans[strategy,3,6,7],dur-(21-time))*clin.qaly[dur]   }
for (dur in 32-time:41-time){
          qs[strategy,time,dur] <- screencases[strategy,time]*pow(1-trans[strategy,1,5,7],10-time)*pow(1-
          trans[strategy,2,5,7],10)*pow(1-trans[strategy,3,5,7],10)*pow(1-trans[strategy,4,5,7],dur-(31-
          time))*screen.qaly[dur]
          qc[strategy,time,dur] <- clincases[strategy,time]*pow(1-trans[strategy,1,6,7],10-time)*pow(1-
          trans[strategy,2,6,7],10)*pow(1-trans[strategy,3,6,7],10)*pow(1-trans[strategy,4,6,7],dur-(31-
          time))*clin.qaly[dur] }
for (dur in 42-time:51-time){
          qs[strategy,time,dur] <- screencases[strategy,time]*pow(1-trans[strategy,1,5,7],10-time)*pow(1-
          trans[strategy,2,5,7],10)*pow(1-trans[strategy,3,5,7],10)*pow(1-
          trans[strategy,4,5,7],10)*pow(1-trans[strategy,5,5,7],dur-(41-time))*screen.qaly[dur]
          qc[strategy,time,dur] <- clincases[strategy,time]*pow(1-trans[strategy,1,6,7],10-time)*pow(1-
          trans[strategy,2,6,7],10)*pow(1-trans[strategy,3,6,7],10)*pow(1-
          trans[strategy,4,6,7],10)*pow(1- trans[strategy,5,6,7],dur-(41-time))*clin.qaly[dur] }}

for (time in 11:20){
for (dur in 1:21-time){
          qs[strategy,time,dur] <- screencases[strategy,time]*pow(1-trans[strategy,2,5,7],dur-
          1)*screen.qaly[dur]
          qc[strategy,time,dur] <- clincases[strategy,time]*pow(1-trans[strategy,2,6,7],dur-
          1)*clin.qaly[dur]}
for (dur in 22-time:31-time){
          qs[strategy,time,dur] <- screencases[strategy,time]*pow(1-trans[strategy,2,5,7],20-time)*pow(1-
          trans[strategy,3,5,7],dur-(21-time))*screen.qaly[dur]
          qc[strategy,time,dur] <- clincases[strategy,time]*pow(1-trans[strategy,2,6,7],20-time)*pow(1-
          trans[strategy,3,6,7],dur-(21-time))*clin.qaly[dur]   }
for (dur in 32-time:41-time){
          qs[strategy,time,dur] <- screencases[strategy,time]*pow(1-trans[strategy,2,5,7],20-time)*pow(1-
          trans[strategy,3,5,7],10)*pow(1-trans[strategy,4,5,7],dur-(31-time))*screen.qaly[dur]
          qc[strategy,time,dur] <- clincases[strategy,time]*pow(1-trans[strategy,2,6,7],20-time)*pow(1-
          trans[strategy,3,6,7],10)*pow(1-trans[strategy,4,6,7],dur-(31-time))*clin.qaly[dur]   }
for (dur in 42-time:51-time){
          qs[strategy,time,dur] <- screencases[strategy,time]*pow(1-trans[strategy,2,5,7],20-time)*pow(1-
          trans[strategy,3,5,7],10)*pow(1-trans[strategy,4,5,7],10)*pow(1-trans[strategy,5,5,7],dur-(41-
          time))*screen.qaly[dur]
          qc[strategy,time,dur] <- clincases[strategy,time]*pow(1-trans[strategy,2,6,7],20-time)*pow(1-
```

```
        trans[strategy,3,6,7],10)*pow(1-trans[strategy,4,6,7],10)*pow(1-trans[strategy,5,6,7],dur-(41-
        time))*clin.qaly[dur] }}

for (time in 21:30){
for (dur in 1:31-time){
        qs[strategy,time,dur] <- screencases[strategy,time]*pow(1-trans[strategy,3,5,7],dur-
        1)*screen.qaly[dur]
        qc[strategy,time,dur] <- clincases[strategy,time]*pow(1-trans[strategy,3,6,7],dur-
        1)*clin.qaly[dur] }
for (dur in 32-time:41-time){
        qs[strategy,time,dur] <- screencases[strategy,time]*pow(1-trans[strategy,3,5,7],30-time)*pow(1-
        trans[strategy,4,5,7],dur-(31-time))*screen.qaly[dur]
        qc[strategy,time,dur] <- clincases[strategy,time]*pow(1-trans[strategy,3,6,7],30-time)*pow(1-
        trans[strategy,4,6,7],dur-(31-time))*clin.qaly[dur]   }
for (dur in 42-time:51-time){
        qs[strategy,time,dur] <- screencases[strategy,time]*pow(1-trans[strategy,3,5,7],30-time)*pow(1-
        trans[strategy,4,5,7],10)*pow(1-trans[strategy,5,5,7],dur-(41-time))*screen.qaly[dur]
        qc[strategy,time,dur] <- clincases[strategy,time]*pow(1-trans[strategy,3,6,7],30-time)*pow(1-
        trans[strategy,4,6,7],10)*pow(1-trans[strategy,5,6,7],dur-(41-time))*clin.qaly[dur]   }}

for (time in 31:40){
for (dur in 1:41-time){
        qs[strategy,time,dur] <- screencases[strategy,time]*pow(1-trans[strategy,4,5,7],dur-
        1)*screen.qaly[dur]
        qc[strategy,time,dur] <- clincases[strategy,time]*pow(1-trans[strategy,4,6,7],dur-
        1)*clin.qaly[dur] }
for (dur in 42-time:51-time){
        qs[strategy,time,dur] <- screencases[strategy,time]*pow(1-trans[strategy,4,5,7],40-time)*pow(1-
        trans[strategy,5,5,7],dur-(41-time))*screen.qaly[dur]
        qc[strategy,time,dur] <- clincases[strategy,time]*pow(1-trans[strategy,4,6,7],40-time)*pow(1-
        trans[strategy,5,6,7],dur-(41-time))*clin.qaly[dur]   }}

for (time in 41:50){
for (dur in 1:51-time){
        qs[strategy,time,dur] <- screencases[strategy,time]*pow(1-trans[strategy,5,5,7],dur-
        1)*screen.qaly[dur]
        qc[strategy,time,dur] <- clincases[strategy,time]*pow(1-trans[strategy,5,6,7],dur-
        1)*clin.qaly[dur] }}          }}

for (cycle in 1:horizon){
        for (time in 1:cycle){
                zs[strategy,cycle,time] <-qs[strategy,time,cycle+1-time]
                zc[strategy,cycle,time] <-qc[strategy,time,cycle+1-time]          }
        timeqs[strategy,cycle] <- sum(zs[strategy,cycle,1:cycle])
        timeqc[strategy,cycle] <- sum(zc[strategy,cycle,1:cycle])}
totqs[strategy] <- sum(timeqs[strategy,1:horizon])
totqc[strategy] <- sum(timeqc[strategy,1:horizon])


# undiscounted utilities
        T2DMqalytot[strategy] <- totqs[strategy] + totqc[strategy] + (total[strategy,4]*screen.ubase)
                qaly[strategy] <- sum(total[strategy,1:3]) + T2DMqalytot[strategy] }
```

```
#discounted utilities (3.5% per year)
for (strategy in 1:4){
for (time in 1:horizon){
        disqaly[strategy,time] <- (sum(number[strategy,1:3,time]) + timeqs[strategy,time] +
        timeqc[strategy,time]  + (number[strategy,4,time]*screen.ubase))*pow(0.965,time-1) }
    totdisqaly[strategy] <- sum(disqaly[strategy,1:horizon]) }
```

Costs per QALY

```
for (strategy in 1:4){
        costperqaly[strategy] <- totalcost[strategy] / qaly[strategy]
        discostperqaly[strategy] <- totdiscost[strategy] / totdisqaly[strategy]
                                                                }
#incremental cost effectiveness ratio (ICER) for each active strategy minus no screening
        diffcost[1] <- (totalcost[1] -  totalcost[4])
        diffcost[2] <- (totalcost[2] -  totalcost[4])
        diffcost[3] <- (totalcost[3] -  totalcost[4])
        diffqaly[1] <- (qaly[1] - qaly[4])
        diffqaly[2] <- (qaly[2] - qaly[4])
        diffqaly[3] <- (qaly[3] - qaly[4])
        inccost[1] <- diffcost[1]  / diffqaly[1]
        inccost[2] <- diffcost[2]  / diffqaly[2]
        inccost[3] <- diffcost[3]  / diffqaly[3]

#incremental cost effectiveness ratio (ICER) for each active strategy minus no screening
        diffdiscost[1] <- (totdiscost[1] -  totdiscost[4])
        diffdiscost[2] <- (totdiscost[2] -  totdiscost[4])
        diffdiscost[3] <- (totdiscost[3] -  totdiscost[4])
        diffdisqaly[1] <- (totdisqaly[1] - totdisqaly[4])
        diffdisqaly[2] <- (totdisqaly[2] - totdisqaly[4])
        diffdisqaly[3] <- (totdisqaly[3] - totdisqaly[4])
        inccostdis[1] <- diffdiscost[1]  / diffdisqaly[1]
        inccostdis[2] <- diffdiscost[2]  / diffdisqaly[2]
        inccostdis[3] <- diffdiscost[3]  / diffdisqaly[3]

#incremental cost effectiveness ratio (ICER) for lifestyle vs. pharmacological
        #inccost[5] <- (totalcost[2] - totalcost[3])  / (diffqaly[2] - diffqaly[3])
        #inccostdis[5] <- (totdiscost[2] - totdiscost[3])  / (diffdisqaly[2] - diffdisqaly[3])

#Cost-effectiveness acceptibility curve
for(k in 1:NK) {
        Q[1,k] <- step(1- ( (totalcost[1] -  totalcost[4]) - K[k] * (qaly[1] - qaly[4]) ))
        Q[2,k] <- step(1- ( (totalcost[2] -  totalcost[4]) - K[k] * (qaly[2] - qaly[4]) ))
        Q[3,k] <- step(1- ( (totalcost[3] -  totalcost[4]) - K[k] * (qaly[3] - qaly[4]) ))
        Qdis[1,k] <- step(1- ( (totdiscost[1] -  totdiscost[4]) - K[k] * (totdisqaly[1] - totdisqaly[4]) ))
        Qdis[2,k] <- step(1- ( (totdiscost[2] -  totdiscost[4]) - K[k] * (totdisqaly[2] - totdisqaly[4]) ))
        Qdis[3,k] <- step(1- ( (totdiscost[3] -  totdiscost[4]) - K[k] * (totdisqaly[3] - totdisqaly[4]) ))
        Qlvd[1,k] <- step(1- ( (totalcost[2] -  totalcost[1]) - K[k] * (qaly[2] - qaly[1]) ))   }

#cost per life year
```

```
for (strategy in 1:3){
        costlifeyr[strategy] <- diffcost[strategy] / difflife[strategy]
        costlifeyrdis[strategy] <- diffdiscost[strategy] / difflifedis[strategy] }
costcase[2] <- diffcost[2] / diffcases[2]
costcase[3] <- diffcost[3] / diffcases[3]
```

#Transition rates

```
#NGTtoIGT
logNGTtoIGT1 ~ dnorm(1.66,0.075)
logNGTtoIGT2 ~ dnorm(2.49,0.109)
NGTtoIGT1 <- exp(logNGTtoIGT1)/100
NGTtoIGT2 <- exp(logNGTtoIGT2)/100              # pooled rates per 1 person year

#IGTtoT2DM                                      #studies of mostly caucasians only
for(i in 1:11){
        prec1[i] <- 1/(SEID[i]*SEID[i])
        LogIRID[i] ~ dnorm(theta1[i],prec1[i])
        theta1[i] ~ dnorm(mu1,tau1)
                                }
        mu1 ~ dnorm(0,0.001)
        tau1 <- 1/(sd1*sd1)
        sd1 ~ dunif(0,2)
        IGTtoT2DM <- exp(mu1)/100                       # pooled rate per 1 person year

#Pharmacological intervention meta (anti-diabetic drugs)
for(i in 1:9){
        prec2[i] <- 1/(SEPH[i]*SEPH[i])
        LogHRPH[i] ~ dnorm(theta2[i],prec2[i])
        theta2[i] ~ dnorm(mu2,tau2)
                                }
        mu2 ~ dnorm(0,0.001)
        tau2 <- 1/(sd2*sd2)
        sd2 ~ dunif(0,2)
        PharmHR <- exp(mu2)

#Lifestyle intervention meta
for(i in 1:12){
        prec3[i] <- 1/(SELI[i]*SELI[i])
        LogHRLI[i] ~ dnorm(theta3[i],prec3[i])
        theta3[i] ~ dnorm(mu3,tau3)
                                }
        mu3 ~ dnorm(0,0.001)
        tau3 <- 1/(sd3*sd3)
        sd3 ~ dunif(0,2)
        LifeHR <- exp(mu3)

#Transition from T2DMu to T2DMd
for(i in 1:2){
        prec4[i] <- 1/(SEud[i]*SEud[i])
        LogT2ud[i] ~ dnorm(theta4[i],prec4[i])
        theta4[i] ~ dnorm(mu4,tau4)
                                }
        mu4 ~ dnorm(0,0.001)
        tau4 <- 1/(sd4*sd4)
        sd4 ~ dunif(0,2)
        T2ud <- 1/exp(mu4)      # pooled rate per person year, mu4 is time spent undiagnosed
```

```
#HR for increased risk of death if diabetic
LogHRdeathdiab ~ dnorm(0.756,132)
HRdeathdiab <- exp(LogHRdeathdiab)

#HR for increased risk of death with increasing HbA1c
LogHRhba1c ~ dnorm(0.10436,658)
HRhba1c <- exp(LogHRhba1c)

#HbA1c values
hba1cu ~ dnorm(hba1cu.m,hba1cu.p)
hba1cu.p <- 1/pow(hba1cu.se,2)
hba1cc ~ dnorm(hba1cc.m,hba1cc.p)
hba1cc.p <- 1/pow(hba1cc.se,2)
hba1cs ~ dnorm(hba1cs.m,hba1cs.p)
hba1cs.p <- 1/pow(hba1cs.se,2)      }
```

Data

```
list((x=structure(.Data=c(1626,138,83,142,51,71,8,7,90),.Dim=c(3,3)),
        preva=c(1847,264,105),Nscreen=1, horizon=50,
        LogIRID=c(1.528,1.386,1.77,0.872, 2.398,1.998,1.187,1.952, 2.533, 3.17,2.93),
        SEID=c(0.087, 0.143, 0.167, 0.242, 0.302, 0.130, 0.258, 0.192, 0.059, 0.447,0.103),
        LogHRLI=c(-0.451,-0.4557,-0.638,-0.8675,-0.916,-0.492,-1.244,-0.1669,-0.6619,-0.288,-
        0.4732,-1.1980),
        SELI=c(0.223,0.298,0.226,0.1084,0.216,0.226,0.603,0.3873,1.225,0.387,0.2005,0.5718),
        LogHRPH=c(-0.3711,-0.7172,-0.506,-0.2877,0.0128,-1.3106,-0.8353,-1.7430,-0.4292),
        SEPH=c(0.0959,0.7071,0.4754,0.091,0.3852,0.5477,0.4944,1.0954,0.1969),
        LogT2ud=c(1.872,1.435),SEud=c(0.225,0.321),
        NK = 10, K = c(100,1000,2500,5000,7500,10000,20000,25000,30000,50000),
        prior_r=structure(.Data=c(1,1,1,1,1,1,1,1,1),.Dim=c(3,3)),
        beta0.m=c(-5.310,-4.977,-8.018,-7.163,-8.718,-6.464),
        beta0.se=c(0.174,0.160,0.408,0.342,0.613,0.326,0.939),
        rho.m=c(1.150,1.257,1.711,1.497,1.451,1.154),
        rho.se=c(0.067,0.060,0.158,0.126,0.232,0.121),
        uy.m=c(-0.141,-0.081,-0.058,-0.131,-0.451,-0.074),
        uy.se=c(0.060,0.052,0.066,0.073,0.131,0.070),
        upy.m=c(-0.079,-0.044,-0.134,-0.199,-0.335,-0.080),
        upy.se=c(0.020,0.021,0.038,0.035,0.068,0.029),
        ufemale.m=0.725,ufemale.se=0.035,
        umale.m=0.092,umale.se=0.009,
        uscreen.m=0.788,uscreen.se=0.020,
        hba1c.m=c(0.125,0.118,0.157,0.128,0.435,0.221),
        hba1c.se=c(0.035,0.025,0.057,0.042,0.066,0.050),
        gender.m=c(-0.471,-0.826,1,-0.516,1,1),
        gender.se=c(0.143,0.103,1,0.171,1,1),
        hba1cu.m=9.0,hba1cu.se=0.056,
        hba1cc.m=7.9,hba1cc.se=0.050,
        hba1cs.m=7.0,hba1cs.se=0.046,
        adjust=c(1,1,0,1,0,0))
```

Inits

```
list(prev=c(0.80,0.15,0.05), r=structure(.Data=c(0.4,0.3,0.3,0.4,0.3,0.3,0.3,0.3,0.4),.Dim=c(3,3)),
mu1=0,sd1=0.5,theta1=c(0,0,0,0,0,0,0,0,0,0),
mu2=0,sd2=0.5,theta2=c(0,0,0,0,0,0,0,0,0),
mu3=0,sd3=0.5,theta3=c(0,0,0,0,0,0,0,0,0,0,0),
mu4=0,sd4=0.5,theta4=c(0,0),
gamma=c(0,0,0,0,0,0),
uy=c(0,0,0,0,0,0),
upy=c(0,0,0,0,0,0))
```

3.7 WinBUGS code for estimating the 3x3 diagnostic table when data is missing

```
model{
# pr( T+ | T+ or T0 or T-, group)        p1       basic parameter
# pr( T0 | T0 or T-)                     p2       basic parameter
# pr( T0 | T+ or T0 or T-)               p3 = p2(1-p1)
# pr( T- | T+ or T0 or T-)               p4 = 1 - p1 - p3
# pr( T+ or T0 | T+ or T0 or T-)         p5 = p1 + p3
# pr(T0 or T- | T+ or T0 or T- )         p6 = p3 + p4

# groups              j=1 NGT,            j=2 IGT,          j =3 T2DM

p[1,1] ~ dunif(0,  p[1,2])
p[2,1] ~ dunif(0,  p[2,2])
p[1,2] ~ dunif(p[1,1],  p[1,3])
p[2,2] ~ dunif(p[2,1],  p[2,3])
p[1,3] ~ dunif(p[1,2],   1)
p[2,3] ~ dunif(p[2,2],  1)

for (j in 1:3) {
p[3,j] <- p[2,j] * (1 - p[1,j])
p[4,j] <- 1 - p[1,j] - p[3,j]
p[5,j] <- p[1,j] + p[3,j]
p[6,j] <- p[3,j] + p[4,j]
}

r41 ~ dbin(p[4,1],n41)
r13 ~ dbin(p[1,3],n13)

pa <- (690 * p[1,1]    +  164 * p[1,2] )  /  854
ra ~ dbin(pa,na)

pb <- (164 * p[4,2]    + 37 * p[4,3] ) / 201
rb ~ dbin(pb,nb)

a <- p[4,1]*690
b <- p[4,2]*164
c <- p[4,3]*37
d <- p[3,1]*690
e <- p[3,2]*164
f <- p[3,3]*37
g <- p[1,1]*690
h <- p[1,2]*164
i <- p[1,3]*37
dg <- p[5,1]*690
eh <- p[5,2]*164
fi <- p[5,3]*37
ad <- p[6,1]*690
be <- p[6,2]*164
cf <- p[6,3]*37
}
```

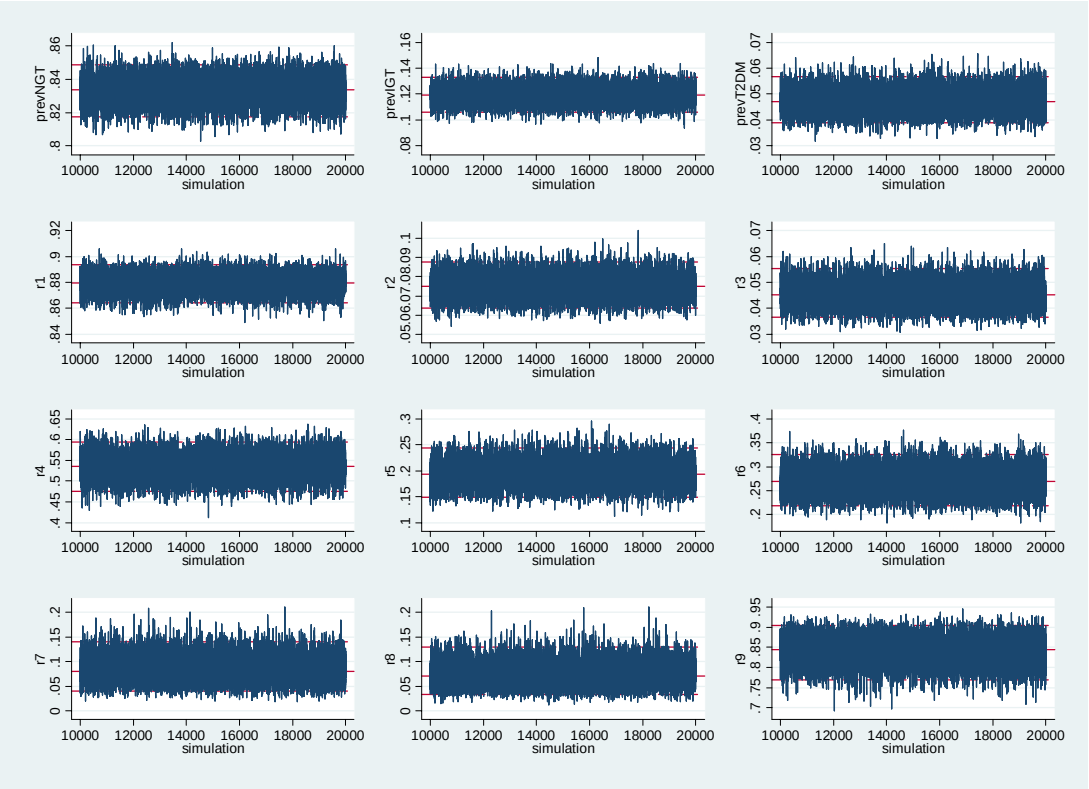
```
list(
r41=421,n41=690,
r13=31,  n13=37,
ra=132,na=854,
rb=75,nb=201)
```

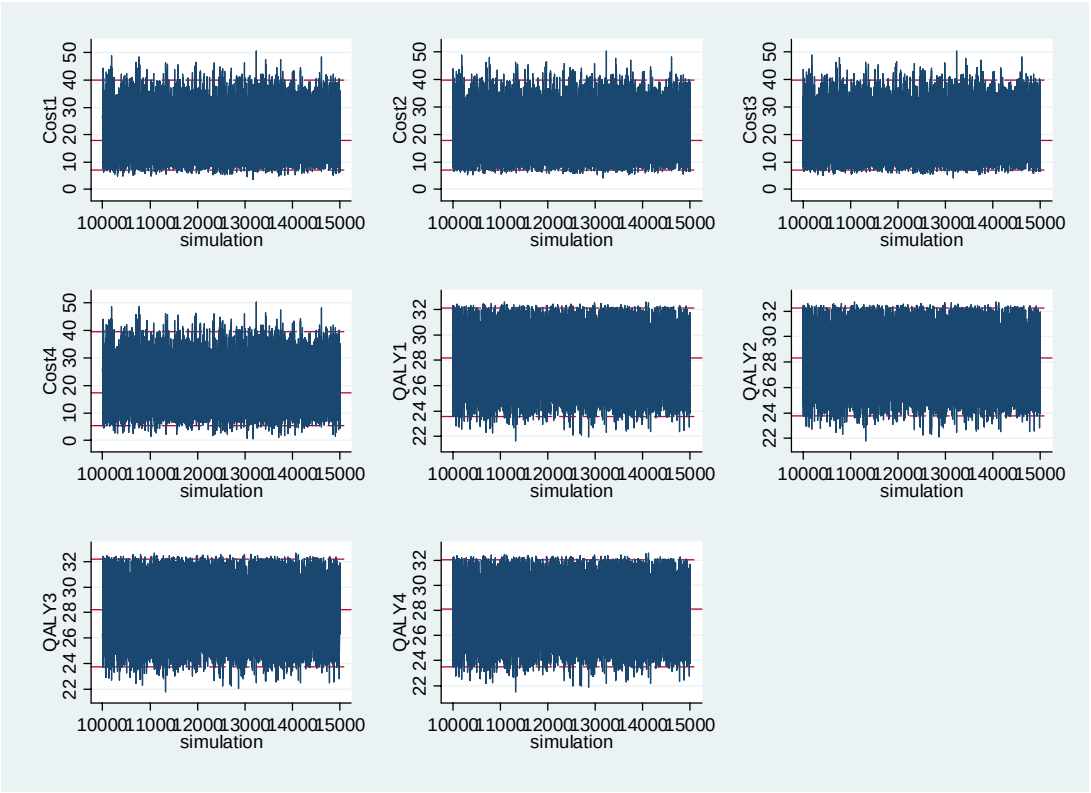
3.8 WinBUGS code incorporated within the full decision model to estimate the expected value of perfect information (EVPI)

```
#differences in costs and QALYs for each active strategy minus no screening
        diffcost[1] <- (totalcost[1] -  totalcost[4])
        diffcost[2] <- (totalcost[2] -  totalcost[4])
        diffcost[3] <- (totalcost[3] -  totalcost[4])
        diffqaly[1] <- (qaly[1] - qaly[4])
        diffqaly[2] <- (qaly[2] - qaly[4])
        diffqaly[3] <- (qaly[3] - qaly[4])


# net benefits and EVPI
        for (strategy in 1:3){
                for(k in 1:NK) {
                        inb[strategy,k] <- (diffqaly[strategy]*k) - diffcost[strategy]
                        evpi[strategy,k] <- max(-inb[strategy,k],0)
                        popevpi[strategy,k] <- evpi[strategy,k]*totadj
}}
for (time in 1:50){
        adj[time] <- 797400/pow((1 + 0.025),(time-1))
                        # 797,400 is the incidence of turning 45, England & Wales, 2007
}                               # discounted at 2.5% per annum
totadj <- sum(adj[1:50])


#where K = c(100,1000,2500,5000,7500,10000,20000,25000,30000,50000),
```

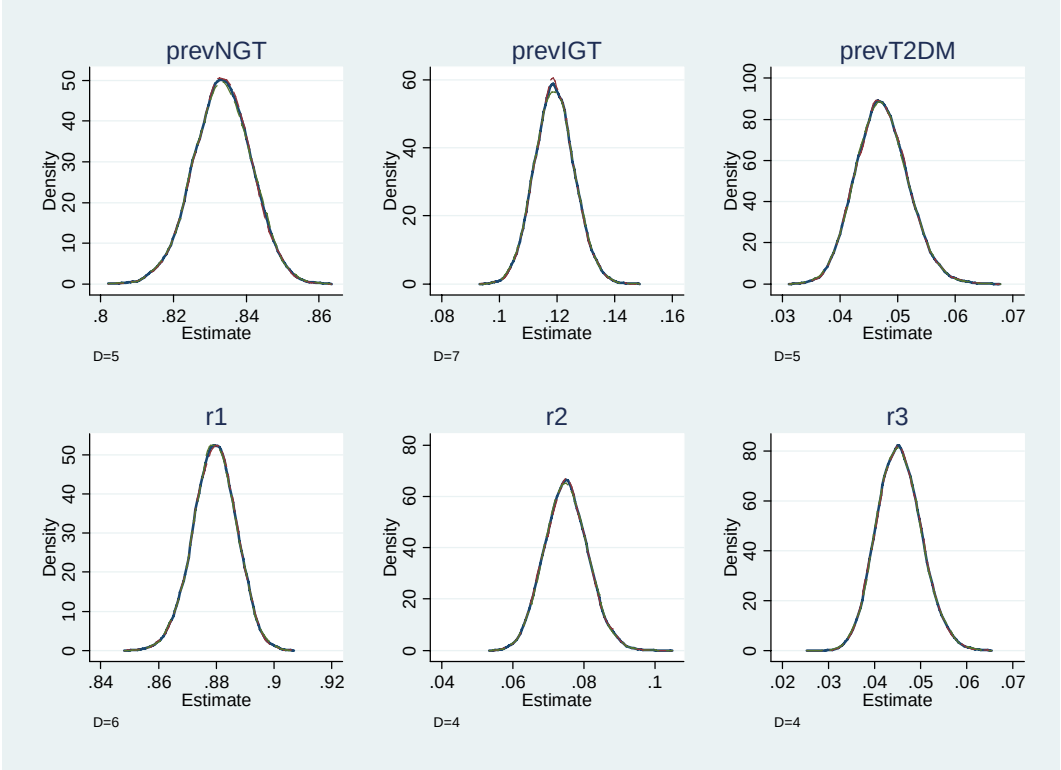
APPENDIX 4: DIAGNOSTIC GRAPHS

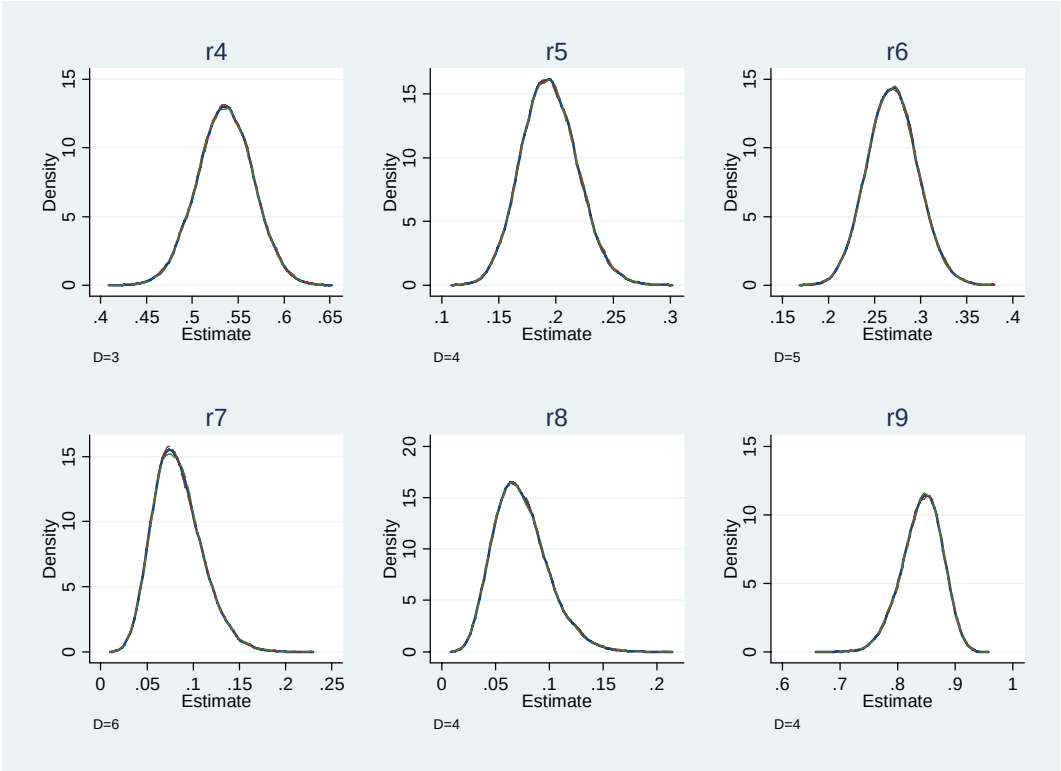
4.1 Trace plots

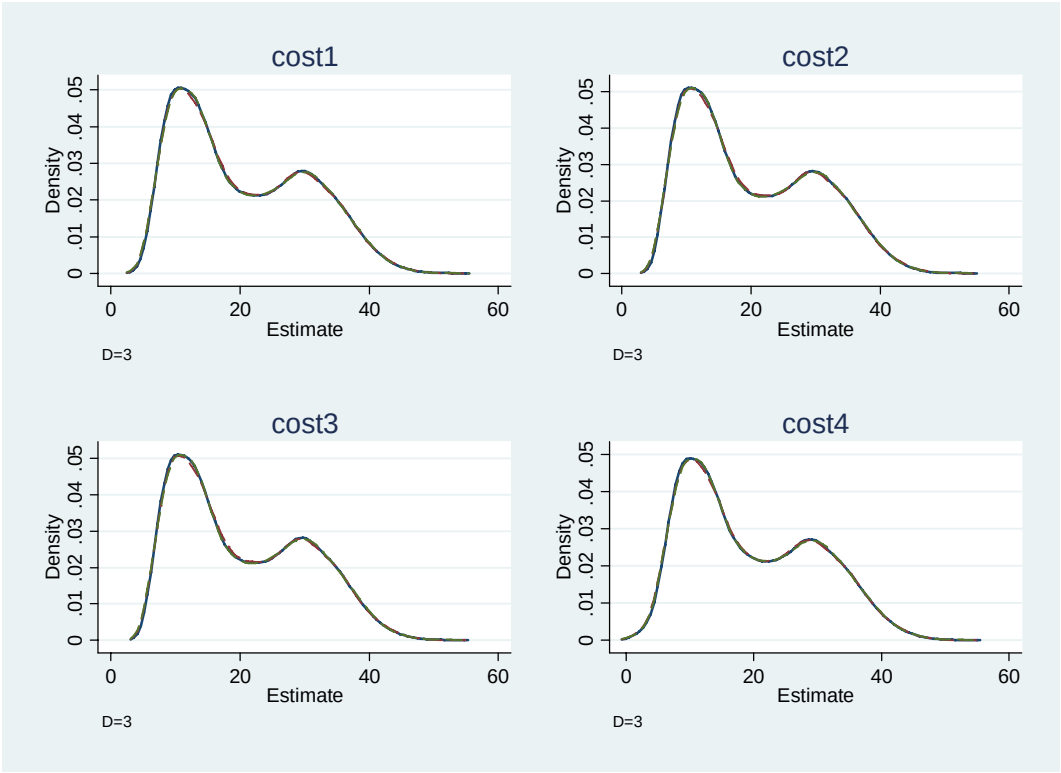


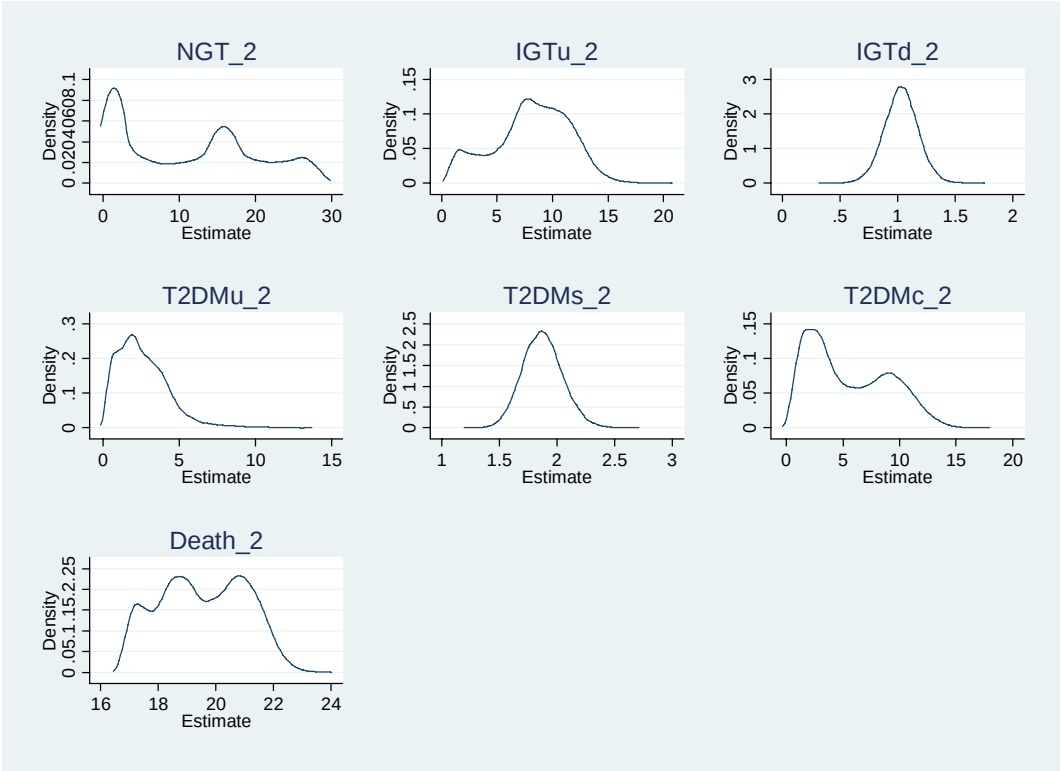

Cost1
Cost2
Cost3
Cost4
QALY1
QALY2
QALY3
QALY4
0 10 20 30 40 50
22 24 26 28 30 32
10000 11000 12000 13000 14000 15000
simulation

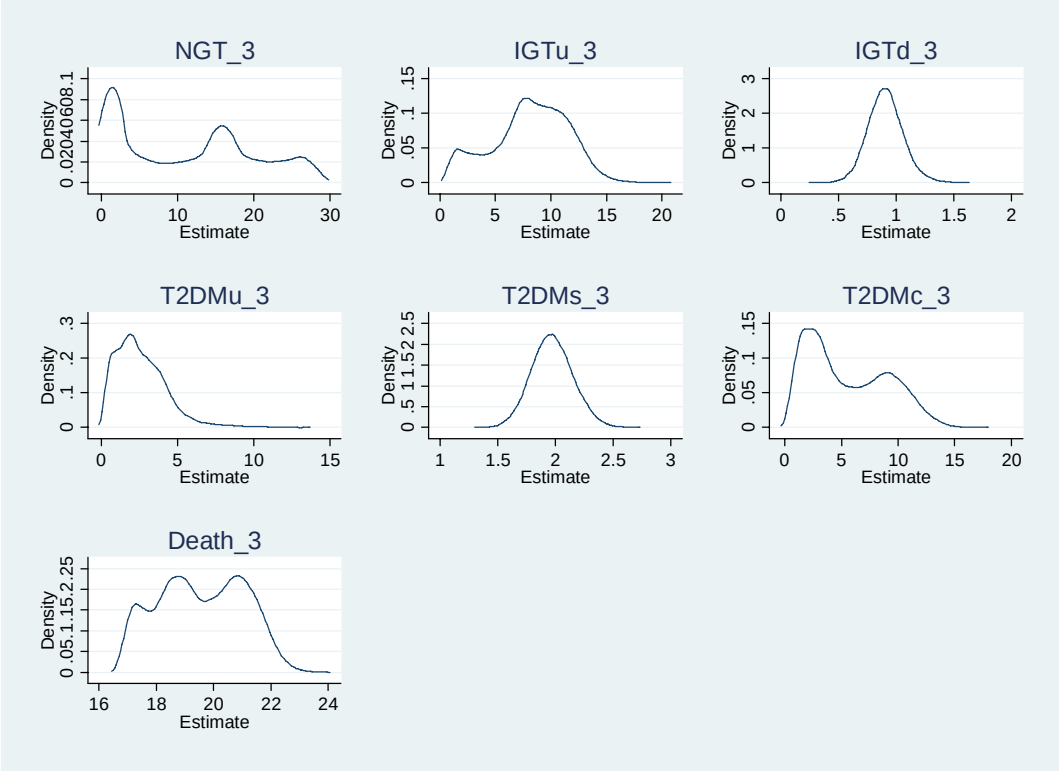
4.2 Density plots

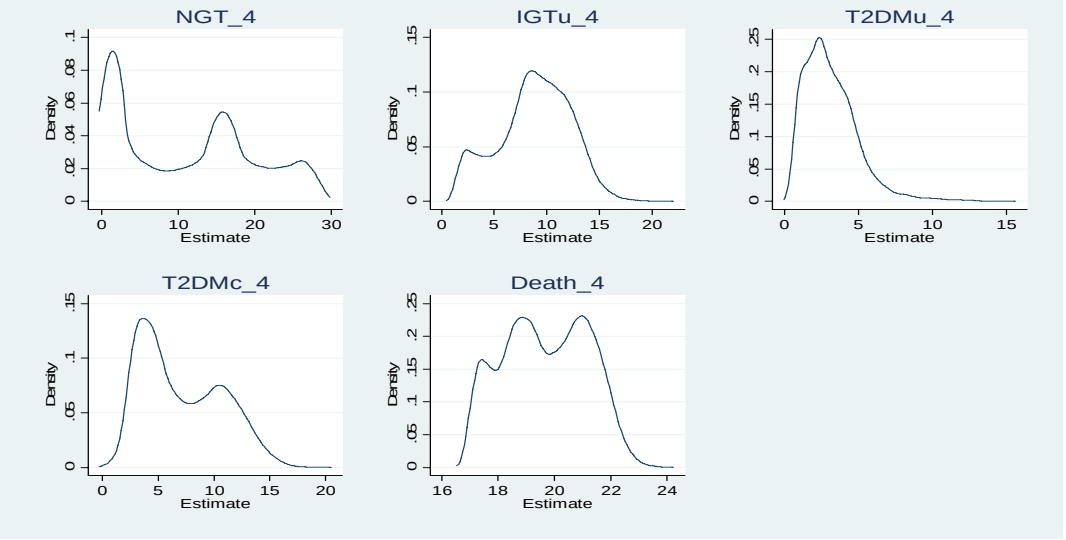




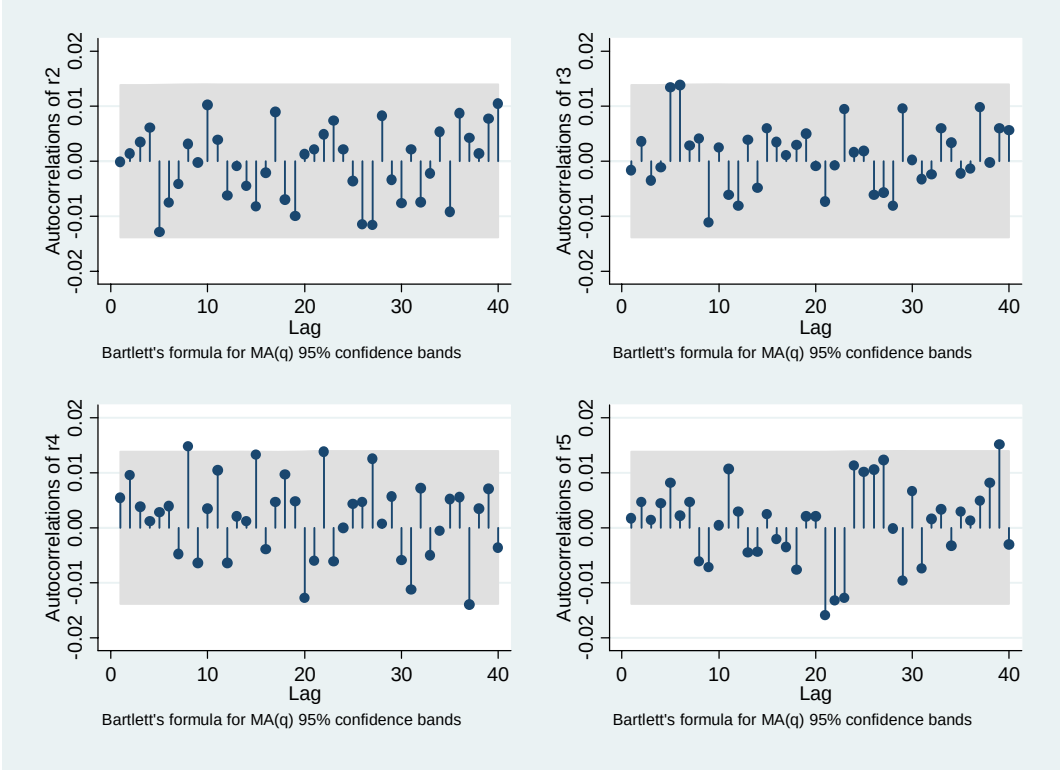

cost1
cost2
cost3
cost4
Density
Estimate
D=3

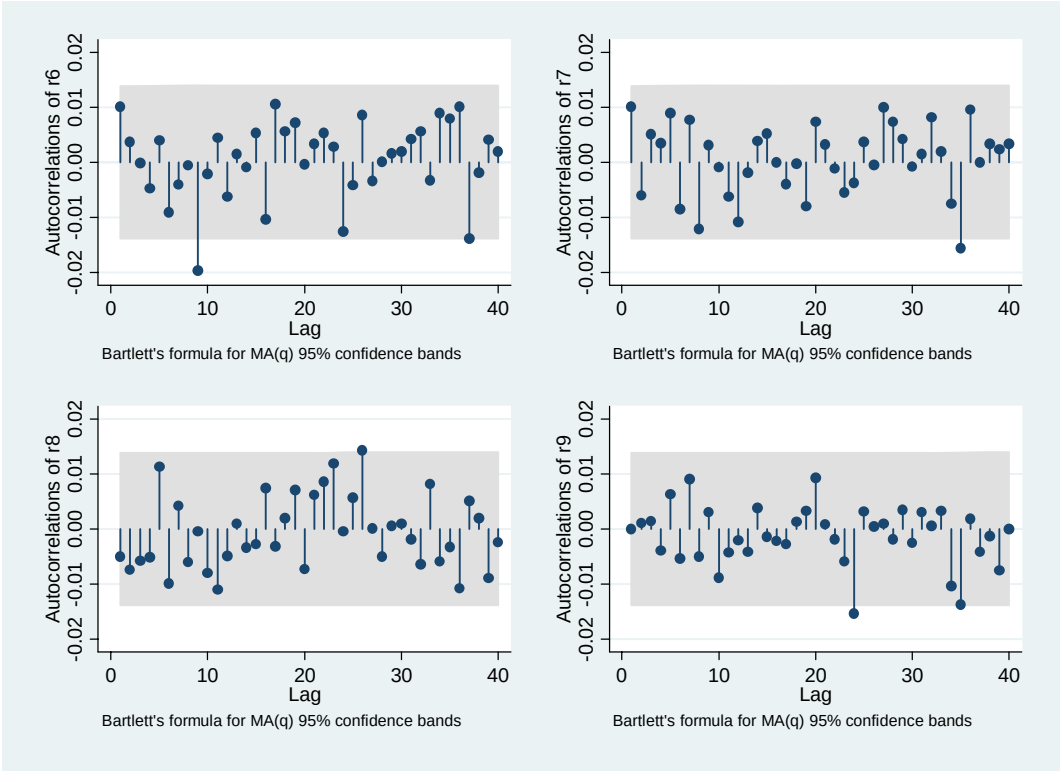

NGT_2
IGTu_2
IGTd_2
T2DMu_2
T2DMs_2
T2DMc_2
Death_2
Density
Estimate

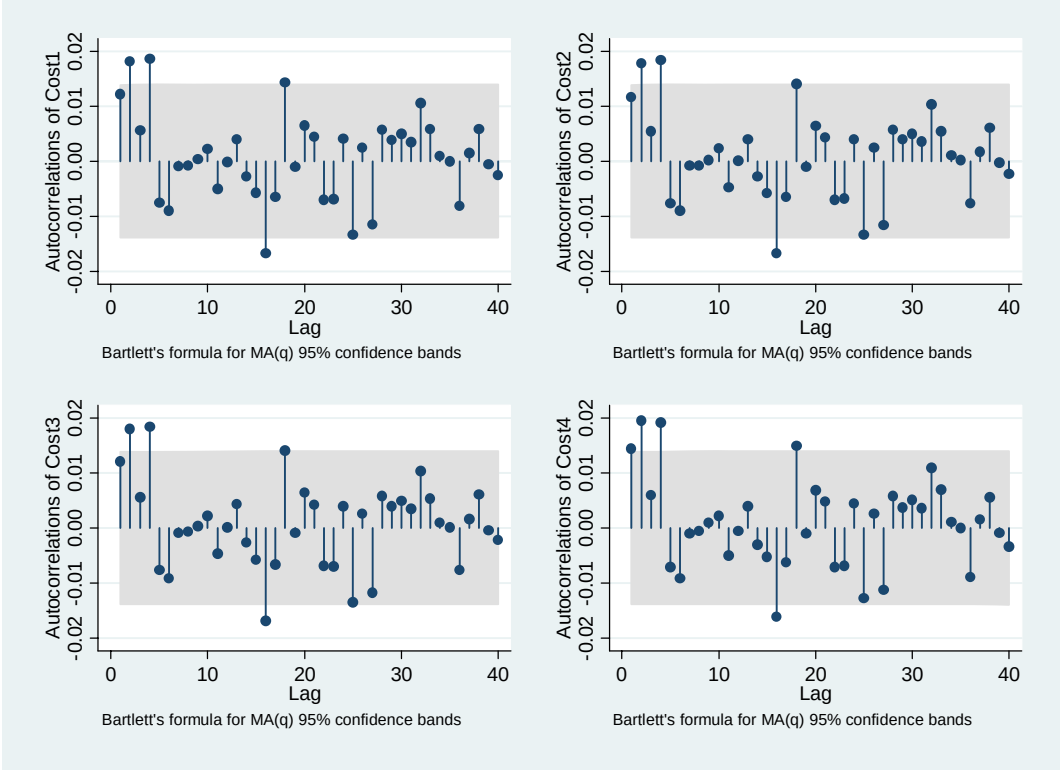

NGT_3
IGTu_3
IGTd_3
T2DMu_3
T2DMs_3
T2DMc_3
Death_3
Density
Estimate

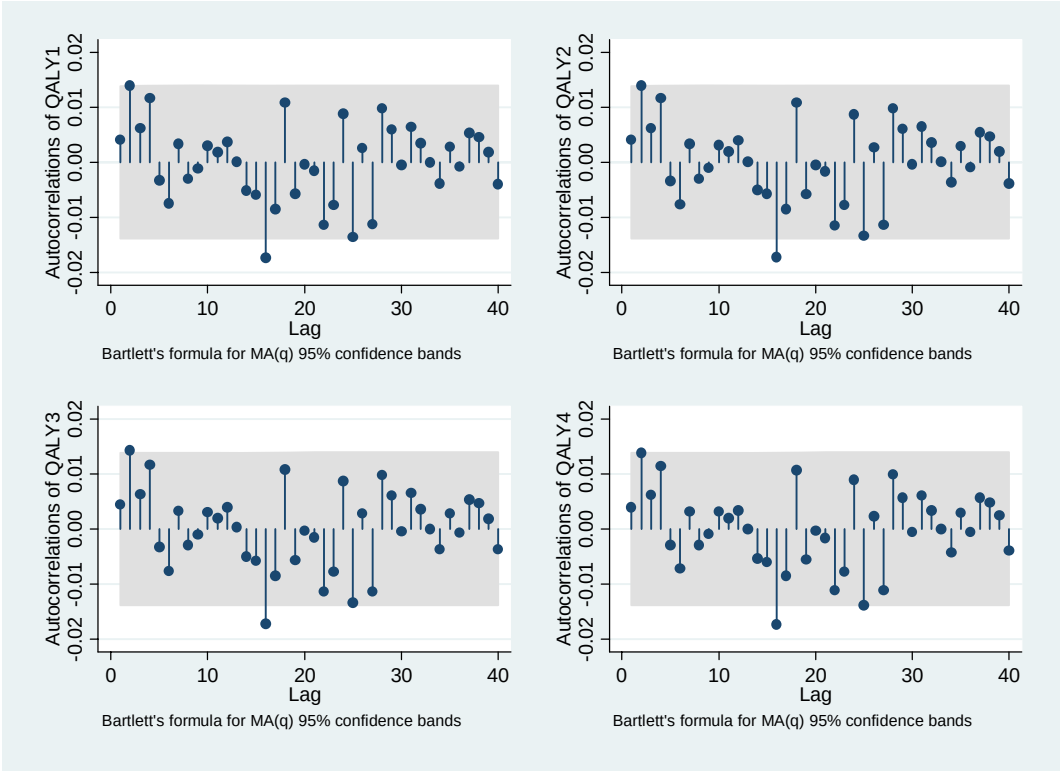

NGT_4
IGTu_4
T2DMu_4
T2DMc_4
Death_4
Density
Estimate

4.3 Autocorrelation plots

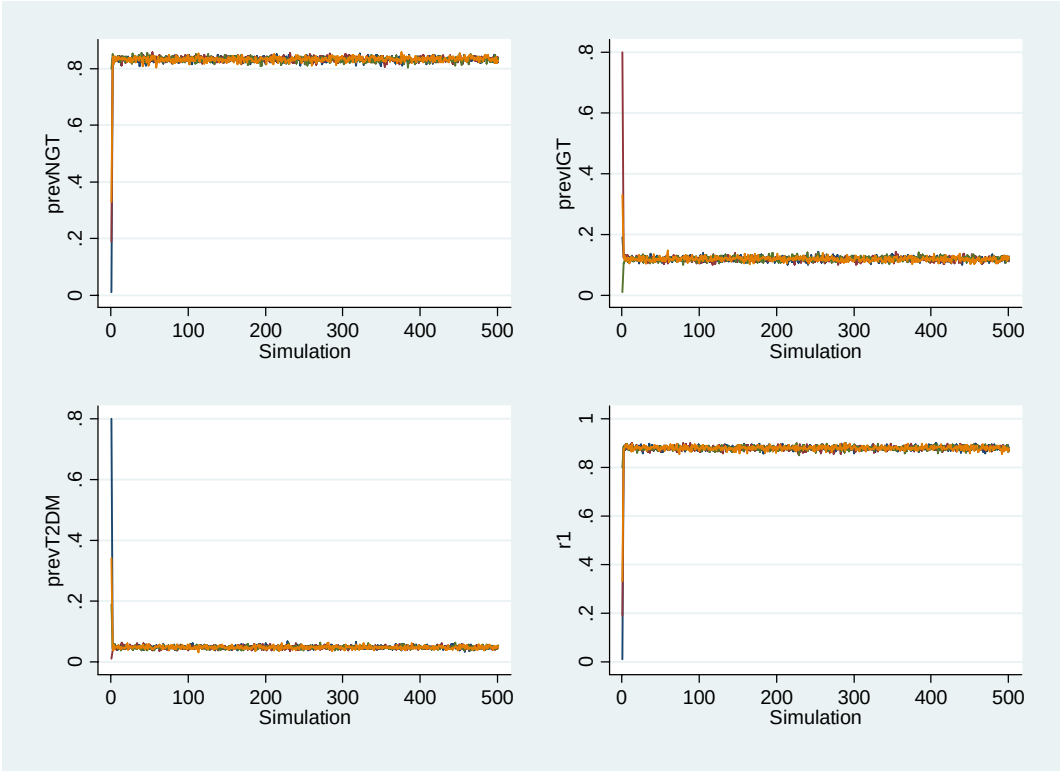


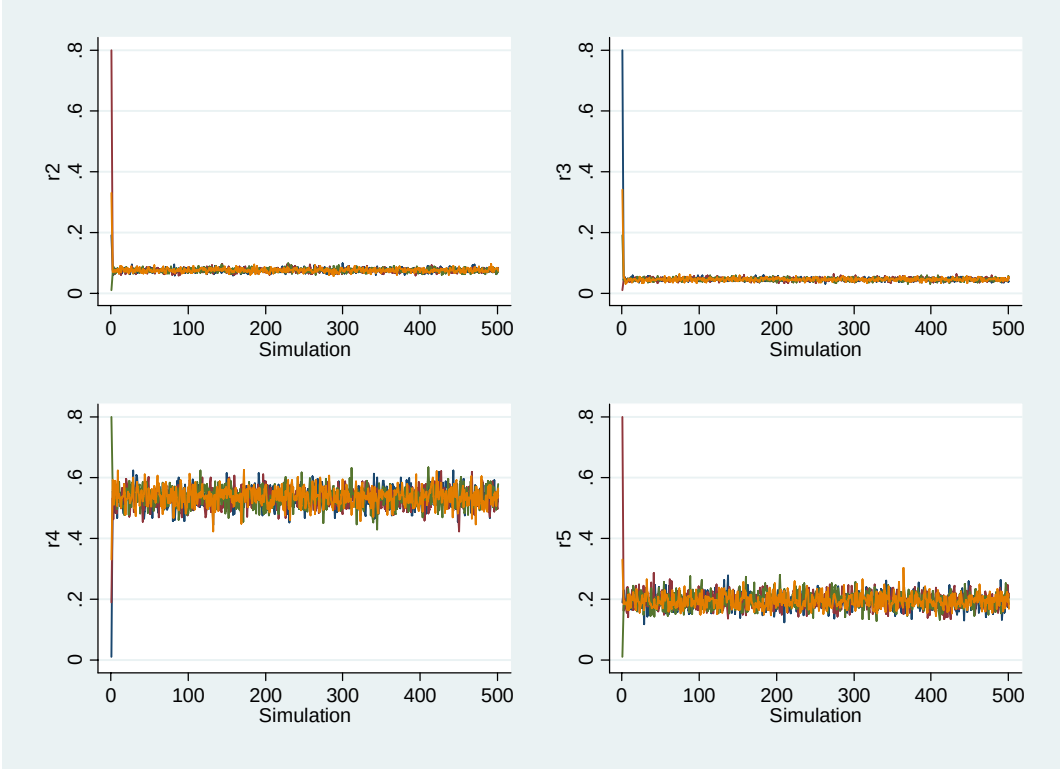


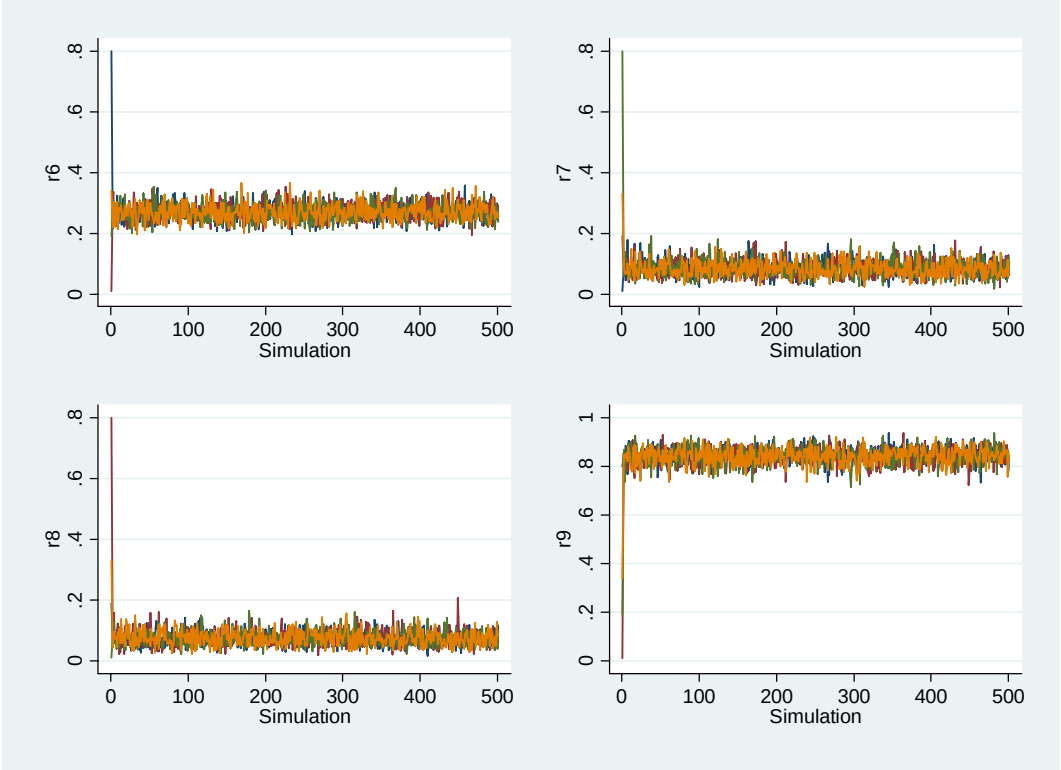

Autocorrelations of Cost1
Autocorrelations of Cost2
Autocorrelations of Cost3
Autocorrelations of Cost4
-0.02 -0.01 0.00 0.01 0.02
0 10 20 30 40
Lag
Bartlett's formula for MA(q) 95% confidence bands


Autocorrelations of QALY1
Autocorrelations of QALY2
Autocorrelations of QALY3
Autocorrelations of QALY4
-0.02 -0.01 0.00 0.01 0.02
0 10 20 30 40
Lag
Bartlett's formula for MA(q) 95% confidence bands

4.4 Overlaid plots of chains with different initial values






r6
r7
r8
r9
0
.2
.4
.6
.8
1
0
100
200
300
400
500
Simulation

APPENDIX 5: PUBLICATIONS, PRESENTATIONS AND POSTERS ASSOCIATED WITH THIS THESIS

This appendix contains details of the dissemination of findings resulting from the work carried out for this thesis, and includes the following:

Clare L Gillies, Keith R Abrams, Paul C Lambert, Nicola J Cooper, Alex J Sutton, Ron Hsu, Kamlesh Khunti. Pharmacological and Lifestyle Interventions to Prevent or Delay Type 2 Diabetes Mellitus in Individuals with Impaired Glucose Tolerance: A Systematic Review and Meta-analysis. *British Medical Journal*, 2007; 334: 299.

Clare L Gillies, Paul C Lambert, Keith R Abrams, Alex J Sutton, Nicola J Cooper Ron Hsu, Kamlesh Khunti. A cost-effectiveness analysis for different strategies for the screening and prevention of type 2 diabetes mellitus. *Invited submission to the BMJ, currently under consideration.*

Clare Gillies, Keith Abrams, Paul Lambert, Kamlesh Khunti. Lifestyle changes and pharmacological interventions are both effective in protecting against type 2 diabetes mellitus but which approach is best? *GP magazine*. May, 2007.

CL Gillies. Screening for impaired glucose tolerance and intervening to delay type 2 diabetes mellitus: Is this an effective health policy? *Cambridge Diabetes Seminar, 2006.*

Clare L Gillies, Keith R Abrams, Paul C Lambert, Nicola J Cooper, Alex J Sutton, Ron Hsu, Kamlesh Khunti. Development of Evidence Synthesis Methods for the Assessment of Health Policies Involving Screening and Intervention. *The International Society of Clinical Biostatistics, Geneva 2006.*

Clare L Gillies, Keith R Abrams, Paul C Lambert, Nicola J Cooper, Alex J Sutton, Ron Hsu, Kamlesh Khunti. Pharmacological and Lifestyle Interventions to Prevent or Delay Type 2 Diabetes Mellitus in Individuals with Impaired Glucose Tolerance: A Systematic Review and Meta-analysis. *Diabetes UK Annual Professional Conference, Birmingham, 2006.*

CL Gillies, KR Abrams, PC Lambert, NJ Cooper, AJ Sutton, K Khunti, R Hsu. Issues in evidence synthesis for comprehensive decision models: An illustration using impaired glucose tolerance. *The International Society of Clinical Biostatistics, Szeged 2005.*

BIBLIOGRAPHY

ADES, A. E. (2002) A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. *Statistics in Medicine,* 22, 359-71.

ADES, A. E., LU, G. & HIGGINS, J. P. (2005) The interpretation of random-effects meta-analysis in decision models. *Medical Decision Making,* 25, 646-654.

ALBERTI, K. G. & ZIMMET, P. Z. (1998) Definition, diagnosis and classification of diabetes mellitus and it's complications. Diagnosis and classification of diabetes mellitus: provisional report of a WHO consultation. *Diabetic Medicine,* 15, 539-553.

ALTMAN, D. G. (1991) *Practical statistics for medical research*, Chapman and Hall.

ALTMAN, D. G. (1998) Confidence intervals for the number needed to treat. *BMJ,* 317, 1309-12.

AMERICAN DIABETES ASSOCIATION (1997) Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care,* 23, 1619-29.

AMERICAN DIABETES ASSOCIATION AND NATIONAL INSTITUTE OF DIABETES AND DIGESTIVE AND KIDNEY DISEASES (2002) The prevention or delay of type 2 diabetes. *Diabetes Care,* 25, 742-749.

ANGELO, J. B., HUANG, J. & CARDEN, D. (2005) Diabetes prevention: A review of current literature. *Advanced Studies in Medicine,* 5, 250-259.

ANONYMOUS (1996) Thioctacid T increases glucose utilization and reduces insulin resistance. *Munchener Medizinische Wochenschrift,* 138, 22-25.

ATKIN, W. (2003) Options for screening for colorectal cancer. *Scandinavian Journal of Gastroenterology,* 237, 13-16.

AVENELL, A., BROOM, J., BROWN, T. J., POOBALAN, A., AUCOTT, L., STEARNS, S. C., SMITH, W. C. S., JUNG, R. T., CAMPBELL, M. K. & GRANT, A. M. (2004) Systematic review of the long-term effects and economic consequences of treatments for obesity and implications for health improvement. *Health Technology Assessment,* 8.

BALKAU, B. (1999) Consequences of the new diagnostic criteria for diabetes in older men and women: The DECODE Study (Diabetes Epidemiology: Collaborative Analysis of Diagnostic Criteria in Europe). *Diabetes Care,* 22, 1667-1671.

BALKAU, B. (2000) The DECODE study. Diabetes epidemiology: collaborative analysis of diagnostic criteria in Europe. *Diabetes metabolism,* 26, 282-286.

BEGG, C. B. & MAZUMDAR, M. (1994) Operating characteristics of a rank correlation test for publication bias. *Biometrics,* 50, 1088-101.

BENNETT, C. M., GUO, M. & DHARMAGE, C. (2007) HbA1c as a screening tool for detection of type 2 diabetes: a systematic review. *Diabetic medicine,* 24, 333-343.

BERNANDO, J. M. & SMITH, A. F. M. (1994) *Bayesian theory,* Chichester, John Wiley & Sons Ltd.

BONORA, E., KIECHL, S., WILLEIT, J., OBERHOLLENZER, F., EGGER, G., MEIGS, J. B., BONADONNA, R. C., MUGGEO, M. & BRUNECK, S. (2004) Population-based incidence rates and risk factors for type 2 diabetes in white individuals: the Bruneck study. *Diabetes,* 53, 1782-9.

BRIGGS, A., SCULPHER, M. & CLAXTON, K. (2006) *Decision modelling for health economic evaluation*, Oxford University Press.

BRIGGS, A. H., ADES, A. E. & PRICE, M. J. (2003) Probabilistic sensitivity analysis for decision trees with multiple branches: Use of a dirichlet distribution in a bayesian framework. *Medical Decision Making,* 23, 341-350.

BRINDLE, P., MAY, M., GILL, P., CAPPUCCIO, F., D'AGOSTINO, R., SR., FISCHBACHER, C. & EBRAHIM, S. (2006) Primary prevention of cardiovascular disease: a web-based risk score for seven British black and minority ethnic groups. *Heart,* 92, 1595-1602.

BROOKS, S. P. & GELMAN, A. (2007) General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics,* 7, 434-455.

BROUWER, W. B., NIESSEN, L. W., POSTMA, M. J. & RUTTEN, F. F. (2005) Need for differential discounting of costs and health effects in cost effectiveness analysis. *British Medical Journal,* 331, 446-448.

BROWNE, W. J. & DRAPER, D. (2000) Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics,* 15, 391-420.

BUCHANAN, T. A., XIANG, A. H., PETERS, R. K., KJOS, S. L., MARROQUIN, A., GOICO, J., OCHOA, C., TAN, S., BERKOWITZ, K., HODIS, H. N. & AZEN, S. P. (2002) Preservation of pancreatic beta-cell function and prevention of type 2 diabetes by pharmacological treatment of insulin resistance in high-risk Hispanic women. *Diabetes,* 51, 2796-2803.

BURTON, P. R. (1994) Helping doctors to draw appropriate inferences from the analysis of medical studies. *Statistics in medicine,* 13, 1699-1713.
CALDWELL, D. M., ADES, A. E. & HIGGINS, J. P. T. (2006) Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *British Medical Journal,* 331, 897-900.
CAO, H. (2004) Effects of life style intervention on early insulin secretion in the patients with impaired glucose tolerance. *Zhongguo Linchuang Kangfu,* 8, 4937-4939.
CARO, J. J., GETSIOS, D., CAROS, I., KLITTICH, W. S. & O'BRIEN, J. A. (2004) Economic evaluation of therapeutic interventions to prevent Type 2 diabetes in Canada. *Diabetic Medicine,* 21, 1229-1236.
CDC DIABETES COST-EFFECTIVENESS STUDY GROUP (1998) The cost-effectiveness of screening for Type 2 diabetes. *JAMA,* 280, 1757-63.
CHEN, T. H. H., YEN, M. F. & TUNG, T. H. (2001) A computer simulated model for cost-effectiveness analysis of mass screening for Type 2 diabetes mellitus. *Diabetes Research and Clinical Practice,* 54, S37-S42.
CHIASSON, J. L., GOMIS, R., HANEFELD, M., JOSSE, R. G., KARASIK, A. & LAAKSO, M. (1998) The STOP-NIDDM Trial: An international study on the efficacy of an alpha-glucosidase inhibitor to prevent type 2 diabetes in a population with impaired glucose tolerance: rationale, design, and preliminary screening data. Study to Prevent Non-Insulin-Dependent Diabetes Mellitus. *Diabetes Care,* 21, 1720-5.
CHIASSON, J. L., JOSSE, R. G., GOMIS, R., HANEFELD, M., KARASIK, A. & LAAKSO, M. (2002) Acarbose for prevention of type 2 diabetes mellitus: The STOP-NIDDM randomised trial. *Lancet,* 359, 2072-2077.
CLARKE, P., GRAY, A. & HOLMAN, R. (2002) Estimating utility values for health states of type 2 diabetic patients using the EQ-5D (UKPDS 62). *Medical Decision Making,* 4, 340-349.
CLARKE, P. M., GRAY, A. M., BRIGGS, A., FARMER, A. J., FENN, P., STEVENS, R. J., MATTHEWS, D. R., STRATTON, I. M. & HOLMAN, R. R. (2004a) A model to estimate the lifetime health outcomes of patients with Type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS 68). *Diabetologia,* 47, 1747-59.
CLARKE, P. M., GRAY, A. M., BRIGGS, A., STEVENS, R. J., MATTHEWS, D. R. & HOLMAN, R. R. (2004b) Cost-utility analyses of intensive blood glucose and tight blood pressure control in type 2 diabetes. *Diabetologia,* 48.
CLAXTON, K. P. & SCULPHER, M. J. (2006) Using value of information analysis to prioritise health research. *Pharmacoeconomics,* 24, 1055-1068.
CLAYTON, D. & HILLS, M. (1993) *Statistical Models in Epidemiology*, Oxford University Press.
COLLETT, D. (1994) *Modelling survival data in medical research*, Chapman and Hall.
COOPER, N. J., ABRAMS, K. R., SUTTON, A. J., TURNER, D. & LAMBERT, P. C. (2003) A Bayesian approach to Markov modelling in cost-effectiveness analyses: application to taxane use in advanced breast cancer. *Journal of the Royal Statistical Society,* 166, 389-405.
COOPER, N. J., SUTTON, A. J. & ABRAMS, K. R. (2002) Decision analytical economic modelling within a Bayesian framework: application to prophylactic antibiotics use for caesarean section. *Statistical Methods in Medical Research,* 11, 491-512.
COOPER, N. J., SUTTON, A. J., ABRAMS, K. R., TURNER, D. & WAILOO, A. (2004) Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. *Health Economics,* 13, 203-26.
COOPER, N. J., SUTTON, A. J., LU, G. & KHUNTI, K. (2006) Mixed comparison of stroke prevention treatments in individuals with nonrheumatic atrial fibrillation. *Archives of Internal Medicine,* 166, 1269-75.
COSTA, B. (2002) Type 2 diabetes prevention: Education, or medication too?. *Medicina Clinica,* 119, 613-615.
COWLES, M. K. & CARLIN, B. P. (1996) Markov Chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association,* 91, 883-904.
CURTIS, L. & NETTEN, A. (2006) Unit costs of health and social care. IN KENT, U. O. (Ed.), Personal Social Services Research Unit.
DAVIES M, TRINGHAM J, FAROOQI A, JARVIS J, RAYMOND N, SKINNER C, THOMSON J & KHUNTI K (2003) STAR: screening those at risk: study design and recruitment. *Diabetic Medicine,* 20, 68 [Published abstracts].
DAVIES, M. J., TRINGHAM, J. R., TROUGHTON, J. & KHUNTI, K. K. (2004) Prevention of Type 2 diabetes mellitus. A review of the evidence and its application in a UK setting. *Diabetic Medicine,* 21, 403-414.

DE VEGT, F., DEKKER, J. M., JAGER, A., HIENKENS, E., KOSTENSE, P. J., STEHOUWER, C. D., NIJPELS, G., BOUTER, L. M. & HEINE, R. J. (2001) Relation of impaired fasting and postload glucose with incident type 2 diabetes in a Dutch population: The Hoorn Study. *JAMA,* 285, 2109-13.

DEEKS, J., GLANVILLE, J. & SHELDON, T. (1996) Undertaking systematic reviews of research on effectiveness: CRD guidelines for those carrying out or commissioning reviews. Report No. 4., Centre for reviews and dissemination. York: York Publishing Services Ltd.

DEEKS, J. J. (2002) Issues on the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine,* 21, 1575-1600.

DEPARTMENT OF HEALTH (2001) National Service Framework for Diabetes: Standards.

DIABETES PREVENTION PROGRAM RESEARCH, G. (2003) Effects of withdrawal from metformin on the development of diabetes in the diabetes prevention program. *Diabetes Care,* 26, 977-80.

DOLAN, P. (1997) Modeling valuations for EuroQol health states. *Medical Care,* 35, 1095–1108.

DONNELLY, R., EMSLIE-SMITH, A. M., GARDNER, I. D. & MORRIS, A. D. (2000) ABC of arterial and venous vascular disease: Vascular complications of diabetes. *British Medical Journal,* 320, 1062-1066.

DREAM TRIAL INVESTIGATORS (2006a) Effect of ramipril on the incidence of diabetes. *New England Journal of Medicine*, 1551-62.

DREAM TRIAL INVESTIGATORS (2006b) Effect of roglitazone on the frequency of diabetes in patients with impaired glucose tolerance or impaired fasting glucose: a randomised controlled trial. *Lancet*, 1096-105.

DRUMMOND, M. F., SCULPHER, M. J., TORRANCE, G. W., O'BRIEN, B. J. & STODDART, G. L. (2005) *Methods for the economic evaluation of health care programmes*, Oxford University Press.

DUVAL, S. & TWEEDIE, R. (2000) Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics,* 56, 455-63.

EBRAHIM, S., SMITH, G. D., MCCABE, C., PAYNE, N., PICKIN, M., SHELDON, T. A., LAMPE, F., SAMPSON, F., WARD, S. & WANNAMTHEE, G. (1998) Cholesterol and coronary heart disease: screening and treatment. *Quality Health Care,* 7, 232-239.

EDDY, D. (2006) Bringing health economic modeling to the 21st century. *Value in Health,* 9, 168-178.

EDDY, D. M. (1989) The confidence profile method: A Bayesian method for assessing health technologies. *Operartions research,* 37, 210-228.

EDDY, D. M. (1990) An Introduction to a Bayesian method for meta-analysis: The confidence profile method. *Medical Decision Making,* 10, 15-23.

EDDY, D. M., SCHLESSINGER, L. & KAHN, R. (2005) Clinical outcomes and cost-effectiveness of strategies for managing people at high risk for diabetes. *Annals of Internal Medicine,* 143, 251-64.

EDELSTEIN, S. L., KNOWLER, W. C., BAIN, R. P., ANDRES, R., BARRETT-CONNOR, E. L., DOWSE, G. K., HAFFNER, S. M., PETTITT, D. J., SORKIN, J. D., MULLER, D. C., COLLINS, V. R. & HAMMAN, R. F. (1997a) Predictors of progression from impaired glucose tolerance to NIDDM: An analysis of six prospective studies. *Diabetes,* 46, 701-710.

EDELSTEIN, S. L., KNOWLER, W. C., BAIN, R. P., ANDRES, R., BARRETT-CONNOR, E. L., DOWSE, G. K., HAFFNER, S. M., PETTITT, D. J., SORKIN, J. D., MULLER, D. C., COLLINS, V. R. & HAMMAN, R. F. (1997b) Predictors of progression from impaired glucose tolerance to NIDDM: an analysis of six prospective studies. *Diabetes,* 46, 701-10.

EGGER, M., SMITH, G. D., SCHNEIDER, M. & MINDER, C. (1997) Bias in meta-analysis detected by a simple graphical test. *British Medical Journal,* 315, 629-34.

ERIKSSON, J., LINDSTROM, J., VALLE, T., AUNOLA, S., HAMALAINEN, H., ILANNE-PARIKKA, P., KEINANEN-KIUKAANNIEMI, S., LAAKSO, M., LAUHKONEN, M., LEHTO, P., LEHTONEN, A., LOUHERANTA, A., MANNELIN, M., MARTIKKALA, V., RASTAS, M., SUNDVALL, J., TURPEINEN, A., VILJANEN, T., UUSITUPA, M. & TUOMILEHTO, J. (1999) Prevention of Type II diabetes in subjects with impaired glucose tolerance: The Diabetes Prevention Study (DPS) in Finland. Study design and 1-year interim report on the feasibility of the lifestyle intervention programme. *Diabetologia,* 42, 793-801.

ERIKSSON, J. G., LEHTOVIRTA, M., EHRNSTROM, B., SALMELA, S. & GROOP, L. (2006) Long-term beneficial effects of glipizide treatment on glucose tolerance in subjects with impaired glucose tolerance. *Journal of Internal Medicine,* 259, 553-60.

ERIKSSON, K. F. & LINDGARDE, F. (1991) Prevention of Type 2 (non-insulin-dependent) diabetes mellitus by diet and physical exercise. The 6-year Malmo feasibility study. *Diabetologia,* 34, 891-898.

ERIKSSON, K. F. & LINDGARDE, F. (1998) No excess 12-year mortality in men with impaired glucose tolerance who participated in the Malmo Preventive Trial with diet and exercise. *Diabetologia,* 41, 1010-6.

FAN, G. J., LUO, G. B. & QIN, M. L. (2004) Effect of jiangtang bushen recipe in intervention treatment of patients with impaired glucose tolerance. *Chinese Journal of Integrated Traditional & Western Medicine,* 24, 317-20.

FANG, Y. S., LI, T. Y. & CHEN, S. Y. (2004) Effect of medicine and non-medicine intervention on the outcomes of patients with impaired glucose tolerance: 5-year follow-up. *Zhongguo Linchuang Kangfu,* 8, 6562-6563.

GAMERMAN, D. (1997) *Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference.*, Chapman and Hall.

GILKS, W. R., RICHARDSON, S. & SPIELGELHALTER, D. J. (1996) *Markov Chain Monte Carlo in Practice,* London, U.K., Chapman and Hall.

GILL, S. & SINICROPE, F. A. (2005) Colorectal cancer prevention: is an ounce of prevention worth a pound of cure? *Seminars in Oncology,* 32, 24-34.

GILLIES, C. L., ABRAMS, K. R., LAMBERT, P. C., COOPER, N. J., SUTTON, A. J., HSU, R. T. & KHUNTI, K. (2007) Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance. *British Medical Journal,* 334, 299-302.

GLASS, G. V. (1976) Primary, secondary and meta-analysis of research. *Educational Researcher,* 5, 3-8.

GLUMER, C., YUYUN, M., GRIFFIN, S., FAREWELL, D., SPIEGELHALTER, D., KINMOUTH, A. L. & WAREHAM, N. J. (2006) What determines the cost-effectiveness of diabetes screening? *Diabetologia,* 49, 1536-44.

GOULD, A. L., ROSSOUW, J. E., SANTANELLO, N. C., HEYSE, J. F. & FURBERG, C. D. (1998) Cholesterol reduction yields clinical benefit: impact of statin trials. *Circulation,* 97, 946-52.

GRAU, S., AGUADO, J. M., MATEU-DE ANTONIO, J., GONZALEZ, P. & DEL CASTILLO, A. (2007) Economic evaluation of linezolid versus teicoplanin for the treatment of infections caused by gram-positive microorganisms in Spain. *Journal of Chemotherapy,* 19, 398-409.

GRIFFIN, S. J., LITTLE, P. S., HALES, C. N., KINMONTH, A. L. & WAREHAM, N. J. (2000) Diabetes risk score: Towards earlier detection or type 2 diabetes in general practice. *Diabetes/Metabolism Research Reviews,* 16, 164-171.

GULLIFORD, M. C., CHARLTON, J. & LATINOVIC, R. (2005) Increased utilization of primary care 5 years before diagnosis of Type 2 diabetes. *Diabetes Care,* 28, 47-52.

HARDY, R. & THOMPSON, S. G. (1996) A likelihood approach to meta-analysis with random effects. *Statistics in Medicine,* 15, 619-629.

HARRIS, M. I., KLEIN, R., WELBORN, T. A., KNUIMAN, M. W., HARRIS, M. I., KLEIN, R., WELBORN, T. A. & KNUIMAN, M. W. (1992) Onset of NIDDM occurs at least 4-7 yr before clinical diagnosis. *Diabetes Care,* 15, 815-9.

HEITJAN, D. F., MOSKOWITZ, A. J. & WHANG, W. (1999) Problems with interval estimates of the incremental cost-effectiveness ratio. *Medical Decision Making,* 19, 102-3.

HELLMICH, M., ABRAMS, K. R. & SUTTON, A. J. (1999) Bayesian approaches to meta-analysis of ROC curves. *Medical Decision Making,* 19, 252-64.

HENEGHAN, C., THOMPSON, M. & PERERA, R. (2006) Prevention of diabetes. *British Medical Journal,* 333, 764-765.

HERMAN, W. H., HOERGER, T. J., BRANDLE, M., HICKS, K., SORENSEN, S., ZHANG, P., HAMMAN, R. F., ACKERMANN, R. T., ENGELGAU, M. M., RATNER, R. E., DIABETES PREVENTION PROGRAM RESEARCH, G., HERMAN, W. H., HOERGER, T. J., BRANDLE, M., HICKS, K., SORENSEN, S., ZHANG, P., HAMMAN, R. F., ACKERMANN, R. T., ENGELGAU, M. M., RATNER, R. E. & GROUP, D. P. P. R. (2005) The cost-effectiveness of lifestyle modification or metformin in preventing type 2 diabetes in adults with impaired glucose tolerance. *Annals of Internal Medicine,* 142, 323-32.

HEYMSFIELD, S. B., SEGAL, K. R., HAUPTMAN, J., LUCAS, C. P., BOLDRIN, M. N., RISSANEN, A., WILDING, J. P. & SJOSTROM, L. (2000) Effects of weight loss with orlistat on glucose tolerance and progression to type 2 diabetes in obese adults. *Archives of Internal Medicine,* 160, 1321-6.

HIGGINS, J. P. T. & GREEN, S. (2005) *Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 [updated May 2005].* Chichester, UK, John Wiley & Sons, Ltd.

HIGGINS, J. P. T. & THOMPSON, S. G. (2002) Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine,* 21, 1539-1558.
HIGGINS, J. P. T., THOMPSON, S. G., DEEKS, J. J. & ALTMAN, D. G. (2003) Measuring inconsistency in meta-analyses. *British Medical Journal,* 327, 557-560.
HIROSE, T. (2005) Effects of nateglinide in impaired glucose tolerance subjects. *Nippon Rinsho - Japanese Journal of Clinical Medicine,* 63 Suppl 2, 438-43.
HOERGER, T. J., HARRIS, R., HICKS, K. A., DONAHUE, K., SORENSEN, S. & ENGELGAU, M. (2004) Screening for Type 2 diabetes mellitus: A cost-effectiveness analysis. *Annals of Internal Medicine,* 140, 756-58.
HOETING, J. A., MADIGAN, D., RAFTERY, A. E. & VOLINSKY, C. T. (1999) Bayesian Model Averaging - A Tutorial (with discussion). *Statistical Science,* 14, 382-417.
HOLMAN, R. R., BLACKWELL, L., STRATTON, I. M., MANLEY, S. E., TUCKER, L. & FRIGHI, V. (2003) Six-year results from the Early Diabetes Intervention Trial. *Diabetic Medicine,* 20, 15.
HOLMAN, R. R., NORTH, B. V. & TUNBRIDGE, F. K. (2000) Possible Prevention of Type 2 Diabetes with Acarbose or Metformin. *Diabetes,* 49, A111-A112.
HOUGAARD, P. (1995) Frailty models for survival data. . *Lifetime Data Analysis,* 1, 255-273.
HUNINK, M. & GLASZIOU, P. (2001) *Decision making in health and medicine: Integrating evidence and values,* Cambridge, U.K., Cambridge University Press.
ICKS, A., HAASTERT, B., GANDJOUR, A., JOHN, J., LOWEL, H., HOLLE, R., GIANI, G. & RATHMANN, W. (2004) Cost-effectiveness analysis of different screening procedures for Type 2 Diabetes. *Diabetes Care,* 27, 2120-28.
ICKS, A., RATHMANN, W., HAASTERT, B., GANDJOUR, A., HOLLE, R., JOHN, J. & GIANI, G. (2007) Clinical and cost-effectiveness of primary prevention of Type 2 diabetes in a 'real world' routing healthcare setting: model based on the KORA Survey 2000. *Diabetic Medicine,* 24, 473-80.
INTERNATIONAL DIABETES FEDERATION (2006) Diabetes Atlas. *http://www.eatlas.idf.org*.
JACKSON, C., BEST, N. & RICHARDSON, S. (2006) Links Improving ecological inference using individual-level data. *Statistics in Medicine,* 25, 2136-59.
JACOBS-VAN DER BRUGGEN, M. A. M., BOS, G., BEMELMANS, W. J., HOOGENVEEN, R. T., VIJGEN, S. M. & BAAN, C. A. (2007) Lifestyle interventions are cost-effective in people with different levels of diabetes risk. *Diabetes Care,* 30, 128-34.
JADAD, A. R., MOORE, A. R., CARROLL, D., JENKINSON, C., REYNOLDS, D. J. M., GAVAGHAN, D. J. & MCQUAY, H. J. (1996) Assessing the quality of reports of randomised clinical trials: Is blinding necessary? *Controlled Clinical Trials,* 17, 1-12.
JARRETT, R. J., KEEN, H., FULLER, J. H. & MCCARTNEY, M. (1979) Worsening to Diabetes in Men with Impaired Glucose Tolerance ("Borderline Diabetes"). *Diabetologia,* 16, 25-30.
JARRETT, R. J., KEEN, H. & MCCARTNEY, P. (1984) The Whitehall study: Ten year follow-up report on men with impaired glucose tolerance with reference to worsening to diabetes and predictors of death. *Diabetic Medicine,* 1, 279-283.
KAPLAN, N. M. (1989) The deadly quartet: upper body adiposity, glucose intolerance, hypertriglyceridaemia and hypertension. *Archives of Internal Medicine,* 149, 1514-20.
KEEN, H., JARRETT, R. J., WARD, J. D. & FULLER, J. H. (1973) Borderline diabetics and their response to tolbutamide. *Advances in Metabolic Disorders,* 2, Suppl 2:521-31.
KING, H., AUBERT, R. E. & HERMAN, W. H. (1998) Global burden of diabetes, 1995-2025: prevalence numerical estimates, and projections. *Diabetes Care,* 21, 1414-31.
KNOWLER, W. C., BARRETT-CONNOR, E., FOWLER, S. E., HAMMAN, R. F., LACHIN, J. M., WALKER, E. A. & NATHAN, D. M. (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *New England Journal of Medicine,* 346, 393-403.

KNOWLER, W. C., HAMMAN, R. F., EDELSTEIN, S. L., BARRETT-CONNOR, E., EHRMANN, D. A., WALKER, E. A., FOWLER, S. E., NATHAN, D. M., KAHN, S. E. & DIABETES PREVENTION PROGRAM RESEARCH, G. (2005) Prevention of type 2 diabetes with troglitazone in the Diabetes Prevention Program. *Diabetes,* 54, 1150-6.
KOSAKA, K., NODA, M. & KUZUYA, T. (2005) Prevention of type 2 diabetes by lifestyle intervention: a Japanese trial in IGT males. 152-62, 2005 Feb.
KUZUYA, H. (2004) Prevention of Type 2 Diabetes in Subjects with Impaired Glucose Tolerance. *Journal of the Japan Diabetes Society,* 47, 14-17.

LAMBERT, P. C., SUTTON, A. J., BURTON, P. R., ABRAMS, K. R. & JONES, D. R. (2005) How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine,* 24, 2401-2428.

LAMBERT, P. C., SUTTON, A. J. & JONES, D. R. (2002) A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology*, 86-94.

LAUPACIS, A., SACKETT, D. L. & ROBERTS, R. S. (1988) An assessment of clinically useful measures of the consequences of treatment. *New England Journal of Medicine,* 318, 1728-33.

LAURITZEN, T., GRIFFIN, S., BORCH-JOHNSEN, K., WAREHAM, N. J., WOLFFENBUTTEL, B. H. R. & RUTTEN, G. (2000) The ADDITION study: proposed trial of the cost-effectiveness of an intensive multi-factorial intervention on morbidity and mortality among people with Type 2 diabetes screen detected by screening. *International Journal of Obesity,* Supplement 3, S6-S11.

LI, C. L., PAN, C. Y., LU, J. M., ZHU, Y., WANG, J. H., DENG, X. X., XIA, F. C., WANG, H. Z. & WANG, H. Y. (1999) Effect of metformin on patients with impaired glucose tolerance. *Diabetic Medicine,* 16, 477-481.

LIAO, D., ASBERRY, P. J., SHOFER, J. B., CALLAHAN, H., MATTHYS, C., BOYKO, E. J., LEONETTI, D., KAHN, S. E., AUSTIN, M., NEWELL, L., SCHWARTZ, R. S. & FUJIMOTO, W. Y. (2002) Improvement of BMI, body composition, and body fat distribution with lifestyle modification in Japanese Americans with impaired glucose tolerance. *Diabetes Care,* 25, 1504-10.

LINDSTROM, J., ERIKSSON, J. G., VALLE, T. T., AUNOLA, S., CEPAITIS, Z., HAKUMAKI, M., HAMALAINEN, H., ILANNE-PARIKKA, P., KEINANEN-KIUKAANNIEMI, S., LAAKSO, M., LOUHERANTA, A., MANNELIN, M., MARTIKKALA, V., MOLTCHANOV, V., RASTAS, M., SALMINEN, V., SUNDVALL, J., UUSITUPA, M. & TUOMILEHTO, J. (2003a) Prevention of diabetes mellitus in subjects with impaired glucose tolerance in the Finnish Diabetes Prevention Study: Results from a randomized clinical trial. *Journal of the American Society of Nephrology,* 14, S108-S113.

LINDSTROM, J., LOUHERANTA, A., MANNELIN, M., RASTAS, M., SALMINEN, V., ERIKSSON, J., UUSITUPA, M. & TUOMILEHTO, J. (2003b) The Finnish Diabetes Prevention Study (DPS): Lifestyle intervention and 3-year results on diet and physical activity. *Diabetes Care,* 26, 3230-3236.

LINDSTROM, J. & TUOMILEHTO, J. (2003) The diabetes risk score: A practical tool to predict type 2 diabetes risk. *Diabetes Care,* 26, 725-731.

LU, G. & ADES, A. E. (2004) Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine,* 23, 3105-3124.

MATCHAR, D. B., SAMSA, G. P., MATTHEWS, J. R., ANCUKIEWICZ, M., PARMIGIANI, G., HASSELBLAD, V., WOLF, P. A., D'AGOSTINO, R. B. & LIPSCOMB, J. (1997) The Stroke Prevention Policy Model: Linking evidence and clinical decisions. *Annals of Internal Medicine,* 127, 704-711.

MEIGS, J. B., MULLER, D. C., NATHAN, D. M., BLAKE, D. R., ANDRES, R., BALTIMORE LONGITUDINAL STUDY OF, A., MEIGS, J. B., MULLER, D. C., NATHAN, D. M., BLAKE, D. R. & ANDRES, R. (2003) The natural history of progression from normal glucose tolerance to type 2 diabetes in the Baltimore Longitudinal Study of Aging. *Diabetes,* 52, 1475-84.

MILLER, D. K. & HOMAN, S. M. (1994) Determining transition probabilities. *Medical Decision Making,* 14, 52-58.

MKRTUMIAN, A. M. (2002) Optimal drug therapy of glucose intolerance is a guarantee of successful prevention of type 2 diabetes mellitus. *Terapevticheskii Arkhiv,* 74, 86-9.

NARAYAN, K. M. V., ZHANG, P., KANAYA, A. M., WILLIAMS, D. E., ENGELGAU, M. M., IMPERATORE, G. & RAMACHANDRAN, A. (2006) Diabetes: The pandemic and potential solutions. *Disease control priorities in developing countries.* Second ed.

NATHAN, D. M., HENRY, R. R., DAVIDSON, M. B., PRATLEY, R., DEFRONZO, R. A., ZINMAN, B. & HEINE, R. J. (2007) Impaired fasting glucose and impaired glucose tolerance. *Diabetes care,* 30, 753-759.

NATIONAL INSTITUTE FOR CLINICAL EXCELLENCE (2004) Guide to the methods of technology appraisal. London, NICE.

NCEP (2001) Executive summary of the third report of the national cholesterol education program (NCEP) expert panel on detection, evaluation and treatment of high blood cholesterol in adults (Adult treatment panel III). *JAMA,* 285, 2486-97.

NHSDIRECT (2007) Coronary heart disease. *http://www.nhsdirect.nhs.uk/articles/article.aspx?articleId=444&sectionId=30373*.

NIKLASON, A., HEDNER, T., NISKANEN, L., LANKE, J. & CAPTOPRIL PREVENTION PROJECT STUDY, G. (2004) Development of diabetes is retarded by ACE inhibition in hypertensive patients--a subanalysis of the Captopril Prevention Project (CAPPP). *Journal of Hypertension,* 22, 645-52.

NIXON, R. M., BANSBACK, N. & BRENNAN, A. (2007) Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Statistics in Medicine,* 26, 1237-54.

NIXON, R. M. & DUFFY, S. W. (2002) Cross-issue synthesis: potential application to breast cancer, tamoxifen and genetic susceptibility. *Journal of Cancer Epidemiology and Prevention,* 7, 205-212.

NORRIS, S. L., ZHANG, X., AVENELL, A., GREGG, E., BOWMAN, B., SCHMID, C. H. & LAU, J. (2005a) Long-term effectiveness of weight-loss interventions in adults with pre-diabetes: A review. *American Journal of Preventive Medicine,* 28, 126-139.

NORRIS, S. L., ZHANG, X., AVENELL, A., GREGG, E., SCHMID, C. H. & LAU, J. (2005b) Long-term non-pharmacological weight loss interventions for adults with prediabetes. *Cochrane Database of Systematic Reviews*, CD005270.

PADWAL, R. & LAUPACIS, A. (2004) Antihypertensive Therapy and Incidence of Type 2 Diabetes: A systematic review. *Diabetes Care,* 27, 247-255.

PALMER, A. J., ROZE, S., VALENTINE, W. J., SPINAS, G. A., SHAW, J. E. & ZIMMET, P. Z. (2004) Intensive lifestyle changes or metformin in patients with impaired glucose tolerance: Modeling the long-term health economic implications of the diabetes prevention program in Australia, France, Germany, Switzerland, and the United Kingdom. *Clinical Therapeutics,* 26, 304-321.

PAN, C. Y., GAO, Y., CHEN, J. W., LUO, B. Y., FU, Z. Z., LU, J. M., GUO, X. H. & CHENG, H. (2003) Efficacy of acarbose in Chinese subjects with impaired glucose tolerance. *Diabetes Research & Clinical Practice,* 61, 183-190.

PAN, X. R., LI, G. W., HU, Y. H., WANG, J. X., YANG, W. Y., AN, Z. X., HU, Z. X., LIN, J., XIAO, J. Z., CAO, H. B., LIU, P. A., JIANG, X. G., JIANG, Y. Y., WANG, J. P., ZHENG, H., ZHANG, H., BENNETT, P. H. & HOWARD, B. V. (1997) Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. *Diabetes Care,* 20, 537-44.

PARMAR, M. K. B., TORRI, V. & STEWART, L. (1998) Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine,* 17, 2815-2834.

PARMIGIANI, G., SAMSA, G. P., ANCUKIEWICZ, M., LIPSCOMB, J., HASSELBLAD, V. & MATCHAR, D. B. (1997) Assessing uncertainty in cost-effectiveness analyses: Application to a complex decision model. *Medical Decision Making,* 17, 390-401.

PHILIPS, Z., BOJKE, L., SCULPHER, M., CLAXTON, K. & GOLDER, S. (2006) Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality of assessment. *Pharmacoeconomics*, 355-371.

PRISANT, L. M. (2004) Preventing Type II Diabetes Mellitus. *Journal of Clinical Pharmacology,* 44, 406-413.

RAMACHANDRAN, A., SNEHALATHA, C., MARY, S., MUKESH, B., BHASKAR, A. D. & VIJAY, V. (2006) The Indian Diabetes Prevention Programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). *Diabetologia,* 49, 289-297.

RAMACHANDRAN, S., FRENCH, J. M., VANDERPUMP, M. J. P., CROFT, P. & NEARY, R. H. (2000) Using the Framingham model to predict heart disease in the United Kingdom: retrospective study. *British Medical Journal,* 320, 676-677.

RAMSOEKH, D., VAN LEERDAM, M. E., VAN BALLEGOOIJEN, M., HABBEMA, J. D. & KUIPERS, E. J. (2007) Population screening for colorectal cancer: Faeces, endoscopes or x-rays? *Cell Oncology,* 29, 185-94.

RASMUSSEN, S. S., GLUMER, C., SANDBAEK, A., LAURITZEN, T. & BORCH-JOHNSEN, K. (2006) Progression from impaired fasting glucose and impaired glucose tolerance to diabetes in a high-risk screening programme in general practice: the ADDITION study, Denmark. *Diabetologia,* 50, 293-7.

RILEY, R. D., SIMMONDS, M. C. & LOOK, M. P. (2007) Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. *Journal of Clinical Epidemiology,* 60, 431-39

ROSSING, P., HOUGAARD, P., BORCH-JOHNSEN, K., PARVING, H. H., ROSSING, P., HOUGAARD, P., BORCH-JOHNSEN, K. & PARVING, H. H. (1996) Predictors of mortality in insulin dependent diabetes: 10 year observational follow up study. *BMJ,* 313, 779-84.

SAKANE, N. (2005) Japan Diabetes Prevention Program. *Nippon Rinsho - Japanese Journal of Clinical Medicine,* 63 Suppl 2, 488-92.

SAMSA, G. P., REUTTER, R. A., PARMIGIANI, G., ANCUKIEWICZ, M., ABRAHAMSE, P., LIPSCOMBE, J. & MATCHAR, D. B. (1999) Performing cost-effectiveness analysis by integrating randomized trial data with a comprehensive decision model: Application to treatment of acute ischemic stroke. *Journal of Clinical Epidemiology,* 52, 259-271.

SANGHA, S., YAO, M. & WOLFE, M. M. (2004) Non-steroidal anti-inflammatory drugs and colorectal cancer prevention. *Postgraduate Medical Journal,* 81, 223-227.

SARTOR, G., SCHERSTEN, B., CARLSTROM, S. & ET AL. (1980) Ten-year follow-up of subjects with impaired glucose tolerance. Prevention of diabetes by tolbutamide and diet regulation. *Diabetes,* 29, 41-49.

SCULPHER, M. & CLAXTON, K. (2005) Establishing the cost-effectiveness of new pharmaceuticals under conditions of uncertainty- when is there sufficient evidence? *Value in Health,* 8, 433-446.

SEGAL, L., DALTON, A. C. & RICHARDSON, J. (1998) Cost-effectiveness of the primary prevention of non-insulin dependent diabetes mellitus. *Health Promotion International,* 13, 197-209.

SENDI, P. P., CRAIG, B. A., MEIER, G., PFLUGER, D., GAFNI, A., OPRAVIL, M., BATTEGAY, M. & BUCHER, H. C. (1999) Cost-effectiveness of azithromycin for preventing *Myobacterium avium* complex infection in HIV-positive patients in the era of highly active antretroviral therapy. *Journal of Antimicrobial Chemotherapy,* 44, 811-817.

SEVERENS, J. L., DE BOO, T. M. & KONST, E. M. (1999) Uncertainty of incremental cost-effectiveness ratios. A comparison of Fieller and bootstrap confidence intervals. *International Journal of Technology Assessment Health Care,* 15, 608-14

SHARP, S. J. & THOMPSON, S. G. (2000) Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. *Statistics in Medicine,* 19, 3251-3274.

SHIRASAYA, K., MIYAKAWA, M., YOSHIDA, K., TAKAHASHI, E., SHIMADA, N. & KONDO, T. (1999) Economic evaluation of alternative indicators for screening for diabetes mellitus. *Preventive Medicine,* 29, 79-86.

SONNENBERG, A., DELCO, F. & INADOMI, J. M. (2000) Cost-effectiveness of colonoscopy in screening for colorectal cancer. *Annals of Internal Medicine,* 133, 573-584.

SPIEGELHALTER, D. J., ABRAMS, K. R. & MYLES, J. P. (2004) *Bayesian approaches to clinical trials and health-care evaluation*, Wiley.

SPIEGELHALTER, D. J. & BEST, N. G. (2003) Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statistics in Medicine,* 22, 3687-3709.

SPIEGELHALTER, D. J., BEST, N. G., A., T. & R., G. W. BUGS: Bayesian Inference Using Gibbs Sampling, Version 1.4.1. Cambridge MRC Biostatistics Unit.

SPIEGELHALTER, D. J., MYLES, J. P., JONES, D. R. & ABRAMS, K. R. (1999) Methods in health service research: An introduction to Bayesian methods in health technology assessment. *British Medical Journal,* 319, 508-512.

SPIEGELHALTER, D. J., THOMAS, A. & BEST, N. (2000) *WinBUGS User Manual*.

SRINIVASAN, B. T., DAVIES, M. J., WEBB, D., HEALEY, E., FAROOQI, A., HILES, S., MANDALIA, P., STOCKMAN, J., GREEN, N. & KHUNTI, K. (2007) Baseline charactersitics and risk of progression from prediabetes to type 2 diabetes in a multiethnic, population based screening. *Diabetic medicine,* 24, 51.

STATACORP (2001) *Stata Statistical Software*, College Station: Stata Corporation.

SUTTON, A. J. & ABRAMS, K. R. (2001) Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research,* 10, 277-303.

SUTTON, A. J., ABRAMS, K. R., JONES, D. R., SHELDON, T. A. & SONG, F. (2000) *Methods for Meta-Analysis in Medical Research*, John Wiley & Sons, Ltd.

SWINBURN, B. A., METCALF, P. A. & LEY, S. J. (2001) Long-term (5-year) effects of a reduced-fat diet intervention in individuals with glucose intolerance. *Diabetes Care,* 24, 619-624.

TAO, L. L., DENG, Y. B., FAN, X. B. & BAO, Q. D. (2004) Effect of exercise training in patients with impaired glucose tolerance. *Zhongguo Linchuang Kangfu,* 8, 2912-2913.

TENENBAUM, A., MOTRO, M., FISMAN, E. Z., SCHWAMMENTHAL, E., ADLER, Y., GOLDENBERG, I., LEOR, J., BOYKO, V., MANDELZWEIG, L. & BEHAR, S. (2004) Peroxisome Proliferator-Activated Receptor Ligand Bezafibrate for Prevention of Type 2 Diabetes Mellitus in Patients with Coronary Artery Disease. *Circulation,* 109, 2197-2202.

THE DECODE STUDY GROUP (1999) Glucose tolerance and mortality: comparison of WHO and American Diabetes Association diagnostic criteria. . European Diabetes Epidemiology Group. Diabetes Epidemiology: Collaborative analysis Of Diagnostic criteria in Europe. *Lancet,* 354, 617-21.

THOMPSON, J., PALMER, T. & MORENO, S. (2006) Bayesian analysis in Stata using WinBUGS. *The Stata Journal,* 6, 530-549.

TODOROVA, K., PALAVEEV, O., PETKOVA, V. B., STEFANOVA, M. & DIMITROVA, Z. (2007) A pharmacoeconomical model for choice of a treatment for pregnant women with gestational diabetes. *Acta Diabetology,* 44, 144-148.

TORGERSON, J. S., HAUPTMAN, J., BOLDRIN, M. N. & SJOSTROM, L. (2004) XENical in the prevention of diabetes in obese subjects (XENDOS) study: A randomized study of orlistat as an adjunct to lifestyle changes for the prevention of type 2 diabetes in obese patients. *Diabetes Care,* 27, 155-61.

UK PROSPECTIVE DIABETES STUDY (UKPDS) GROUP (1991) Study design, progress and performance (UKPDS VIII). *Diabetologia,* 34, 877-90.

UK PROSPECTIVE DIABETES STUDY (UKPDS) GROUP (1998a) Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet,* 352, 837-853.

UK PROSPECTIVE DIABETES STUDY (UKPDS) GROUP (1998b) Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. UK Prospective Diabetes Study Group. *BMJ,* 317, 703-13.

UNWIN, N., SHAW, J., ZIMMET, P. & ALBERTI, K. (2002) Impaired glucose tolerance and impaired fasting glycaemia: The current status on definition and intervention. *Diabetic Medicine,* 19, 708-723.

VAN HOUT, B. A., AL, M. J., GORDON, G. S. & RUTTEN, F. F. H. (1994) Costs, effects and cost-effectiveness ratios alongside a clinical trial. *Health Economics,* 3, 309-319.

VIJAN, S., HWANG, E. W., HOFER, T. P. & HAYWARD, R. A. (2001) Which colon cancer screening test? A comparison of cost-effectiveness and compliance. *American Journal of Medicine,* 111, 593-601.

WARD, S., LLOYD JONES, M., PANDOR, A., HOLMES, M., ARA, R., RYAN, A., YEO, W. & PAYNE, N. (2007) A systematic review and economic evaluation of statins for the prevention of coronary events. *Health Technology Assessment,* 11.

WAREHAM, N. J., BYRNE, C. D., WILLIAMS, R., DAY, N. E. & HALES, C. N. (1999) Fasting proinsulin concentrations predict the development of type 2 diabetes. *Diabetes Care,* 22, 262-70.

WAREHAM, N. J. & GRIFFIN, S. J. (2007) Should we screen for type 2 diabetes? Evaluation against National Screening Committee criteria. *British Medical Journal,* 322, 986-988.

WAUGH, N., SCOTLAND, G., MCNAMEE, P., GILLETT, M., BRENNAN, A., GOYDER, E., WILLIAMS, R. & JOHN, A. (2007) Screening for type 2 diabetes: literature review and economic modelling. *Health Technology Assessment,* 11.

WEIN, P., BEISCHER, N., HARRIS, C. & PERMEZEL, M. (1999) A trial of simple versus intensified dietary modification for prevention of progression to diabetes mellitus in women with impaired glucose tolerance. *Australian & New Zealand Journal of Obstetrics & Gynaecology,* 39, 162-166.

WELTON, N. J., WHITE, I. R., LU, G., HIGGINS, J. P., HILDEN, J. & ADES, A. E. (2007) Correction: Interpretation of random effects meta-analysis in decision models. *Medical Decision Making,* 27, 212-4.

WILD, S., ROGLIC, G., GREEN, A., SICREE, R. & KING, H. (2004) Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care,* 27, 1047-1053.

WORLD HEALTH ORGANISATION (1999) Definition, diagnosis and classification of diabetes mellitus and its complications. IN DEPARTMENT OF NONCOMMUNICABLE DISEASE SURVEILLANCE (Ed.) Geneva.

WORLD HEALTH ORGANISATION EXPERT COMMITTEE (1980) Diabetes Mellitus, Second Report, Technical Report Series No. 646.

WORLD HEALTH ORGANISATION EXPERT COMMITTEE (1994) Prevention of Diabetes Mellitus. Technical Report Series No. 844. 17-18.

WORLD HEALTH ORGNISATION EXPERT COMMITTEE (1985) Diabetes Mellitus. Technical Report Series No. 742.

YAMAOKA, K. & TANGO, T. (2005) Efficacy of lifestyle education to prevent type 2 diabetes: a meta-analysis of randomised controlled trials. *Diabetes Care,* 28, 2780-7.

YUSUF, S., GERSTEIN, H., HOOGWERF, B., POGUE, J., BOSCH, J., WOLFFENBUTTEL, B. H., ZINMAN, B. & INVESTIGATORS, H. S. (2001) Ramipril and the development of diabetes. *JAMA,* 286, 1882-5.

ZHANG, P., ENGELGAU, M. E., VALDEZ, R., BENJAMIN, S. M., CADWELL, B. & NARAYAN, K. M. V. (2003) Costs of screening for pre-diabetes among U.S. adults. *Diabetes Care,* 26, 2536-2542.

ZHANG, P., ENGELGAU, M. E., VALDEZ, R., CADWELL, B., BENJAMIN, S. M. & NARAYAN, K. M. V. (2005) Efficient cutoff points for three screening tests for detecting undiagnosed diabetes and pre-diabetes. *Diabetes Care,* 28, 1321-1325.